A Qualitative Numerical Study of High Dimensional Dynamical Systems

by
David James Albers

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
(Physics)

at the

University of Wisconsin-Madison
2004

To my parents for their love and support.

Abstract

Since Poincaré, the father of modern mathematical dynamical systems, much effort has been exerted to achieve a qualitative understanding of the physical world via a qualitative understanding of the functions we use to model the physical world. In this thesis, we construct a numerical framework suitable for a qualitative, statistical study of dynamical systems using the space of artificial neural networks. We analyze the dynamics along intervals in parameter space, separating the set of neural networks into roughly four regions: the fixed point to the first bifurcation; the route to chaos; the chaotic region; and a transition region between chaos and finite-state neural networks. The study is primarily with respect to high-dimensional dynamical systems. We make the following general conclusions as the dimension of the dynamical system is increased: the probability of the first bifurcation being of type Neimark-Sacker is greater than ninety-percent; the most probable route to chaos is via a cascade of bifurcations of high-period periodic orbits, quasi-periodic orbits, and 2-tori; there exists an interval of parameter space such that hyperbolicity is violated on a countable, Lebesgue measure 0, "increasingly dense" subset; chaos is much more likely to persist with respect to parameter perturbation in the chaotic region of parameter space as the dimension is increased; moreover, as the number of positive Lyapunov exponents is increased, the likelihood that any significant portion of these positive exponents can be perturbed away decreases with increasing dimension. The maximum Kaplan-Yorke dimension and the maximum number of positive Lyapunov exponents increases linearly with dimension. The probability of a dynamical system being chaotic increases exponentially with dimension. The results with respect to the first bifurcation and the route to chaos comment on previous results of Newhouse, Ruelle, Takens, Broer, Chenciner, and Iooss. Moreover, results regarding the high-dimensional chaotic region of parameter space is interpreted and related to the closing lemma of Pugh, the windows conjecture of Barreto, the stable ergodicity theorem of Pugh and Shub, and structural stability theorem of Robbin, Robinson, and Mañé.

Contents

Abstract **ii**

List of Figures **ix**

1 Introduction and background **1**
1.1 Introduction 1
1.2 Introduction to our construction 2
1.3 Routes to Turbulence — Regions I and II 2
1.4 The Dynamic Stability Dream — region IV 4
1.5 Mathematics versus computation 5

2 Notions of dynamic stability and previous results **7**
2.1 Notions of dynamic stability and equivalence 7
2.1.1 Hyperbolicity and Lyapunov exponents 7
2.1.1.1 General conditions needed for the existence of Lyapunov exponents 9
2.1.1.2 SRB measures 9
2.1.2 Topological equivalence 9
2.1.3 Robust dynamics - k-degree LCE stability 10
2.1.4 Ergodicity 11
2.2 Functional Versus Parameter Perturbation 11
2.2.1 C^r perturbations 12
2.2.2 Parameter perturbations 12
2.2.3 Parameter perturbations in neural networks 13
2.3 Previous results 13
2.3.1 Non-density of structural stability 13
2.3.2 The closing lemma 14
2.3.3 Pugh-Shub stable ergodicity 15
2.3.4 Windows conjecture 16

3 Computation of Lyapunov exponents **19**
3.1 Ordinary differential equations background 19
3.2 Gram-Schmidt orthonormalization 20
3.3 Explicit Lyapunov characteristic exponent computation 20
3.3.1 The case for discrete time 21
3.3.2 Examples 22
3.3.2.1 A linear ordinary differential equation 22
3.3.2.2 The Hénon map 22
3.3.2.3 Delayed logistic map 23

4 Artificial Neural Networks **25**
4.1 Abstract Neural Networks . . . 25
4.1.1 General neural network formulation . . . 25
4.1.2 Approximation Capabilities . . . 26
4.1.3 Marriage with Takens theorem . . . 28
4.2 Our Neural Network Construction . . . 29

5 General dynamics **31**
5.1 Probability of Chaos . . . 31
5.2 Lyapunov spectrum diagnostics . . . 34
5.2.1 Dimensional Variation . . . 34
5.2.2 Entropy . . . 34
5.2.3 Dissipation . . . 36

6 Splitting up the Parameter Space **38**
6.1 Region I: fixed points to the first bifurcation . . . 38
6.2 Region II: the route to chaos - between the first bifurcation and chaos . . . 38
6.3 Region III: between the onset of chaos and the bifurcation chains . . . 41
6.4 Region IV: Bifurcation chain region . . . 41
6.5 Region V: Bifurcation chains to finite state dynamics . . . 41

7 Region I - the First Bifurcation **42**
7.1 Conjectures . . . 42
7.1.1 Random Matrix Theory . . . 42
7.1.1.1 The Circular Law . . . 42
7.1.1.2 Expected Value of Real Eigenvalues and Related Results . . . 43
7.1.2 A conjecture regarding the first bifurcation from a fixed point . . . 45
7.2 Numerical Results . . . 46
7.2.1 The First Bifurcation Qualitatively . . . 46
7.2.2 First Bifurcation Probabilities . . . 48
7.3 Future directions . . . 49

8 Region II - between the First Bifurcation and Chaos **52**
8.1 Conjectures . . . 52
8.2 Numerical Results . . . 54

9 Region IV - The Bifurcation Chain Region **62**
9.1 Definitions for numerical arguments . . . 62
9.2 Conjectures with respect to dynamics in region IV along a 1-d interval . . . 64
9.2.1 Conjectures with respect to a 1-d interval in parameter space . . . 65
9.3 Conjectures with respect to dynamics in region IV with respect to general parameter perturbation in $R^{N(d+2)+1}$ . . . 68
9.3.1 Comments regarding the windows conjecture . . . 68
9.3.1.1 The windows conjecture and the closing lemma. . . . 68
9.3.1.2 The windows conjecture and stable ergodicity. . . . 69
9.3.1.3 The windows conjecture and k-degree LCE stability. . . . 69
9.3.1.4 On the existence of spine loci . . . 69
9.3.2 Robust chaos in large dynamical systems . . . 70
9.4 Numerical preliminaries with respect to conjectures on a 1-d curve in parameter space . . . 72
9.4.1 *num*-continuity . . . 72
9.4.1.1 Qualitative analysis . . . 72
9.4.1.2 Quantitative and numerical analysis . . . 73

9.4.1.3 Relevance . . . 75
9.4.2 $a-$density of zero crossings . . . 76
9.4.2.1 Qualitative analysis . . . 76
9.4.2.2 Quantitative and numerical analysis . . . 77
9.4.2.3 Relevance . . . 78
9.4.3 Arbitrarily large number of positive exponents . . . 78
9.4.3.1 Qualitative analysis . . . 78
9.4.3.2 Quantitative analysis . . . 79
9.4.3.3 Relevance . . . 79
9.5 Numerical arguments with respect to conjectures on a 1-d curve in parameter space . . . 79
9.5.1 Decreasing window probability . . . 79
9.5.1.1 Qualitative analysis . . . 80
9.5.1.2 Quantitative and numerical analysis . . . 80
9.5.1.3 Relevance . . . 81
9.5.2 Hyperbolocity violation . . . 81
9.5.2.1 Qualitative analysis . . . 81
9.5.2.2 Quantitative and numerical analysis . . . 82
9.5.2.3 Relevance . . . 83
9.5.3 Hyperbolocity violation versus parameter variation . . . 83
9.5.3.1 Qualitative analysis . . . 84
9.5.3.2 Quantitative and numerical analysis . . . 84
9.5.3.3 Relevance . . . 84
9.6 Outline of Analytical Arguments for conjectures (14) and (12) on a $R^{N(d+2)+1}$ surface in parameter space. . . . 85
9.6.1 Jacobian matrix of our dynamical systems . . . 85
9.6.2 Linear algebra . . . 85
9.6.3 Discussion of parameter space . . . 86
9.6.4 Relationship to the a_i elements to stability and Lyapunov exponents . . . 86
9.6.5 Outline of a proof of no spines in neural networks . . . 87
9.6.6 Outline of a proof of robust chaos in neural networks . . . 88
9.7 Numerical arguments with respect to conjectures on a $R^{N(d+2)+1}$ surface in parameter space. 89
9.7.1 The experimental methodology . . . 89
9.7.1.1 Searching for periodic windows . . . 89
9.7.1.2 k-degree LCE stability . . . 89
9.7.2 Robust Chaos: Existence of periodic windows . . . 90
9.7.2.1 1-dimensional interval . . . 90
9.7.2.2 $N(d+1)$-ball . . . 90
9.7.3 Robust chaos of degree k . . . 93
9.7.3.1 1-dimensional interval . . . 93
9.7.3.2 $N(d+1)$-ball . . . 94
9.7.4 Integrating data from sections (9.7.2) and (9.7.3) . . . 98
9.7.5 Existence of Spines: Distribution of the a_k's . . . 99

10 Regions III and V **100**
10.1 Region III — Chaos to Bifurcation Chains . . . 100
10.2 Region V — Bifurcation Chains to Binary Dynamics . . . 101

11 Summary **103**
11.1 Summary of major points 103
11.2 Relationship between results from region IV and mathematical dynamics results 104
11.2.1 Windows conjecture 104
11.2.2 Structural stability 104
11.2.3 Closing lemma 105
11.2.4 Stable ergodicity 105
11.3 Lore 106
11.4 Future work 107

Bibliography **113**

A Acknowledgments **114**

List of Figures

3.1 The bifurcation diagram for the delayed logistic map. 24
3.2 The Lyapunov spectrum for the delayed logistic map over the region from a fixed point to chaos, including a Neimark-Sacker bifurcation at $a = 2$. 24

4.1 Schematic diagram of the Takens embedding theorem and how it applies to our construction. 29

5.1 Probability of chaos contour plot for $s = 8$ and various d and N. 32
5.2 Probability of chaos contour plot for $n = 8$ and various d and s. 33
5.3 Mean maximum Kaplan-Yorke dimension versus dimension, d. For the set of networks analyzed, $d_{KY} \sim 0.46d$. 34
5.4 Variation of Kaplan-Yorke dimension versus s for a single network with $N = 32$ and $d = 128$. 35
5.5 Mean maximum entropy versus dimension, d. 35
5.6 Variation in the entropy of a single network with s. The network considered has 32 neurons and 128 dimensions. 36
5.7 The mean maximum dissipation in neural networks versus log() dimension. The dissipation is proportional to $-12.8\log(d)$. 37
5.8 Variation in the dissipation of a single network with s. The network considered has 32 neurons and 128 dimensions. 37

6.1 Bifurcation diagram with the largest Lyapunov exponent for $N = 4$ and $d = 4$ with regions I, II, III, and V. 39
6.2 Bifurcation diagram with the largest Lyapunov exponent for $N = 64$ and $d = 4$ with regions I, II, III, a possible region IV, and V. 39
6.3 Bifurcation diagram with the largest Lyapunov exponent for $N = 4$ and $d = 64$ with regions I, II, III, IV, and V. 40
6.4 Bifurcation diagram with the largest Lyapunov exponent for $N = 64$ and $d = 4$ with regions I, II, IV, region V is not displayed. 40

7.1 Bifurcation diagram with the largest Lyapunov exponent for $N = 64$ and $d = 4$. 47
7.2 Bifurcation diagram with the largest Lyapunov exponent for $N = 4$ and $d = 4$. 47
7.3 Bifurcation diagram with the largest Lyapunov exponent for $N = 4$ and $d = 64$. 48
7.4 Bifurcation diagram with the largest Lyapunov exponent for $N = 64$ and $d = 64$. 48
7.5 Percent first bifurcation for $N = 16$, error bars represent the error in the probability. 49
7.6 Percent first bifurcation for $N = 4$, error bars represent the error in the probability. 50
7.7 Percent first bifurcation for $N = 256$, error bars represent the error in the probability. 50

8.1 Bifurcation diagram for a typical network; $n = 32$, $d = 64$. 54
8.2 The largest Lyapunov exponent for a typical network; $n = 32$, $d = 64$. 55
8.3 The Largest Lyapunov at the onset of chaos. 56

8.4 Phase plots on either side if the 2^{nd} bifurcation, $s = 0.0275$ and $s = 0.029$ respectively. The bifurcation occurs at $s \sim 0.02754$. . . . 56
8.5 Phase plots on either side if the 3^{rd} bifurcation, $s = 0.046$ and $s = 0.047$ respectively. The bifurcation occurs at $s \sim 0.04667$ 57
8.6 Phase plots on either side of the 4^{th} bifurcation, $s = 0.0512$ and $s = 0.0514$ respectively. The bifurcation occurs at $s \sim 0.05124$ 57
8.7 Phase space plot near the 6^{th} bifurcation — $s = 0.052$ and $s = 0.053$ respectively. The bifurcation occurs at $s \sim 0.05294$. 58
8.8 Lyapunov spectrum the (typical) network; $n = 32$, $d = 64$. 58

9.1 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. 65
9.2 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. 66
9.3 LE spectrum: 32 neurons, 4 dimensions 72
9.4 LE spectrum: 32 neurons, 64 dimensions. 73
9.5 *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. 74
9.6 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 over the number of networks considered. 74
9.7 k-scaling: $\log_2$ of dimension versus $\log_2$ of *num*-Lipschitz constant of the largest Lyapunov exponent. 75
9.8 Positive LE spectrum for typical individual networks with 32 neurons and 16 (left) and 64 (right) dimensions. 76
9.9 Number of positive LE's for typical individual networks with 32 neurons and 32 (left) and 128 (right) dimensions. 77
9.10 Mean distance (δs) between each of the first 10 zero crossings of LE's for many networks with 32 neurons and 16, 32, 64, and 128 dimensions. 78
9.11 Mean maximum number of positive LE's versus dimension, all networks have 32 neurons (slope is approximately $\frac{1}{4}$). 79
9.12 $\log_2$ 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 to the plotted data. 81
9.13 Mean fraction of LE's near zero (0 ± 0.01) for networks with 32 neurons and 32 or 64 dimensions (averaged over 100 networks). 82
9.14 Log probability of the existence of periodic windows versus log of dimension for 500 cases per d. Each case has all the weights perturbed on the order of 10^{-3}, 100 times per case. The line of best is $\sim \frac{1}{d^2}$. 91
9.15 Log fraction of networks with periodic windows versus log of dimension for 500 cases per d. Each case has all the weights perturbed on the order of 10^{-3}, 100 times per case. The line of best fit is $\sim \frac{1}{d}$. 92
9.16 Log probability of the existence of periodic windows in networks with windows versus log of dimension for 500 cases per d. Each case has all the weights perturbed on the order of 10^{-3} 100 times per case. The line of best fit is $\sim \frac{1}{d}$. 92
9.17 Number of positive LE's for typical individual networks with 128 dimensions. 94
9.18 The mean number of positive Lyapunov exponents versus dimension for networks with 32 neurons at $s = 3$. The line of best fit is $\sim 0.15d$. 95
9.19 Histogram of the number of positive Lyapunov exponents for $d = 8$. 95
9.20 Histogram of the number of positive Lyapunov exponents for $d = 64$. 96

9.21 The range between the maximum and minimum number of positive Lyapunov exponents versus the mean of the number of positive Lyapunov exponents per case for 100, 64-dimensional networks. . . . 96
9.22 The mean deviation from the mean number of positive exponents versus the mean of the number of positive Lyapunov exponents per case for 100, 64-dimensional networks. The error bars represent the standard deviation in the mean deviation. . . . 97
9.23 The histogram of the number of positive Lyapunov exponents for a typical 64-dimensional network perturbed 100 times. . . . 97
9.24 Histogram of the modulus of eigenvalues along the orbit for a typical network with 64 neurons and 64 dimensions; mean= 0.964, standard deviation= 0.1135 . . . 99

10.1 Bifurcation diagram with the largest Lyapunov exponent for $N = 2$ and $d = 2$ with $\beta_0 = 0$. . . 100
10.2 Close-up of the bifurcation diagram of region III of the example given in figure (10.1). . . . 101
10.3 Plot of the state diagram for the network given in figure (10.1) at $s = 12.2301$ . . . 102
10.4 Plot of the state diagram for the network given in figure (6.3) at $s = 1.10079$ . . . 102

Chapter 1

Introduction and background

1.1 Introduction

Since Poincare, the father of modern mathematical dynamical systems, much effort has been exerted to achieve a qualitative understanding of the physical world via a qualitative understanding of the functions we use to model the physical world. This work is a computational study with the same motivation.

The interaction between computational and mathematical worlds has waxed and waned over the last 40 years depending on the problems and emphasis of the respective fields. This work is intended to fit directly between the world of mathematical dynamics and computational dynamics. This framework is constructed such that there is enough common language such that results from both perspectives can be interpreted and hopefully, eventually integrated.

However, in this work, we limit ourselves to a qualitative numerical study of high-dimensional dynamical systems. We will have little to say about low-dimensional dynamical systems, and implicitly we claim that there are few generalizations that can be made about high-dimensional dynamics from low-dimensional dynamics.

To achieve the desired end, we perform the following steps:

1. we select a very general, concrete set of mappings that can approximate many relevant problems; this is done in chapter (4);

2. we develop a construction such that we can study such cases in a statistical manner, this is also done in chapter (4);

3. via a single parameter that has been hard-coded into the model equations, we stratify the parameter space by different qualitative dynamic types, this is discussed in chapter (6);

4. we study the various parameter regions attempting to link our results to the computational world while comparing with the mathematical world, this is contained in chapters (7), (8), and (9).

Although we will limit our study to high-dimensional dynamical systems, we do consider low-dimensional systems en-route to high-dimensional systems. The high-dimensional systems can be stratified into roughly four regions; in chapter (6) we will introduce five regions, but the fifth region appears only in low-dimensional systems. The four main regions we will investigate are comprised of: region I, the region between a fixed point and the first bifurcation; region II, the route to chaos region or the region between the first bifurcation and the onset of chaos; region IV, the chaotic region; and region V, a transition region between chaos and finite-state dynamics.

Often throughout this work we will refer to dynamic stability of a dynamical system. We will, where necessary, qualify this term precisely; however, it will be useful to discuss our meaning of these words here briefly. By dynamic stability, we mean stability of a particular dynamic type with respect to parameter, initial condition, or functional variation. Thus, a chaotic map can be dynamically stable if it is difficult to make the map non-chaotic by altering the initial conditions, parameters, or the functional form in non-drastic ways. This is contrasted to words used by others discussing chaos as a unstable type of dynamics and periodic orbits being a stable type of dynamics. In this work, we will not use "dynamical stability" or "dynamically stable" to differentiate between chaotic and periodic types of dynamics. A dynamically stable map is one that preserves a given type of dynamics upon perturbations.

The rest of the introduction will discuss background material for chapters (7), (8), and (9) as well as a rough comparison between many computational and mathematical constructions and mind sets.

1.2 Introduction to our construction

We will formally define the mappings considered in this thesis in Chapter (4) and specifically section (4.2), however, for ease of understanding, it is useful to briefly describe our construction as well as make a few clarifications.

All the maps considered in this thesis are of the form:

$$x_t = \beta_0 + \sum_{i=1}^{N} \beta_i \tanh\left(s\omega_{i0} + s\sum_{j=1}^{d} \omega_{ij} x_{t-j} \right) \tag{1.1}$$

where β is a parameter vector and ω is a parameter matrix, both have elements that are real and finite. The s parameter will adjust the amplitude of the argument of the hyperbolic tangent function thus we will stratify our mappings with the s parameter. These discrete-time maps are technically referred to as neural networks. Neural networks are an obvious choice of function space for our type of study because they are "universal approximators" as will be discussed in Chapter (4). Neural networks map compact sets in R^d to compact sets in R. We will only consider bounded mappings. Further, because we will use Lyapunov exponents as a primary tool of analysis, we will always refer to a network with a fixed set of initial conditions, unless otherwise specified, to avoid multiple attractor concerns. Lastly, we will consider all attractors unique if, for a given set of initial conditions, the Lyapunov exponents converge to a distinct and particular values given enough iterations.

$$x_t = \beta_0 + \sum_{i=1}^{N} \beta_i G\left(s\omega_{i0} + s\sum_{j=1}^{d} \omega_{ij} x_{t-j} \right) \tag{1.2}$$

1.3 Routes to Turbulence — Regions I and II

In their first edition of Fluid Mechanics (71), Landau and Lifschitz proposed a route to turbulence in fluid systems. Since then, much work, both in dynamical systems, experimental fluid dynamics, and many other fields has been done concerning the routes to turbulence. In this thesis, we present early results from the first large statistical study of the route to chaos in a very general class of high-dimensional, C^r, dynamical systems. Our results contain both some reassurances based on a wealth of previous results and some surprises. We conclude that, for high-dimensional discrete-time maps, the most probable route to chaos (in our general construction) from a fixed point is via at least one Neimark-Sacker bifurcation, followed by persistent zero Lyapunov exponents, and finally a bifurcation into chaos. We observe both the Ruelle-Takens scheme as well as persistent n-tori, where $n > 2$ before the onset of chaos.

To understand a routes to chaos type construction, begin with an ordinary differential equation in R^k with a single real parameter μ, $\frac{dv}{dt} = F(\mu, U)$ where F is as smooth as we wish and $U \subset R^k$ is compact. At μ_0 there exists a fixed point, and at μ_c, $\mu_0 < \mu_c$, F is chaotic. The bifurcation sequence proposed by Landau (71) and Hopf (57) is the following: as μ is increased from μ_0 there will exist a bifurcation cascade of quasi-periodic solutions existing on higher and higher dimensional tori until the onset of "turbulence."

In other words, the solutions would be of the following type, $x_{\mu_1}(t) = f(\omega_1, \omega_2)$, $x_{\mu_2}(t) = f(\omega_1, \omega_2, \omega_3)$, ..., $x_{\mu_{k-1}}(t) = f(\omega_1, \omega_2, \ldots, \omega_k)$ for $\mu_i < \mu_{i+1}$, and where none of the frequencies are rationally related. However, Landau and Hopf's notion of turbulence was high-dimensional, quasi-periodic flow. Ruelle and Takens (114) proposed both an alternative notion of turbulence (the strange attractor) and an alternative route to turbulence in a now famous paper. Ruelle and Takens claimed that the Landau path was highly unlikely from a topological prospective. The basis for their claim originates in the work of Peixoto (92) who has shown that quasi-periodic motion on T^2 (the 2-torus) is non-generic[1] in the set of C^r vector fields. However, Peixoto's theorem applies only to flows on T^2 and not T^k for $k > 2$. For diffeomorphisms of the circle, irrational rotations make up the full Lebesgue measure set of rotations. Suspensions of such diffeomorphisms correspond to quasi-periodic motion on T^2. It would seem that quasi-periodic motion of flows on T^2 would be high measure, however, there is not a one-to-one correspondence between flows on T^2 and discrete-time maps of the circle (e.g. the Reebs foliation or (89)). Further, quasi-periodic orbits of diffeomorphisms on the circle are structurally unstable. For flows on T^2, the structurally stable, hyperbolic periodic orbits are topologically generic; however, it is likely that quasi-periodic orbits are common in a measure theoretic sense on T^2. There remain many open questions regarding bifurcations of periodic orbits, for example, a list of codimension 2 bifurcations and their status can be found at the end of section (9.1), page 397 of (70). How this will all play out in practice is unclear and comprises a good portion of motivation for our study. Ruelle and Takens then went on to prove two results for flows relevant to this report. The first is a normal form theorem for the "second" Hopf bifurcation for vector fields, or the "first" Hopf bifurcation for maps (often referred to as the Neimark-Sacker bifurcation (115) (81)). This theorem gives a normal form analysis of the bifurcation of an invariant circle of a flow, but it does not state the type of dynamic that will exist upon the loss of stability of the invariant circle. The second relevant result was that, given a quasi-periodic solution $f(\omega_1, \ldots, \omega_k)$ on T^k, $k \geq 4$, in every C^{k-1} small neighborhood of $(\omega_1, \ldots, \omega_k)$, there exists an open set of vector fields with a strange attractor (prop. 9.2 (114)). These results were extended by Newhouse, Ruelle and Takens (85) who proved that a C^2 perturbation of a quasi-periodic flow on T^3 can produce strange (axiom A) attractors, thus reducing the dimension to 3 for tori with quasi-periodic solutions for which an open set of C^2 perturbations yield strange attractors. The basic scheme used by Takens, Ruelle and Newhouse was to first prove a normal form theorem from periodic orbits to 2−tori in vector fields and then prove something about how the 2-tori behave under perturbations — showing that bifurcations of m-tori to $(m+1)$-tori will yield chaos since 3-tori can be perturbed away to axiom A chaotic attractors.

If the story were only as simple as a disagreement between topological and measure theoretic viewpoints, we would be in good shape. However, out of the complexity of the dynamics and the difficulties posed by bifurcation theory regarding what happens to bifurcations of resonant periodic orbits and quasi-periodic orbits, the field of quasi-periodic bifurcation theory was born (24) (20). We will not attempt to discuss the history from where the above leaves off to the current state. For those interested, see (22), (29), (62), (37), (23), (63), and (61). The question regarding the most common route to turbulence is, in any but a very select set of specific examples, still an open and poorly defined question. Analytically piecing together even what types of bifurcations exist en route to chaos has been slow and difficult. We contribute to the existing partial solution in the following ways: we will provide a framework for numerical analysis that does not have a priori tori built in; we will provide evidence that the dominant scenario with respect to the cascade of bifurcations leading to chaos in many high-dimensional dynamical systems will consist of various bifurcations between periodic orbits of high period and quasi-periodic orbits. Moreover, flip and fold bifurcations (due to real eigenvalues) occur with less and less frequency as the dimension of the dynamical system is increased.

The work of Newhouse, Ruelle and Takens was followed by a significant amount of work by G. Iooss with various collaborators. The most relevant result is of Chenciner and Iooss (29) where bifurcations from flows with two frequencies to flows with three frequencies are shown to be non-generic. Iooss addresses some of the mathematics of the dynamics after a Neimark-Sacker bifurcation. However, we will refrain from an explanation here, and the interested reader is directed to (62) and (70).

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

There have of course been many fluid experiments, but since our work is related more to the theoretical work of the aforementioned researchers, we will refrain from a summary and instead encourage the reader to consider, as a starting point, (37).

1.4 The Dynamic Stability Dream — region IV

To present a full background with respect to the topics and motivations for our study would be out of place; we will instead discuss the roots of our problems and a few relevant highlights, leaving the reader with references to the survey papers of Burns et. al (65), Pugh and Shub (106) and Palis (91) for a more thorough introduction.

The origin of our work, as with all of dynamical systems, lies with Poincaré who split the study of dynamics in mathematics into two categories, conservative and dissipative systems; we will be concerned with the latter. We will refrain from beginning with Poincaré and instead begin in the $1960's$ with the pursuit of the "lost dream."

The "dream" was the conjecture that structurally stable dynamical systems would be dense among all dynamical systems. For mathematicians, the dream was motivated primarily by a desire to classify dynamical systems via their topological behavior. For physicists and other scientists, this dream was two-fold. First, since dynamical systems (via differential equations and discrete time maps) are usually used to model physical phenomena, a geometrical understanding of how these systems behave in general is, from an intuitive standpoint, very insightful. However, there is a more practical motivation for the stability dream. Most experimental scientists who work on highly nonlinear systems (e.g. plasma physics and fluid dynamics) are painfully aware of the dynamic stability. By dynamic stability we mean that upon normal or induced experimental perturbations, dynamic types are highly persistent. Experimentalists have been attempting to control and eliminate turbulence and chaos since they began performing experiments. The hope lies in that, if the geometric characteristics that allow chaos to persist can be understood, it might be easier to control or even eliminate those characteristics. The dream was "lost" in the late 1960's via many counterexamples ((30)), leaving room for a very rich theory. Conjectures regarding weaker forms of the dream for which a subset of "nice" diffeomorphisms would be dense were put forth; many lasted less than a day and none worked. The revival of the dream in the 1990's involved a different notion of "nice" — stable ergodicity.

Near the time of the demise of the "dream," the notion of structural stability together with Smale's notion of hyperbolicity were used to formulate the Stability Conjecture (the connection between structural stability and hyperbolicity - now a theorem) (90). The Stability Conjecture says that "a system is C^r stable if its limit set is hyperbolic and, moreover, stable and unstable manifolds meet transversally at all points." (91)

To attack the Stability conjecture, Smale had introduced axiom A. Dynamical systems that satisfy axiom A are strictly hyperbolic and have dense periodic points on the non-wandering set[2]. A further condition that was needed is the strong transversality condition - f satisfies the strong transversality condition when, on every $x \in M$, the stable and unstable manifolds W^s_x and W^u_x are transverse at x. That axiom A and strong transversality imply C^r structural stability was shown by Robbin (107) for $r \geq 2$ and Robinson (108) for $r = 1$. This settled one half of the stability conjecture. The other direction of the stability conjecture has proven much more elusive, yet in 1980 this was shown by Mañé (75) 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. Newhouse (83) 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 (74), (78), (122), (66). 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 the "dream" (via topological notions), yet they opened the door for a wonderful and diverse theory. Nevertheless, despite the fact that structural stability

[2] $\Omega(f) = \{x \in M | \forall \text{ neighborhood } U \text{ of } x, \exists n \geq 0 \text{ such that } f^n(U) \cap U \neq 0\}$

does not capture all we wish it to capture, 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 destabilize chaos and turbulence in high dimensional systems for several hundred years. 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 modeling nature, structural stability says that if one selects (fits) a model equation, small errors will be irrelevant since small C^r perturbations will yield topologically equivalent models. It is the topological equivalence that is too strong a characteristic for structural stability to apply to the broad range of systems to which we wish. Structural stability is difficult to use in a very practical way because it is very difficult to show (or disprove the existence of) topological (C^0) equivalence of a neighborhood of maps. Hyperbolicity can be much easier to handle numerically, yet it is not always common. Luckily, to quote Pugh and Shub (104), "a little hyperbolocity goes a long way in guaranteeing stably ergodic behavior." This thesis has driven the partial hyperbolicity branch of dynamical systems and is our claim as well. We will define precisely what we mean by partial hyperbolicity and will discuss relevant results *a la* stable ergodicity and partial hyperbolicity.

Our work will, in a practical, computational context, investigate the extent to which ergodic behavior and topological variation (versus parameter variation) behave given a "little bit" of hyperbolicity. Further, we will investigate one of the overall haunting questions: 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. Stated more generally, how does hyperbolicity (and thus structural stability) "behave" in a space of bounded C^r dynamical systems.

It is important to note a difference between our perspective and the perspective of mathematicians such as Palis. In Palis's paper (91), he notes that the chaotic nature of the Lorenz equations (74) is still an open and interesting question. While we do would not argue that the Lorenz equations have not been proven chaotic, nor that further study along these lines is interesting; from our perspective the Lorenz equations are chaotic. From an experimental perspective, using the "detectors" available to us (computer simulations), the Lorenz equations were shown to be chaotic over twenty years ago. We do not wish to incite an argument over ideologies, as we acknowledge that both are worthy of respect. Rather, we use the above example to demonstrate the difference in perspective between strict mathematics and computational or experimental physics.

1.5 Mathematics versus computation

There are two large themes in this work highlighting the difference in perspective between computational dynamical systems and mathematical dynamical systems:

> *The number of dimensions of the dynamical system matters; there is a stark difference between common dynamics in high and low-dimensional dynamical systems.*
>
> *Parameters matter, the practical effects of parameters with respect to dynamical systems is significant. Instead of selecting a manifold to impose dynamics on about which one can classify a generic type behavior, the existence of parameters asks the opposite question: begin with a dynamical system, vary the parameters, and observe the types of manifolds that are present, and study the types of dynamics that are predominant on those manifolds.*

The first notion, that dimensions matter, is somewhat obvious from the practical standpoint of the physical sciences. Highly complicated, high-dimensional physical systems usually have very complicated dynamics for most relevant parameter values. Often the trick with a good physical model is to strip away nearly all of the variables such that you have dynamics that are very simple, low dimensional, and yet retain many of the key stylized facts of the original system. For many physical systems, such is not possible, and highly complicated models with dynamics that are difficult to interpret form the heart of many numerical studies of physical systems. From the mathematics standpoint however, there are usually significant distinctions between only 1, 2 and 3-dimensional manifolds with respect to the dynamics they can support; little distinction is made for manifolds with $d \geq 3$. In fact, many high-dimensional dynamical systems from the mathematics prospective are multiple three-dimensional prototypical examples "glued" together in some interesting and insightful way. From the perspective of a physical scientist, there is a large difference between a 3 dimensional manifold and a 500-dimensional manifold. That fundamental, practical difference in the perspective on 3 and 500-dimensional manifolds is highlighted by the difference between fixing the manifold and imposing the dynamics, and fixing the dynamical system and varying the manifold. The former is a common mathematics mind-set, and the later is a common computational dynamical systems perspective. When one constructs a model to study a physical system, such as molasses in a mixer, one builds the model around the ideas of viscosity, temperature, the propeller/mixer geometry, etc. All of these parameters are hard coded in a fixed dynamical system. The parameters are then varied, and the dynamics are studied and observed. Relative to this dynamical system, the generic or common dynamics depend heavily on the portions of parameter space studied, e.g. if the viscosity is very high, the dynamics will likely be trivial, whereas moderately high viscosity dynamics are often turbulent. A natural course of study in this framework is to analyze how the dynamics change with parameters. This is analogous to changing the manifolds under consideration and studying a system on these variety of manifolds. Many results from, say, topological dynamics, are often difficult to interpret in this framework whereas results from bifurcation theory are directly applicable in such a framework. We intend to study, in a general way, dynamical systems by selecting a large set of fixed dynamical systems by varying their parameters. We often arrive at different, but complementary conclusions as those in mathematical dynamics.

Chapter 2

Notions of dynamic stability and previous results

2.1 Notions of dynamic stability and equivalence

There exist many notions of stable dynamical behavior in dynamical systems and differential equations, and the development and subsequent investigation has led to a very beautiful theory. We will begin discussing hyperbolicity and Lyapunov exponents, followed by topological notions. We will then discuss robustness of dynamical behavior, a notion central to our analysis. Last, we will briefly discuss ergodicity.

2.1.1 Hyperbolicity and Lyapunov exponents

Since strict hyperbolicity is a bit restrictive, we will begin with a standard definition of partial hyperbolicity:

Definition 1 (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:*

$$T_xM = E^u(x) \oplus E^c(x) \oplus E^s(x) \tag{2.1}$$

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[1]*. Thus there will exist numbers $0 < a < b < 1 < c < d$ such that, for all $x \in M$:*

$$v \in E^u(x) \Rightarrow d||v|| \leq ||D_xf(v)|| \tag{2.2}$$

$$v \in E^c(x) \Rightarrow b||v|| \leq ||D_xf(v)|| \leq c||v|| \tag{2.3}$$

$$v \in E^s(x) \Rightarrow ||D_xf(v)|| \leq a||v|| \tag{2.4}$$

More specific characteristics and definitions can be found in references (21), (56), (104), (65), and (27). The key provided by definition 1 is the allowance of center 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 (1), see (105)). Thus what is left is to again attempt to show the extent to which stable ergodicity persists, and topological variation is not pathological, under parameter variation with non-trivial center bundles present. Again, we note that results in this area will be discussed in a later section.

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

In numerical simulations we will never observe an orbit on the unstable, stable, or center manifolds. 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.

We will initially define Lyapunov exponents formally, followed by a more practical, computational definition.

Definition 2 (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 $v \in T_xM$ Lyapunov exponent at x is denoted:*

$$\chi(x, v) = \lim \sup_{t\to\infty} \frac{1}{t} log||Df^t v|| \tag{2.5}$$

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

$$\mathrm{Sp}\chi(x) = \{\chi_j^k(x) | 1 \leq i \leq m\} \tag{2.6}$$

(For more information regarding Lyapunov exponents and spectra see (15), (69), or (80).

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

$$\chi_j = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \ln(\langle (Df_k \cdot \delta x_j)^T, (Df_k \cdot \delta x_j)\rangle) \tag{2.7}$$

where $\langle , \rangle$ is the standard inner product, δx_j is the j^{th} component of the x variation[2] and Df_k is the "orthogonalized" Jacobian of f at the k^{th} iterate of $f(x)$. Through the course of our discussions we will dissect equation (2.7) further. It should also be noted that Lyapunov exponents have been shown to be independent of coordinate system and thus the specifics of the above definition do not affect the outcome of the exponents.

The existence of Lyapunov exponents is established by a multiplicative ergodic theorem (for a nice example, see theorem (1.6) (111)). There exist many such theorems for various circumstances. The first multiplicative ergodic theorem was proven by Oseledec (87); many others - (68), (112), (111), (97), (96), (98), and (21) - 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 a C^r ($r > 0$) map of a compact manifold M to itself and an $f-$invariant probability measure ρ, on M. The particular f-invariant measure we would use is not a trivial issue, however this measure would likely be a standard Sinai-Ruelle-Bowen (SRB) measure (131) (19). It would be useful to rigorously prove that our set of dynamical systems have SRB measures, we will refrain from discussing this topic here however. For specific treatments we leave the curious reader to study the aforementioned references beginning with (112), (21), and (27).

There is an intimate relationship between Lyapunov exponents and global stable and unstable manifolds. In fact, each Lyapunov exponent corresponds to a global manifold. 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 the existence theorems for these global manifolds, 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 (112).

Finally, for a nice, sophisticated introduction to the above topics see (69).

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

2.1.1.1 General conditions needed for the existence of Lyapunov exponents

Lyapunov exponents are one of our principal diagnostics, thus we must briefly justify their existence for our construction. We will begin with a standard construction for the existence and computation of Lyapunov exponents as defined by the theories of Katok (68), Ruelle (112), (111), Pesin (97), (96), (98), Brin and Pesin (21), and Burns, Dolgopyat and Pesin (27). We will then note how this applies to our construction. (For more practical approaches to the numerical calculation of Lyapunov spectra see (17), (18), (118), and (42).)

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 $\sigma-$algebra of sets, and ρ is a probability measure, $\rho(X) = 1$ (see (73) for more information). Now consider a C^r $(r > 1)$ map $f_t : X \mapsto X$ which preserves ρ (ρ is $f-$invariant) defined for $t \geq T_0 \geq 0$ 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

$$\Lambda = \{ \bigcap_{t>T_0} f_t(X) | f_t(\Lambda) \subseteq \Lambda \} \tag{2.8}$$

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 4.2) 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 (112) (similar, requirements are made by Brin and Pesin (21) in a slightly more general setting).

Now we must quickly justify our use of Lyapunov exponents. Clearly, we can take X in the above construction to be the R^d of section (4.1.3). As our neural networks map their domains to compact sets, and 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. All we need yet deal with is the measure preservation of which previously there is no mention. This issue is partially addressed in (33) in our neural network context. 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 (15). The specific measure theoretic properties of our networks (i.e. issues such as absolute continuity, uniform/non-uniform hyperbolicity, basin structures, etc) is a topic of current investigation.

2.1.1.2 SRB measures

As will be clear upon consideration of figures (5.7) and (5.8) of section (5.2.3), the dynamical systems we will consider are clearly dissipative. Further, upon considering figure (6.2) of chapter (6), the neural networks we are using also have at least mild multiple basin issues. Thus, for a rigorous proof of the existence of Lyapunov exponents for our dynamical systems we will likely need some type of SRB measure (131). We will not undergo an attempt to prove the existence of such a measure here as this is a very difficult task. Rather we will take the standpoint that if we can achieve convergence of a Lyapunov exponent for a particular set of initial conditions, we will consider the Lyapunov exponent as a relevant, existent quantity. Showing that there exist SRB measures for dynamical systems like ours is a source of future work.

2.1.2 Topological equivalence

Topological equivalence is an extremely common notion; thus we will be brief. Two C^r $(r > 0)$ discrete time maps f and g are topologically conjugate if there exists a homeomorphism h such that $f = h^{-1} \circ g \circ h$. The persistence of topological equivalence locally with respect to the Whitney topology is captured by the notion of structural stability.

Definition 3 (Structural Stability) *A C^r discrete time map, f, is structurally stable if there is a C^r neighborhood, V of f, such that any $g \in V$ is topologically conjugate to f.*

In other words, a map is structurally stable if, for all other maps g in a C^r neighborhood, there exists a homeomorphism that will map f to g, and the inverse respectively. This is a purely topological notion. That topological conjugacy is "robust" with respect to C^r perturbations is structural stability. That structural stability is "robust" is a topic we will discuss in a later section.

The theories of hyperbolicity, Lyapunov exponents and structural stability have had a long, wonderful, and tangled history (for good starting points see (121) or (9)).

2.1.3 Robust dynamics - k-degree LCE stability

Another notion of dynamic stability is that of dynamic robustness with respect to parameter variation. There exist various definitions and categorizations of dynamics and several interpretations of robustness.

A property X of some operation or morphism (e.g. a mapping), Y is robust if X is a persistent property in an open neighborhood of Y. Thus, whenever one uses the term robust, the following notions must be made precise: the property, the mapping, and the open set (and thus a topology which we will take to be the stand topology on R^n unless otherwise noted) upon which the property is persistent. Every definition of robustness we formulate will be made along the above lines. Robustness can be differentiated from structural stability, hyperbolicity, and stable ergodicity in that the three former notions are properties which can be robust, robustness is not itself a property - rather a property can be robust.

Definition 4 (degree-k Lyapunov exponent equivalent) *Assume two discrete time maps f and g as in section (4.2) from compact sets to themselves. The mappings f and g are called k Lyapunov equivalent if f and g have the same number of Lyapunov exponents and if f and g have the same number of positive Lyapunov exponents.*

The definition could be used where f and g differ by some C^r perturbation, however, since we are considering explicit parameter perturbations, we will think of f and g as the same explicit mapping with different parameter values.

Definition 5 (Robust chaos of degree k) *Assume a discrete time map f as in section (4.2) mapping a compact set to itself. The map f has robust chaos of degree k if there exists a p-dimensional subset of parameter space $U \in R^p$, $p \in N$, such that, for all $\xi \in U$, and a given set of initial conditions of the map $f|_U$, f retains k positive Lyapunov exponents. In other words, f is k-Lyapunov equivalent on the subset U.*

We will refer to such a map as having robust chaos of degree k or having k-degree LCE stability. If $k = 1$, for f at $U \in R^p$, we will refer to f as having robust chaos. This is a very weak condition for dynamic stability; it will be fruitful to discuss circumstances that fit under the umbrella of robust chaos of degree k. The first characteristic worth noticing is that since this notion of stability is based solely on the preservation of k of the positive Lyapunov exponents, the Lyapunov exponents need not be the same magnitude for any $\xi \in U$ or for any initial conditions in the domain. Further, the actual number of positive exponents need not be preserved, only k positive exponents. This allows for the possibility of a complicated basin containing many different attractors (e.g. Milnor attractors (78)). In fact, in a practical sense, the Lyapunov exponents need not even converge very well; for instance, the largest one could oscillate between 1 and 1.3 over time, as long as it is convergent in some weak sense,[1] we can still apply this notion — as long as some meaning can be drawn from a Lyapunov spectrum. Milnor attractor problems can be quite severe, and as long as all of the various attractors in our basin have k positive Lyapunov exponents, we will retain robust chaos of degree k. It almost goes without mentioning that structural stability can be freely violated and robust chaos of degree k can remain. In a strict sense, whether stable ergodicity can be violated is not clear; it is likely however that there will exist many attractors that are robustly chaotic of degree k and not stably ergodic.

Definition 6 (Robust chaotic attractor) *Assume a discrete-time map as in section (4.2) with a compact, invariant set Λ where $f(\Lambda) \subset \Lambda$ for which there exists a positive Lyapunov exponent. Λ is a robust*

[1] By weak sense, we mean a very slow variation in time, however we will leave this notion vague since it will not be formally addressed in this thesis.

chaotic attractor if there exists a $p-$dimensional subset $U \subset R^p$ (again $p \in N$), upon which Λ persists with respect to all $\xi \in U$, and a given set of initial conditions in the domain of f.

With respect to definition (5), definition (6) is considerably more restrictive. If the attractor is unique, the Lyapunov exponents must agree on U, there cannot exist multiple attractors for a various initial conditions, there just exists a single invariant measure, almost surely f will be stably ergodic (however a positive Lyapunov exponent does not imply stable ergodicity), and it would seem that the only manner in which structural stability could be broken is via a C^r perturbation that is not a perturbation of the parameters. Definition (6) likely excludes cases with zero Lyapunov exponents whereas definition (5) does not. Further, it is considerably more difficult to demonstrate the existence of a robust chaotic attractor than it is to demonstrate k-degree LCE stability.

Proving relations between these definitions and those of stable ergodicity and the like are goals of future work.

2.1.4 Ergodicity

Stable ergodicity has been the subject of a wonderful and exciting area of mathematics over the past several years. We will, in a later section, state one of the keynote stable ergodicity results; for a more full understanding to the interested reader see (65), (27), (105), or (106). We will begin with a definition of ergodic and follow this with a definition of a stably ergodic dynamical system.

Definition 7 (Ergodic dynamical system) *A dynamical system is ergodic if it preserves a measure and each measurable invariant set is a zero set or the complement of a zero set.*

We will refrain from more precise definitions and instead note a few standard characteristics. Assume $g : M \to R$ where M is some manifold and g is continuous, if g is ergodic, the spatial and time averages are the same. Ergodic behavior implies that, upon breaking the attractor into measurable sets, $V_i \neq 0$, for g applied to each measurable set for enough time, $g^n(V_i)$ will intersect every other measurable set, V_j. This implies a weak sense of recurrence; for instance, quasi-periodic orbits, chaotic orbits, and some random processes, are at least colloquially ergodic. For an ergodic dynamical system, most orbits visit all parts of the phase space. Ergodicity is a precise mathematical notion meant to capture Boltzmann's ergodic hypothesis. More insight along with a more thorough presentation can be found in standard texts (99), (119), or (127). Let us now quickly define stable ergodicity:

Definition 8 (Stable ergodicity (106)) *Let M be a compact differentiable manifold without boundary and let m be a smooth probability measure on M. Define $f \in Diff^2_m(M)$ as an element of C^2 diffeomorphisms of M which are measure preserving with respect to m. f is said to be stably ergodic if there exists a neighborhood U of f in $Diff^2_m(M)$ consisting entirely of $m-$ergodic diffeomorphisms.*

There exist many dynamical systems that are stably ergodic, but not structurally stable. Likewise, there exist many dynamical systems that are not stably ergodic, but will be robustly chaotic of degree k since the existence of a positive Lyapunov exponent does not imply ergodicity. In other words, of the three notions of equivalence we have presented, topological equivalence, $k-$degree LCE stability, and ergodicity; although each implies different conclusions, robust chaos of degree k may be the weakest notion (although clearly not in all circumstances).

2.2 Functional Versus Parameter Perturbation

As the main point of this report addresses various types of dynamic stability of dynamical systems versus perturbation, it is imperative that we be precise and clear by what we mean by a perturbation. Further, as we are attempting to interpret our results in the context of other results, we must spend a brief time discussing various notions of perturbations.

2.2.1 C^r perturbations

In most mathematical dynamics papers, the notion of perturbation of a function that is used is the notion of a C^r perturbation in the C^k Whitney topology. We will not put forth a precise definition of the Whitney topology here (for those curious, see (55) or (50)), rather we will present a heuristic notion of the Whitney topology. We will follow this with an explanation of C^r perturbations.

In the C^r topology, two functions f and g are C^r close if their partial derivatives up to and including order r are close. Locally, one can picture this as the coefficients of the Taylor series (the $r-jets$) being close up to the r^{th} order. Globally, one considers the graph of such functions. This business is nicely depicted in (34) Fig. 9. page 55 — we will not reproduce those figures here.

When one thinks of perturbations in the Whitney topology, one must begin by thinking of perturbing the graphs of functions and their derivatives. If, for a C^k function f, a property is said to persist under C^r ($k \geq r$) perturbations, this means that the property holds for any variation in the graph of the function as long as that change does not change the derivative or the function of order not equal to r. If a property is said to be persistent to C^r perturbations for $0 \leq r \leq 10$, then small perturbations of the function of order 0 to 10 will not affect the existence of the property. Moreover, a property can be persistent under C^r perturbations for $r = 5$, but the same property can be easily destroyed for $r \neq 5$. When one thinks of the perturbation of the r^{th} order, one is allowed to "move" the graph of the r^{th} derivative of the function a little bit, in all possible ways.

2.2.2 Parameter perturbations

In the world where researchers are working with explicit functions, dynamical systems, etc, the actual equations have explicit parameters which can be varied. In a practical sense it is this scenario, where the parameters are concrete numbers or vectors which can be changed, that is often more relevant than general, C^r perturbations.

Parameter perturbations are almost always considerably different than general C^r perturbations, for when speaking of all C^r perturbations, one is not thinking of any parameter change in a practical sense (although such a perturbation could be a parameter change). Rather, one is thinking of altering the graph of the function. Without an infinite number of parameters, one cannot, for a particular dynamical system, realize every possible C^r perturbation (even if r is restricted to one). This fundamental difference is not particularly well understood in many systems — many abstract dynamical systems pathologies may not be able to be realized in a practical system as parameter variation can't bring about such pathologies. Likewise, it is possible that certain pathologies, which can be destroyed by many C^0 perturbations, may not be able to be eliminated in a particular dynamical system because the parameters are restrictive enough that the C^0 perturbations that will destroy the pathologies are not possible. Furthermore, C^r perturbations can be, and are often thought of, as local perturbations while parameter perturbations are often global. For example, changing the viscosity of a fluid affects the entire fluid system and can be quite different from perturbing a vector field on a manifold at a point like one might do in the construction of the DA map (129).

Thus it makes a lot of sense to study parameter perturbations. However, studying parameter variation in the general qualitative theory of dynamical systems is often lacking useful descriptive language, is too restrictive, or does not make any sense in the important contexts. For instance, when studying the structural stability of solutions of ordinary differential equations, one often needs the approximations to be dynamically similar with respect to all possible perturbations for the approximation theorems to work. In the practical world of solving a differential equation on a computer, however, this could be much less relevant as one is usually toying with, say, the coefficients of a polynomial solution. In this report we review several conjectures and stability theorems regarding both parameter and general C^r perturbations. Moreover, we put forth several conjectures and subsequent numerical arguments regarding parameter perturbation. For a correct interpretation of our work and how it fits with that of others, the difference between parameter perturbations and general C^r perturbations must be well understood.

2.2.3 Parameter perturbations in neural networks

Neural networks provide a unique format for parameter perturbations. First, the parameter space of neural networks is simply $R^{n(d+2)+1}$, and there is no a priori meaning to any one of the parameters. Of course, this is both an advantage and a disadvantage when modeling an actual physical system. Upon having the model equation, one cannot tweak a parameter for, say the viscosity, or some other relevant physical parameter. However, because no one parameter controls the viscosity, the idea of parameter perturbation in neural networks lies someplace between perturbing a "physical" parameter and the abstract C^r perturbations of mathematical dynamics.

For our networks, in the limit where $n \to \infty$, we are allowed the full power of being able to explore all C^r perturbations. However, as we will always be working well below that limit, we cannot assume that any of our conjectures can speak precisely to those of abstract dynamical systems theory. Since we are more interested in the bridge between the abstract theory and practical matters of neural networks as approximation functions and as representatives of physical/social/biological models in general, our inability to perform all possible C^r perturbations is of little concern.

2.3 Previous results

Since the background for this thesis is diverse, sometimes obscure, and from a variety of fields, and we are attempting to reach a diverse readership, we will briefly discuss results of others that are necessary to understand and interpret our work. We will begin with some non-density of structural stability results of Smale, Pugh and Peixoto, and a density of structural stability result of Shub. We will then move on to the closing lemma of Pugh and Robinson, followed by a discussion of the Pugh-Shub stable ergodicity theorem. This list is far from complete, but it corresponds well to our work. We will finish with a discussion of the windows conjecture of Barretto et. al. Finally, we will make some statements and discuss other conjectures relevant to the results discussed below. However, as they will be intertwined with results and conjectures of our own, we will refrain from discussing these results until a later section.

2.3.1 Non-density of structural stability

The original dynamic stability dream of Poincaré was that some sort of dynamically[5] stable dynamical systems were dense in the set of all dynamical systems. The notion of structural stability was created in an attempt to provide an answer to the stability dream. The dream was crushed in the mid $1960's$. We will discuss four results, the first two are non-density of structural stability, the third being a positive result highlighting the difference between stability in the C^0 versus C^r, $r > 0$ topologies, and the final result is the theorem representing the structural stability conjecture of Palis and Smale.

Theorem 1 (Structurally stable systems are not dense (120)) *There exists a compact* 4 *dimensional manifold* M*, an open set* U *in the space of* C^r *vector fields,* C^r *topology* $r > 0$*, on* M *such that no vector field* $X \in U$ *is structurally stable*

This famous theorem of Smale can be easily generalized to dimensions greater then 4. Pugh and Peixoto provided this generalization with the following theorem:

Theorem 2 (Structurally stable systems on open manifolds are never dense (93)) *On any open,* $n-$*dimensional differentiable manifold* $M = M^n$*,* $n \geq 2$*, the set of structurally stable vector fields is not dense in* $\mathcal{X} = \mathcal{X}(M)$*.*

Note that $\mathcal{X}$ is the space of vector fields with the C^r Whitney topology for $r > 0$.

Despite the fact that the examples constructed for the above theorems that cannot be perturbed away with an arbitrary small perturbations, how frequently these examples exist in practice is a question that is yet open and difficult to formulate.

[5]The exact definition of "dynamically stable" is flexible.

However, which topology one is considering can make a fundamental difference in the results one can prove — for the above results do not hold with respect to the C^0 topology.

Theorem 3 (Structurally stable diffeomorphisms are dense) *Let M be a C^∞ compact, m-dimensional manifold without boundary. Next, let $1 \leq r \leq \infty$. Then the set of structurally stable diffeomorphisms are dense in $Diff^r(M)$ with the C^0 topology.*

This result in a sense highlights a difference, which might be thought of as irrelevant to some, between what can be said about diffeomorphisms with respect to the C^0 and C^r, $r > 0$ topologies. When defining open sets with respect to only the continuity, structural stability is relatively common. However, the C^1 topology makes much more sense if we wish to compare local expansion rates. Likewise, the C^∞ topology makes more sense if one wants to compare all the orders of the derivatives. In other words, what topology one picks depends upon what one wants to compare. With respect to numerics, this is not always a clear-cut choice, but comparison of numerical results with these rigorous analytical results is important. With respect to our dynamical systems, we are clearly concerned with the C^r topology for $r > 0$ (more realistically $1 \ll r$), but $r \ll \infty$.

A next question might be, what properties are necessary for a dynamical system to be structurally stable. This question has a rather simple answer.

Theorem 4 (Mañé (75), Robbin (107), Robinson (109)) *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 strictly hyperbolic with dense periodic points on its non-wandering set Ω ($p \in \Omega$ is non-wandering if for any neighborhood U of x, there is an $n > 0$ such that $f^n(U) \cap U \neq 0$). The diffeomorphsim f satisfies the strong transversality condition if every x in its stable and unstable manifolds intersect transversally at x, i.e. if $T_xW^s(x) + T_xW^u(x) = T_xM$ for every $x \in M$. This theorem was originally conjectured by Palis and Smale (90) and proven in various stages by Robbin (107), Robinson (109), and Mañé (75). It is likely that many relevant subsets of the C^r dynamical systems are structurally stable.

2.3.2 The closing lemma

The closing lemma, or the idea that any map from a bounded (non-trivial) set to itself can be perturbed via a conjugation to a different, but arbitrarily close point in the domain that corresponds to a stable periodic orbit was conjectured by Thom and was originally thought to be somewhat obvious. The strongest result along these lines was proven by Pugh and was of considerable difficulty. We will state a friendly and slightly less precise version of Pugh's lemma, followed by some slight clarifications to aid in the interpretation.

Theorem 5 (Closing Lemma (101) (103)) *Suppose that $X \in \mathcal{X}^{j,k}$ has a nontrivial recurrent trajectory through $p \in M^n$, and suppose that $\mathcal{U}$ is a neighborhood of X in $\mathcal{X}^{j,k}$. Then there exists $Y \in \mathcal{U}$ such that Y has a closed orbit through p.*

Since we have refrained from stating Pugh's precise lemma, a few comments are in order. First, by nontrivial recurrent trajectory, we mean a non-recurrent trajectory of p where p is a non-wandering point which is also not a equilibrium point. Given that $\mathcal{X}^{j,k}$ is the set of C^k tangent vector fields on M under the C^j topology with $1 \leq j \leq k \leq \infty$, $j < \infty$, Pugh's proof of the closing lemma applies to the case where $j = 1$ and $1 \leq k \leq \infty$ on manifolds with dimension greater than 2. The proof is known to fail for the C^j topology for $j \geq 2$. This doesn't mean it isn't true for $j \geq 2$, but no one has been able to show it. The current consensus is that the above theorem is not true for $j \geq 2$ (102).

For a better understanding of the statement of the closing lemma, let us contrast it with the structural stability theorem from the previous section. The structural stability theorem says that, given a diffeomorphism f on a compact manifold, if f satisfies the strong transversaility condition and axiom A on the non-wandering set, there will not exist a C^r, $r > 0$ perturbation on a small, C^1 neighborhood of f such that

f will undergo a topological change. Axiom A requires that periodic points be dense on the non-wandering set, thus with respect to the closing lemma, there will exist a diffeomorphism, h, where $g = h^{-1}fh$, such that $h(p) = q$ where q is a periodic point for g and p is a non-trivial recurrent point of f. There might only be one such q. Further, the periodic orbit at p is not necessarily a stable periodic orbit. Such perturbations obtained or supplied by the closing lemma are likely rare.

For more information, see (110), (101), (103), and (79).

2.3.3 Pugh-Shub stable ergodicity

Currently the most general version of a dynamic stability theorem is that of the Pugh-Shub theorem. In this theorem, they abandon the topological notions of dynamic equivalence (e.g. various degrees of conjugacy) for the notion of ergodicity. A dynamical system f is said to be stably ergodic with respect to a probability measure μ if, given $f \in \mathit{Diff}^2_\mu(M)$ (M compact), there is a neighborhood, $f \in Y \subset \mathit{Diff}^2_\mu(M)$ such that every $g \in Y$ is ergodic with respect to μ.

Pugh and Shub (105) put forth the following conjecture regarding partial hyperbolicity and stable ergodicity:

Conjecture 1 (Pugh and Shub (105) Conjecture 3) *Let $f \in \mathit{Diff}^2_\mu(M)$ where M is compact. If f is partially hyperbolic and essentially accessible, then f is ergodic.*

In that same paper they also proved the strongest result that had been shown to date regarding their conjecture:

Theorem 6 (Pugh-Shub theorem (theorem *A* **(105)))** *If $f \in \mathit{Diff}^2_\mu(M)$ is a center bunched, partially hyperbolic, dynamically coherent diffeomorphism with the essential accessibility property, then f is ergodic.*

A diffeomorphism is partially hyperbolic if it satisfies the conditions of definition (1) and ergodic behavior as discussed in section (2.1.4). 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. The diffeomorphism f is center bunched if the spectra of Df corresponding to the stable ($D^s f$), unstable ($D^u f$), and ($D^c f$) central directions lie in thin, well separated annuli (see (104), page 131 for more detail; the radii of the annuli is technical and is determined by the Holder continuity of the diffeomorphism.) We will refrain from divulging an explanation of dynamical coherence see (106) for more details.

This result of Pugh, Shub, and many others is highly relevant to our current work. First, the Pugh-Shub theorem applies to any C^2 dynamical system, including the very high-dimensional ones we are concerned with in this thesis. Second, for diffeomorphisms, periodic orbits are not ergodic, hence if a dynamical system is stably ergodic, there will not exist periodic windows for any (sufficiently small) perturbation. We would like to be able to claim that our dynamical systems are stably ergodic; this, however, has proven to not be very computationally feasible. We instead have to settle for a notion of dynamic stability based on the existence of positive Lyapunov exponents — a more coarse version of dynamic stability[6]. In this thesis, one claim we make is that periodic windows do not exist or are very rare among high-dimensional chaotic dynamical systems. If we could show that our systems are stably ergodic, we would know for sure that windows don't exist. Currently, it is not known "how many" dynamical systems satisfy the conditions for stably ergodicity in general — it is known in some cases, however (e.g. (28)). Numerical evidence of a lack of periodic windows amongst a general set of high-dimensional chaotic (and possibly ergodic) dynamical systems reinforces the belief that stable ergodicity will indeed be at least prevalent, and likely dominant amongst high-dimensional chaotic dynamical systems. In other words, this thesis, and the work of stable ergodicity in the mathematics dynamics community is complementary and in agreement.

[6]It is worth noting that positive Lyapunov exponents do not imply the existence of stable ergodicity, specifically, they do not imply accessibility.

2.3.4 Windows conjecture

This discussion of the windows conjecture follows the work of Barreto, Hunt, Grebogi and Yorke (16). Note that we will make very slight modifications to their original construction; the comments are ours. Let us begin with a few definitions.

Definition 9 (Windows) *Assume an $f \in C^r$ such that $f : R^p \times R^n \to R^n$ where R^p is the parameter space ($p \in N$), and R^n is the state space ($n \in N$). A window is a subset $V_w \in R^p$ such that f has Lyapunov exponents that are strictly less than zero for all $a \in V_w$ for all initial conditions.*

The subset V_w could be quite diverse, for instance, it could be connected or disjoint, it could span R^p or it could span R^m where $m < p$. A convenient requirement might be that $m(V \cap V_w)$ occupies full measure (m is the standard Lebesgue measure) for $V \subset V_w$. It is important to note that this definition allows for windows with different periods to exist in V_w.

We will now define the condition on which the windows conjecture is built:

Definition 10 (Spine Locus) *A spine locus (or just a spine) for a window $V_w \subset R^p$ consists of the set of points in parameter space such that the matrix of partial derivatives - i.e. the Jacobian, is nilpotent (i.e. there exists a super-stable equilibrium point).*

The spine locus consists of the super stable orbits and is located at the center of the windows. There are two important issues to glean from this concept. The first issue is that the geometry of the spine determines, to some extent, the geometry of the window of periodic orbits (for a two-dimensional example see (16)). This brings up the issue of the codimension of a spine. Let us define codimension for a manifold first. Given a manifold Z and a submanifold $X \subset Z$, the codimension of X in Z is defined as $codim(X) = dim(Z) - dim(X)$. The codimension of the submanifold X is really a measure of how much of Z cannot be spanned by X, thus giving a measure of the avoid-ability of X while moving around in Z. The codimension of a spine[7] is the number of independent conditions defining the spine. For instance, if we have the parameter space R^3, a point is a codimension three spine, a line is a codimension two spine, a surface is a codimension one spine, and a volume is a codimension zero spine. The dimension of the spine is the dimension of the window, thus as the codimension of the spine increases, the portion of the space occupied by the window decreases. From this it is clear that the number of parameters needed to avoid a transversal intersection with the spine is p (the dimension of the spine). What may not be clear is how we define parameters in practice. In a sense, the important circumstances are those for which there exist codimension zero or codimension one spines (however, we will also consider codimension p spines in this thesis). In these cases, non-null transversal intersections of a parameterized curve and the window might not only be common, they might be unavoidable. (For rationale along these lines see Wiggins' (128) discussion of bifurcations (section 1.5), or for a more technical version see Sotomayor (123).) The above arguments imply the other important issue which is of course that near spines, windows will be prevalent (see (16) for more information).

Let us define the conditions on the windows needed for the windows conjecture:

Definition 11 (Limited and extended windows) *A window is limited if $dim(V_w) = 0$. A window is extended if $dim(V_w) > 0$.*

The right way to think of this concept is the codimension of the window relative to the parameter space, which, as discussed above, is dominated by the geometry of the spine. A 2-dimensional window (and hence a spine that spans R^2) in a 560,000-dimensional parameter space, i.e. a codimension 559,998 spine, is likely of little concern. However, a 2-dimensional window in a 3-dimensional parameter space is likely to be observable.

Let us now define the condition that will later be claimed to be common near spines of high codimension.

Definition 12 (Dispelled attractor) *Given a chaotic attractor Λ of f at $a \in R^p$, Λ is said to be dispelled at $a \in R^p$ if there exists an $\tilde{a}$ such that $f_{\tilde{a}}$ is C^1 close to f_a, and almost all (with respect to Lebesgue measure) points in a neighborhood of Λ belong to basins of attracting periodic points of $f_{\tilde{a}}$.*

[7]Note, a spine might not be a submanifold in general, thus we cannot assume the rigorous definition of the codimension of a submanifold. It is more likely that a spine will be an algebraic variety.

Note that all perturbations in a neighborhood of a is quite different from all C^r ($r \geq 0$) perturbations at a of which there are uncountably more (recall the discussion from section (2.2)). In some ways, this is a somewhat difficult definition to handle as there are subtleties that can be encountered. For instance, as was seen in the real quadratic family, ((51), (64), (77)), for $a \in (0, 4)$, attracting periodic points are dense and occupy a set of positive measure, but there are chaotic attractors that exist on neighborhoods $U \subset (0, 4)$ such that $0 < m(U|\lambda < 1) < 1$ and $0 < m(U|\lambda > 1) < 1$. In fact, there are chaotic attractors such that for a neighborhood U, $m(U|\lambda > 1) = 1$. Moreover, Anosov diffeormophisms have dense periodic points on the non-wandering set, but these periodic points are measure zero. The idea of a dispelled attractor may not be as black and white as being dispelled on full measure or not.

We will now formulate a definition of fragile behavior:

Definition 13 (Fragile) *If there exist (possibly rare) parameter values of f_a arbitrarily close to a for which the attractor Λ is dispelled, we say that Λ is fragile.*

It would be nice to generalize this notion of closeness to closeness in function space (C^1 close in the Whitney topology for instance) in a meaningful way. Let us define the set upon which attractors are dispelled:

Definition 14 (Window set) *We define the window set W to be the set of $a \in R^p$ such that Λ is dispelled; i.e. $W = \bigcup V_w$.*

We are now prepared to state the windows conjecture of Barreto et. al (16):

Conjecture 2 (Windows Conjecture) *Given $f \in C^1$ mapping of a compact set $S \in R^n$ to itself that exhibits a fragile chaotic attractor Λ with $k \geq 1$ positive Lyapunov exponents where all invariant measures on Λ yield the same number of positive Lyapunov exponents. Let W be the window set of a typical (generic in the standard topology) family f_ξ, where $\xi \in R^p$ and $f_0 = f$.*

(1) *If $p < k$, there exists a neighborhood of $\xi = 0$ entirely outside of W.*

(2) *If $p = k$, W is dense in a neighborhood of $\xi = 0$ and the components of W are limited (as per definition (11)).*

(3) *If $p > k$, W is dense in a neighborhood of $\xi = 0$ and the components of W are extended (as per definition (11)).*

The point being that if f has a fragile chaotic attractor and there are more parameters than positive Lyapunov exponents, then we will be able to move one of those parameters in a very small way such that the chaotic attractor is dispelled. It is important to notice that there is a difference between a fragile chaotic attractor and the chaotic attractor actually being robust. A chaotic attractor can be non-robust, or even structurally unstable, and yet not be dispelled. All the attractor need be is non-unique in the neighborhood of a yet all (full measure) the attractors in the neighborhood of a are likewise chaotic for all initial conditions. This conjecture speaks to what type, and how severe a *parameter* perturbation is required for the chaotic attractor to be destroyed.

A useful comparison is made by noting that in the stability conjecture (now a theorem) of Palis and Smale, their notion of perturbation was any C^1 perturbation and the dynamical change had to be a C^0 change — a chaotic attractor bifurcating to a periodic orbit is definitely a C^0 change. There exist many C^1 topological changes of many attractors that will preserve the chaotic dynamics — simply alter the number of positive Lyapunov exponents while leaving the largest Lyapunov exponent positive. One crucial practical issue that is abstracted away by mathematical constructs (such as the stability conjecture) is the specific definition of a parameter. The windows conjecture specifically addresses parameter variation, thus, to address this conjecture, we must, in a very practical way, define what we mean by parameters and what parameter variations we consider. One can imagine difficulties that might arise in determining the dependencies of parameters on one another in dynamical systems with complicated and interdependent parameters (e.g. our construction as given in sections (1.2) and (4.2)).

There are four questions that we will address related to this conjecture in the sections that follow: we will address the probability of the observed existence of fragile attractors; we will address the existence and probability of existence of spine loci, a chief condition around which the conjecture is based; we will briefly address the necessity, sufficiency, and relevance of spines to periodic windows; and we will address the relationship between the windows conjecture and both the stability conjecture and the Pugh-Shub stable ergodicity theorem.

Chapter 3

Computation of Lyapunov exponents

In section(2.1.1) we introduced Lyapunov exponents and the Lyapunov spectra in an abstract framework (for more information along these lines, see (15)). Now we will discuss, in a very practical way, how to compute the Lyapunov spectra. We will begin with a discussion of the Gram-Schmidt orthonormalization technique. We will then briefly introduce some background theory of ordinary differential equations that will make understanding the discrete-time computation easier. A specific description of the Lyapunov spectrum calculation will follow. We will finish with two standard examples applying the outlined algorithm.

3.1 Ordinary differential equations background

Before we proceed, we must introduce the basic concepts of ordinary differential equations necessary for our discussion. Our presentation will not be exhaustive or abstract and it will not always be in precise mathematical language (for such a treatment, see (15)). Rather, we will formulate the above concepts in the most explicit and practical manner possible to aid in understanding the computation of Lyapunov exponents. A more complete and precise formulation can be found in many standard ordinary differential equations texts (see (94), (82), (11)).

We will begin with the linear, homogeneous case. Let A be an real $n \times n$ matrix and x_t as a time dependent n-dimensional vector. Define the initial value problem as:

$$\dot{x_t} = f(x_t) = Ax_t \tag{3.1}$$

where the solution for the autonomous (time-independent) case with initial value at a given initial condition $x_0 \in R^n$ is unique and is given by $x_t = e^{At}x_0$ where $e^{At} = I + \frac{At}{1!} + \frac{A^2t^2}{2!} + \ldots$ is the exponential of a matrix. We approximate the general non-autonomous case as a finite product whose terms are local autonomous approximations of the non-autonomous equation over short time intervals. Such would be achieved using the following equation

$$\delta\dot{x} = Df(x_t)\delta x \tag{3.2}$$

over small variations in $\dot{x}$, $\delta\dot{x}$.

Let Φ_t be a real, $n \times n$ matrix that depends upon time. We call Φ_t the fundamental matrix solution of f if it satisfies:

$$\frac{d\Phi}{dt} = A\Phi_t = Df(x_t)\Phi_t \tag{3.3}$$

Equation (3.3) is often referred to as the variational differential equation (see Chapter 2 of (15) for a beautiful explanation of this theory in a very general context).

Now, if we want to extract Lyapunov exponents from the above construction, we calculate:

$$\chi_{v_j} = \limsup_{t\to\infty} \frac{1}{t} log(|\Phi_t v_j|) \tag{3.4}$$

where $v_j = \frac{\delta x_0}{|\delta x_0|}$ where δx_0 is the small variation vector.[1]

In an intuitive sense, calculation of the Lyapunov exponents corresponds to an averaging of the distortion of the variation "ball" integrated over the very short time intervals. The "growth rates," of each v_j, $0 < j \leq d$ are then summed over the orbit for a fixed set of initial conditions.

3.2 Gram-Schmidt orthonormalization

There are many computational means of calculating the Lyapunov exponents from the fundamental matrix $U_t(x_0)$. We will discuss the simplest of these algorithms, the Gram-Schmidt orthonormalization .

The Gram-Schmidt orthonormalization procedure for replacing a set of linearly independent vectors with a set of orthonormal vectors is a standard technique of linear algebra; we include it here only for completeness, for more information see (58).

Let $\{x_1, x_2, \ldots, x_n\}$ be a basis for a vector space. We can convert $\{x_i\}$ into an orthonormal basis $\{z_1, z_2, \ldots, z_n\}$ via the following algorithm. Let $y_1 = x_1$, and set

$$z_1 = \frac{y_1}{\langle y_1, y_1 \rangle^{1/2}} \tag{3.5}$$

such that z_1 is normalized[2]. Next, set

$$y_2 = x_2 - \langle x_2, z_1 \rangle z_1 \tag{3.6}$$

thus making y_2 orthogonal to z_1. To finish with this step, again normalize y_2 to get

$$z_2 = \frac{y_2}{\langle y_2, y_2 \rangle^{1/2}} \tag{3.7}$$

This can be recursively repeated using the formula

$$y_k = x_k - \langle x_k, z_{k-1} \rangle z_{k-1} - \langle x_{k-1}, z_{k-2} \rangle z_{k-2} - \cdots - \langle x_k, z_1 \rangle z_1 \tag{3.8}$$

assuming that $z_1, \ldots, z_{k-1}$ have already been calculated. At each step, z_i must be normalized as has already been demonstrated. The outcome of this procedure will yield Z, an orthonormal matrix with a basis for the same space as the original matrix X. In practice, there are many other computation schemes available for such a computation of Z. One typical such computation is know as the modified Gram-Schmidt procedure. See (49) and (54) for more information regarding such techniques.

3.3 Explicit Lyapunov characteristic exponent computation

As in the above scenario, begin with the ordinary differential equation $\dot{x} = f(x)$ where $f \in C^1(U)$ and where U is a compact subset of R^n. The differential equation f can be integrated numerically using the standard fourth order Runga-Kutta integrator. Similarly, once the Jacobian matrix of partial derivatives, $Df_t(x)$, has been constructed, each entry can be integrated with the same Runga-Kutta integrator as the differential equation f. The explicit algorithm for the calculation of the Lyapunov exponents is as follows:

1. fix the initial deviation matrix Φ_0;
2. integrate f one time-step;
3. integrate $Df_t(x)\Phi_t$ for one time-step to get the solution $\tilde{\Phi_t}$;
4. apply the Gram-Schmidt orthogonalization to the fundamental solution $\tilde{\Phi_t}$ retaining both the orthogonal matrix Y and the orthonormal matrix Z (Y and Z are exactly as given in section (3.2));

[1] The components of the vector δx_0 consist of a small scalar "perturbation" Δx_0.

[2] A notational side note, $\langle y_1, y_1 \rangle^{1/2} = \sqrt{\sum_{i=1}^{n} y_{1,i}^2}$ where y_1 is a column vector with elements $y_{1,i}$

5. calculate the exponents in the following way:

   i. calculate the exponent at the time t and position x

$$\lambda_j(t) = \frac{\langle y_j, y_j \rangle^{1/2}}{|y_j(0)|} = \frac{\sqrt{\sum_{i=1}^{n}(y_{ij})^2}}{|y_j(0)|} \tag{3.9}$$

   where y_j is the j^{th} column vector of the orthogonal but not orthonormal matrix Y and $|y_j(0)|$ is the initial "perturbation" as given in Φ_0;[3]

   ii. perform the following calculation: $\tilde{\chi}_j(t) = \log(\lambda_j(t)) + \tilde{\chi}_j(t-1)$;

   iii. repeat for each exponent;

6. exchange the old Φ_t with the newly orthonormalized set Z;

7. repeat until convergence of the exponent is achieved (this is known via trial and error) for T time steps;

8. at the T^{th} iteration, average the sum of Lyapunov exponents along the orbit, $\chi_j = \frac{\tilde{\chi}_j}{T}$;

In theory, one could simply just integrate $Df(x_t)\Phi_t$ along the orbit and at time T, calculate characteristic exponents. However, because elements of Df are growing and shrinking at an exponential rate, this is not numerically feasible. Instead, after each time-step[4] the Jacobian needs to be orthonormalized and reset to the initial perturbation size δx_0. The components that form the characteristic exponents must be taken off of the orthogonal but non-normalized matrix Y, for it is this ratio of the non-normalized $|y_i|$ to $y_i(0)$ that yields the expansion and contraction rates.

If the Jacobian exists and can be easily written down analytically, the above discussion provides a rather simple algorithm to program, as it requires only a Runga-Kutta integrator and a Gram-Schmidt orthonormalization scheme. If an explicit Jacobian does not exist, there are many alternatives that we will not discuss here; for algorithms to calculate Lyapunov exponents from a time-series see (42) (117) (25) (116) (38). For the situation where a differential equation exists but the Jacobian is difficult to formulate, there exist several numerical differentiation schemes that will supply the information necessary to perform the Lyapunov exponent calculation[5].

It is worth noting that the Gram-Schmidt orthonormalization is not the only technique that can be used to generate an orthogonal and orthonormal matrix, for instance the modified Gram-Schmidt algorithm often gives far better numerical results. The QR decomposition is another alternative to the Gram-Schmidt orthonormalization that also often yields good results (126). There exist many flavors of Lyapunov exponent calculation algorithms. Many can be found in the references (17) (18) (118) (33) (130) (126) (41).

3.3.1 The case for discrete time

The discrete time version of the Lyapunov spectrum algorithm described above requires an extremely simple alteration. Instead of integrating the Jacobian matrix of partial derivatives, the Jacobian simply needs to be iterated one time step (the averaging done at the end of the Lyapunov exponent calculation must be adjusted to take this into account). Aside from this trivial alteration, the algorithm is identical to the continuous-time case. For a nice discussion see (88).

---

[3] Note, with respect to the previous two sections the following notations are equivalent: $v_j = \frac{\langle y_j, y_j \rangle^{1/2}}{|y_j(0)|} = \frac{\langle \delta x_j, \delta x_j \rangle^{1/2}}{|\delta x_j(0)|}$.

[4] In practice, the renormalization does not need to be done very time step, see (18) for more details.

[5] One early method of calculating the Jacobian without actually calculating the derivative can be found in equations $A.3$ and $A.4$, page 1615 of (118). Another alternative is to calculate the derivative using various computational means; one interesting alternative is provided by the automatic differentiation library, ADOLC, http://www.math.tu-dresden.de/wir/project/adolc/

3.3.2 Examples

We will now discuss two very basic examples demonstrating the computation of the Lyapunov spectra. The first example will be a very simple, linear, ordinary differential equation. The second example will be that of the Hénon map, one of the standard and most studied maps in computational and mathematical dynamical systems.

3.3.2.1 A linear ordinary differential equation

Begin with the differential equation, $\dot{x} = f(x) = Ax$ where A is a diagonal matrix with eigenvalues a and b. In this case, the Jacobian matrix is $Df = A$. Because f is linear, we already know the Lyapunov exponents a and b.

Begin by assigning the initial deviation at Δx_0. Now integrate f from $t = 0$ to some small time δt to get the fundamental matrix times the initial variation, $Df_t(x)\Phi_t$. Next apply the Gram-Schmidt orthonormalization to $Df_t(x)\Phi_t$ to calculate Y and Z. In this case we have:

$$Y = \begin{bmatrix} e^{aT}\Delta x_0 & 0 \\ 0 & e^{bT}\Delta x_0 \end{bmatrix} \text{ and } Z = \begin{bmatrix} \Delta x_0 & 0 \\ 0 & \Delta x_0 \end{bmatrix}$$

Continuing with the calculation, the spectrum $\sigma(Y) = \{e^{a\delta t}\Delta x_0, e^{b\delta t}\Delta x_0\}$, with $\lambda_1 = e^{a\delta t}$ and $\lambda_2 = e^{b\delta t}$. This yields the characteristic exponents up to time δt of $\tilde{\chi_1} = |a|\delta t$ and $\tilde{\chi_2} = |b|\delta t$. Replace Φ_t with $Z\Delta x_0$ and repeat the above procedure for T iterates, keeping track of the sum of the exponents at each iterate. After the T time-steps have been performed, all that remains is to calculate the average of the Lyapunov exponents at each time-step. To do this, divide the sum of the exponents by $T\delta t$ to get the Lyapunov exponents $\chi_1 = |a|$ and $\chi_2 = |b|$.

3.3.2.2 The Hénon map

The standard Hénon map (53) is defined:

$$\begin{aligned} x_{t+1} &= 1 - ax_t^2 + by_t \\ y_{t+1} &= x_t \end{aligned} \tag{3.10}$$

with $a = 1.4$ and $b = 0.3$ as the usual parameters. Calculating the Jacobian of (3.10) analytically we arrive at

$$Df = \begin{bmatrix} -2ax_t & b \\ 1 & 0 \end{bmatrix} \tag{3.11}$$

From this, we can calculate Φ_t:

$$Df(x_t)\Phi_t = \begin{bmatrix} -2ax_1 & b \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta x_0 & 0 \\ 0 & \Delta x_0 \end{bmatrix} \tag{3.12}$$

Carrying out the matrix multiplication explicitly yields Y. From this we can calculate Z (using the Gram-Schmidt procedure), and also the Lyapunov exponents at time $t = 1$ explicitly. Begin by computing the inner product of y_1,

$$\langle y_1, y_1 \rangle^{1/2} = (4a^2x_t^2\Delta x_0^2 + \Delta x_0^2) \tag{3.13}$$

which, given y_1, we can calculate z_1:

$$z_1 = \begin{bmatrix} \frac{-2ax_t\Delta x_0}{<y_1,y_1>^{1/2}} \\ \frac{\Delta x_0}{<y_1,y_1>^{1/2}} \end{bmatrix} \tag{3.14}$$

Now, to calculate z_2 we note $y_2 = x_2 - < x_2, z_1 > z_1$, where $x_2 = (b\Delta x_0, 0)$ and z_1 as given in equation (3.14). Carrying out the computation, we arrive at

$$y_2 = \begin{bmatrix} \frac{2abx_t\Delta x_0^3}{|y_1|} \\ 0 \end{bmatrix} \tag{3.15}$$

from which we can calculate the normalization factor for z_2,

$$< y_2, y_2 >= (b\Delta x_0 - \frac{4a^2bx_t^2\Delta x_0^3}{|y_1|})^2 + (\frac{2abx_t\Delta x_0^3}{|y_1|})^2 \tag{3.16}$$

Finally this yields

$$z_2 = \frac{1}{< y_2, y_2 >^{1/2}} \begin{bmatrix} b\delta x_0 - \frac{4a^2bx_t^2\Delta x_0^3}{|y_1|} \\ \frac{2abx_t\Delta x_0^3}{|y_1|} \end{bmatrix} \tag{3.17}$$

The Z matrix is now used to replace the old variation Φ_t, $Z = \Phi_t$. Lastly, the Lyapunov exponents at time 1 are given by

$$\begin{aligned} \chi_1(1) &= \log \frac{|\lambda_1(t)|}{\Delta x_0} = \log \frac{|y_1|}{\Delta x_0} \\ \chi_2(1) &= \log \frac{|\lambda_2(t)|}{\delta x_0} = \log \frac{|y_2|}{\delta x_0} \end{aligned} \tag{3.18}$$

This completes the first iteration of the calculation of the Lyapunov spectra for the Hénon map. This process is repeated until the exponents have converged within a desired target error. A long time average of such for the Hénon map gives $\chi_1 = 0.417212$ and $\chi_2 = -1.62119$, yielding a sum of the exponents as -1.203978.

3.3.2.3 Delayed logistic map

One last example, involving a bifurcation of prime importance, the Neimark-Sacker bifurcation, is that of the delayed logistic map. The delayed logistic map is given by the following set of equations:

$$\begin{aligned} x_{t+1} &= ax_t(1 - y_t) \\ y_{t+1} &= x_t \end{aligned} \tag{3.19}$$

which yields a Jacobian:

$$Df = \begin{bmatrix} a(1 - y_t) & ax_t \\ 1 & 0 \end{bmatrix} \tag{3.20}$$

We will refrain from a dissection of the Lyapunov spectrum calculation here and simply show the results for the delayed logistic map upon variation of the a parameter. Considering figure (3.1), the bifurcation diagram of equation (3.19), it is clear that there exists a Neimark-Sacker bifurcation at $a = 2.0$. Noting figure (3.2) at the bifurcation point, both of the Lyapunov exponents are zero, however, as a is increased further, one Lyapunov exponent remains zero representing the neutral direction, and one exponent become negative representing the contraction to the one-dimensional quasi-periodic orbit. The persistent limit cycle is followed by a bifurcation to a periodic orbit followed by a bifurcation into chaos. In later chapters where we discuss the route to chaos, we will spend a considerable amount of time studying high-dimensional analogs to this example.

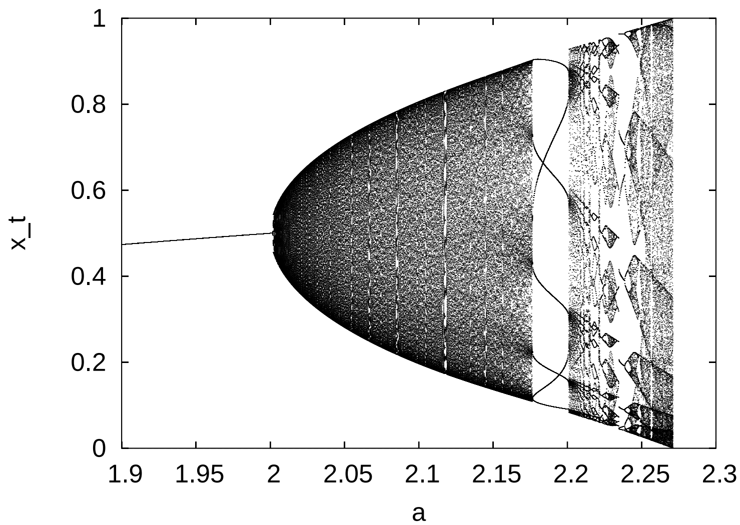

Figure 3.1: The bifurcation diagram for the delayed logistic map.

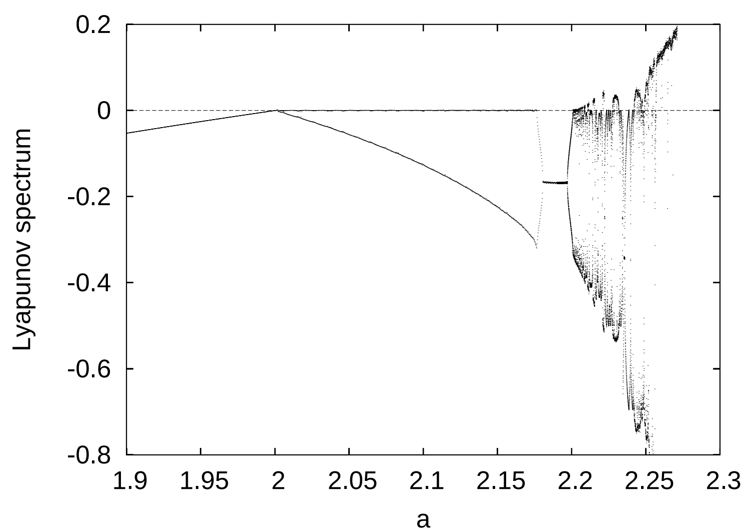

Figure 3.2: The Lyapunov spectrum for the delayed logistic map over the region from a fixed point to chaos, including a Neimark-Sacker bifurcation at $a = 2$.

Chapter 4

Artificial Neural Networks

4.1 Abstract Neural Networks

The motivation and construction of the set of mappings we will use for our investigation of dynamical systems follows via two directions, the embedding theorem of Takens ((125), (86)) and the neural network approximation theorems of Hornik, Stinchomebe, and White (60). We will use the Takens embedding theorem 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 $F : R^d \to R^d$. This is important as we will be using time-delayed scalar neural networks for our study. The neural network approximation theorems show that neural networks of a particular form are open and dense in several very general sets of functions and thus can be used to approximate any function in the allowed function spaces.

There is overlap, in a sense, between these two constructions. The embedding theory shows an equivalence or the approximation capabilities of scalar time-delay dynamics with standard, $x_{t+1} = F(x_t)$ ($x_i \in R^d$) dynamics. There is no mention of, in a practical sense, the explicit functions in the Takens construction. The neural network approximation results show in a precise and practical way, what a neural network is, and what functions it can approximate. It says that neural networks can approximate the $C^r(R^d)$ mappings and their derivatives, but there is no mention of the time-delays we wish to use. Thus we need to discuss both the embedding theory and the neural network approximation theorems.

Those not interested in the mathematical justification of our construction may skip to section (4.2) where we define, in a more concrete manner, our neural networks.

4.1.1 General neural network formulation

Begin by noting that, in general, a neural network is a C^r mapping $\gamma : R^n \to R$. More specifically, the set of feedforward networks with a single hidden layer, $\Sigma(G)$, can be written:

$$\Sigma(G) \equiv \{\gamma : R^d \to R | \gamma(x) = \sum_{i=1}^{N} \beta_i G(\tilde{x}^T \omega_i)\} \tag{4.1}$$

where $x \in R^d$, is the $d-$vector of networks 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 $G : R \to R$ is the hidden layer activation function (or neuron). The partial derivatives of the network output function, γ, are

$$\frac{\partial g(x)}{\partial x_k} = \sum_{i=1}^{N} \beta_i \omega_{ik} DG(\tilde{x}^T \omega_i) \tag{4.2}$$

where x_k is the k^{th} component of the x vector, ω_{ik} is the k^{th} 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 for precisely this reason that we choose the time-delayed formulation.

4.1.2 Approximation Capabilities

We will begin with a brief description of spaces of maps useful for our purposes and conclude with the keynote theorems of Hornik et al. (60) necessary for our work. Hornik et al. 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 (59), (60).

The ability of neural networks to approximate functions which are of particular interest, can be most easily seen via a brief discussion of Sobolev function space, S_p^m. We will be brief, noting references Adams (1) and Hebey (52) 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. $p \in R$, $1 \leq p < \infty$; 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 $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_k$, $|\alpha| \leq m$;

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 $|\alpha|$

$$\frac{\partial^{|\alpha|}}{\partial x^\alpha} \equiv \frac{\partial^{|\alpha|}}{(\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_d^{\alpha_d})} \tag{4.3}$$

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 upon 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 $1 \leq p < \infty$, we define the Sobolev norm, $\| \cdot \|_{m,p}$, as follows:

$$||u||_{m,p} = \left(\sum_{0 \leq |\alpha| \leq m} (\| D^\alpha u \|_p^p) \right)^{1/p} \tag{4.4}$$

where $u \in L^1_{loc}(U)$ is a locally integrable, real valued function (with differentiability prescribed by α)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:

$$\rho^m_{p,\mu}(f, g) \equiv ||f - g||_{m,p,U,\mu} \tag{4.5}$$

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,

$$C^m(U) = \{f \in C(U) | D^\alpha f \in C(U), ||D^\alpha f||_p < \infty \forall \alpha, |\alpha| \leq m\} \tag{4.6}$$

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

Definition 15 *For any positive integer m and $1 \leq p < \infty$, we define a Sobolev space $S_p^m(U, \lambda)$ as the vector space on which $|| \cdot ||_{m,p}$ is a norm:*

$$S_p^m(U, \lambda) = \{f \in C^m(U) | \; ||D^\alpha f||_{p,U,\lambda} < \infty \text{ for all } |\alpha| \leq m\} \tag{4.7}$$

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

Two functions in $S_p^m(U, \lambda)$ are close in the Sobolev metric if all the derivatives of order $0 \leq |\alpha| < m$ are close in the L_p metric. As will be evident considering figure (4.1), we are attempting to approximate $\tilde{F} = g \circ F \circ g^{-1}$ where $\tilde{F} : R^{2d+1} \to R$, $F : M \to M$ (M is a d-dimensional, compact manifold), and g is an embedding; for this task the functions from $S_p^m(U, \lambda)$ will serve us 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 16 (m-uniformly dense) *Assume m and l are non-negative integers $0 \leq m \leq l$, $U \subset R^d$, 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:*

$$\max_{|\alpha| \leq m} \sup_{x \in K} |D^\alpha f(x) - D^\alpha g(x)| < \epsilon \tag{4.8}$$

then $\Sigma(G)$ is m-uniformly dense on compact, metrizable set 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 17 (l-finite) *Let l be an non-negative integer. G is said to be l-finite for $G \in C^l(R)$ if:*

$$0 < \int |D^l G| d\lambda < \infty \tag{4.9}$$

i.e. the l^{th} derivative of G must be both bounded away from zero, and finite for all l (recall $d\lambda$ 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 (60)) *If G is l-finite, $0 \leq m \leq l$, 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 $1 \leq p < \infty$.*

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 (60) and (1).

4.1.3 Marriage with Takens theorem

We wish, in this report, to investigate dynamical systems on compact sets. Specifically, begin with a compact manifold M of dimension d and a diffeomorphism $F \in C^r(M)$ for $r \geq 2$ defined as:

$$x_{t+1} = F(x_t) \tag{4.10}$$

with $x_t \in M$. However, for computational reasons, we will be investigating this space with neural networks that can approximate (see section (4.1.2)) dynamical systems $f \in C^r(R^d, R)$ that are time-delay maps given by:

$$y_{t+1} = f(y_t, y_{t-1}, \ldots, y_{t-(d-1)}) \tag{4.11}$$

where $y_t \in R$. Both systems (4.10) and (4.11) form dynamical systems. However, since we intend to use systems of the form (4.11) to investigate the space of dynamical systems as given in equation (4.10), we must show that a study of mappings of the form (4.11) is somehow equivalent to mappings of the form (4.10). We will demonstrate this by employing an embedding theorem of Takens to demonstrate the relationship between time-delay maps and non-time-delay maps in a more general and formal setting.

We call $g \in C^k(M, R^n)$ an embedding if $k \geq 1$ and if the local derivative map (the Jacobian - the first order term in the Taylor expansion) is one-to-one for every point $x \in M$ (i.e. g must be an immersion) and if g maps M homeomorphically into its image. The idea of the Takens embedding theorem is that given a d-dimensional dynamical system and a "measurement function," $E : M \to R$ (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 $x \in M$ 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 M. Note that the $2d+1$ tuple is a time-delay map of x. We can now state the Takens embedding theorem:

Theorem 7 (Takens' embedding theorem (125) (86)) *Let M be a compact manifold with dimension d. There is an open dense subset $S \subset Diff(M) \times C^k(M, R)$ with the property that the Takens map*

$$g : M \to R^{2d+1} \tag{4.12}$$

given by $g(x) = (E(x), E \circ F(x), E \circ F^2(x), \ldots, E \circ F^{2d}(x))$ is an embedding of C^k manifolds, when $(F, E) \in S$.

Here *Diff*(M) is the space of C^k diffeomorphisms from M to itself with the subspace topology from $C^k(M, M)$. Thus, there is an equivalence between time-delayed Takens maps of "measurements" and the "actual" dynamical system operating in time on $x_t \in M$. This equivalence is that of an embedding (the Takens map), $g : M \to R^{2d+1}$.

To demonstrate how this applies to our circumstances, consider figure (4.1) in which F and E are as given above and the embedding g is explicitly given by:

$$g(x_t) = (E(x_t), E(F(x_t)), \ldots, E(F^{2d}(x_t))) \tag{4.13}$$

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 $2d+1$ 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:

$$(y_1, \ldots, y_{2d+1}) \mapsto (y_2, \ldots, y_{2d+1}, g(F(g^{-1}(y_1, \ldots, y_{2d+1})))) \tag{4.14}$$

where, again, $\tilde{F} = g \circ F \circ g^{-1}$. 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).

Let us summarize what we are attempting to do: we wish to investigate dynamical systems given by (4.10) but for computational reasons we wish to use dynamical systems given by (4.11); the Takens embedding theorem says that dynamical systems of the form (4.10) can be generically represented (via the Takens embedding map g) by time-delay dynamical systems of the form (4.14). Since neural networks will approximate dynamical systems of the form (4.14) on a compact and metrizable set, it will suffice for our investigation of dynamical systems of the form (4.10) to consider the space of neural networks mapping compact sets to compact sets as is given in section (4.1.1).

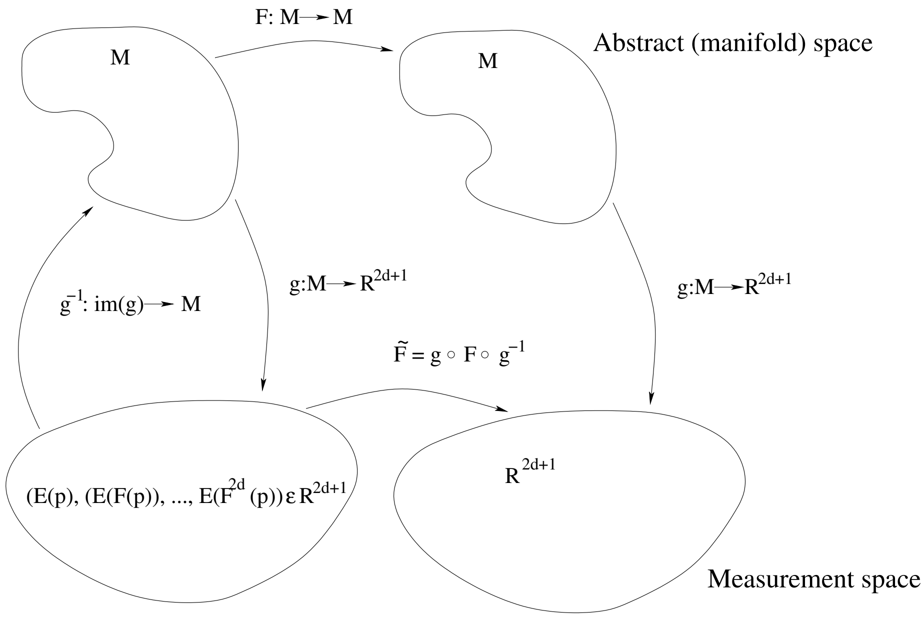

Figure 4.1: Schematic diagram of the Takens embedding theorem and how it applies to our construction.

4.2 Our Neural Network Construction

The single layer feed-forward neural networks (γ's from the above section) we will consider are of the form

$$x_t = \beta_0 + \sum_{i=1}^{N} \beta_i G\left(s\omega_{i0} + s\sum_{j=1}^{d} \omega_{ij} x_{t-j}\right) \tag{4.15}$$

which is a map from R^d to R. The activation function[1] G, for our purpose will be tanh(). In (4.15), 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 ($\beta_i, w_{ij}, x_j, s \in R$) and the β_i's and w_{ij}'s are elements of weight vectors and matrices respectively (which we hold fixed for each case). The initial conditions are denoted as $(x_0, x_1, \ldots, x_d)$, and $(x_t, x_{t+1}, \ldots, x_{t+d})$ represents the current state of the system at time t.

We assume that the β's are *iid* uniform over $[0, 1]$ and then re-scaled to satisfy the condition $\sum_{i=1}^{N} \beta_i^2 = N$. The w_{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's are chosen *iid* uniform on the interval $[-1, 1]$. All the weights and initial conditions are selected randomly using a pseudo-random number generator (72), (100).

We would like to make a few notes with respect to our 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 toward the finite set $(\beta_0 \pm \beta_1 \pm \beta_2 \cdots \pm \beta_n)$, 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 $< \beta >$ 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's with $\sum_{i=1}^{N} |\beta_i| = k$ fixing k for all N and d. This is a bad idea however since, if the β_i's are restricted to a sphere of radius k, as N is increased, $\langle {\beta_i}^2 \rangle$ goes to zero (7). The other extreme of our squashing also yields a very specific behavior type. For x very near 0, the tanh(x) function is nearly

[1]The activation function is sometimes referred to as a squashing function.

linear. Thus choosing s small will force the dynamics to be mostly linear, again yielding fixed point and periodic behavior (no chaos). Thus the scaling parameter s will provide 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.

Note that in a very practical sense, the measure we are imposing on the set of neural networks is our means of selecting the weights that define the networks. This will introduce a bias into our results that is unavoidable in such experiments; the very act of picking networks out of the space will determine, to some extent, our results. Unlike actual physical experiments, we could, in principle, prove an invariance of our results to our induced measure. This is difficult and beyond the scope of this thesis. Instead it will suffice for our purposes to note specifically what our measure is (our weight selection method), and how it might bias our results. Our selection method will include all possible networks, but clearly not with the same likelihood. In the absence of a theorem with respect to an invariance of our induced measure, we must be careful in stating what our results imply about the ambient function space.

Chapter 5

General dynamics

Before we begin with our analysis of the various regions, let us first briefly present some general background material regarding our neural networks.

5.1 Probability of Chaos

The results in this section have been previously published in (8), (32), and (7). When discussing probability of an event, in this case a chaotic neural network, one must always define a notion of a measure. For our construction it is not trivial to define such a measure in a rigorous manner, and we will not formalize a notion of a measure here. We do have, in a practical sense, a measure on our set of neural networks, however, and that is the distributions of the weights. The neural networks are defined by their weight distributions and the weight distributions have explicit probability measures placed on them. The ω_{ij} weight matrix has the standard normal distribution with mean zero and variance s and the β_i weight vector is uniform on the interval $(0, 1)$. Thus, the neural networks have a probability measure associated with their selection method. All of our results must be interpreted with this implicit measure in mind.

To achieve a general feeling for large-scale variation of the neural networks relative to variation of N,[1] d[2] and s, let us consider the following figures. Figure (5.1) shows the percentage of chaos in networks for varying N and d with s fixed at 8. Notice that as d is increased, the probability of chaos approaches unity; the same is true of N. The nature of the attractors at high d and low N, and low d and high N is quite different as we will see in later chapters. The relatively straight contour lines in figure (5.1) are related to the way we scale the β's. In previous work (7) we assigned the β_i's differently, resulting in different contour lines and a different N dependence.

Figure (5.2) shows the percentage of chaotic networks for varying s and d with N fixed at 8. Notice the "C"-shaped contour lines. This would suggest that the s parameter can be optimized for maximum chaos. This "C"-shaped s dependence is a result of the s parameter moving the networks through the five dynamics regions discussed in chapter 6. The s parameter moves the argument of the squashing function, tanh(), from the linear region (s near zero), to nonlinear region ($s \in (a, b)$ where a and b depend upon N and d), and into the binary region ($s > b$). A large portion of the intuition for the rest of our work comes from this figure. When we discuss 1-d intervals in parameter space in chapter (9), we will be discussing vertical slices of figure (5.2).

[1]Recall N is the number of neurons and in a sense corresponds to the order of derivative approximation capable.
[2]Recall that d is the number of embedding dimensions for dynamics on a $\frac{d-1}{2}$ dimensional manifold.

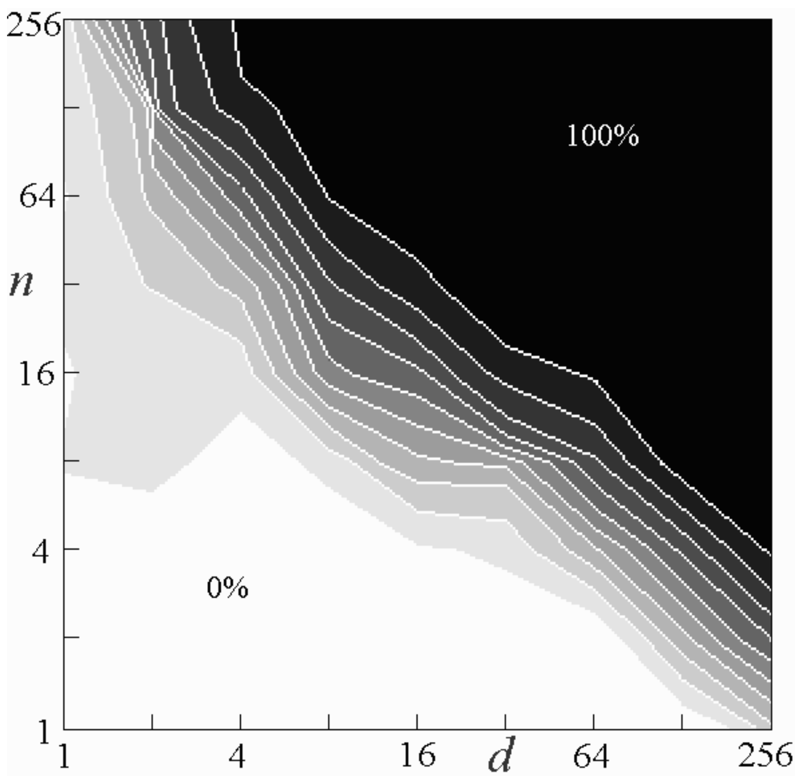

Figure 5.1: Probability of chaos contour plot for $s = 8$ and various d and N.

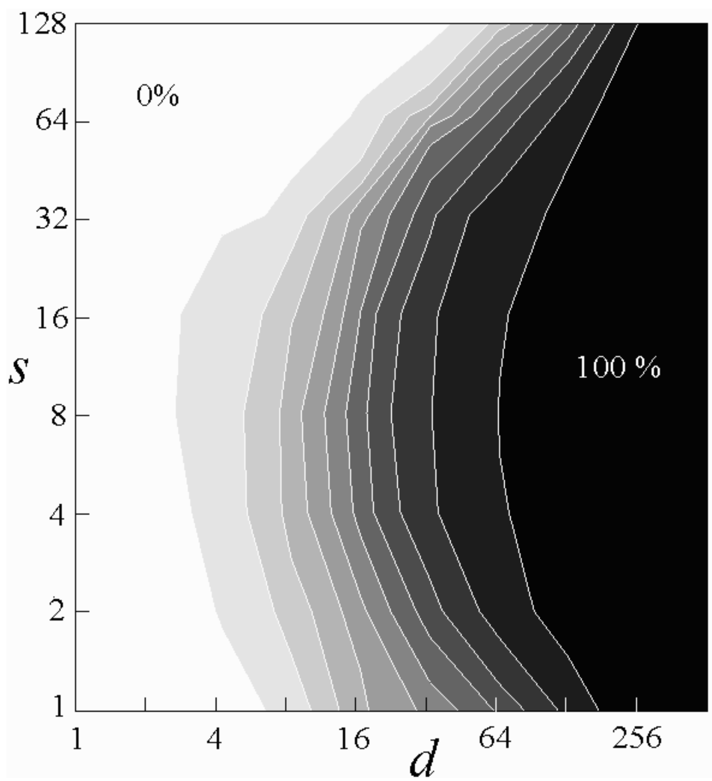

Figure 5.2: Probability of chaos contour plot for $n = 8$ and various d and s.

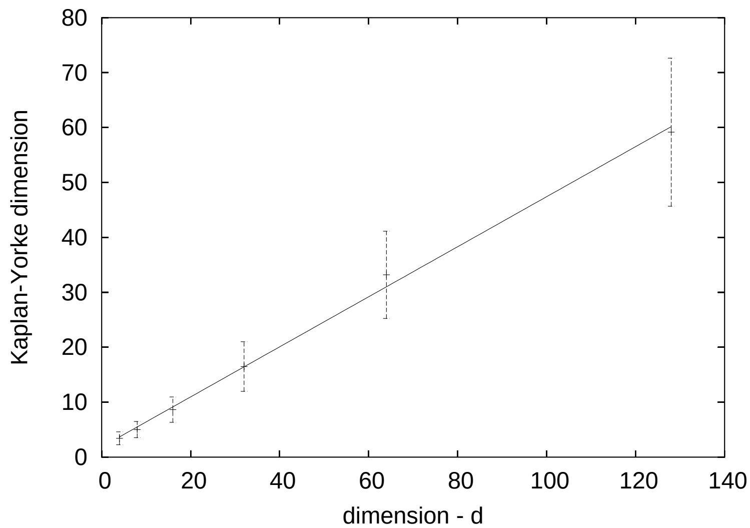

Figure 5.3: Mean maximum Kaplan-Yorke dimension versus dimension, d. For the set of networks analyzed, $d_{KY} \sim 0.46d$.

5.2 Lyapunov spectrum diagnostics

We will now briefly discuss three diagnostics that can be extracted from considering the Lyapunov spectrum. Such diagnostics are both interesting in their own right, and aid in the intuition for later discussion.

5.2.1 Dimensional Variation

The Kaplan-Yorke dimension of an attractor (67) is one of the most commonly used notions of dimension. Given the Lyapunov characteristic exponents, χ_i, the Kaplan-Yorke dimension is defined:

$$d_{KY} = j + \frac{\chi_1 + \cdots + \chi_j}{|\chi_{j+1}|} \tag{5.1}$$

where j is the largest integer such that:

$$\chi_1 + \cdots + \chi_j \geq 0 \tag{5.2}$$

For our networks and our weight structure, the maximum Kaplan-Yorke dimension scales like $0.46d$, as can be seen in figure (5.3). However, there is a strong s dependence on the Kaplan-Yorke dimension. Figure (5.4) depicts the dependence of the Kaplan-Yorke dimension on the s parameter. Important features of figure (5.4) include the single maximum which is achieved quickly upon increasing s from zero followed by a exponential decay in dimension after the maximum dimension is reached. As we shall see, this trend is preserved in both the entropy variation and in the number of positive Lyapunov exponents.

5.2.2 Entropy

For our purposes, we will define the entropy for our networks as the metric entropy, or the sum of the positive Lyapunov exponents. The entropy variation with dimension is distinctly different the Kaplan-Yorke dimension. The entropy grows with dimension until $d = 8$ at which point it begins to decrease as can be seen in figure (5.5). Trends from the data in figure (5.5) are somewhat difficult to discern since the error (variance) about the mean maximum entropy is large. However, the maximum entropy is decreasing slowly with increasing dimension. This decrease in mean maximum entropy is an artifact due to an increase in

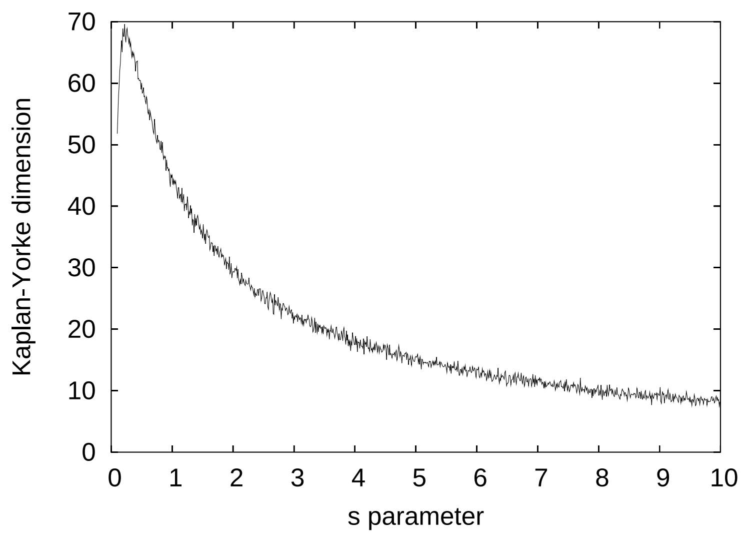

Figure 5.4: Variation of Kaplan-Yorke dimension versus s for a single network with $N = 32$ and $d = 128$.

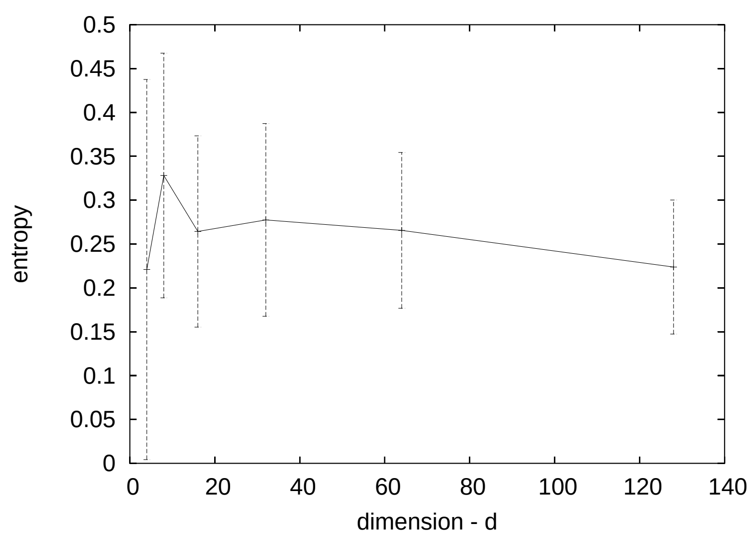

Figure 5.5: Mean maximum entropy versus dimension, d.

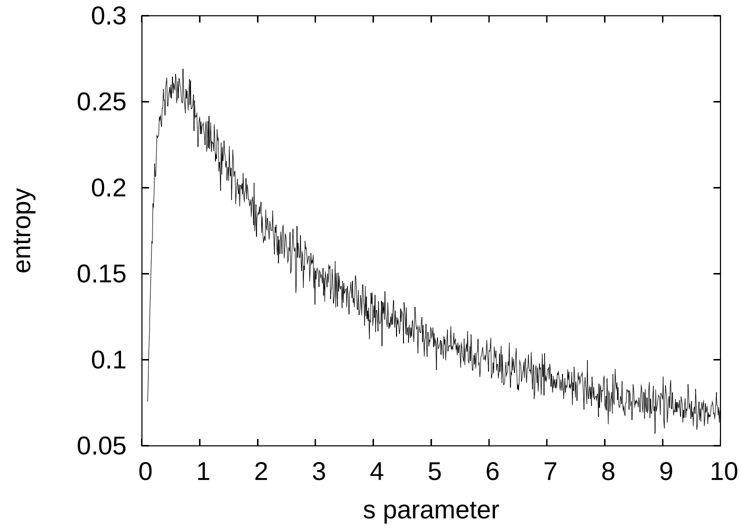

Figure 5.6: Variation in the entropy of a single network with s. The network considered has 32 neurons and 128 dimensions.

the time scale in our networks — our networks are time-delay networks, and thus an increase in d imposes an increase in the time scale. The increase in time scale decreases the Lyapunov exponents. There is an s dependence in the maximum entropy just as with the Kaplan-Yorke dimension, as can be seen from figure (5.6), the mean maximum entropy occurs for a slightly larger value of s than the mean maximum Kaplan-Yorke dimension.

5.2.3 Dissipation

We define the dissipation of our networks as the sum of all the Lyapunov exponents. As the dimension of the network is increased, the maximum dissipation increases like $-12.8 \log(d)$. This result is not surprising in light of the relatively slow increase in the Kaplan-Yorke dimension. Looking ahead, this result can be further understood considering figure (9.11) which shows that the maximum number of positive Lyapunov exponents is growing at $\sim d/4$, or about a third as fast as the maximum number of negative exponents at the same s value. Finally, considering figures (9.3) and (9.4), the negative exponents become both larger in number and more negative as the dimension is increased. It is worth noting that dissipation is also clearly dependent on s. The mean maximum dissipation we present for each case is reached at the maximum s value considered, $s = 10$. To see this, consider figure (5.8) where we plot the dissipation with variation in s. Unlike the entropy and Kaplan-Yorke dimension, the dissipation decreases linearly with s (dissipation $\sim -7s$).

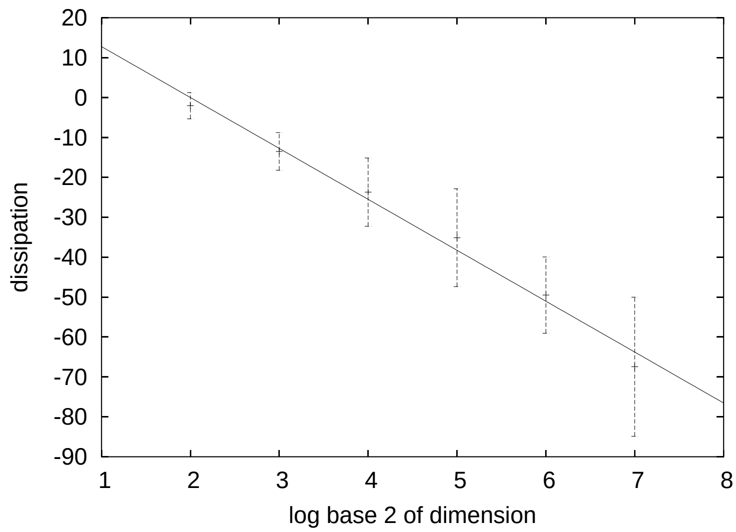

Figure 5.7: The mean maximum dissipation in neural networks versus log() dimension. The dissipation is proportional to $-12.8\log(d)$.

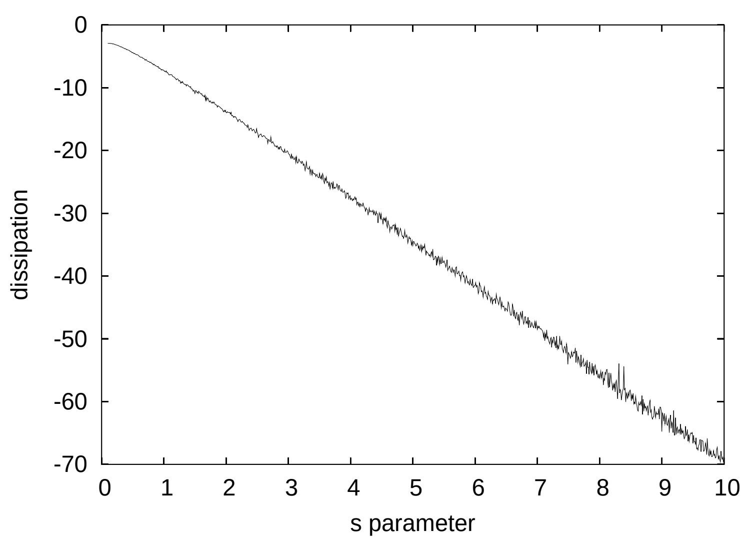

Figure 5.8: Variation in the dissipation of a single network with s. The network considered has 32 neurons and 128 dimensions.

Chapter 6

Splitting up the Parameter Space

We will use s as our primary bifurcation parameter, and thus we will use a path along s to stratify the parameter space of our networks into five regions. However, this does not mean that we will only consider perturbations of a single parameter s in each region. The stratification afforded by the s parameter is a mere convenience for a more structured analysis of the types of dynamics encountered in our set of neural networks. A schematic of the vague stratification is given in figures (6.1), (6.2), (6.3), and (6.4). It must be emphasized that this stratification was created to aid in the organization of our arguments, and nothing more. Recall the regions include the first bifurcation region (I), the routes to chaos region (II), the region between the onset of chaos and bifurcation chains (III), the bifurcation chain region (IV), and the transition from bifurcation chains to finite state dynamics.

Figure (6.1), which is both low-dimensional and has a low number of parameters has regions I, II, III, and V. In this figure region IV is difficult to discern from region III. Region V is really trivial as the transition from chaos to periodic behavior is quite sharp.

Figure (6.2)is low-dimensional but has a large number of parameters. This network again has regions I, II, and III, region IV is difficult to identify. Region V however is quite present.

Figure (6.3) is high-dimensional but has relatively few parameters. This network is the only network we will show having all five regions. In this network, region V is quite distinct, however the existence of region III is questionable. Regions I, II, and IV are clearly evident.

Figure (6.4) is both high-dimensional and has a large number of parameters. This network has only regions I, II, IV, and V, however region V is not displayed. This case is prototypical for high-dimensional networks with many parameters. Much of our analysis will comprise studying networks such as this one.

6.1 Region I: fixed points to the first bifurcation

All of our networks, for small enough values of s, are linear and thus give stable fixed points. Because bifurcations from fixed points of C^r maps with one parameter are relatively well understood, the simplicity of this bifurcation affords us an opportunity that does not exist for the later bifurcations. We will attempt to quantify the probability of the first bifurcation types of our maps and provide a conjecture for why such a probability should be the case based on the random matrix theory of Girko (44), Edelman (39) and Bai (14). The beginnings of our understanding of this region is put forth in (8).

6.2 Region II: the route to chaos - between the first bifurcation and chaos

Region II comprises the region along the s curve between the first bifurcation and the onset of chaos. It is in this region that we address questions regarding the "routes to turbulence." Some of our conjectures from

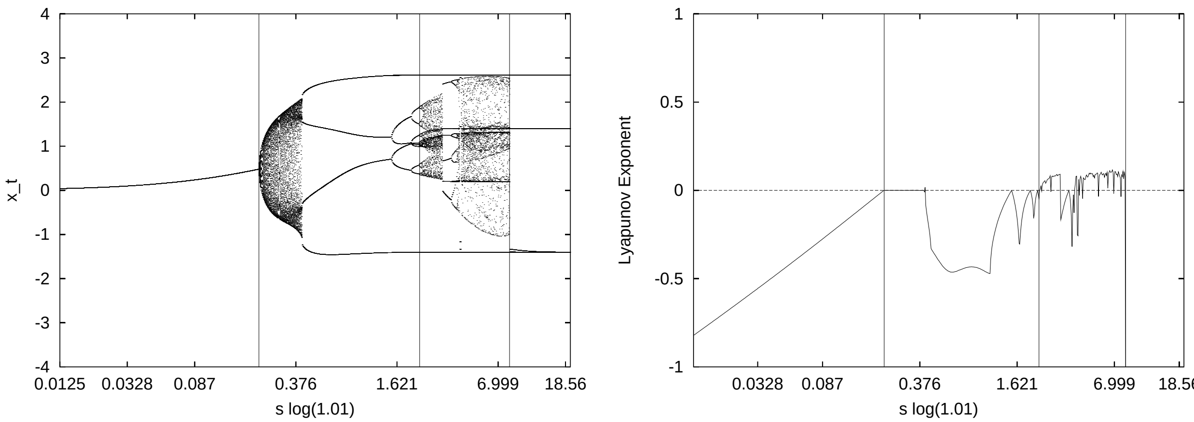

Figure 6.1: Bifurcation diagram with the largest Lyapunov exponent for $N = 4$ and $d = 4$ with regions I, II, III, and V.

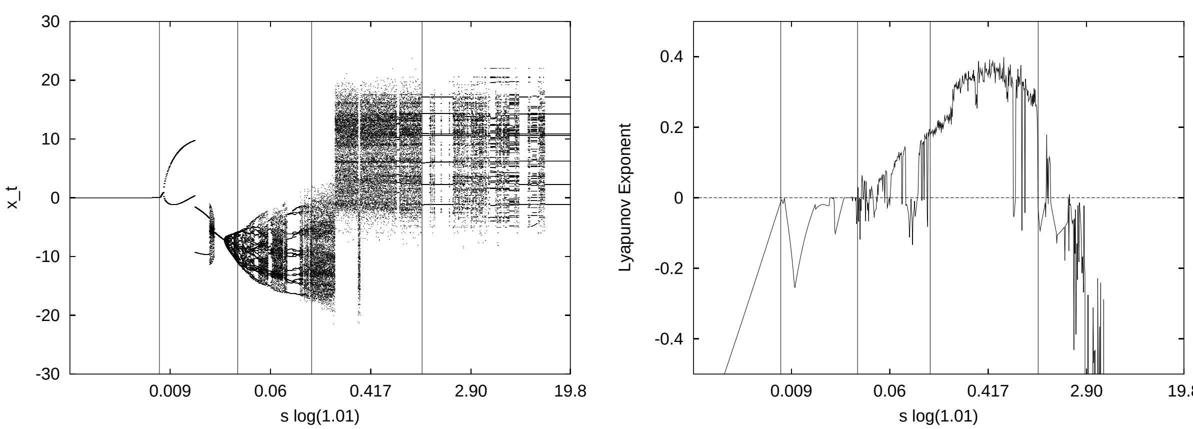

Figure 6.2: Bifurcation diagram with the largest Lyapunov exponent for $N = 64$ and $d = 4$ with regions I, II, III, a possible region IV, and V.

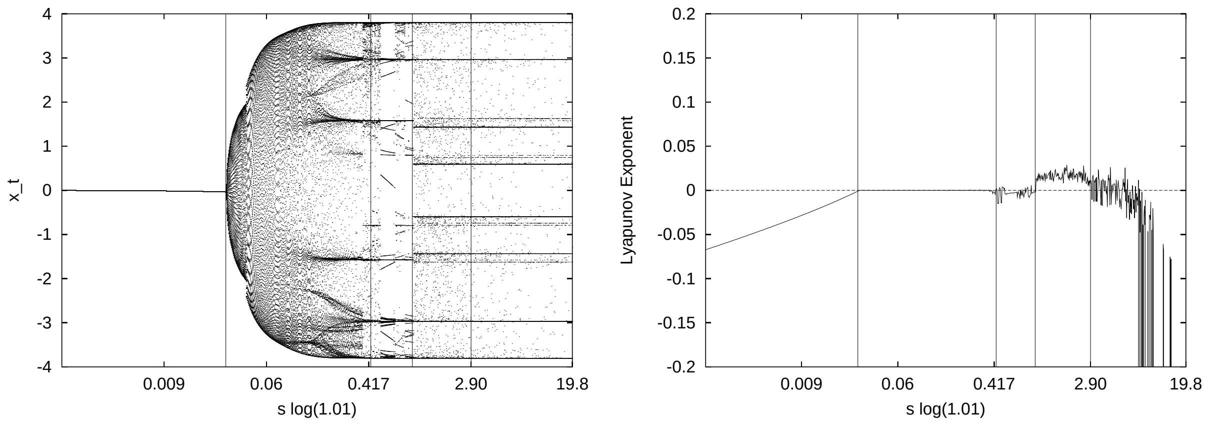

Figure 6.3: Bifurcation diagram with the largest Lyapunov exponent for $N = 4$ and $d = 64$ with regions I, II, III, IV, and V.

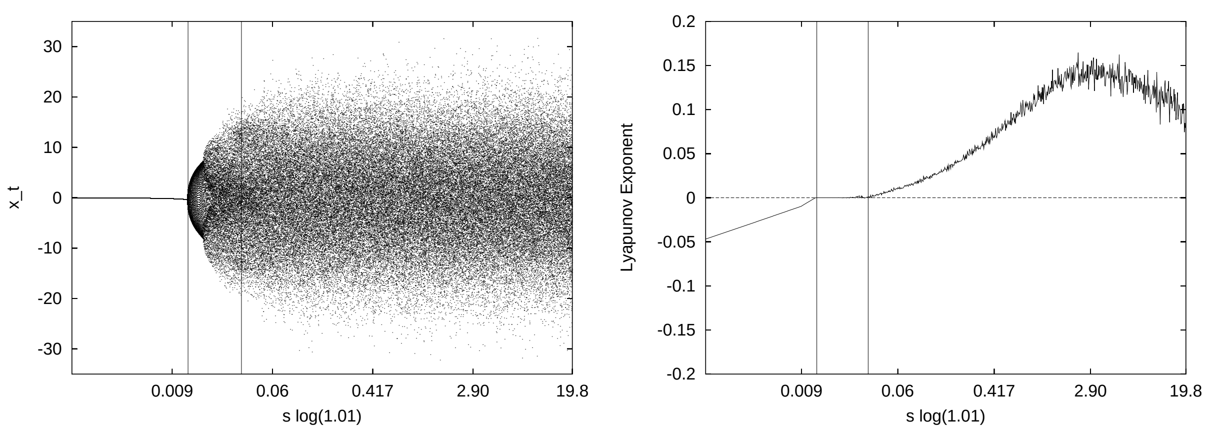

Figure 6.4: Bifurcation diagram with the largest Lyapunov exponent for $N = 64$ and $d = 4$ with regions I, II, IV, region V is not displayed.

region I vaguely carry over into region II. However, good precise statements are yet lacking. Some of this is due to difficulties in the bifurcation theory, others are holes yet to be filled by our computational investigations. At this time, we have a substantial collection of cases, but the current results remain anecdotal. Thus, in this section we will present our current conjectures as well as a brief synopsis of our results.

6.3 Region III: between the onset of chaos and the bifurcation chains

Region III is the region between the onset of chaos and the bifurcation-chain region. For high-dimensional neural networks, this region is quite well defined when it exists, however, for networks with many parameters region III is nonexistent. For low-dimensional networks this region is difficult to resolve and often comprises the entirety of the chaotic region up to the transition where the network becomes binary. Region IV seems only to exist when the dimension of the network is high enough. Thus, if the dimension of the network is low enough such that the existence of region IV does not clearly exist, we will not make a distinction between regions III and IV, and rather combine them into one region, referring to that region as region III. However, the primary concern of this work is high dimensional dynamical systems. Thus this ambiguity will cause us little strife. Further, our stratification was somewhat vague from the onset, and since its only purpose is for organization, we will not concern ourselves with this issue further.

All that being said, we have a very slight understanding of region III. For many systems with a large number of neurons and a large number of dimensions, both region II and III seem to scarcely exist — region II is only obvious when considering variation in variation of $\log_{10}(s)$. In networks of intermediate dimensions, region III can be extremely complicated and contain a great deal of diversity between different networks. A similar situations can arise for high-dimensional networks with a low number of neurons or parameters.

Very little is currently understood about region III, and in this work we will refrain from a particularly detailed analysis. One possible direction for a better understanding of this region is an analysis using the numerical bifurcation software developed by Yuri Kuznetsov (70). However, at this time we include this region only for completeness of presentation, noting it largely consists of problems for future work.

We will include a brief discussion of the diversity in this region, along with some examples.

6.4 Region IV: Bifurcation chain region

Region IV comprises the bulk of our analysis, and exists primarily in neural networks of high dimension with a large number of neurons. This region provides, in a sense, some dynamic stability, but also many subtleties and opportunities for analysis relating to theory by the mathematics dynamics community.[1]

In this region we will address the following issues: existence of periodic windows in parameter space; the observability of the structural stability conjecture (now proven) of Palis and Smale (90); the Pugh-Shub stable ergodicity conjecture (106); the existence of center bunching; and we will put forth some dynamic stability conjectures of our own. Many ongoing questions will remain, but we have obtained a convincing set of evidence to support our claims.

6.5 Region V: Bifurcation chains to finite state dynamics

Region V corresponds to the region along the s curve where the neurons are nearly saturated. Like region III, this region is not well understood and is included for completeness. However, unlike region III, the dynamics are likely much more uniform across various networks with different numbers of neurons and dimensions. Further, we have a much better idea of both what is occurring dynamically in this region, as well as how to proceed with an analysis. We will present the current understanding, along with a brief outline of how we plan on proceeding with our analysis.

[1] We will refrain from a formal definition until Chapter (9)

Chapter 7

Region I - the First Bifurcation

In section (6.1), region I was introduced as the region along a one dimensional interval in parameter space between $s = 0$ up to and including the s value at the first bifurcation. More specifically, region I consists of the dynamics along the interval $[0, s_{fb}]$ where s_{fb} is the location of the first bifurcation from a fixed point along the interval in parameter space. The dynamics for $s > s_{fb}$ are included in regions II-V. Many of the results in this section have been published in (8).

This chapter will be organized as follows: we will begin with a discussion of random matrix theory, the primary driving force behind our intuition and a key ingredient for our arguments; we will formulate a conjecture motivated by the random matrix theory results regarding the most probable first bifurcation from a fixed point in our set of maps; the conjecture will be followed by some numerical results based upon a very large statistical study that supports our conjecture; we will conclude this chapter with a brief list of future directions we wish to continue with this work.

7.1 Conjectures

We will begin the formulation of our conjecture regarding the first bifurcation of a dynamical system from a fixed point with a discussion of the theoretical foundation upon which it was based, random matrix theory. We will follow this with the formulation of our conjecture and a simple corollary applying the random matrix results to some dynamical systems.

7.1.1 Random Matrix Theory

Our discussion of random matrix theory will be limited to the circular law of Girko (44), Bai (14), and Edelman (39). In general, circular laws in random matrix theory relate the distributions of elements of a random matrix to the distribution of those matrices' eigenvalues on a disk, usually centered at the origin, in complex plane. We will begin by discussing the circular law outright and follow this with a discussion of the expected value of real eigenvalues of a random matrix, and various related results. In both of the sections that follow, all of our matrices will be $n \times n$ matrices with real elements drawn from a random distribution yet to be specified.

7.1.1.1 The Circular Law

The study of the circular law has a long, somewhat colorful, and debated history. In the early 50's it was conjectured that the empirical spectral distribution (i.e. the distribution of eigenvalues) of $n \times n$ matrices with independent and identically distributed elements that were normalized by $\frac{1}{\sqrt{n}}$ converged to a uniform distribution on the unit disk in the complex plane. This is what is refered to as the circular law. Ginbre (43) proved this conjecture in the case where the random matrix is complex and has elements whose real and imaginary parts are independent and normally distributed (i.e. the real and imaginary parts are independent normals) in 1965. V.I. Girko published, in 1984, 1994, and again in 1997, papers proving the circular law for real, random, Gaussian, matrices. Girko's circular law states that as $n \to \infty$, the distribution of $\frac{\lambda}{\sqrt{n}}$ tends to uniformity on the unit disk. This result, which implies that as $n \to \infty$, the probability of an eigenvalue being real must go to zero, is a key ingredient towards showing that local bifurcations from fixed points due to purely real eigenvalues will be unlikely. It is this result that limits the kinds of generic local bifurcations from fixed points we can observe in the infinite dimensional limit. One particularly unfortunate problem with Girko's measure (as well as Edelman's and Bai's) is that it is not absolutely continuous with respect to Lebesgue measure with increasing but finite dimension (the infinite dimensional limit is however absolutely continuous). In particular, the probability of an eigenvalue being real with respect to Edelman is higher than one might expect for finite dimensions. Thus convergence in distribution to uniformity on the unit disk becomes an issue. Lucky for us, Edelman (40) derived a formula for the expectation value of real eigenvalues in Girko's measure in finite dimensions (we will discuss that result in section 7.1.1.2) which we will discuss since it is very important with respect to our results. In the process of deriving this expection formula, Edelman also proved Girko's result. In 1997, Bai (14) provided an alternate proof of the circular law for real random matrices with a significantly weaker hypotheses than either Edelman or Girko. Bai's result requires that the elements of the matrix be from a distribution with only a finite sixth moment. We will state Bai's result here since is it is both the simplest, and most general.

Theorem 8 (Circular law (14) (page 496)) *Suppose that the entries of a $n \times n$ matrix M have finite sixth moment and that the joint distribution of the real and imaginary part of the entries has a bounded density. Then, with probability 1, the empirical distribution $\mu_n(x, y)$ tends to the uniform distribution over the unit disk in two-dimensional space.*

7.1.1.2 Expected Value of Real Eigenvalues and Related Results

The circular law will provide the intuition for the conjecture we will state shortly. The circular law is difficult to use for our purposes, and provides little practical understanding of how the distribution of eigenvalues evolves and converges to uniformity as the dimension of the matrix is increased. Luckily, Edelman essentially evaluated the integral formula of Girko (in spirit at least) and arrived at a formula for the expected number of real and complex eigenvalues as a function of the dimension of the matrix. Edelman has proved the following results which will be useful and relevant for our work: a formula for the density of real eigenvalues in the complex plane as a function of the dimension of the matrix; a formula for the density of non-real eigenvalues on the complex plane as a function of the dimension of the matrix; a formula for the expectation value of real eigenvalues of a matrix as a function of the dimension of that matrix; and a theorem that states that the real eigenvalues converge in distribution to that of a uniform random variable on $[-1, 1]$ in the limit of an infinite-dimensional matrix. For completeness, we will reproduce the aforementioned results, noting that all the statements that follow are relevant for matrices such that $a_{ij} \in A \in R^{n^2}$ where $a_{ij} \in N(0, 1)$.

We will begin with two definitions.

Definition 18 (True Real Eigenvalue Density (40)) *Assume a real eigenvalue, λ, of a fixed, real, $n \times n$ matrix A. The true density of real eigenvalues, or the expected number of real eigenvalues per unit length can be defined:*

$$\rho_n(\lambda) = (\frac{1}{\sqrt{2\pi}}[\frac{\Gamma(n-1, \lambda^2)}{\Gamma(n-1)}] + \frac{|\lambda^{n-1}|e^{-\frac{\lambda^2}{2}}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}[\frac{\Gamma(\frac{(n-1)}{2}, \frac{\lambda^2}{2})}{\Gamma(\frac{(n-1)}{2})}]) \tag{7.1}$$

or, in a different light:

$$\rho_n(x) = \frac{d}{dx} E_A \#_{(-\infty,x)}(A) \tag{7.2}$$

where $\#_{(-\infty,x)}(A) \equiv$ number of real eigenvalues of $A \leq x$, E_A denotes the expectation value for a random A and Γ is the standard gamma function.

Definition 19 (Probability Density of $\lambda_n \in R$ (40)) *Assume a real eigenvalue λ_n of an $n \times n$ random matrix, then it's probability density, $f_n(\lambda)$ is given by:*

$$f_n(\lambda) = \frac{1}{E_n}(\frac{1}{\sqrt{2\pi}}[\frac{\Gamma(n-1,\lambda^2)}{\Gamma(n-1)}] + \frac{|\lambda^{n-1}|e^{-\frac{\lambda^2}{2}}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}[\frac{\gamma(\frac{(n-1)}{2},\frac{\lambda^2}{2})}{\Gamma(\frac{(n-1)}{2})}]) \tag{7.3}$$

or more simply:

$$f_n(\lambda) = \frac{1}{E_n}\rho_n(\lambda) \tag{7.4}$$

where E_n denotes the expected number of real eigenvalues of the $n \times n$ random matrix.

Integrating ρ_n along the real line provides the expected number of real eigenvalues. Edelman provides several formulas from such a calculation, the simplest being summarized by the following corollary:

Corollary 2 (Corollary 5.2 (40)) *We have the aysmptotic series:*

$$E_n = \sqrt{\frac{2n}{\pi}}(1 - \frac{3}{8n} - \frac{3}{128n} + \frac{27}{1024n^2} + \frac{499}{32768n^4} + O(\frac{1}{n^5})) \tag{7.5}$$

as $n \to \infty$.

Again, E_n is the expected number for a real, $n \times n$ random matrix. The manner in which the convergence in measure is not absolutely continuous (with respect to Lebesgue measure) is highly relevant to our results because the issue lies with a high then expected density of real eigenvalues. For a full discussion, see (40), however, this is not a pathological problem since the distribution of real eigenvalues on the real line is uniform. Moreover, for our purposes, the expected value of the real eigenvalues is not enough since, if all the real eigenvalues are located at ± 1, then clearly there will exist many local bifurcations from fixed points due to purely real eigenvalues — and these bifurcations will likely be of high codimension. That such is not the case, is given in the following corollary:

Corollary 3 (Corollary 4.5 (40)) *If λ_n denotes a real eigenvalue of an $n \times n$ random matrix, then as $n \to \infty$, the normalized eigenvalue $\frac{\lambda_n}{\sqrt{n}}$ converges in distribution to a random variable uniformly distributed on the interval $[-1, 1]$*

Besides the results regarding the real eigenvalues, Edelman also provides information regarding the density of non-real eigenvalues:

Theorem 9 (Density of Non-Real Eigenvalues: Theorem 6.2 (39)) *The density of a random comples eigenvalue of a normally distributed matris is:*

$$\rho_n(x,y) = \sqrt{\frac{2}{\pi}} y e^{y^2 - x^2} erfc(y\sqrt{2}) e_{n-2}(x^2 + y^2) \tag{7.6}$$

where $e_n(z) = \sum_{k=0}^{n} \frac{z^k}{k!}$ and $erfc(z) = 2/\pi \int_z^\infty exp - t^2)dt$, the complementary error function. . Integrating this over the upper half plane gives the number of non-real eigenvalues.

All of these results can be nicely concluded with the following two theorems regarding the circular law.

Theorem 10 (Theorem 6.3 (39)) *The density function $\hat{\rho}$ converges pointwise to a very simple form as $n \to \infty$:*

$$\lim_{n\to\infty} \frac{1}{n}\hat{\rho}(\hat{x},\hat{y}) = \begin{cases} \frac{1}{\pi} & \hat{x}^2 + \hat{y^2} < 1 \\ 0 & \hat{x}^2 + \hat{y^2} > 1 \end{cases} \tag{7.7}$$

where $\hat{\rho}_n$ is simply ρ as a function of $\hat{x} = \frac{x}{\sqrt{n}}$ and $\hat{y} = \frac{y}{\sqrt{n}}$. Note that $\frac{\hat{\rho}(\hat{x},\hat{y})}{n}$ is a randomly chosen normalized eigenvalue in the upper half plane.

Finally, Edelman's version of the circular law can be proved using theorem 10 and a central limit theorem.

Theorem 11 (Circular Law: Convergence in distribution (39)) *Let z denote a random eigenvalue of A chosen with probability $\frac{1}{n}$ and normalized by dividing by $\sqrt{n}$. As $n \to \infty$, z converges in distribution to the uniform distribution on the disk $|z| < 1$. Furthermore, as $n \to \infty$, each eigenvalue is almost surely non-real.*

7.1.2 A conjecture regarding the first bifurcation from a fixed point

There are three generic, codimension one, local bifurcations from a fixed point in maps of dimension two or greater (70) (26). These three bifurcations depend on symmetries of the dynamical system, but generally they consist of: the flip bifurcation, corresponding to the largest eigenvalue being -1; the fold, corresponding to the largest eigenvalue being 1; and the Neimark-Sacker (81) (115), corresponding to a complex conjugate pair of eigenvalues with non-zero real part having modulus one. Edelman, Girko, and Bai have all shown that in the infinite dimensional limit, a real matrix with elements selected from a real, Gaussian distribution, the normalized eigenvalues will be distributed uniformly on the unit disk in the complex plane. Since the Neimark-Sacker bifurcation corresponds to the bifurcation via a complex conjugate pair of eigenvalues, a logical application of the circular law is to infinite dimensional dynamical systems whose Jacobian matrix has elements whose distribution has a finite sixth moment. In this circumstance, the probability one bifurcation would seem to be a Neimark-Sacker bifurcation. If the real eigenvalues pile up near 1 and -1, we will run into problems, but Edelman (corollary (3)) has shown that this circumstance will not occur. Instead, the real eigenvalues will be distributed uniformly on the real axis. In the standard bifurcation sequence constructions, one would be concerned with a parameterized curve of matrices. In such a scenario the matrices would not be independent along the curve in general. Surmounting this obstacle is yet an open problem, however, in some special cases like in the case where the parameterized curve is linear, the difficulties are greatly reduced. Thus we can make the following statement:

Corollary 4 (First bifurcation probability) *Given the dynamical system F*

$$F(x_{t-1}) = x_t = \epsilon A x_{t-1} + \epsilon G(x_{t-1}) \tag{7.8}$$

where $x_t \in R^n$, $\epsilon \in R$, $A \in R^{n^2}$ where $a_{ij} \in N(0,1)$, and where $G(x_{t-1})$ is a nonlinear, C^r $(r > 0)$, function of x_{t-1} which is of order 2 or higher. Thus $F(x_{t-1}) \cong \epsilon A x_{t-1}$ for ϵ small. Assume F has a fixed point at $\epsilon = 0$ and upon the increase of ϵ, F undergoes a local, codimension one bifurcation. As the dimension of the dynamical system F goes to infinity (i.e. given $A \in R^{n^2}$, $n \to \infty$), the probability that the first bifurcation will be of type Neimark-Sacker will converge to one.

Proof: This result follows trivially from the results of Edelman (39), (40), Girko (44) (46) (48) (47) (45), and Bai (14).

We can, with a little work, impose a measure on the set of dynamical systems for which this result holds via results of Edelman, the neural networks, and some standard arguments using measure theory. Upon doing so, one nontrivial issue is understanding what such a set of dynamical system would "look" like. We will refrain from further discussion of this extension here. It is likely that we can extend this result such that the elements of the A matrix can be selected from any distribution with a finite sixth moment in line with the circular law of Bai (14).

Corollary (4) falls far short of satisfying our desires, however. First, corollary (4) does not speak to the probability of it's hypothesis being satisfied in a general space of dynamical systems. Again, we can construct a measure such that the relevant hypothesis are always satisfied, but how general such a measure is, which is the same problem, is still a problematic issue. Moreover, corollary (4) is not cast in the general parameterized curves of the bifurcation theory framework we desire — linear "curves" are vary limiting. See (123) or (128) for a construction of the general bifurcation framework we reference here.

Providing an answer to such a question would likely require establishing some type of measure on the space of general C^r dynamical systems, and establishing such a measure is often very difficult to say the least. If such a measure could be defined, some meaningful notion of equivalence in measure would need to be shown. This issue however, would likely be the least of the problems. Moreover, corollary (4) does not provide any information regarding how large but finite dimensional dynamical systems that satisfy the all the hypotheses aside from the infinite dimensionality behave. Lastly, corollary (4) does not provide any insight into how the convergence to such a result might occur as the dimension of the dynamical system is increased. Thus, we will present a conjecture that we believe captures more of what we wish corollary would capture.

Conjecture 3 (Genericity of Neimark-Sacker bifurcations in high-dimensional dynamical systems) *Begin with the space of C^r $(0 < r)$ dynamical systems with bounded first derivatives on compact sets with a single, real parameter and such that there exists a fixed point on a measurable interval of the parameter space, and at least one local bifurcation upon a continuous variation of the parameter. There exists a probability measure on the such that, as the dimension of the dynamical system is increased, the probability of the bifurcation from fixed points via the Neimark-Sacker bifurcation will increase, and approach probability one as $d \to \infty$.*

Since this is a conjecture, the remaining portion of this chapter will be dedicated to a presentation of numerical evidence supporting the above conjecture in our space of neural networks.

7.2 Numerical Results

We will present our numerical results in two steps. First, we will present four typical cases with different numbers of neurons and dimensions. This will be followed by a set of figures that demonstrate how, as the dimension of our neural networks is increased, the probability of the flip and fold bifurcations are nearly equal and decrease with increasing dimension and the Neimark-Sacker bifurcation becomes by far the most probable first bifurcation.

7.2.1 The First Bifurcation Qualitatively

Figures (7.1)-(7.4) are bifurcation diagrams and corresponding largest Lyapunov exponent plots for various values of N and d. In these figures at each value of s, the first $120,000$ y_t values are discarded and the next 128 are plotted. The initial conditions and weights are the same for each s value in each figure. Note that the meaning of typical or characteristic for one set of N and d values can be very different from that of another set of N and d values. For instance, there is more apparent variability in the dynamics and the bifurcation sequence at low d than at high d. At high d most of the diagrams look similar, whereas at low d the diagrams are erratic and differ in detail.

From these plots it is possible to distinguish super-critical fold, flip, and Neimark-Sacker bifurcations. We rarely observe sub-critical bifurcations and many would be evident considering the bifurcation diagrams and searching for a discontinuity at the bifurcation point in the largest Lyapunov exponent. Although we cannot rule out the existence of sub-critical bifurcations, they will always result in bounded dynamics since our neural networks cannot become unbounded. Chaos can be clearly differentiated from limit cycles and periodic orbits since a positive largest Lyapunov exponent differentiates chaos from periodic or quasi-periodic dynamics.

Figure (7.1) is a bifurcation diagram for $N = 4$ and $d = 4$. In this figure the first bifurcation happens to be that of a Neimark-Sacker. Figure (7.2) is a bifurcation diagram for $N = 64$, $d = 4$, and shows a typical fold first bifurcation. Figure (7.3), $N = 4$, $d = 64$, shows a Neimark-Sacker first bifurcation, as does Fig. (7.4), ($N = 64$, $d = 64$).

Notice that as N is increased, the range of x_t values increases. This is due to the method used to choose the β's. In general, as the dimension increases, the bifurcation diagrams are more centered and symmetric about the origin, almost as if there were no bias term (the bias terms break the symmetry about the origin). This is because, as the number of w_{ij}'s is increased, the importance of the individual weights is decreased, thus the importance of each bias term is decreased. This does not effect the results since there is no dynamical

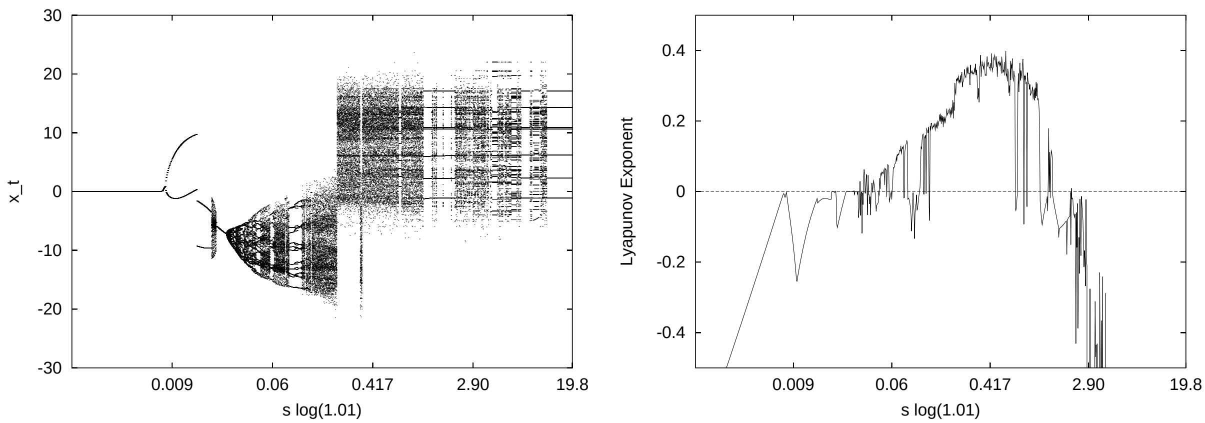

Figure 7.1: Bifurcation diagram with the largest Lyapunov exponent for $N = 64$ and $d = 4$.

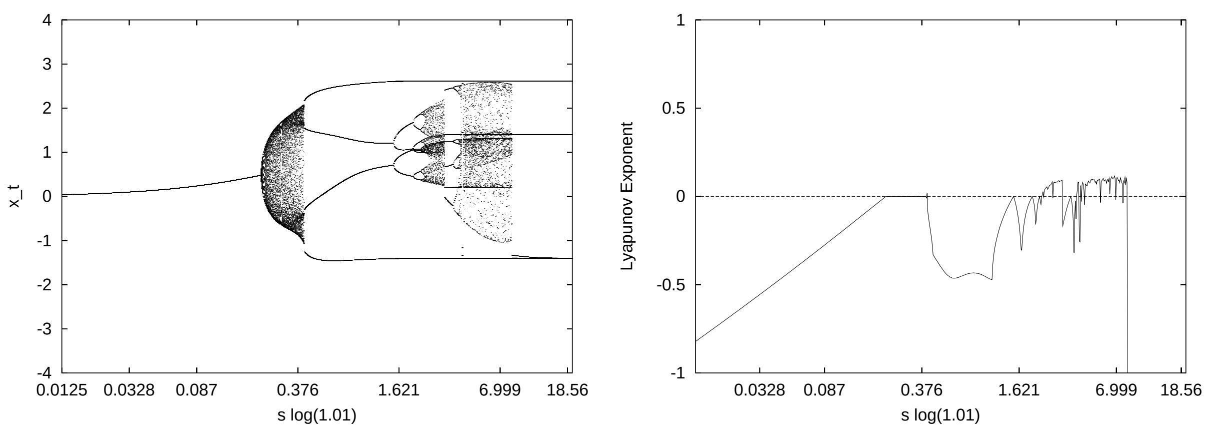

Figure 7.2: Bifurcation diagram with the largest Lyapunov exponent for $N = 4$ and $d = 4$.

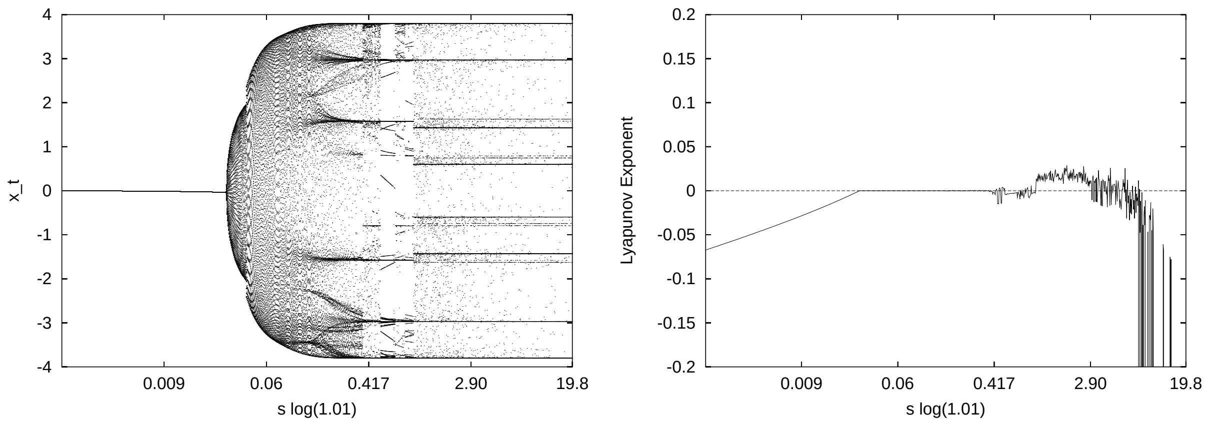

Figure 7.3: Bifurcation diagram with the largest Lyapunov exponent for $N = 4$ and $d = 64$.

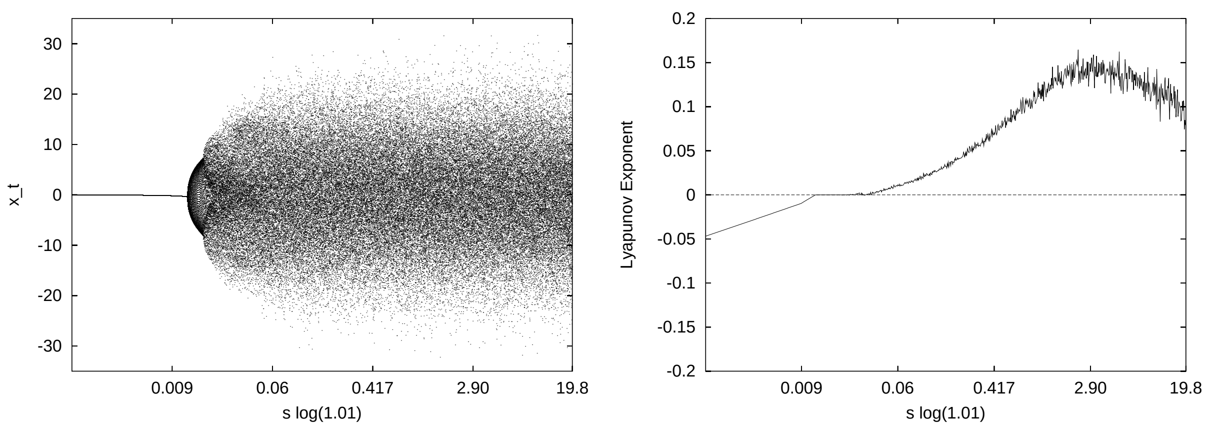

Figure 7.4: Bifurcation diagram with the largest Lyapunov exponent for $N = 64$ and $d = 64$.

difference between networks with and without bias terms aside from affecting the fold bifurcation which is replaced by a pitchfork bifurcation in the case where the network is symmetric about the origin. It is worth noting that changing the mean of the w_{ij}'s from zero to some small, non-zero value destroys the bifurcation sequence, and in general, destroys all but periodic dynamics. The non-zero mean of the w_{ij}'s that destroys the existence of the first bifurcation (and all the other non-periodic dynamics) depends on both N and d, but in general, the larger d is, the smaller the non-zero mean has to be to destroy the non-periodic dynamics.

7.2.2 First Bifurcation Probabilities

We will now present results from a statistical study regarding the first bifurcation from a fixed point of our networks. For each case, we pick and fix the weights and initial conditions, run the case for an s value, calculate the eigenvalues and largest Lyapunov exponent and then increase s by a constant multiple (usually close to one), re-initialize with the same weights and initial conditions, and repeat the process. When the modulus of the largest eigenvalue reaches the unit circle we decide what kind of bifurcation has occurred and move on to the next set of weights and initial conditions. This process was done over N and d values both ranging from 2 to 256.

We begin each network with an s value small enough such that the dynamics are that of a stable fixed

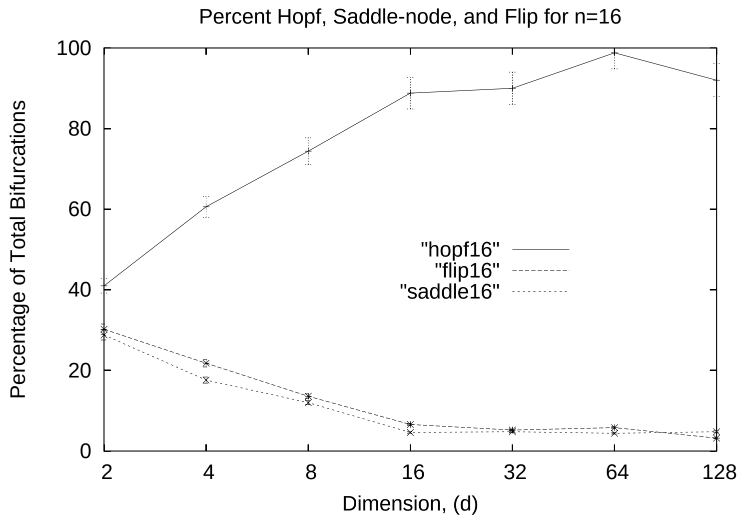

Figure 7.5: Percent first bifurcation for $N = 16$, error bars represent the error in the probability.

point. All cases we observed had a first bifurcation, but not all cases were chaotic over some portion of the parameter interval observed.

Figure 7.5 shows the percentage of each bifurcation as the dimension is increased for an intermediate number of neurons, $N = 16$. Much like the prediction in conjecture (3), the percentage of Neimark-Sacker bifurcations start at about 40 percent for the first bifurcations at $d = 2$ and increase above 90 percent at $d = 64$. Also, notice that the percentage of flips and folds are, on average, equal throughout the range.

Figures (7.6) and (7.7) show the percentage of first bifurcations over an increasing range of d at low and high N, ($N = 4$ and $N = 256$). Note at high N and low d, the percentage of each bifurcation is nearly equal. For the low-N, low d cases, the percent of each bifurcation is not nearly as close as at high N. As d is increased, the percentage of Neimark-Sacker bifurcations rapidly increases so that, at $d = 8$, the percentages of each bifurcation is about equal to those of all N. This N dependence is an artifact of how we choose the β vector. Consider a two-dimensional system. For a Neimark-Sacker bifurcation to occur, the discriminant must be negative. As we increase N, we are increasing the variance of the coefficients of the matrix, thus pushing the expected value of the discriminant positive. The result is a decrease in Neimark-Sacker bifurcations as N is increased at low d. As d is increased, this effect becomes negligible. Contrasting figures (7.5) and (7.6) you will notice that at $d = 2$ the percentages of first bifurcations are quite different, but for $d \geq 8$, figures (7.5) and (7.6) are almost identical.

In the high-d limit, we found that the Neimark-Sacker bifurcation was the dominant first bifurcation. We looked at cases with d as high as 1024 and found the percent of first Neimark-Sacker bifrucations approached unity. This confirms a similar result of Doyon *et. al.* (36), that in the limit of high d, the first bifurcation will be Neimark-Sacker.

7.3 Future directions

There are many problems we wish to address in the future, beginning with a better formulation and proof of conjecture (3). Such a step has, to the present, seemed difficult, but there are many intermediate and highly relevant computational questions remaining.

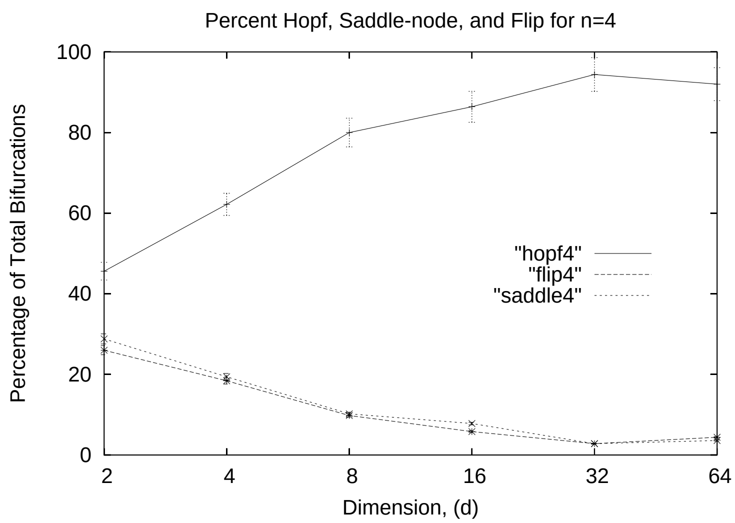

Figure 7.6: Percent first bifurcation for $N = 4$, error bars represent the error in the probability.

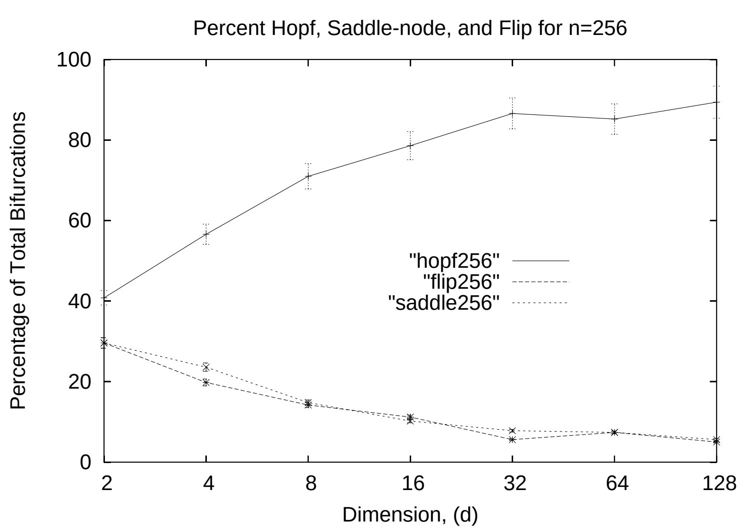

Figure 7.7: Percent first bifurcation for $N = 256$, error bars represent the error in the probability.

Numerically, an analysis of the distribution of the elements of the Jacobians of our networks at the first bifurcation could be much more thorough, and an understanding could be more complete. Despite the fact that we calculate the full spectrum of eigenvalues and Lyapunov exponents, we have really only considered the largest eigenvalue(s). An analysis of the first six moments of the distribution of eigenvalues would likely lead to much insight. Also, statistical tests such as the K-S statistic would be useful for comparing the required theoretical hypothesizes with the distributions observed numerically. A brute force analysis comparing the density of eigenvalues in our networks at the bifurcation points with Edelman's density formulas for both real and complex eigenvalues would also be of considerable interest and value. A comparison with Edelman's formula for the expected number of real eigenvalues as the dimension of the dynamical system is increased would help provide an understanding of how well the random matrix results might reflect in the first bifurcations in some sets of dynamical systems. Finally, a rigorous comparison of the curves observed in figures (7.5)-(7.7) with what might be expected based upon Edelman's results would be highly valuable.

Chapter 8

Region II - between the First Bifurcation and Chaos

Region II is the region along the interval in parameter space between the first bifurcation and the appearance of a positive Lyapunov exponent. The arguments and results in this section consist of, and are based on, intuition from the first bifurcation, numerical investigation, and a potpourri of results of other researchers regarding the route to turbulence and general bifurcations of limit cycles and tori. It is in this region that the routes to chaos will be observed and analyzed.

8.1 Conjectures

Based on random matrix theory, the construction of our neural network framework, and intuition based on some numerical experiments which we will present in the section (8.2), we have some conjectures regarding region II and the route to chaos in our set of mappings.

Extending corollary (4) or any of the results or claims from section (7.1.2) analytically for the k^{th} bifurcation before the onset of chaos where $k > 1$ is made difficult by the fact that the normal form and center manifold theory for quasi-periodic bifurcation theory is a long way from providing a general form(s) about which the results of the random matrix theory could be applied (the codimension 2 situation is not complete yet, the codimension 3 case is even further from completion). However, it is assumed that, in the end, most bifurcations of periodic and quasi-periodic orbits can be captured by some sort of Taylor series expansion (via a vector field approximation or a suspension). Though the linear term of the Taylor expansion will be degenerate and the outcome of the bifurcation will be determined by contributions of higher-order terms, the degeneracies in the linear term of the Taylor polynomial will nevertheless be of complex eigenvalues — leading to some sort of bifurcation (yet to be understood) from quasi-periodic or periodic orbits to other quasi-periodic or periodic orbits.

Once a discrete-time map f has undergone a supercritical Neimark-Sacker bifurcation, one of three types of dynamics will ensue:

i. f will have a quasi-periodic orbit;

ii. f will be periodic with period ≥ 5;

iii. f will have strong resonance and thus have periods of 3 or 4;

To create a bifurcation theory regarding the above three options, some sort of linearization must be made. If the dynamics are as per item (i), one must begin by making a suspension of the map and then take a vector field approximation, and finally take a Poincaré section to construct the linearization (see (10) (13) (20), (23), (12), (70) or (84) for more information regarding these methods). If the dynamics are per items (ii)-(iii) with period n, then the time-n map can be made from which a Jacobian can easily be constructed. In the best of scenarios, connecting the random matrix results requires constructing results after many matrix multiplications. At worst, one must be able to carry the distributions of elements through normal form calculations. We will not attempt any of these tasks at this time, but instead we will note that the requirement for a circular law is as weak as the elements having a finite sixth moment, and put forth a conjecture that we will attack numerically.

Conjecture 4 (Most probable route to chaos) *Given a high dimensional, C^r dynamical system, if a Jacobian can be constructed, and all the eigenvalues remain inside the unit circle on the complex plane, upon parameter variation, if an eigenvalue(s) (complex conjugate pair) reaches the unit circle, that eigenvalue(s) will be complex with probability approaching unity as the dimension approaches infinity. In other words, the most probable route to chaos from a fixed point for high-dimensional dynamical system is via a cascade of (supercritical) bifurcations of periodic, quasi-periodic orbits and tori, and the probability of this route increases with dimension.*

By high dimensional, we mean $d > 100$, but we are being purposefully vague. The convergence to our claim will likely follow some Edelman-like expected value formula, eventually converging to a circular law a la Bai or Girko. The basic point is that, after the first bifurcation of a high-dimensional system has occurred, if the dynamical system is not chaotic, successive bifurcations will not be bifurcations due to real eigenvalues, but rather bifurcations due to complex conjugate pairs of eigenvalues[1].

One strength of our methods and results is, unlike topologically based results, our neural network framework allows not only a practical means of analyzing topological results, but it contains, in a manageable way, the supposed pathological examples. Due to the mappings that neural networks can approximate, if the spectrum of Lyapunov exponents of a d-dimensional network is contained in $[\chi_1, \chi_d]$, then likely there exists at least one path through parameter space such that any network can be transformed into a T^d torus with all Lyapunov exponents being zero (we haven't proved such a result, however). If one were to stratify the networks by their spectra, the aforementioned torus would be but a point along the interval $[\chi_1, \chi_d]$, and thus, in this sense, unusual. There are other stratifications of the networks that can be made, such as the map $\phi : R^{n(d+2)} \rightarrow \Sigma(\tanh())$, where $\Sigma(\tanh())$ is the set of neural networks with tanh() as the squashing function. The point is, we are presenting a practical framework that yields numerical observations regarding common routes to chaos. For example, the number of constraints required to achieve a d-torus might be high[2], and the T^d torus would likely exist on a surface in parameter space of much lower dimension than the ambient parameter space. However, since the neural networks can approximate m-tori for $m \leq d$, it is possible to study the transitions, and the likelihood of such transitions, and the persistence of high-dimensional tori in a practical way. From our experience perturbing away 2-tori and limit cycles requires drastic parameter variation, however, we never observe 3-tori consistent with the prediction of Newhouse, Ruelle and Takens.

A full statistical study is hampered by numerical accuracy issues which may be fundamental problems associated with the existence of neutral directions. We will present the most clear picture of the prototypical route to chaos in high dimensions for our set of functions via a prototypical example. For the understanding reached in this report, we will employ bifurcation diagrams, phase-space plots, the largest Lyapunov exponent, and the full Lyapunov spectrum. Numerical issues related to the Lyapunov spectrum are the main problem with respect to the statistical study, and we will discuss and display such problems. For information regarding why the numerical problems (such as truly infinite convergence times) we encounter may be unavoidable see (106) and (27) for the latest with regard to the existence of zero Lyapunov exponents and issues with computing them. At the end of the Chapter we will note the numerous extensions our current line of study suggests.

[1] Chenciner and Iooss (29) proved bifurcations of flows with two frequencies to flows with three frequencies is non-generic in a certain sense, we claim that in our systems, the observed secondary bifurcations will be of this non-generic type.

[2] The number of constraints for a d-torus might also be low, as requiring area preservation, or a local Jacobian to have an average determinant equal to ± 1 is, in a sense, a singe constraint.

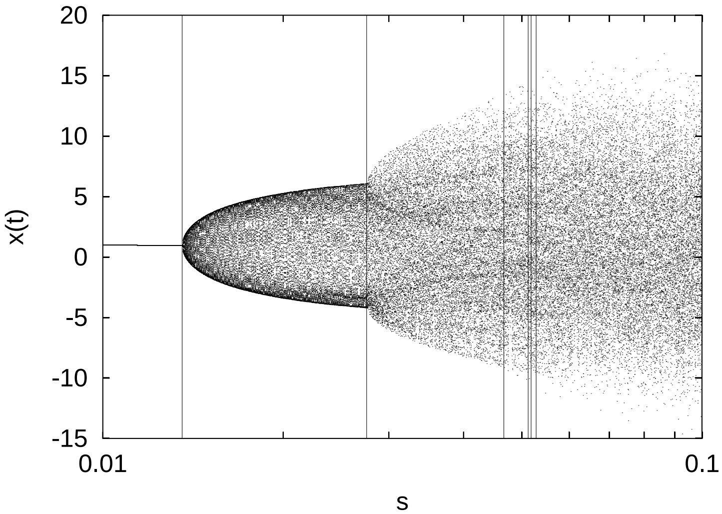

Figure 8.1: Bifurcation diagram for a typical network; $n = 32$, $d = 64$.

8.2 Numerical Results

A statistical study for this region of parameter space is yet lacking. Nevertheless, we have observed a rather sizeable number of neural networks in this region (on the order of hundreds over a large number of neurons and dimensions). Thus, we will present a typical example in hopes of painting a picture of the common transitions to chaos and providing the intuition and justification for our claims in the previous section.

The example we will present is typical (we will discuss what is not typical about our example during the analysis) amongst the 500 or so cases (with this number of parameters and dimensions) we have observed in the sense that between the first bifurcation and the onset of chaos, the only type of orbit that exists is either quasi-periodic or periodic orbits with periods high enough such that they are indiscernible from quasi-periodic orbits. It is worth noting that the route to chaos we discuss in this report, which we believe is the typical route in high dimensions, is considerably different from what is observed at low dimensions. For an intuitive feel for the lower-dimensional cases, see (8).

A primary tool of analysis will be the Lyapunov exponent spectrum ((18) (112)), since it is a good measure of the tangent space of the mapping along its orbit. Negative Lyapunov exponents correspond to global stable manifolds or contracting directions; positive Lyapunov exponents correspond to global unstable manifolds or expanding directions and are ((112)), in a computational framework, the hallmark of chaos. A zero exponent corresponds to a neutral direction (although the story in this case is significantly more complicated, see (27)); when the largest Lyapunov exponent for a discrete-time map is zero and all other exponents are less than zero, then there exists a neutral direction. Lyapunov exponents relate to quasi-periodic orbits and rotations in the following way, neutral rotating directions correspond to a zero Lyapunov exponent, and high-period orbits correspond to pairs of negative Lyapunov exponents. If the largest exponent is zero (while all others are negative), then there exists a quasi-periodic orbit on the circle (on T^2 in the flow). If two exponents are zero, then there is a full 2-torus in discrete-time (a 3-torus in the flow), and so on (for specific examples of high-dimensional tori see (23)). With respect to the codimension 2 bifurcations, there can also be as many as 10 possibilities for zero Lyapunov exponents at bifurcation points (again see page 397 of (70)), some of which correspond to real and complex conjugate pairs of eigenvalues. Besides the full Lyapunov spectrum algorithm, we will also use an independent version of a calculation of the largest Lyapunov exponent a la Wolf et. al (130). Upon analysis of the example, it will be obvious why we need such a calculation.

To circumvent the numerical stability issues that will be apparent shortly, we are required to consider each network individually with four types of figures: the largest Lyapunov exponent independently computed

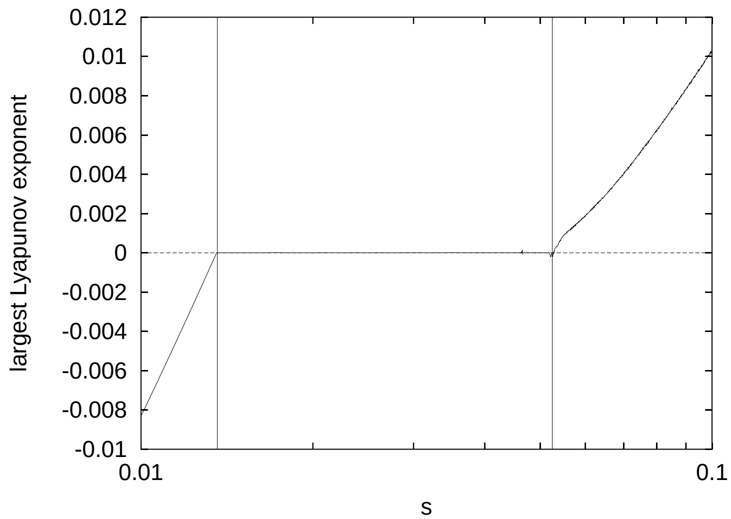

Figure 8.2: The largest Lyapunov exponent for a typical network; $n = 32$, $d = 64$.

of the full spectrum (it is considerably more numerically stable than the algorithm for the full spectrum); a standard bifurcation diagram; the full Lyapunov spectrum computed in the standard manner; and phase-space diagrams.

Our choice of the number of neurons and the number of dimensions is based on Fig. (1) of (8) and the compromise required by computational time limits. Considering Fig. (1) of (8), 32 neurons puts our networks deeply in the region of the set of neural networks that correspond to highly complicated and chaotic dynamics. The dimension, 64 was chosen because it was the highest dimension for which we could reliably compute the nearly 500 cases we surveyed. The compute time increases as a power of the dimension. Thus as this time, we had too few cases of $d = 128$ and $d = 256$ to make conclusive statements.

Beginning with Fig. (8.1), the standard bifurcation diagram, there are four important features to notice. The first feature is the first bifurcation, which occurs at $s = 0.0135$ from a fixed point to some type of limit cycle or torus. A secondary bifurcation is clearly visible at $s = 0.02755$, the nature of this bifurcation is entirely a mystery from the perspective of Fig. (8.1). Chaos seems to onset near $s = 0.05$, and has definitely onset by $s = 0.06$, however the exact location is difficult to discern. Lastly, all of the dynamics between the fixed point and chaos are some sort of n-torus ($n \geq 0$) type behavior.

Next, let us consider Fig. (8.2) — the largest Lyapunov exponent versus variation in s for the same case as Fig. (8.1). Again, as expected, we see the first bifurcation at $s = 0.0135$, in agreement with the bifurcation diagram. Figure (8.2), however, gives a clear picture of the onset of chaos, which occurs at $s = 0.05284$. The largest exponent is near zero between the first bifurcation and the onset of chaos, providing evidence for the existence of at least one persistent complex conjugate pair of eigenvalues with modulus one (assuming a Jacobian can be constructed) — e.g. a persistent quasi-periodic orbit.

Considering Fig. (8.2), near the onset of chaos the exponent becomes negative over a very short s interval. Ignoring all the intermediate bifurcations, let us briefly consider the onset of chaos via Fig. (8.3). Considering this figure, there is an apparent periodic orbit, followed but what might be a period doubling bifurcation, followed by what could be a complicated bifurcation structure. Besides noting this for general interest and completeness, we will refrain from a further discussion of this small interval since this behavior seems to disappear for high-dimensional networks and is not particularly related to the point of this thesis.

We will begin our presentation of phase-space figures at the second bifurcation. The second bifurcation is the obvious in the bifurcation diagram, corresponding to the rapid change in the attractor size near $s \sim 0.0275$. The locations and nature of these bifurcations can't be determined by a consideration of the

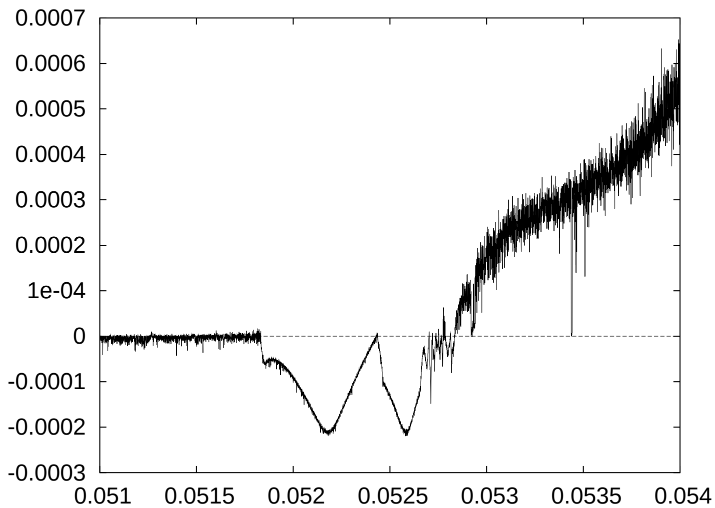

Figure 8.3: The Largest Lyapunov at the onset of chaos.

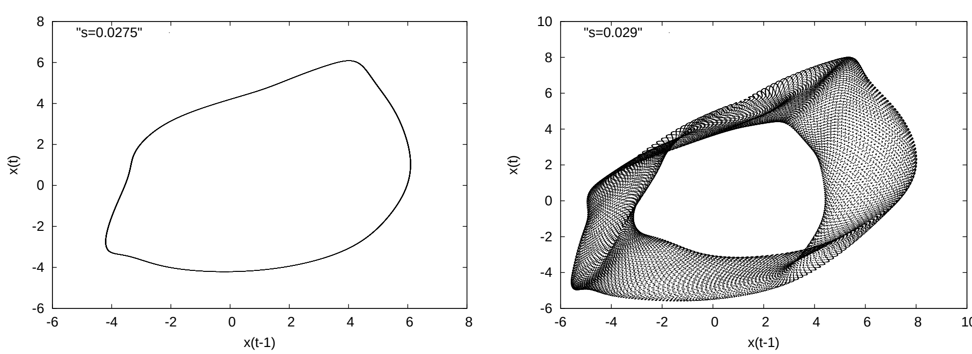

Figure 8.4: Phase plots on either side if the 2^{nd} bifurcation, $s = 0.0275$ and $s = 0.029$ respectively. The bifurcation occurs at $s \sim 0.02754$.

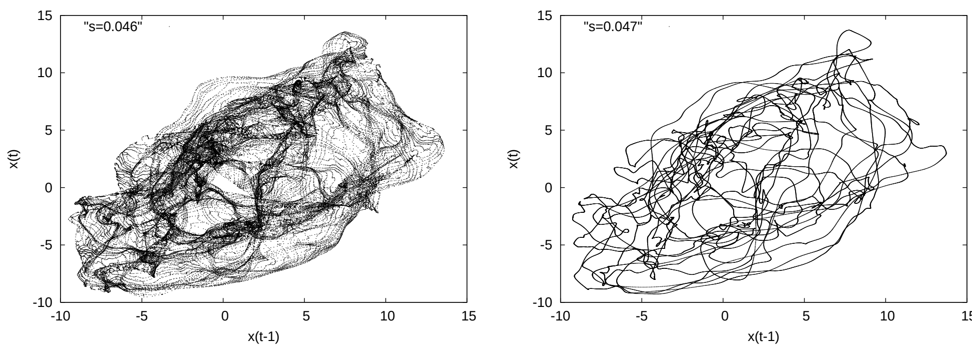

Figure 8.5: Phase plots on either side if the 3^{rd} bifurcation, $s = 0.046$ and $s = 0.047$ respectively. The bifurcation occurs at $s \sim 0.04667$

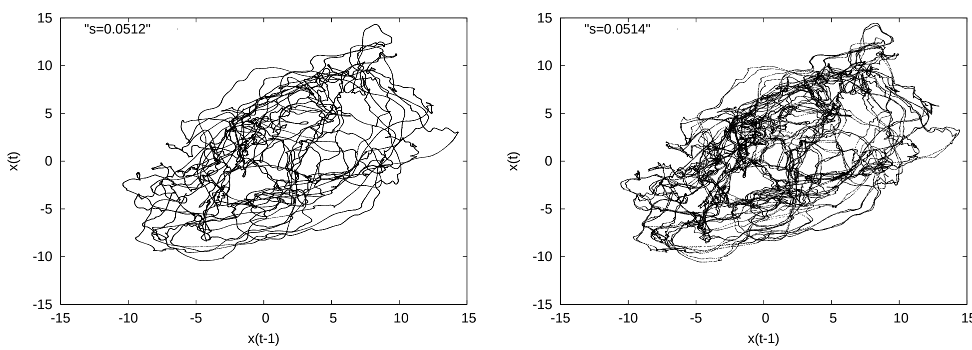

Figure 8.6: Phase plots on either side of the 4^{th} bifurcation, $s = 0.0512$ and $s = 0.0514$ respectively. The bifurcation occurs at $s \sim 0.05124$

Lyapunov spectrum or the largest Lyapunov exponent, although considering the Lyapunov spectrum it is clear that such a bifurcation does occur. The second bifurcation appears to be from a 1-torus to a 2-torus as shown in Fig. (8.4).

The phase-space plots on either side of the third bifurcation, which is not clearly apparent in the bifurcation diagram or in the Lyapunov spectrum as we will see, is depicted in Fig. (8.5). By this point, the smooth looking torus-like object from Fig. (8.4) has become "kinked" significantly before the third bifurcation. At the third bifurcation, the torus-like, two-dimensional object becomes a one-dimensional object. Thus the third bifurcation is from a 2-torus to a 1-torus and occurs at $s \sim 0.0466$. The 1-torus is a severely "kinked" quasi-periodic orbit.

Considering the phase-space plots on either side of the fourth bifurcation in Fig. (8.6), one might conclude that there our example has undergone a period doubling of the quasi-periodic orbit. An analytical explanation of such a bifurcation is yet an open problem, but it is likely a "Neimark-Sacker-Flip" bifurcation. We will refrain from a further discussion of this bifurcation, directing the interested reader to chapter 9 of (70).

We will refrain from showing figures for the fifth bifurcation, simply noting that it is a bifurcation from this quasi-periodic orbit on the 1-torus to a high-period, periodic orbit. Rather, we will skip to what we will call the sixth bifurcation. Figure (8.7) demonstrates the final bifurcation into chaos from a high-period, periodic orbit. However, in our particular example, considering Fig. (8.3), just before the onset of chaos, there is a likely a sequence of bifurcations. We will not belabor this further, as little insight is gained from a further consideration, and this sequence of bifurcations just before the onset of chaos appears to be

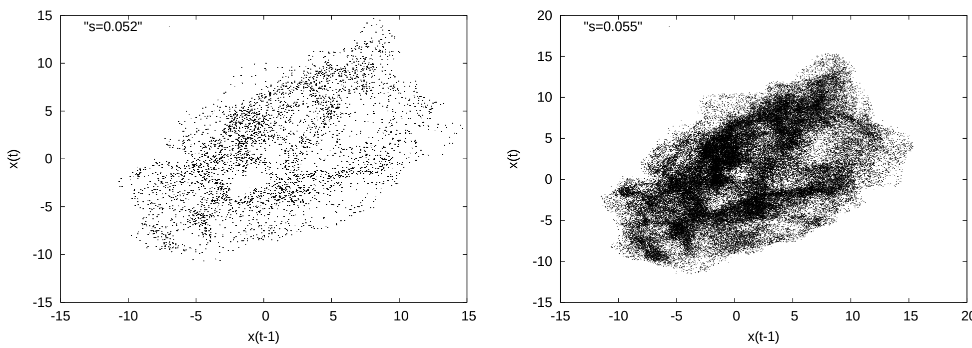

Figure 8.7: Phase space plot near the 6^{th} bifurcation — $s = 0.052$ and $s = 0.053$ respectively. The bifurcation occurs at $s \sim 0.05294$.

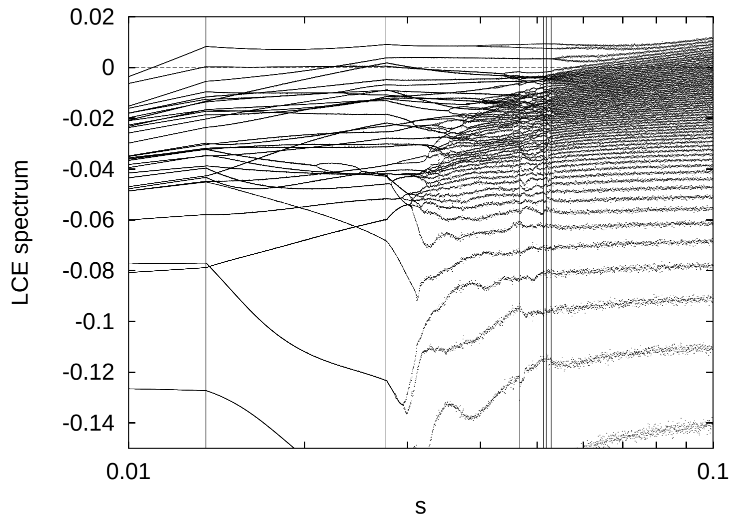

Figure 8.8: Lyapunov spectrum the (typical) network; $n = 32$, $d = 64$.

increasingly rare as the dimension increases.

Before discussing what can be gleaned from the Lyapunov spectrum data, we must make a few comments with respect to the data plotted in Fig. (8.8). If one were to pick an s value along the s interval $(0, 0.03)$ and count the number of Lyapunov exponents, it is likely that one could often only identify roughly half the number of exponents that might be expected. There is a simple reason for this, if one were to consider the eigenvalues of the variational differential equation,[3] nearly all the eigenvalues are complex conjugate pairs, and hence the pairs of nearly equal Lyapunov exponents. This in a sense, implies a lot of rotation, but as this can depend upon the coordinate system, little can be said in this regard. Nevertheless, from a practical point of view, this implies sink behavior in all directions. The point is that the "largest" exponent is actually pair of nearly equal exponents. This is true of at least the ten largest "exponents," and often this is true of all of the "exponents" on $(0, 0.03)$. The Neimark-Sacker style bifurcations to periodic orbits and tori are due to those pairs of exponents. If the bifurcations were of real eigenvalues, there would not be pairs of Lyapunov exponents of equal magnitude.

[3]See page 36 of (15) for a definition of the variational differential equation. In practical terms, we are referring to something like the eigenvalues of the local "Jacobian" of the network at a specific time.

On the s interval $\sim (0, 0.03)$, nearly all the exponents come in pairs, all the lines represent two exponents. Much of this is related to the numerical stability issues related to the linear algebra engines used to calculate the exponents. This is a fundamental and common problem in matrices with many nearly equal, and hence degenerate eigenvalues. This issue lies in the standard algorithm's ability to accurately orthogonalize the eigenvalues and eigenvectors, for more information see (49) and (54). At $s \sim 0.03$ the negative exponents begin to split apart. This cascade continues to occur from the most negative exponents to the largest exponents, in that order, up to the onset of chaos. By the onset of chaos, there exists 64 distinct exponents, one for each dimension. In fact, by $s \sim 0.052$, all the exponent pairs are observably split. It is possible that this is the first time this phenomena has been observed. There are several interpretations for such behavior; we will sketch our interpretation. In the interval where the exponents exist in nearby pairs, the dynamics are dominated by rotation in nearly all eigen-directions, and the many nearly equal local eigenvalues cause numerical stability problems.

As the s parameter is increased, the, rotating directions become less degenerate, and the lower exponents begin to separate. This phenomena has very little to do with the overall dynamics when the eigenvalues are strongly negative since these directions are so strongly contracting that they are not observed on the attractor after transients die out. As the s parameter is further increased, more pairs of Lyapunov exponents separate until the neutral directions begin to separate — this is the break up of the tori and the onset of turbulence.[4] We are displaying an example of the Ruelle-Takens scenario which contrasts the Landau and Hopf turbulence model. Strange attractors are not a collection of interacting quasi-periodic orbits or a rotating soup, but rather, distinct directions of expansion, contraction *a la* axiom A, and a little bit of rotation (neutral directions). This example, and Fig. (8.8) display this distinction in a very nice way. This cascade from all coincident to all non-coincident Lyapunov exponents was unexpected.

The first bifurcation is relatively easy to identify in Fig. (8.8), it occurs at a relative maximum of the largest overlaying set of exponents in the spectrum at $s = 0.0135$. This signals the Neimark-Sacker type first bifurcation, as expected. In our experience, the first bifurcation can be determined from the full spectrum reliably. The second bifurcation is also, relatively obvious at $s = 0.02755$, and corresponds to the relative maximum of the six (three pairs) largest exponents. This would signal the existence of a torus of some type, but due to the numerical stability issues it is difficult to make a definitive statement regarding the dynamics. Nevertheless, by the second bifurcation point, there clearly exists at least a 2-torus. Near the third bifurcation the Lyapunov spectrum becomes significantly more complicated — due to the numerical accuracy issues we dare not make a claim with respect to how many exponents are actually zero; however, it appears as if there are several. The onset of chaos is vague, but chaos can clearly be identified by $s = 0.8$. Likewise, considering Figs. (8.8) and (8.2), the respective largest Lyapunov exponents begin to be indistinguishable at $s = 0.08$. The secondary bifurcations cannot be identified, however, and it is this problem that has hindered the full statistical study.

The previous information can be summarized in the following table for which the location of the various bifurcations is given as observed via the bifurcation diagram, the largest Lyapunov exponent, the phase space diagrams, and the Lyapunov spectrum:

[4]With the break up of the commensurate directions, our numerical stability returns, see (6) or (4) for more regarding the dynamics after this transition

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

Putting this all together into a unified picture, it is likely that the first bifurcation, of type Neimark-Sacker, occurs at $s = 0.0135$. The second bifurcation occurs at $s = 0.02754$ is a bifurcation from T^1 to T^2. The third bifurcation, which occurs at $s = 0.04667$ is a bifurcation from T^2 back to T^1. The fourth bifurcation appears to be a period-doubling like bifurcation from T^1 to T^1. The fifth bifurcation is a bifurcation from the quasi-periodic orbit to a very high-period orbit. Following this is a very subtle sequence of bifurcations followed by the onset of chaos at $s = 0.05289$.

All the other observed cases with $d = 64$ were variations on a theme. We have observed more and fewer bifurcations between the first bifurcation and chaos, but all the bifurcations are relatively similar. Period-doubling cascades and such routes are rare — they occur in less than one percent of 64-dimensional networks with our weight distribution (i.e. we did not observe a period doubling route to chaos in the 500+ networks we considered for this report). Only five percent of the first bifurcations are due to real eigenvalues, and there never exist cascades of multiple real bifurcations.[5] The most common route to chaos we observed is a cascade of bifurcations between T^1, T^2, and high-period, periodic orbits; 3-tori are rare in our experience, we never observed one in nearly 500 cases.

As this discussion is intended as a survey of our space of neural networks along the route to chaos, let us list current and future directions of work:

i. a specific numerical analysis of each of the bifurcations in this example that can be compared to the theory; specifically the second and third bifurcations which are likely of low codimension;

ii. an analytical normal form calculation — again starting with the early bifurcations along the route;

iii. a study of Lyapunov spectrum calculation technique a la (42), (41), (126), (18), or (118) — these networks form a nice set of high-dimensional mappings to study Lyapunov spectrum calculation schemes since our mappings are high dimensional, not pathological, and are not, as high-dimensional maps go, computationally intensive to use; very often, upon the presentation of a new Lyapunov spectrum computation algorithm the test cases are very low dimensional dynamical systems for which numerical stability is rarely a problem;

iv. a brute force — by hand — statistical study of the bifurcation sequences in these networks;

v. a more numerically accurate Lyapunov spectrum calculation routine that can be used for a full statistical analysis of the routes to chaos in this set of dynamical systems;

vi. a systematic investigation of the dependence on the number of neurons;

vi. a systematic investigation of the sensitivity of the weight distribution;

[5] By cascades of real bifurcations we are, in general, referring to the period doubling cascade. Cascades of fold bifurcations are always unlikely do to the very nature of the fold bifurcation.

vii. constraints on weight distributions which lead to dynamics similar to a particular physical phenomenon, leading to a better understanding for how special a physical phenomena is relative to a general function space.

Our results help extend the current analytical results in the sense that we have a practical way of observing transitions along an interval in parameter space. Further, we present a framework such that the parameter set for which the claimed non-generic tori and persistent quasi-periodic behavior can be more concretely understood. Based on our observations, in high dimensions, the quasi-periodic route to chaos, often with a cascade of bifurcations is the dominant route. We observe what is predicted to occur according to the Ruelle-Takens route, but the bifurcation cascade is significantly more complicated than $T^0 \to T^1 \to T^2 \to chaos$. The bifurcation cascade often involves successive bifurcations between tori with dimension≤ 2.

Chapter 9

Region IV - The Bifurcation Chain Region

9.1 Definitions for numerical arguments

Since we are conducting a numerical experiment, we will present some notions needed 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 neighborhood about a point in the domain is shrunk, this implies a shrinking of a neighborhood 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 20 (ϵ_{num}) *ϵ_{num} is the numerical accuracy of a Lyapunov exponent, χ_j.*

Definition 21 (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 22 (*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 for a given set of initial conditions. The map f is said to have num-continuous Lyapunov exponents at $(\mu, x) \in R^1 \times R^d$ if for $\epsilon_{num} > 0$ there exists a $\delta_{num} > 0$ such that if:*

$$|s - s'| < \delta_{num} \tag{9.1}$$

then

$$|\chi_j(s) - \chi_j(s')| < \epsilon_{num} \tag{9.2}$$

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

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

Definition 23 (*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 at a fixed set of initial conditions, 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*

$$|\chi_j(s) - \chi_j(s')| < k_{\chi_j}|s - s'| \tag{9.3}$$

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

Note that neither of these definitions imply strict continuity since neither ϵ_{num} nor δ_{num} are infinitesimals, 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 don't 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 24 (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 thesis, 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 region IV, they will be seen to be extremely rare. As previously seen however, along the route-to-chaos for our networks, such non-transversal intersections are common. 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 corresponding center manifolds E^c as we do with the positive and negative exponents and their respective stable and unstable manifolds.

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 25 (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, $1 \leq i < n$, such that $dist(x, c_i) < \epsilon$.*

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

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

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. Formally —

Definition 26 (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 $j \geq N$.*

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

[1] 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.

course, many ways we can have a countably infinite set that is not dense, and, since 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 $(1, \frac{1}{2}, \frac{1}{3}, \cdots, \frac{1}{n}, \cdots)$. We will, in the section that addresses $a-$density, state the necessary conditions for an $a-$dense set for our purposes.

9.2 Conjectures with respect to dynamics in region IV along a 1-d interval

The point of this exercise is verifying three properties of C^r maps along a one-dimensional interval in parameter space. The first property is the existence of a collection of points along an interval in parameter space such that hyperbolicity of the mapping is violated. The second property, which is really dependent on the first and third properties, is the existence of an interval in parameter space of positive measure such that topological change (in the sense of changing numbers of unstable manifolds) with respect to slight parameter variation on the aforementioned interval is common. The final property we wish to show, which will be crucial for arguing the second property, is that on the aforementioned interval in parameter space, the topological change will not yield periodic windows in the interval if the dimension of the mapping is sufficiently high. More specifically, we will show that the ratio of periodic window size to parameter variation size (δ_s) goes to zero on our chosen interval.

Condition 1 *Given a map (neural network) as defined in section (4.2), the Lyapunov exponents are (locally) num-continuous with respect to s.*

There are many counterexamples to this condition, so many of our results will rest upon our ability to show how generally the above condition applies in high-dimensional systems.

Definition 27 (Chain link set) *Assume f is a mapping (neural network) as defined in section (4.2). A **chain link set** is denoted:*

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

If $\chi_j(s)$ for $j = 1, \ldots, d$ are continuous at its Lyapunov exponent zero-crossing, 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 countable (if the Lyapunov exponents are continuous), disjoint, adjacent open intervals of the form $\bigcup_i (a_i, a_{i+1})$.

Definition 28 (Bifurcation link set) *Assume f is a mapping (neural network) as defined in section (4.2). Denote a **bifurcation link set** of $C_k \cap V$ as:*

$$V_i = (a_i, a_{i+1}) \tag{9.4}$$

The number of positive Lyapunov exponents is constant on $V_i \subset V$ for all i. If 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 V_i's are the points where there exist Lyapunov exponent zero crossings. We are not particularly interested in these sets, however; rather we are interested in the collection of endpoints of these sets.

Definition 29 (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:*

$$U_k = \{a_i\} \tag{9.5}$$

or equivalently:

$$U_k = \partial(C_k \cap V) \tag{9.6}$$

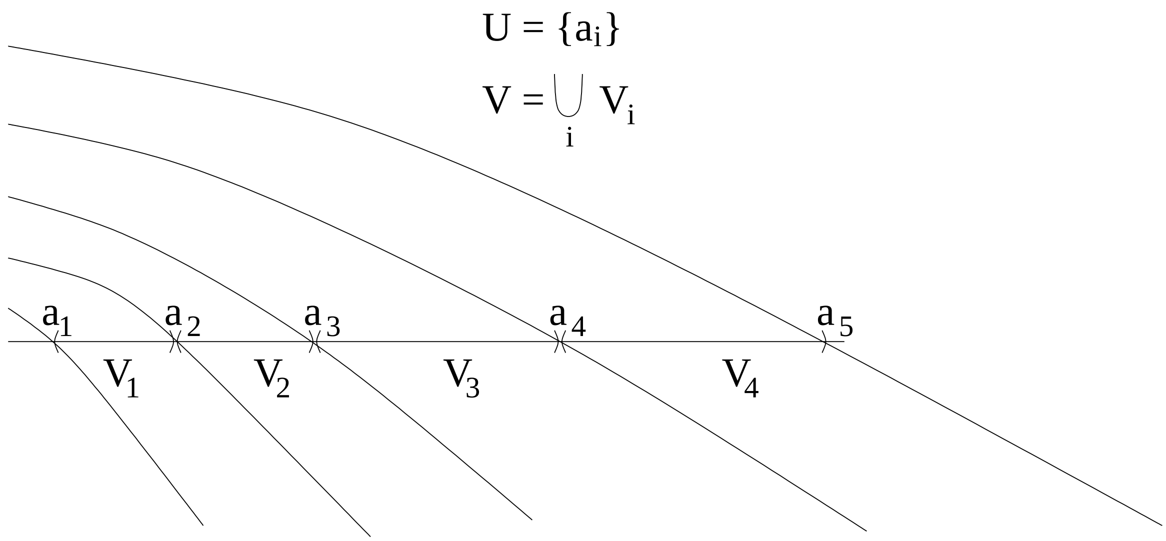

Figure 9.1: An intuitive diagram for chain link sets, V, bifurcation link sets, V_i, and bifurcation chain sets, U. for an LCE decreasing chain link set V.

For our purposes in 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 b_n will depend upon the specific case and dimension. In a practical sense, $a_1 \sim 0.5$ and $b_n \sim 6$. For higher-dimensional networks, $b_n \sim 6$ 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 (9.1).

9.2.1 Conjectures with respect to a 1-d interval in parameter space

We will now state the conjectures, followed by some definitions and an outline of what we will test and why those tests will verify our claims.

Conjecture 5 (Hyperbolicity violation) *Assume f is a mapping (neural network) as defined in section (4.2) with a sufficiently high 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 (4.2). From 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 behavior 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 parameterized curve such that there will exist neutral directions. The ideas listed above provide the framework for computational verification of conjecture (6). We must investigate conjecture (5) 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 6 (Existence of a Codimension ϵ bifurcation set) *Assume f is a mapping (neural network) as defined in section (4.2) with a bifurcation chain set U as per conjecture (5). The two following (equivalent) statements hold:*

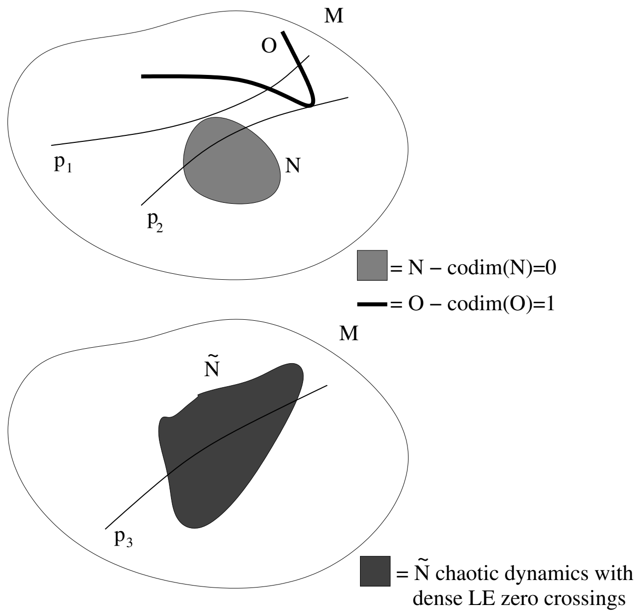

Figure 9.2: The top drawing represents various standard pictures from transversality theory. The bottom drawing represents an idealized version (in higher dimensions) of transversality catering to our arguments.

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,* $\overline{V}$.

ii. Given an arbitrarily small perturbation δ_s *of* s*, there will exist a dimension (in the asymptotic limit of high dimension) such that* f *will undergo a topological change on the interval* $[s - \delta_s, s + \delta_s]$ *where by topological change we mean a different number of global stable and unstable manifolds for* f *at* s *compared to* f *at* $s \pm \delta$.

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 $codim(N) = dim(M) - dim(N)$. If there exists a curve p through M such that p is transverse to N and the $codim(N) \leq 1$, then there will not exist an arbitrarily small perturbation to p such that p will become non-transverse to N. Moreover, if $codim(N) = 0$ and $p \bigcap N \subset int(N)$, then there does not even exist an arbitrarily small perturbation of p such that p intersects N at a single point of N.

The former paragraph can be more easily understood via figure (9.2) where we have drawn four different circumstances. This 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 via 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 avoid-ability 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.

To understand roughly why we believe conjecture (6) 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 zero 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.

Conjecture 7 (Periodic window probability decreasing) *Assume f is a mapping (neural network) as defined in section (4.2) and a bifurcation chain set U as per conjecture (5). 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.*[2] *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 (and length) of the existence of a stable 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 dimensions (inputs) completely changes the function. Thus the curve through the function space is an abstraction we are not afforded by our construction. In this work, conjecture (15) 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 concerned 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 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 were actually making statements we could rigorously prove, conjectures (5), (6), and (15) would function as lemmas for conjecture (8).

Conjecture 8 *Assume f is a mapping (neural network) as defined in section (4.2) with a sufficiently high number of dimensions, d, a bifurcation chain set U as per conjecture (5), 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 proven long ago. These examples were, however, very specialized and carefully constructed circumstances and do not speak to the commonality of structural stability failure. Along the road to investigating conjecture (8) 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). To some, 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 section (11.2).

The larger question that remains, however, is whether conjecture (8) is valid on high-dimensional surfaces in parameter space. We believe this is a much more difficult question with a much more complicated

[2] In other words, for every i, V_i and V_{i+1} can be made into a connected set by adding their common boundary point

answer. We can, however, speak to a highly related problem, the problem of whether chaos persists in high-dimensional dynamical systems. Thus, let us now make a very imprecise conjecture that we will make more concise in a later section.

Conjecture 9 *Chaos is a robust, high-probability behavior for high-dimensional, bounded, nonlinear dynamical systems.*

This is not a revelation (as previously mentioned, many experimentalists have been attempting to break this robust, chaotic behavior for the last hundred years), nor is it a particularly precise statement. We have studied this question using neural networks much like those described in section (4.2), and we found that for high-dimensional networks with a sufficient degree of nonlinearity, the probability of chaos was near unity (32). Over the course of investigation of the above claims, we will see a qualitative verification of conjecture (9). A more complete study will come from combining results from this study with a statistical perturbation study and combined with a study of windows proposed by (16) and the closing lemma of Pugh (101).

9.3 Conjectures with respect to dynamics in region IV with respect to general parameter perturbation in $R^{N(d+2)+1}$.

We now briefly discuss aspects of the windows conjecture related to the conjectures and results of stable ergodicity and the closing lemma. We will then state the conjectures we will be investigating for our chosen set of dynamical systems.

9.3.1 Comments regarding the windows conjecture

In a sense, the windows conjecture of Barreto et. al. (16) is a parameter perturbation version of the closing lemma of Pugh (101) and Pugh and Robinson (103). It is useful to relate the windows conjecture to the work of others as it will help highlight why we believe our ideas are relevant.

9.3.1.1 The windows conjecture and the closing lemma.

The closing lemma and the attempts to prove the closing lemma say much about the windows conjecture and the notion of fragile dynamical systems. We will briefly discuss this here, saving an in-depth discussion for a different report. Begin by limiting parameter variations of the diffeomorphism f to those that yield C^1 close approximations of f. Given an $f \in C^1$ on a compact manifold, the closing lemma says there will exist a g at q that is C^1 close to f at p, where $q = h(p)$, h is a diffeomorphism, p is a non-trivial recurrent point of the non-wandering set of f, and q is a periodic point of the non-wandering set of f. The point q might be very rare in a measure-theoretic sense, but dense in the non-wandering set — regardless of the dimension of the dynamical system. The windows conjecture says that, given $f \in C^1$ and a compact invariant set Λ, if there exists a periodic $g = h \circ g \circ h^{-1}$ where h is a diffeomorphism that maps f to a nearby set of initial conditions via a parameter change, and h can achieve this on an open set (maybe small) in parameter space, then f is fragile. Based on the closing lemma, a dynamical system being fragile is likely rare since a full measure set (e.g. an open set) near Λ that yields stable periodic orbits is likely uncommon (e.g. axiom A diffeormorphisms). There do exist ergodic attractors with no periodic points, but we will refrain from constructing such an example here. The g however, is likely rare. The windows conjecture claims that if more parameters are changed on f generating the g than there are positive Lyapunov exponents, then the set of parameter changes that will lead to windows is dense. The closing lemma says nothing about this issue directly as there are an infinite number of parameters involved with C^1 perturbations with which to construct the g, but even with an infinite number of parameters, the g's are difficult to construct. If we consider perturbations that yield C^r ($r > 1$) close approximations, the windows conjecture is unlikely to hold, since the g is unlikely to exist (102). The main differences between the closing lemma and the windows conjecture are: in the closing lemma g is achieved by a general C^1 perturbation versus in the windows conjecture g is achieved via a parameter change; in the closing lemma there only needs to exist a single periodic point near the non-trivial recurrent point versus in the windows conjecture, where the set of periodic points for the g must be full measure near the invariant set.

9.3.1.2 The windows conjecture and stable ergodicity.

In relating stable ergodicity to the windows conjecture, we will begin with a proposition from which the intuition for two conjectures will follow.

Proposition 1 (Axiom A and fragile attractors) *Assume an $f \in C^r(M)$ $r > 0$ such that f satisfies axiom A. The diffeormophism is not fragile.*

Proof: This result follows clearly from the definitions — axiom A attractors have dense but measure zero periodic points on the non-wandering set.

It is likely that stable ergodic attractors are not fragile. There exist stable ergodic attractors that have no periodic points, and axiom A attractors are stably ergodic (106) (95). However, in general whether stable ergodic attractors can be fragile is unknown.

Conjecture 10 (Fragile and stable ergodic attractors) *Assume a mapping $f \in C^2(M, M)$ where M is a compact, boundaryless manifold, with a parameter, $a \in R^n$. If f is fragile, f is not stably ergodic. Likewise, if f is stably ergodic, f is not fragile.*

The strongest result one might hope to achieve with respect to the commonality of fragile attractors would be that they are non-residual amongst partially hyperbolic dynamical systems. Adding the existence of a neutral direction seems to add little with respect to the measure of periodic points; we will refrain from discussing a proof, and simply make the following conjecture:

Conjecture 11 (Fragility is not residual) *Fragility is a non-residual property among C^r ($r > 0$), partially hyperbolic diffeomorphisms.*

9.3.1.3 The windows conjecture and k-degree LCE stability.

Finally, we would like to make the connection between fragile attractors, which will we do with the following proposition.

Proposition 2 (k-degree LCE stability and fragility) *Assume a (dissipative) mapping $f \in C^r(M, M)$, $r \geq 2$, with a parameter, $a \in R^n$, which maps a compact, d-dimensional, boundaryless manifold M to itself. The mapping f is robustly chaotic (we are assuming the Lyapunov exponents can be defined and exist) on a subset $V \subset R^p$ where $dim(V) = p$ if and only if f is not fragile on V.*

Proof: follows clearly from the definitions and by noting that M is compact — in our definition of robust chaos, the chaos must hold on the full Lebesgue measure set of initial conditions.

This proposition says nothing of the commonality of either k-degree LCE stability or fragile attractors; that is a topic that will be discussed over the course of the rest of this thesis. The relationship between k-degree robust chaos and a fragile attractor is quite distinct. We claim that both phenomena exist — that there are regions of parameter space in our networks for which the attractors are fragile and regions of the parameter space for which the attractors have k-degree robust chaos. We will only be addressing the latter in this thesis. One key claim we would like to make is that the dimension of the dynamical system matters when it comes to stability with respect to perturbations.

9.3.1.4 On the existence of spine loci

We will claim that for our neural networks, there exists a substantial portion of parameter space such that periodic windows either will not exist, or are extremely rare as the dimension of the network is increased. However, we will need to formulate some questions which yield themselves to our construction. Thus, we will focus on the existence and relevance of spines, since spine loci are the mechanism upon which the existence of windows is based.

Conjecture 12 (Non-existence of spine loci in high dimensional dynamical systems) *Assume a dynamical system f as given in section (4.2), for sufficiently high dimension, there will exist a connected subset (with more than one element) of parameter space such that there will exist no spine locus points.*

Why we believe this conjecture is reasonable is based upon arguments we will present in sections (9.6) and (9.7). In the previously mentioned sections we will also define the subset of parameter space on which we believe this conjecture will hold.

9.3.2 Robust chaos in large dynamical systems

One major point of making a conjecture such as conjecture (2) is to be able to discuss whether variation in dynamical behavior versus parameter variation (as opposed to general C^r perturbation) is, in some sense, common. The previous sections have either concentrated on the mechanisms relevant to conjecture (2) or touched on the relationship between the windows conjecture and other stability conjectures and results. In this section we will formally put forth four conjectures of our own (three new and one old) that we will later provide a large body of computational evidence supporting.

Let us begin by making a statement with respect to the robustness of chaos for our dynamical systems along an interval in parameter space.

Conjecture 13 (Robust chaos in large dynamical systems) *Assume f is a mapping (neural network) as defined in section (4.2) with a sufficiently large number of dimensions, d. There will exist an open interval in parameter space (R^1) for which chaos will be a robust dynamic.*

In (4) we put forth the above conjecture in various forms and provide numerical evidence supporting it. Here, however, we wish to extend this conjecture — we have considerable numerical evidence that suggests that robust chaos does not just exist along a particular interval in parameter space, but rather on large open sets in parameter space. In other words, instead of the robust chaos only being with respect to one parameter that varies all the other parameters in a specific and dependent manner (our s parameter), chaos is robust with respect to independent variation of all parameters up to a certain maximum perturbation. Hence the following conjecture.

Conjecture 14 (Robust chaos with respect to parameter perturbation in R^p) *Assume f is a mapping (neural network) as defined in section (4.2) with a sufficiently large number of dimensions, d, and let $p = N(d+2)+1$. There will exist an open set V in R^p (with significant Lebesgue measure) of parameter space for which chaos will be a robust dynamic.*

Let us make two comments with respect to this conjecture. First, we claim that the existence of chaos as a robust dynamic depends on dimension; the set of parameter space such that chaos becomes robust increases in size (with respect to Lebesgue measure) as the dimension of the dynamical system increases. The second comment, which runs somewhat contrary to the windows conjecture, is that the stability of chaos as a dynamic is related to the number of (linearly independent) parameters in the dynamical system. In fact, if we have very few parameters, our neural networks will have considerably less robust chaos.

The intuition for the first comment originates from (32) where we note that for a certain portion of parameter space, we observe nearly all networks are chaotic and the portion of the parameter space for which chaos is the dominant dynamic increases with dimension. Further, for a given d and s, increasing N produces an increased probability of chaos. This does not imply that perturbations of a particular network will not yield periodic solutions, however.

The intuition for the second comment is somewhat more complicated. First, note the universal approximation of arbitrary order of derivatives is made possible with an arbitrary number of parameters — i.e. an arbitrary number of neurons. As we increase the number of parameters in our neural networks, we can approach the possible perturbations available with respect to the C^r perturbations in the C^r topology. Recall, if we are considering C^1 diffeomorphisms there will always exist a perturbation on the nonwandering set that will yield a periodic orbit as per the closing lemma. However, for C^r perturbations and approximations where $r \geq 2$, there does not exist a result for the existence of a periodic orbit near every non-wandering point and the periodic orbits may not be stable. Since our dynamical systems are likely more than C^1, and the perturbations are likely C^r for $r > 1$, the closing lemma results might not apply and furthermore the periodic orbits guaranteed by the closing lemma might not be stable. (It is hard to say what topology we are in, and what derivative we are perturbing, thus it is difficult to relate our perturbations to the closing lemma). If there ever is a closing lemma for C^r for $r \geq 2$, surely with an infinite number of parameters our networks will have periodic windows reachable via parameter perturbation. It is more likely that such periodic orbits are unstable. How many such windows would still be an open question — we would claim that even in such a circumstance, finding such a periodic window for a high-dimensional map would still be extremely unlikely since stable periodic orbits are likely rare. Within our d-dimensional networks, nearly all types of dynamics are possible, nevertheless, we do not believe perturbations of high-dimensional chaotic systems will, in general, yield periodic solutions. It is likely that the chaotic behavior will not be perturbed away by 1^{st} order perturbations - hence it will take altering many more than k parameters (given the k positive Lyapunov exponents) to destroy chaos. We can easily have 100 times as many parameters as dimensions (and hence positive Lyapunov exponents) $\sim N + 2\frac{N}{d}$. Nearly all imaginable d-dimensional dynamics exist, and there exists a path in parameter space between then — hence there will exist a path in parameter space that will destroy all the positive Lyapunov exponents. However, such a path will likely be drastic (e.g. require perturbations that are large compared to the size of the weights.)

As promised, we will make a conjecture regarding periodic window existence and structure of our own. We will claim that, for our very general set of functions, as the dimension becomes large, periodic windows over certain (sizable in Lebesgue measure) sets of parameter space will tend to zero or vanish.

Conjecture 15 (Periodic window probability diminishing) *Assume f is a mapping (neural network) as defined in section (4.2) with a sufficiently large number of dimensions, d. There will exist a set $V \in R^p$ (again let $p = N(d+2)+1$) of parameter space such that there will not exist periodic windows on a positive Lebesgue measure set within V.*

We think of conjecture (15) as a corollary (or a restatement) of conjecture (14).[3] However since conjecture (15) is much easier to study numerically, we will argue for conjecture (14) based on evidence for conjecture (15). We perform a statistical study analyzing the probability of the existence of periodic windows given random perturbations of all the parameters with respect to increasing dimension.

We are not claiming that for our set of dynamical systems periodic windows do not exist — certainly they do.[8]. Rather we are claiming that we can find a large portion of parameter space such that given any parameter variation below some threshold, there will not exist a periodic window in parameter space.

It is worth noting that there exists a large class of C^1 dynamical systems for which robust chaos exists (e.g. Anosov diffeomorphisms) - whether the attractors are always unique is a different question. The Milnor attractor phenomena ((78) (66)) raises serious doubts about the existence of unique attractors. However, there is a current conjecture by Palis that claims that the number of attractors for a dynamical system at a set of parameter values is finite (91). We will reserve the topic of the existence of Milnor attractors and uniqueness of chaotic attractors over open sets of initial conditions for another report.

We can do better than conjecture (14) which only claims 1-degree robust chaos. In general, if the dimension of the network is greater than about 64, we can find a region of parameter space such that the network exhibits $\frac{\mathcal{L}}{4}$-degree robust chaos where $\mathcal{L}$ is the maximum number of positive Lyapunov exponents. We will make a conjecture regarding k-degree robust chaos as $d \to \infty$; saving an analysis specific to our circumstance for section (9.7.3).

Conjecture 16 ($k-$robust chaos in large dynamical systems) *Assume f is a mapping (neural network) as defined in section (4.2) with a sufficiently large number of dimensions, d, and an arbitrarily large number of parameters (neurons) $p = N(d+2)+1$. There will exist an open set in parameter space (R^p) for which chaos will be robust of degree k with $k \to \infty$ as $d \to \infty$.*

[3] That this conjecture is a corollary of conjecture (14) assumes that the only possibility for the destruction of chaos is via periodic and quasi-periodic orbits.

[8] In the region where the s parameter becomes large, there surely exists a complicated periodic window structure (3)

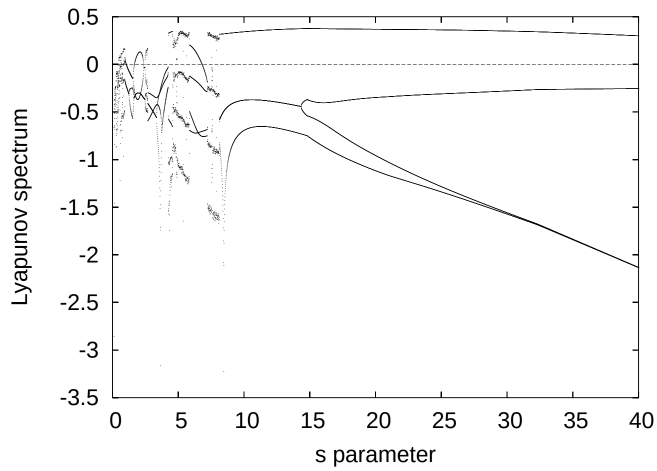

Figure 9.3: LE spectrum: 32 neurons, 4 dimensions

9.4 Numerical preliminaries with respect to conjectures on a 1-d curve in parameter space

Before we present our arguments supporting our conjectures we must present various preliminary results. Specifically we will discuss the *num*-continuity of the Lyapuonv exponents, the *a*-density of Lyapunov exponent zero-crossings, and argue for the existence of an arbitrarily high number of positive exponents given an arbitrarily large number of dimensions. With these preliminaries in place, the arguments supporting our conjectures will be far more clear.

9.4.1 *num*-continuity

Testing for the *num*-continuity of Lyapunov exponents formally will be two-fold. First, we will need to investigate, for a specific network, f, the behavior 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 parameter variation. 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 neighborhoods upon which continuity of the Lyapunov exponents are examined are over ranges of plus and minus one parameter increment. This is all 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 $[0.1, 10]$.

9.4.1.1 Qualitative analysis

Qualitatively, our intuition for *num*-continuity comes from examining hundreds of Lyapunov spectrum plots versus parameter variation. In this vein, Figs. (9.3) and (9.4) present the difference between low and higher dimensional Lyapunov spectra.

In Fig. (9.4), the Lyapunov exponents look continuous within numerical errors (usually ± 0.005). Figure (9.4) by itself provides little more than an intuitive picture of what we are attempting to argue. As we will be making arguments that the Lyapunov spectrum will become more smooth, periodic windows will disappear, etc, with increasing dimension, Fig. (9.3) shows a typical graph of the Lyapunov spectrum versus parameter variation for a neural network with 32 neurons and 4 dimensions. The contrast between Figs. (9.4) and (9.3) illustrates the increase in continuity we are claiming.

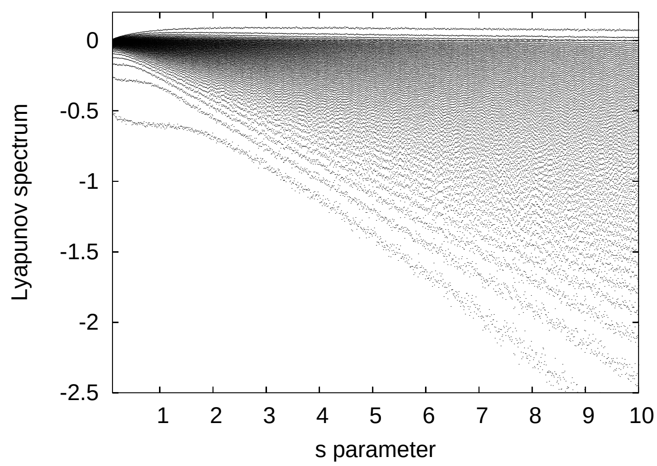

Figure 9.4: LE spectrum: 32 neurons, 64 dimensions.

Although a consideration of Figs. (9.3) and (9.4) yields an observation that, as the dimension is increased, the Lyapunov exponents appear to be more continuous functions of the s parameter, the above figures alone do not verify $num-$continuity. In fact, it should be noted 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 frequency of pathologies appears to become vanishingly small (this is noted over our observation of several thousand networks with dimensions ranging from 4 to 256).

9.4.1.2 Quantitative and numerical analysis

Our quantitative analysis will follow two lines. The first will be a specific analysis along the region of parameter change for three networks with dimensions 4 and 64, respectively. This will be followed with a more statistical study of a number of networks per dimension where the dimensions will range from 4 to 128 in powers of 2.

Consider the num-continuity of two different networks while varying the s parameter. Figure (9.5) is a plot of 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 has many large spikes. As the dimension is increased, considering the 64-dimensional case, the large spikes disappear. The spikes in the 4-dimensional case can be directly linked to the existence of periodic windows and bifurcations that result in dramatic topological change. This is one verification of $num-$continuity of Lyapunov exponents. These two cases are quite typical, but it is clear that the above analysis, although quite 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 focus on the parameter region between $s = 0$ and 64, the region in parameter space over which the maxima of entropy, Kaplan-Yorke dimension, and the number of positive Lyapunov exponents exists. Figure (9.6) considers the num-continuity along parameter values ranging from 0 to 6. The points on the plot correspond to the mean (over a few hundred networks) of the mean exponent change between parameter values, or:

$$\mu^d = \frac{1}{Z}\sum_{k=1}^{Z} \frac{\sum_{i=1}^{d} |\chi_i^k(s) - \chi_i^k(s+\delta s)|}{d} \tag{9.7}$$

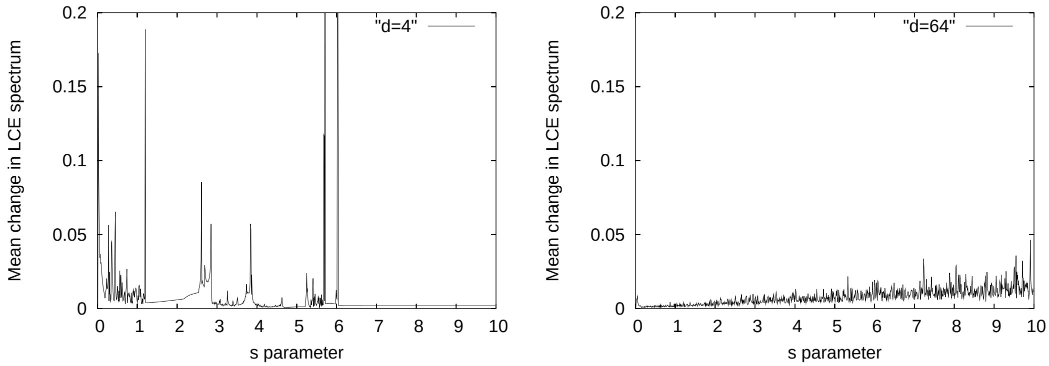

Figure 9.5: *num*-continuity (mean of $|\chi_i(s) - \chi_i(s+\delta s)|$ for each i) versus parameter variation: 32 neurons, 4 (left) and 64 (right) dimensions.

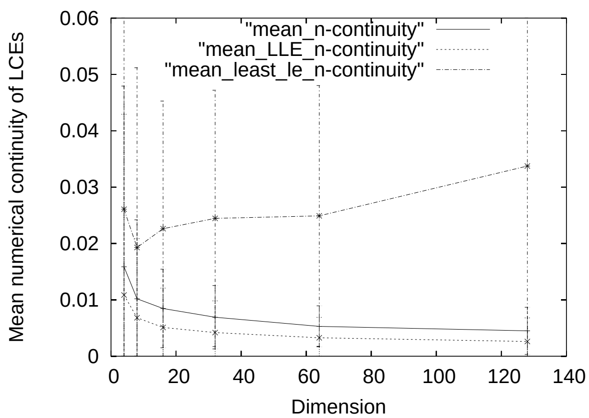

Figure 9.6: Mean *num*-continuity, *num*-continuity of the largest and the most negative Lyapunov exponent of many networks versus their dimension. The error bars are the standard deviation about the mean over the number of networks considered.

summed over an interval of the s parameter where Z is the total number of networks of a given dimension considered.

Figure (9.6) clearly shows that as the dimension is increased, for the same computation time, both the mean exponent change versus parameter variation per network and the standard deviation of the exponent change decrease substantially as the dimension is increased.[4] 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 in our numerical techniques (specifically the Gram-Schmidt orthogonalization). The mean error in the most negative exponent increases with dimension, but is a numerical artifact. 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

[4]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.

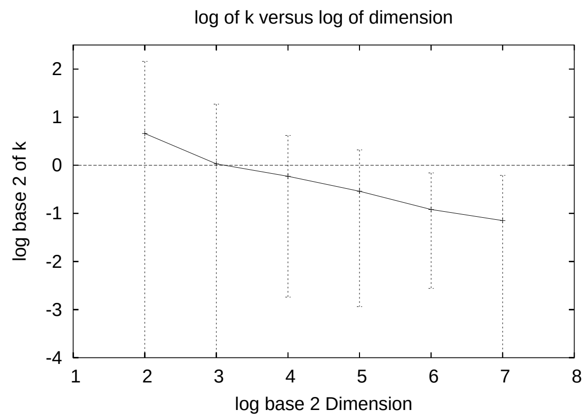

Figure 9.7: k-scaling: $\log_2$ of dimension versus $\log_2$ of *num*-Lipschitz constant of the largest Lyapunov exponent.

to having a few positive exponents, the variation in any specific Lyapunov exponent is very small.

There is a specific relation between the above data to definition 23; *num*-Lipschitz is a stronger condition than *num*-continuity of Lyapunov exponents. The mean $num-$continuity at $n = 32$, $d = 4$

$$|\chi_j(s + \delta_{num}) - \chi_j(s)| < k\delta_{num} \tag{9.8}$$
$$|0.02| < k|0.01| \tag{9.9}$$

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

$$|\chi_j(s + \delta_{num}) - \chi_j(s)| < k\delta_{num} \tag{9.10}$$
$$|0.004| < k|0.01| \tag{9.11}$$

which yields $k = 0.4$ which is less than one and thus 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

$$|\chi_j(s + \delta_{num}) - \chi_j(s)| < k\delta_{num} \tag{9.12}$$
$$|0.015| < k|0.01| \tag{9.13}$$

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

$$|\chi_j(s + \delta_{num}) - \chi_j(s)| < k\delta_{num} \tag{9.14}$$
$$|0.002| < k|0.01| \tag{9.15}$$

which gives $k = 0.2$. As the dimension is increased, k decreases, and thus *num*-continuity increases. As can be seen from Fig. (9.6), 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 in an asymptotic limit of high dimension, consider Fig. (9.7). As the dimension is increased the $\log_2$ of the dimension versus the $\log_2(k_{\chi_1})$ yields the scaling $k \sim \sqrt{(\frac{2}{d})}$; 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. This completes our evidence for the *num*-continuity in high-dimensional networks.

9.4.1.3 Relevance

Conjectures (5), (6), and (8) are all fundamentally based on condition (1). For the neural networks, all we need to establish conjecture (5) is the *num*-continuity of the Lyapunov exponents, the existence of the fixed

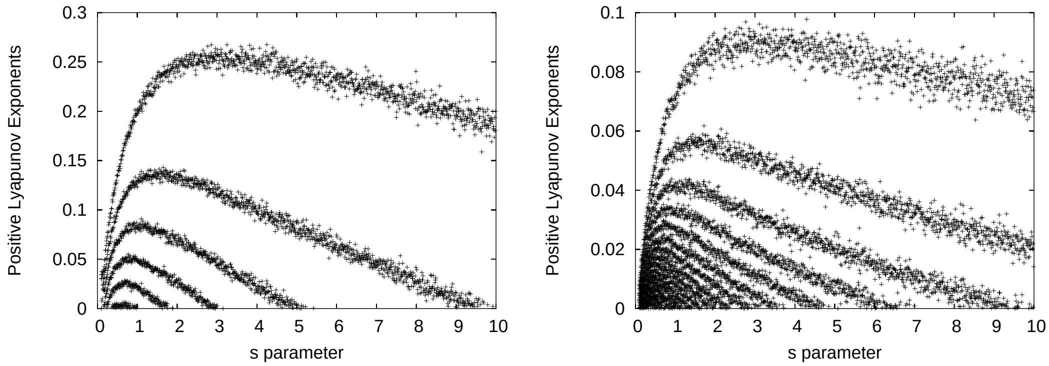

Figure 9.8: Positive LE spectrum for typical individual networks with 32 neurons and 16 (left) and 64 (right) dimensions.

point for s near zero, the periodic orbits for $s \to \infty$, and three exponents that are, over some region of parameter space, all simultaneously positive. The *num*-continuity of Lyapunov exponents implies, within numerical precision, that Lyapunov exponents both pass through zero (and don't jump from positive to negative without passing through zero) and are, within numerical precision, zero.

9.4.2 $a-$density of zero crossings

Many of our arguments will revolve around varying s in a range of 0.1 to 6 and studying the behavior 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 χ_{max}. Since we are dealing with a countable set, we will refrain from any type of measure theoretic notions, and instead rely on a-density of the set of positive exponents as the dimension is increased. Recall the definition of a-dense (definition (26)), the definition of a bifurcation chain subset (definition (29)), which corresponds to the set of Lyapunov exponent zero crossings, and the definition of a chain link set (definition (27)). Our conjectures will make sense if and only if, 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 dimension is increased provides us with the convergence to density of non-hyperbolic points we need to satisfy our goals.

9.4.2.1 Qualitative analysis

The qualitative analysis will focus on pointing out what characteristics we are looking for and why we believe $a-$density of Lyapunov exponent zero-crossings (a-dense bifurcation chain sets in the closure of the chain link set) over a particular region of parameter space exists. A byproduct of this analysis will be a picture of one of the key traits needed to support our conjectures. We will begin with figures showing the positive Lyapunov spectrum for 16 and 64 dimensions.

Considering the 16-dimensional case, and splitting the s parameter variation into two regions, region a - $R_a = [0, 0.5]$, and region b - $R_b = [0.5, 10]$. We then partition up R_b 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 the s parameter on the chain link set will lead to a Lyapunov exponent zero-crossing (and a transition from V_i to $V_{i\pm1}$)[5]. Considering region b, 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

[5]Recall, the bifurcation chain sets will not exist when the zero crossings are not transverse.

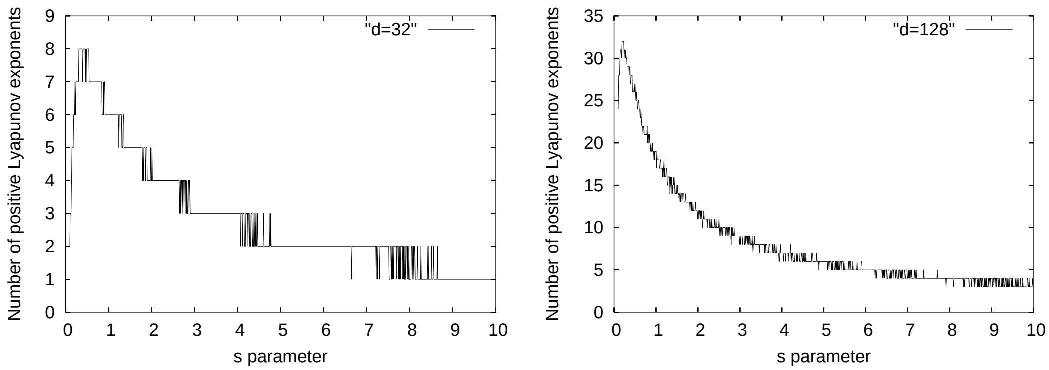

Figure 9.9: Number of positive LE's for typical individual networks with 32 neurons and 32 (left) and 128 (right) dimensions.

cease to be at least somewhat uniformly distributed, the rest of our arguments will surely fail. For instance, in region b of the left plot of Fig. (9.8), 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 Fig. (9.8), our desires seem to be met since the spacing between exponent zero-crossings is clearly decreasing 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 quite distinct and look *num*-continuous as previously asserted. The above figures are, of course, only a picture of two networks; if we wish for a more conclusive statement, we will need arguments of a statistical nature.

9.4.2.2 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.

Qualitatively, the two examples given in Fig. (9.9) (both of which typify the behavior 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[6], while the width of the peak becomes more narrow. Thus, the slope of the number of positive exponents versus s (i.e. the graph in figure (9.9)) 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 $d = 128$. Noting that the more negative the slope, the less variation in s is required to force a zero-crossing, it is clear that 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.

It is important to note, however, that upon rescaling the plot with 32 dimensions to the plot with 128 (i.e. the 32-dimensional case seems about $1/4^{th}$ of the 128-dimensional case), the two plots nearly overlay. This suggests a universal type scaling, and in this sense, the 32-dimensional plot is no more smooth than the 128-dimensional plot.

From Fig. (9.10), 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 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, on the

[6] 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 Fig. (9.9), regardless of the number of exponents.

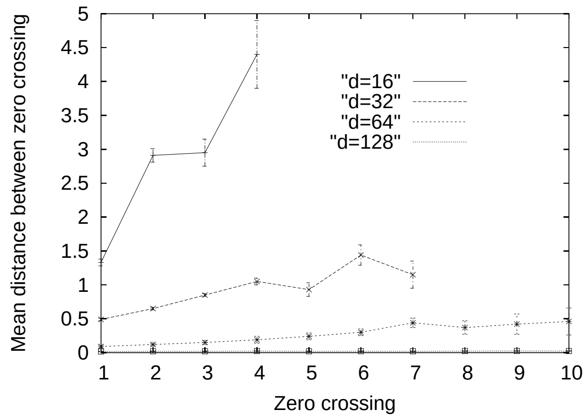

Figure 9.10: Mean distance (δs) between each of the first 10 zero crossings of LE's for many networks with 32 neurons and 16, 32, 64, and 128 dimensions.

order of a hundred for $d = 128$) of networks. For the cases where the dimension is 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 a resolution of 0.01 in the s variation 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 point for the zero crossing. All the zero crossings were determined after the smallest positive exponent that became positive hit its peak value, i.e. after approximately $s \sim 0.75$ in the $d = 16$ case of Fig. (9.8).

9.4.2.3 Relevance

The a-density of zero crossings of Lyapunov exponents provides the most important element in our arguments of conjectures (5) and (6); combining *num*-continuity with a-density will essentially yield our desired results. If continuity of Lyapunov exponents increases, and if the density of zero crossings of exponents increases over a set $U \in R^1$ in parameter space, it seems 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 those will be dealt with in the final arguments.

9.4.3 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.

9.4.3.1 Qualitative analysis

The qualitative analysis of this can be seen in Fig. (9.9); as the dimension is increased, the maximum number of positive Lyapunov exponents clearly increases. We wish to quantify that the increase in the number of positive exponents versus dimension occurs for a statistically relevant set of networks.

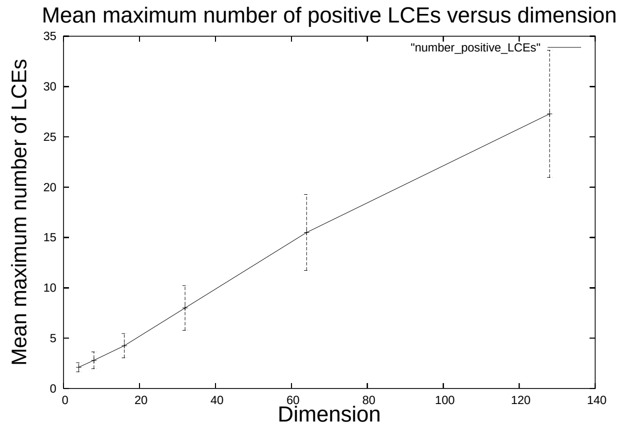

Figure 9.11: Mean maximum number of positive LE's versus dimension, all networks have 32 neurons (slope is approximately $\frac{1}{4}$).

9.4.3.2 Quantitative analysis

We will use a brute-force argument to demonstrate the increase in positive Lyapunov exponents with dimension; we will simply plot the number of positive exponents at the maximum number of exponents as dimension is increased. We claim that the number of Lyapunov exponents increases and, in fact, diverges to infinity as the limit as the dimension of the network is taken to infinity. Figure (9.11) shows the number of positive Lyapunov exponents versus dimension.

From Fig. (9.11) it is clear that as the dimension is increased, the number of positive exponents increases in a nearly linear fashion.[7] Further, this plot is linear to as high a dimension as could be computed for reasonable statistics. If the maximum number of exponents versus dimension remains linear beyond the range we could compute, we will have the countably infinite number of positive exponents we require.

9.4.3.3 Relevance

The importance of the increasing number of positive exponents with dimension is quite simple. For the a-density of exponent zero-crossings 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.

9.5 Numerical arguments with respect to conjectures on a 1-d curve in parameter space

9.5.1 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 a bit of 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.

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

9.5.1.1 Qualitative analysis

Qualitative evidence for the dissappearance of periodic windows amidst chaos is evident from Figs. (9.3), (9.4) and (9.8); the periodic windows that dominate the 4-dimensional network over the parameter range $s = 0$ to 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 to 10. We will avoid, for the most part, the "route to chaos" region (s near zero), since it yields many complex issues that will be saved for another report. 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 (8); further analysis is necessary for a more complete understanding (3).

9.5.1.2 Quantitative and numerical analysis

The quantitative analysis we wish to perform will involve arguments of two types; those that are derived from data given in sections (9.4.1) and (9.4.2), and those that follow from statistical data regarding the probability of a window existing for a given s along an interval in R. We begin by recalling what we are attempting to claim and what conditions we need to verify the claim. We will then present the former argument and conclude with the latter.

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, dependent on the region of parameter space one is observing — and there is likely no way to rid ourselves of such an issue. 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 to behave 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 on the evidence of *num*-continuity of Lyapunov exponents. From Fig. (9.6) 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 our 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 (9.12) 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 $\sim \frac{1}{d}$. Moreover, observe that in high-dimensional dynamical systems, when periodic windows are observed on the interval $(2, 4)$, they are usually large in length. 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 ((6)), but hints to why this phenomena is occuring can be found in (16).

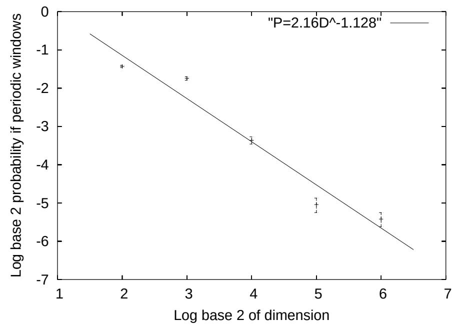

Figure 9.12: $\log_2$ of the probability of periodic or quasi-periodic windows versus $\log_2$ of dimension. The line $P_w = 2.16d^{-1.128}$ is the least-squares fit to the plotted data.

9.5.1.3 Relevance

Decreasing window probability inside the chaotic region provides direct evidence for conjectures (15) and (9) along a one-dimensional interval in parameter space. We will, in a more complete manner, attack those conjectures in a different report. We will use the decreasing periodic window probability to help verify conjecture (6) 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 our 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. $l = |b_n - a_1|$ is increasing yet will remain finite.

9.5.2 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 fraction of Lyapunov exponents near zero over an ensemble of networks versus variation in the s parameter. 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 (9.4.2). To argue for this conjecture, we only need the existence of a neutral direction[8], or, more accurately, at least two bifurcation link sets, which is not beyond reach.

9.5.2.1 Qualitative analysis

A qualitative analysis of hyperbolocity violation comes from combining the *num*-continuity of the exponents in Fig. (9.4) and the evidence of exponent zero crossings from Figs. (9.9) and (9.6). 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.

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

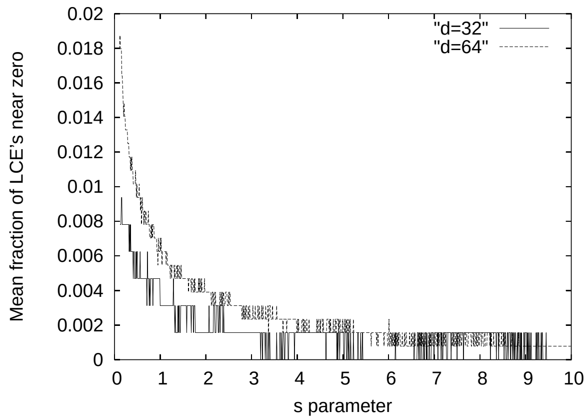

Figure 9.13: Mean fraction of LE's near zero (0 ± 0.01) for networks with 32 neurons and 32 or 64 dimensions (averaged over 100 networks).

9.5.2.2 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 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 Fig. (9.9) 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; these oscillations are due largely to numerical fluctuations in the calculations[9]. 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 need to note the significance of the exponents near zero. To do this, we calculate the relative number of Lyapunov exponents numerically at zero compared to the ones away from zero. All this information is summarized in Fig. (9.13), which addresses the mean fraction of exponents that are near zero versus parameter variation.

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 Fig. (9.13): there are no sharp discontinuities in the curves; 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; the curves are concave — implying that exponents are somehow leaving the interval $(-0.01, 0.01)$; and 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 (6). To satisfy conjecture (5) we only need three exponents to be near zero and undergo a zero crossing for the minimal bifurcation chain subset[10] 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.

[9]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.

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

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 variation 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, the number of exponent crossings on the interval A will monotonically increase with dimension. As can be seen from Fig. (9.10), the exponent zero-crossings are relatively uniform, with the distance between crossings decreasing with increasing dimension. Considering Fig. (9.8), 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, are becoming dense. This is overkill for the verification of the existence of a minimal bifurcation chain set. This is strong evidence for both conjectures (5) and (6) . It is worth noting that hitting these points of hyperbolocity violation upon parameter variation is extremely unlikely under any uniform measure on R since they are a countable collection of points.[11] Luckily, this does not matter for either the conjecture at hand or for any of our other arguments.

9.5.2.3 Relevance

The above argument provides direct numerical evidence of hyperbolocity violation over a range of the parameter space. This is strong evidence supporting conjecture (5). It does not yet verify conjecture (6), but it sets the stage as we have shown that there is a significant range over which hyperbolocity is violated. The former statement speaks to conjecture (8) also; a full explanation of conjecture (8) requires further analysis, which is the subject of a discussion in the final remarks.

9.5.3 Hyperbolocity violation versus parameter variation

We are finally in a position to consider the final arguments for conjecture (6). To complete this analysis, we will need the following pieces of information:

i. we need the maximum number of positive exponents to go to infinity

ii. we need 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. we need *num*-continuity of exponents to increase as the dimension increases

iv. a major simplification can be provided with the existence of one global maximum in the number of positive exponents and entropy, and along any portion of parameter space where s is greater than the s at the maximum number of positive exponents, the maximum and minimum number of exponents occur on the graph at the end points of the parameter range (within numerical accuracy)

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 (9.4.3). 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 $d \geq 64$), there exists a single, global maximum in the number of positive exponents versus parameter variation.

[11] When considering parameter values greater than the point where the smallest exponent that becomes positive, 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.

9.5.3.1 Qualitative analysis

The qualitative picture we are using for intution is that of Fig. (9.8). 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 Fig. (9.8) (and hundreds of similar plots), we claim that qualitatively this parameter range exists for at least $0.5 \leq s \leq 6$.

9.5.3.2 Quantitative and numerical analysis

Let us now complete our arguments for conjecture (6). 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 $B = \cup V_i = V$, 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 (6).

Assuming a sufficiently large number of dimensions, verification of conjecture (5) gives us the existence of the bifurcation chain set and the existence of the adjoining bifurcation link sets. The existence of an arbitrary number of positive Lyapunov exponents, and thus an arbitrarily large number of zero crossings follows from section (9.4.3). That the bifurcation chain set has an arbitrarily large number of elements, $\#U \to \infty$ is established by conjecture (15), because, without periodic windows, every bifurcation link set will share a limit point with another bifurcation link set. From section (9.4.1), the *num*-continuity of the exponents persists for a sufficiently large number of dimensions, thus the Lyapunov exponents will cross through zero. Finally, section (9.4.2) 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.

Specifically for our work, we can identify U such that $U \subset [0.5, 6]$. We could easily extend the upper bound to much greater than 6 for large dimensions ($d \geq 128$). How high the upper bound can be extended will be a discussion in further work.

Finally, 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 fall off in the number of positive Lyapunov exponents.

9.5.3.3 Relevance

The above arguments provide direct evidence of conjectures (6) and (8) for a one-dimensional curve (specifically an interval) in parameter space for our networks. This evidence also gives a hint with respect to the robustness 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.

9.6 Outline of Analytical Arguments for conjectures (14) and (12) on a $R^{N(d+2)+1}$ surface in parameter space.

This being a paper full of conjectures, we wish to present a brief outline of proofs of conjectures (12) (the no-spines conjecture) and (14) (the robust-chaos conjecture) whose proofs have been unsuccessful to date.

9.6.1 Jacobian matrix of our dynamical systems

We will begin with the calculation of the matrix of partial derivatives for our scalar neural network which we will use throughout this section.

The scalar networks have an extremely convenient property that aids in the calculation of the Jacobian and in the computation of the Lyapunov spectra; the Jacobians of our scalar networks are companion matrices:

$$Df_x = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{d-2} & a_{d-1} & a_d \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & & & & \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

The a_i's are given as:

$$\frac{\partial x_t}{\partial x_{t-k}} = a_k = \sum_{i=1}^{n} \beta_i s w_{ik} sech^2(s w_{i0} + s \sum_{j=1}^{d} w_{ij} x_{t-j}) \tag{9.16}$$

Thus, we will always have a particularly simple matrix of partial derivatives — it is this simple form that gives us hope for the existence of an eventual proof of the arguments that follow.

9.6.2 Linear algebra

Let us begin by noting that all the matrices we will be considering are square matrices with real elements. We will briefly recall several standard results and definitions from linear algebra.

Theorem 12 *Given a matrix $A \in R^{n^2}$, A is nilpotent if and only if A is strictly upper triangular, that is, $\forall \lambda_i \in \sigma(A)$, $\lambda_i = 0$.*

Proof: See (58).

Next, the simple form of the Jacobian of our neural networks will bear fruit:

Definition 30 (Companion matrix) *A real valued matrix $A \in GL(n, R)$ is said to be a companion matrix if it has the following form:*

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{d-2} & a_{d-1} & a_d \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & & & & \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

The companion matrix is the companion of it's characteristic polynomial

$$\lambda^{d+1} + a_d \lambda^d + \cdots + a_2 \lambda + a_1 = 0 \tag{9.17}$$

Proposition 3 *A d-dimensional companion matrix is nilpotent if $a_i = 0$ for all i such that $0 < i \leq d$.*

Proof: This follows from a brief analysis of equation (9.17).

The above results give the conditions on the Jacobian (at an equilibrium point) for the existence of spine locus points that are required by the windows conjecture of Barreto et. al. (16). In other words, for spines to occur in our systems, we need the following, $a_k = 0$ for all $1 \leq k \leq d$. Hence, all we need accomplish for a proof of conjecture (12) is to show that, in general, $a_k \neq 0$.

9.6.3 Discussion of parameter space

We are now forced to address a topic whose discussion we have delayed; definitions of parameters. In the windows conjecture, it is stated that if there are p parameters and only $k < p$ conditions for stability, small parameter variation will lead to the discovery of stable orbits. In our neural networks, we have $N(d+2)+1$ parameters; thus our networks nearly always have many more parameters than stability conditions. If we wished to rigorously show how varying the parameters affected the neural network with respect to the space of neural networks, $\Sigma(G)$, we would first have to create the following map:

$$\phi : R^{N(d+2)+1} \to \Sigma(G) \tag{9.18}$$

and then analyze how open sets of $R^{N(d+2)+1}$ are related to open sets of $\Sigma(G)$ (for such a construction see (123)). Since ϕ is C^1 (just differentiate the neural network with respect to the parameters), as long as $\phi(p)$ ($p \in R^{N(d+2)+1}$) is a regular point at p (see (76) for a discussion of regular points) $D\phi_p$ will have full rank and doesn't collapse any dimensions. With a full analysis of the set of critical points of $\phi()$ in place we could make rigorous statements regarding how and what open $N(d+2)+1$-balls centered at $p \in R^{N(d+2)+1}$ would be represented in $\Sigma(G)$. In a general sense, it is likely that all our $N(d+2)+1$ parameters are linearly independent and perturbing all the parameters does not collapse dimensions of $\gamma \in \Sigma(G)$ since an infinite number of parameters is required for derivative approximation and critical points in the parameter space are "uncommon." Without studying the critical points of ϕ however, we cannot rigorously claim that our $N(d+2)+1$-ball study is representative of an open set in $\Sigma(G)$. Instead we will assume that by varying all our parameters will sweep out a significant open ball in the function space $\Sigma(G)$. Based on the windows conjecture, one might be led to believe that the neural networks might have terrible periodic window instabilities. However, we must not forget that the windows conjecture requires the attractor to be fragile, which will be a topic of discussion in later sections. Nevertheless, there are two practical issues that must be addressed. The first issue concerns the independence of the parameters varied. The second issue deals with the number of parameters varied and in that vein, the means of parameter variation. For our numerical experiments, the parameters must vary over the entire span of the parameter space.

The point of this is, based on the windows conjecture of Barretto et. al. (16), if the parameter space is of much higher dimension than the phase space, such as is the case for our networks, then if the attractor is fragile, which they claim is common, there will be a neighborhood in parameter space such that the chaotic attractor can be dispelled. If we consider a 4-dimensional network with 4 neurons, the parameter space is R^{25}, which is much larger than d. In this scenario, often there is a maximum of two positive exponents, thus there should exist a "codimension" 23 "manifold" in parameter space for which there exists a periodic window. In the neural networks, the high number of parameters is necessary to fit high-order derivatives, thus, it seems somewhat unlikely that varying many of the parameters, which in a sense amounts to varying high-order derivatives, will have the large, first-order effects necessary to dramatically alter the dynamical behavior. However, the former statement is an oversimplification and is clearly incorrect in many circumstances.

9.6.4 Relationship to the a_i elements to stability and Lyapunov exponents

We wish to put conditions on the a_i terms and imply something about Lyapunov exponents or stability in general, thus is will be useful to briefly describe the relationship between the a_i's and the Lyapunov exponents.

$$U = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{d-2} & a_{d-1} & a_d \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & & & & \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \delta x_{11} & \delta x_{12} & \cdots & \delta x_{1d} \\ \delta x_{21} & \delta x_{22} & \cdots & \delta x_{2d} \\ \delta x_{31} & \delta x_{32} & \cdots & \delta x_{3d} \\ \vdots & \ddots & \vdots & \\ \delta x_{d1} & \delta x_{d2} & \cdots & \delta x_{dd} \end{bmatrix}$$

The Lyapunov exponents of f at time t correspond to $\log(< u_i, u_i >)$ where u_i is the i^{th} vector of the matrix U. It is of course these numbers which are summed and averaged to generate the Lyapunov exponents for the orbit of f. At time $t = 0$, $\delta - ij = 1$ when ever $i = j$, δ_{ij} is zero otherwise. As f is iterated forward, the variation matrix elements are all, generally, non-zero.

There is a direct connection between the a_i's and the existence of spines — there will exist a spine if and only if Df is nilpotent (this is evident upon a consideration of (9.17)). Thus if we could show some set of conditions such that at least one a_i will have positive probability of being non-zero, then we are well on our way to proving conjecture (12).

We are, however, not so fortunate with respect to the conjecture regarding robust chaos (conjecture (14). It would be nice if finding conditions such that at least one $a_i > 1$ with positive probability would imply at least on positive Lyapunov exponent. This however seems much more elusive. Nevertheless, we will present the outlines of both arguments, as these arguments will contribute considerably to the intuition and understand of our claims.

9.6.5 Outline of a proof of no spines in neural networks

We will begin by restating conjecture (17) for our specific circumstance.

Conjecture 17 (No spines in dimensional neural networks) *There exists a connected subset of dynamical systems (as defined in section (4.2)) for which spines will not exist.*

This conjecture will essentially be a corollary of the following lemma

Lemma 1 *Assume the set $\{a_k\}$ were each a_k is as given in equation (9.16). As $d \to \infty$, or equivalently as the number of a_k's goes to infinity, there will exist an s interval such that the variance of $\{a_k\}$ is not zero for all but a set of measure zero in the weight space (i.e. the space of w_{ij} and β_i) for our weight distributions.*

Outline of a proof: To prove the above lemma, it will suffice to show that the variance of the a_i's is nonzero — as all we need is for some of the a_i's to be non-zero. Achieving this goal requires two steps: first a mean must exist and be found; and second, the variance about the mean must be shown not to converge to zero as either N or d approach infinity.

To see why this is reasonable, begin by noting

$$a_k = \sum_{i=1}^{N} \beta_i s w_{ik} sech^2(s w_{i0} + s \sum_{j=1}^{d} w_{ij} x_{t-j}) \tag{9.19}$$

which is a sum of the product of two random variables and $sech()$ of random variables and the state space variables. What we would like to be able to say is the following:

$$E[a_k] = E[\beta]E[\omega]E[sech^2()] \tag{9.20}$$

If the above was true, $E[\omega] = 0$ would imply $E[a_k] = 0$ and we would be half way done with the first step. Since $sech^2()$ is not independent of β or ω, and since $sech^2()$ is not i.i.d. equation (9.20) may not be reasonable. However, noticing that $sech^2() > 0$ and considering the power series expansion of $sech^2()$

$$sech^2(c) = 1 - c^2 + \frac{2}{3}c^4 - \frac{17}{45}c^6 + \frac{62}{315}c^8 - \frac{1382}{14175}c^{10} + \dots \tag{9.21}$$

Clearly, $csech^2cx$ will produce only odd powers of c

$$csech^2(c) = c - c^3x^2 + \frac{2}{3}c^5x^4 - \frac{17}{45}c^7x^6 + \frac{62}{315}c^9x^8 - \frac{1382}{14175}c^{11}x^10 + \dots \tag{9.22}$$

Thus,

$$E[csech^2(c)] = E[c] - E[c^3]E[x^2] + \frac{2}{3}E[c^5]E[x^4] - \frac{17}{45}E[c^7]E[x^6] + \frac{62}{315}E[c^9]E[x^8] + \dots \tag{9.23}$$

Now, $E[x^k] = 0$ for k odd, thus $E[csech^2(c)] = 0$, and we would be done if c and x were i.i.d. random variables.

If $E[a_k] \to 0$ with respect to N, d, or both, we still only need the variance of a_k to be non-zero. Clearly if $E[a_k] \neq 0$ we do not need to deal with the variance at all — we will not attempt this tract however. If the terms in the sum of the a_k's are mutually independent, then the variance of a_k will simply be the sum of the variance of each of the terms in the sum which makes up a_k. However, the terms in the sum which is a_k are definitely not mutually independent, thus the heart of the problem that remains is deriving a sensible formula for the variance of the a_k's. At any rate, if the variance of the a_k's is non-zero, which it almost certainly is, then in the limit where the statistics apply (i.e. when N and d are big enough that the a_k's begin to converge in distribution), spines will not exist.

9.6.6 Outline of a proof of robust chaos in neural networks

The conjecture regarding the existence of spines seems, at this time, considerably easier to prove than the conjecture regarding robust chaos of neural networks. It would be convenient if there was a statement that periodic windows in parameter space exist if and only if there are spines, but this makes little sense. Nevertheless, we will provide an outline which will hopefully lead to a successful proof.

Let us begin by re-stating the conjecture:

Conjecture 18 (Robust chaos with respect to parameter perturbation in R^p) *Assume f is a mapping (neural network) as defined in section (4.2) with a sufficiently high number of dimensions, d. There will exist a open set V in R^p (with significant Lebesgue measure) in parameter space (R^1) for which chaos will be a robust dynamic.*

Outline of a proof: A proof might proceed in the following steps:

i. establish explicit formulas for the mean and variance of the a_k's

ii. establish a connection between the distribution of the a_k's and the Lyapunov exponents

Let us begin by assuming that, via a proof of conjecture (12) from the previous section, we have a formula for the mean and variance of a_k. From here it would be nice if we could make a simple claim along the lines of, if $a_i > 1$ for all $t > T$, then there will exist a positive Lyapunov exponent. However, this is clearly not always true[10]. The heart of this problem is the old problem of establishing a connection between the distribution of the coefficients of the characteristic polynomial and the eigenvalues of a matrix. If the explicit formulas for the mean and variance of the a_k's is such that the connection between the distribution of the characteristic exponents and the eigenvalues is known, then we will be in luck and our proof will not be difficult. However, if such is not the case, such a proof will likely be difficult. Even if we do have a connection between the eigenvalues of our local Jacobian and the distribution of the a_k's, we then need to take this one step further, and establish a connection between the local eigenvalues and the global Lyapunov exponents. It is likely that some type of assumption akin to uniform (or non-uniform) partial hyperbolicity (see (15) for more information regarding uniform and non-uniform hyperbolicity) will be needed for such a link to be established, but such does not seem out of reach nor unreasonable. However, since we do not have the first step completed, we will not press this issue any further.

[10] A simple counter-example for the two-dimensional case would be $a_1 = 3/2$ and $a_2 = -9/16$.

9.7 Numerical arguments with respect to conjectures on a $R^{N(d+2)+1}$ surface in parameter space.

We will now put forth the numerical justification for conjectures (15), (13), (14), and (12). We will proceed with three different techniques, one for conjectures (15) and (13), a similar technique for conjecture (14), and a single-case analysis for conjecture (12), noting that this case is highly representative. The conclusions we reach are: the probability of periodic windows decreases like $\frac{1}{d^2}$; the degree of the k-degree LCE stability becomes positive and once positive, increases monotonically; and the existence of spines in networks with high N and d becomes vanishingly small.

9.7.1 The experimental methodology

In our presentation of evidence supporting conjectures (15), (13), and (14), we will use two different methods of computing. Regarding conjectures (15) and (13) we will use the largest Lyapunov exponent, and for conjecture (14) we will have to compute the entire spectrum. We have chosen such a course in an attempt to produce a study on the largest set of neural networks, using the least computationally costly technique for the desired task. We will now discuss our computational methods.

9.7.1.1 Searching for periodic windows

We examined networks with dimensions over 6 powers of 2 — 4, 8, 16, 32, 64 and 128 dimensions. All the networks had 32 neurons; the number 32 was chosen because in our experience (e.g. figure (1) of (8)) for this range of dimensions, addition of neurons changed little but the computation time. For each dimension, we randomly selected 500 networks, but we required that every network be chaotic before being perturbed. This is an important distinction because, for $d = 4$, chaotic neural networks are quite rare, and thus we are sampling little of the space, whereas for $d = 64$, nearly all the networks are chaotic and thus nearly every case we select is adequate (for more information in this regard, consider figure (2) of (8)). Each network was then perturbed 100 times (from it's original set of weights), and each perturbation involved perturbing all $N(d+2)$ parameters. The perturbation size for the data shown in this thesis was on the order of 10^{-3}. A significant amount of data has been accumulated for other perturbation sizes, and we observe no change until the perturbation size is on the order of unity and no significant change until the perturbation size is on the order of 10. The largest Lyapunov exponent was computed in the standard way *a la* Bennetin et. al. (18) or Wolf et. al. (130).

It is worth noting that there can exist very long transients for many orbits, especially for the quasi-periodic orbits. To combat this problem, we allow the order of $80,000$ to $300,000$ iterations of the network before we begin calculating the Lyapunov exponents.

9.7.1.2 k-degree LCE stability

Regarding the k-degree LCE stability, we considered networks with dimensions over 5 powers of 2 — 4, 8, 16, 32, and 64 dimensions. All the networks again had 32 neurons. For each dimension, we randomly selected 100 networks, but we required that every network be chaotic before being perturbed. Each network was then perturbed 100 times, and like the periodic windows case, each perturbation involved perturbing all $N(d+2)$ parameters. The perturbation size for the data shown in this report was on the order of 10^{-3}. A significant amount of data has been accumulated for other perturbation sizes, and we observe no change until the perturbation size is on the order of unity and no significant change until the perturbation size is on the order of 10. The Lyapunov exponent spectrum was computed in the standard way *a la* Bennetin et. al. (18).

To rid ourselves of issues with transients, we allow $100,000$ to $500,000$ iterations of the network before we begin calculating the Lyapunov exponent spectrum.

9.7.2 Robust Chaos: Existence of periodic windows

There is an intimate relationship between robust chaos (conjecture (14)) and the existence of periodic windows (conjecture (15)) — if a dynamical system has periodic windows after an arbitrarily small parameter perturbation, then the dynamical system cannot be robustly chaotic. Likewise, a robustly chaotic dynamical system cannot have periodic or quasi-periodic windows. In this section, we will not differentiate between quasi-periodic and periodic orbits. We will refer to all orbits without a positive largest Lyapunov exponent as being "periodic."

Recall conjecture (15) in practical terms: begin with a neural network as per our prescription, for most networks (determined by their weights), there will exist a particular interval in the s parameter such that as the dimension of the network is increased, there will not exist an arbitrarily small parameter perturbation in $R^{N(d+2)+1}$ such that the network can be made non-chaotic. We clearly cannot prove this with our computational methods. However, we will show strong evidence for our claim as d becomes large. It is worth noting that conjecture (15) and conjecture (13) imply the same dynamics and are both corollaries of conjecture (14).

9.7.2.1 1-dimensional interval

As the dimension in increased, the probability of existence of periodic windows in parameter space for our networks on the interval $s \in (0.5, 20)$ decreases roughly inversely with d (4). For networks of 32, 64, and 128 dimensions we have carefully searched for periodic windows with s variation from 10^{-3} to 10^{-12} for many individual networks. Network having near the canonical number of positive Lyapunov exponents rarely have periodic windows. Furthermore, when periodic windows have been observed, they are always relatively large (i.e. they can be observed with an s variation of 10^{-3}). Finally, if a periodic window is not observed with parameter increments on the order of 10^{-3}, then finding a periodic window along the interval in parameter space in question is difficult (i.e. we have not been able to observe a periodic window in such circumstances).

Variation of s varies Nd of the parameters. However, this carves out a particular surface in parameter space. Furthermore, variation of all the parameters via the s parameter is clearly a linearly dependent parameter variation and will not suffice for attacking the conjectures in the report. The windows conjecture of Barretto at. el. (16) states that if a dynamical system is fragile and has k positive exponents, then there will exist a variation of k parameters that will yield a periodic orbit. Thus, an s variation cannot really speak to this conjecture. With our study, we can never prove that there will not exist a periodic window in parameter space for all possible parameter variations. However, we can extend the results of (4) a considerable amount by simply varying all of our parameters at once for a given network at a given value of s. Surely there will exist portions of the interval $s \in (-\infty, \infty)$ such that the windows conjecture will be valid. We are arguing that for $s \in (0.5, 20)$, which is the region of s space such that tanh() is nonlinear, but not nearly binary, as the dimension of the dynamical system becomes large, the neural networks will cease to be fragile.

For more information regarding dynamics of our neural networks along one-dimensional intervals in parameter space, see (2), (4), or (8).

9.7.2.2 $N(d+1)$-ball

As previously mentioned, in the windows conjecture, given k positive Lyapunov exponents, there will exist a k-dimensional surface in parameter space such that there will exist a periodic window if the dynamical system is fragile. In this section we argue for conjecture (15) along three lines:

1. the probability of existence of periodic windows averaged over all the networks considered (500) with 100 perturbations each decreases like $\frac{1}{d^2}$ with increasing dimension;

2. the fraction of the (500) networks that can be perturbed to a periodic orbit decreases like $\frac{1}{d}$ with increasing dimension;

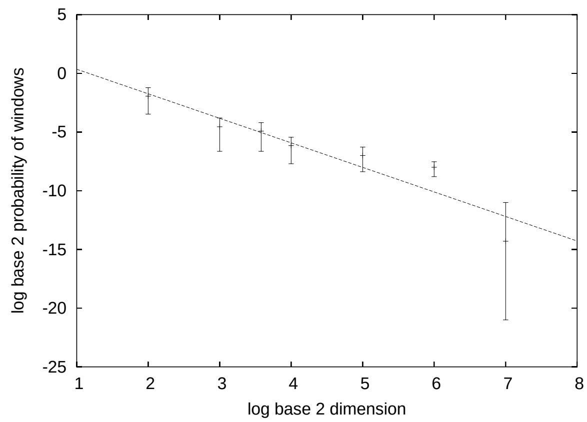

Figure 9.14: Log probability of the existence of periodic windows versus log of dimension for 500 cases per d. Each case has all the weights perturbed on the order of 10^{-3}, 100 times per case. The line of best is $\sim \frac{1}{d^2}$.

3. the probability of existence of periodic windows averaged over only networks with windows per perturbation decreases with increasing dimension, but much more slowly than the overall probability of windows averaged over all the networks — this establishes that if a network has a window, it is more likely to have many windows and the majority of the positive probability of periodic windows at high dimensions is due to relatively few networks.

In figure (9.14), each data point represents the probability of finding a periodic network among the 500 networks each perturbed 100 times for each dimension. The probability of the existence of networks with periodic orbits decreases like $\frac{1}{d^2}$. Thus, as the dimension is increased, our networks become more robustly chaotic or are far less likely to display periodic windows. However, it is important to make the distinction between probability of the existence of periodic windows per network averaged over many networks and the probability of a network having a window. For claim (3) to have any validity, the probability of a given network having a perturbation that yields a periodic window must decrease in such a way that the probability of networks with perturbations that yield periodic orbits occurs for a smaller and smaller set of networks. This information cannot be discovered simply by considering figure (9.14).

Figure (9.15) (from the same data set) represents the fraction of networks for which a periodic network was observed. The fraction of networks with periodic windows decreases as $\frac{1}{d}$, which is not as steep a decrease as the decrease in the overall probability of the existence of periodic windows. If the fraction of networks with periodic orbits decreased faster than the probability of the existence of periodic windows, our claim (3) would be self evident. Nevertheless, the decrease in the fraction of networks that have nearby networks with periodic orbits is substantial — networks at $d = 4$ are ~ 20 times more likely to have perturbations that lead to periodic windows than networks with 64 dimensions. Figure (9.15) supports claim (2), but we need yet more information for verification of claim (3).

Combining the data from figures (9.14) and (9.15) we arrive at figure (9.16), the probability of periodic windows averaged over only networks where periodic windows were observed. In this figure, the probability of finding a window decreases with dimension, but considerably more slowly than the overall probability *a la* figure (9.14). With an increase in 4 decades (powers of 2 in dimension), the probability of finding a periodic window in networks with observed periodic windows has only decreased by little more than a factor of 3.5. Thus, even though the probability of finding a network that yields a periodic window is decreasing with dimension, even amongst networks with periodic windows, the mass of the probability of perturbations that yield periodic windows accumulates on a faster decreasing fraction of the networks. In other words, given the information in figure (9.16), claim (3) is highly reasonable.

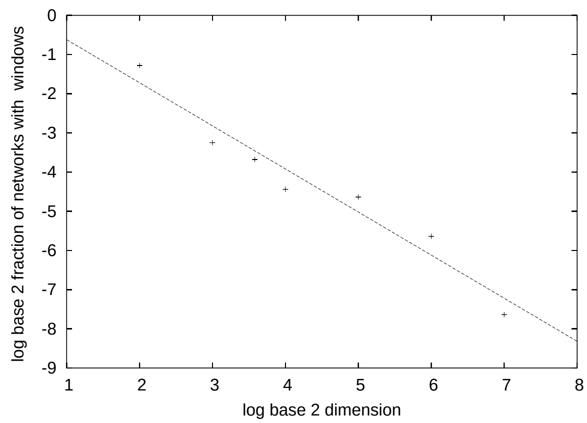

Figure 9.15: Log fraction of networks with periodic windows versus log of dimension for 500 cases per d. Each case has all the weights perturbed on the order of 10^{-3}, 100 times per case. The line of best fit is $\sim \frac{1}{d}$.

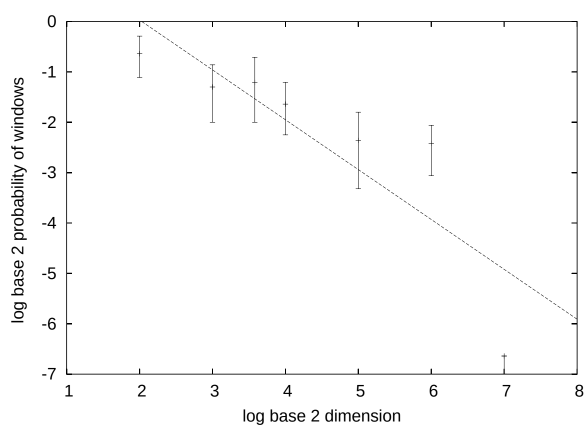

Figure 9.16: Log probability of the existence of periodic windows in networks with windows versus log of dimension for 500 cases per d. Each case has all the weights perturbed on the order of 10^{-3} 100 times per case. The line of best fit is $\sim \frac{1}{d}$.

Combining the information from these figures, for $d = 4$, if we assume the probability of windows is uniformly distributed, then for each of the 500 networks, 26 of the 100 perturbations should yield networks with periodic orbits. Adjusting for the actual fraction of networks that have been observed to yield periodic windows, 64 of the 100 perturbations yield periodic windows. In networks that can be perturbed to have periodic windows, windows are observed roughly 2.5 times as frequently as they would be if the existence of windows was uniformly distributed across all the networks. This situation becomes much more extreme as the dimension is increased. For $d = 64$, assuming the probability of windows is distributed uniformly, about 0.4 percent of the perturbations will yield periodic windows — one window per 2 networks perturbed 100 times each, or possibly one periodic window every 200 perturbations per network (in the data in this thesis each network was only perturbed 100 times). Adjusting for the actual fraction of networks with parameter perturbations that yield periodic orbits, 18 of every 100 perturbations yield periodic windows. Hence, the probability per perturbation for networks with windows to have perturbations that yield windows only decreases by about a factor of 3.5, whereas the overall probability of the existence of windows decreases by nearly two orders of magnitude.

Networks are twenty times more likely to have periodic windows at $d = 4$ versus $d = 64$, however, when a network has at least one observed perturbation that yields a window, a given perturbation is only 3.5 times more likely to yield a window at $d = 4$ than at $d = 64$. In practical terms, given a 64-dimensional network, there is a 2 percent chance that it will have a perturbation that yields a network with a periodic orbit — if the network has a periodic orbit, there is an 18 percent chance of finding it. Given a 4-dimensional network, there is a 41 percent chance that the network will have a perturbation that yields another network that has a periodic orbit — if such is the case, then there is a 64 percent chance of finding a network with a periodic orbit upon perturbing the network (with our perturbation scheme).

Given the above data, networks for which perturbations yield periodic orbits are not only much more likely to have many more periodic windows, but as the dimension is increased, the networks that have perturbations that yield similar (with respect to the parameters) networks with periodic orbits, contain a higher and higher percentage of the mass of the probability of observed periodic windows. As to the nature of the networks that can be perturbed to nearby networks that have periodic orbits, one might hypothesize that the periodic windows exist in networks with a significantly lower mean number of positive Lyapunov exponents. We cannot answer such questions with only the largest Lyapunov exponent, and thus we must move on to an analysis along the lines of k-degree LCE stability.

9.7.3 Robust chaos of degree k

Recall the claim from conjecture (16) in a practical sense: beginning with a dynamical system (network) with a sufficiently large number of dimensions ($d \geq 64$), for nearly every network there will exist an interval with respect to the s parameter such that upon perturbations (on orders less than unity) of all the parameters, the perturbed mapping will retain at least k positive exponents where $k > 0$. In other words, there exists a portion of parameter space for our high-dimensional networks such that nearly all the networks are k LCE stable. The specific k varies with the s value, and thus we must first discuss the dynamics along the s interval, and follow this with a general perturbation study.

9.7.3.1 1-dimensional interval

We would like to be able to make the statement, assuming a chaotic 128-dimensional network and $s \in (0.5, 20)$, that the dynamical system will have $14-$degree robust chaos, or something along these lines. However, there is a strong dependence with s on the number of positive Lyapunov exponents. Figure (9.17) is the plot of the variation in the number of positive Lyapunov exponents versus s for a typical 64-dimensional network. Let us demonstrate what we mean when we claim that there is an s dependence on the degree of k LCE stability. For $s \in (0.5, 1)$, small perturbations can lead to the loss of several positive Lyapunov exponents. Nevertheless, the network in figure (9.17) on the interval $s \in (0.5, 1)$ likely has $10-$degree robust chaos. Alternatively, considering the region $s \in (8, 25)$, it is likely that the network has $3-$degree robust chaos but the variation in the number of positive number of Lyapunov exponents for small perturbations

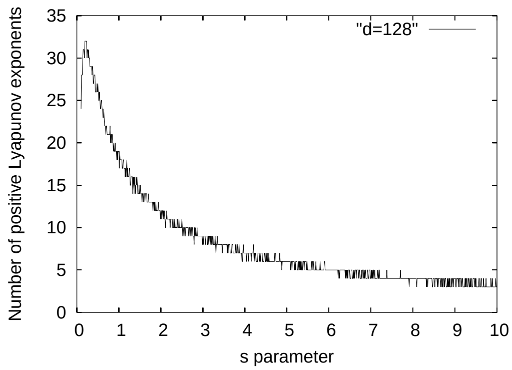

Figure 9.17: Number of positive LE's for typical individual networks with 128 dimensions.

will likely be much less than for the interval $s \in (0.5, 1)$. Of course, variation of the s parameter is not the same as an independent variation of all of the parameters, but the region along the s interval will affect the k-robust chaos. Moreover, figure (9.17) also suggests that perturbations leading to all the positive Lyapunov exponents becoming negative seems uncommon for networks such as the one depicted in figure (9.17).

9.7.3.2 $N(d+1)$-ball

Noting that any argument we make with respect to k-degree LCE stability will be dependent on s, we will select $s = 3$ since it is at the transition between the spike in the number of positive Lyapunov exponents and the region where there is relatively little change in the number of positive exponents. To make our point, we would like to emphasize four observations:

1. as the dimension of the network is increased, the mean number of positive exponents increase monotonically;

2. the higher the mean number of positive exponents a network has, the higher the degree of k-degree LCE stability;

3. for a given d, s, and perturbation size, the distribution of the number of positive Lyapunov exponents has compact support (i.e. has a finite minimum and maximum) — this is trivial for finite d, and thus in a practical sense, we claim that $\frac{R}{d}$ is decreasing with dimension where R is the range of the number of positive Lyapunov exponents;

4. for dynamical systems per our construction, there exists a sizable s interval such that as the dimension is increased, the minimum number of positive exponents for all the networks will increase above zero (i.e. the lower bound on the support of the distribution of the number of positive Lyapunov exponents will become positive given enough dimensions) — in a practical sense, we claim that $\frac{\bar{L}}{R} > 1$ where $\bar{L}$ is the mean number of positive Lyapunov exponents for a given d and s if d is high enough;

Figure (9.18) shows that at $s = 3$, the number of positive Lyapunov exponents is increasing linearly with dimension. This is not surprising given figure (15) of a previous report (4). In figure (9.18), we are considering 100 networks with 100 perturbations each, and noting what the mean number of positive exponents is for each network; the error bars represent the standard deviation in the mean of the mean number of positive exponents. If we did not have this feature, claims (2)-(4) would make little sense.

To provide intuition for why claims (2), (3), and (4) make sense, consider the respective histogram of two of the data points in figure (9.18), the point at $d = 8$ and the point at $d = 64$. In figure (9.19), 8.5 percent of the networks have a largest exponent less than or equal to zero; i.e. 8.5 percent of the distribution has zero positive exponents. A full 50 percent of the networks only have one exponent. Of course this is not

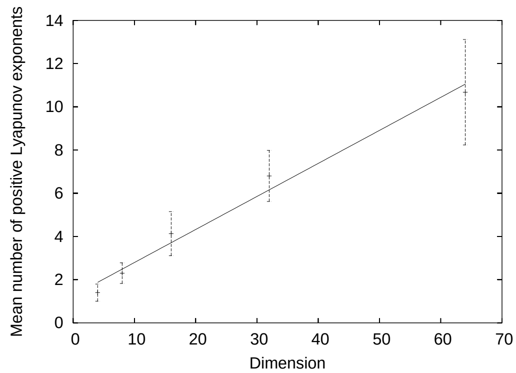

Figure 9.18: The mean number of positive Lyapunov exponents versus dimension for networks with 32 neurons at $s = 3$. The line of best fit is $\sim 0.15d$.

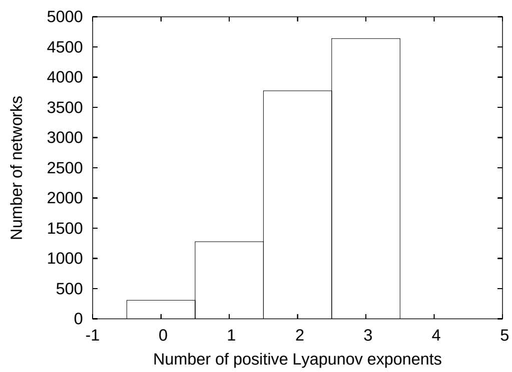

Figure 9.19: Histogram of the number of positive Lyapunov exponents for $d = 8$.

surprising given the expected number of exponents is slightly greater than one. In contrast, figure (9.20), the histogram of the number of positive exponents for each network with 64 dimensions, shows no network for which all the positive Lyapunov exponents have been perturbed away. Networks with five or more exponents comprise 99 percent of the networks considered, and no network has zero positive exponents. This is again not particularly surprising given the expected number is eleven positive exponents. The point of this is to show how, as the dimension is increased, and as the mean number of positive exponents is increased, that the number of networks that can be perturbed such that all the windows can be destroyed will vanish. Simply considering the distribution is not enough to verify claim (2), i.e. with just the histogram of positive exponents we cannot discern the origin of the cases with few positive exponents. However, figure (9.20) does provide evidence of claims (3), and (4). Networks with five or more exponents comprise 99 percent of the networks considered, and no network has zero positive exponents. The ratio of the range of maximum to minimum number of exponents, $\frac{R}{d} = 1$ for the $d = 4$ and decreases to $\frac{R}{d} = 0.28$ for $d = 64$; $\frac{R}{d}$ is clearly decreasing with increasing dimension. Finally, since the lower bound on the number of positive exponents is positive for $d = 64$, figure (9.20) is evidence for claim (4).

Evidence for claim (2) is provided by figures (9.21) and (9.22). Figure (9.21) is a collection of 100 networks, each perturbed 100 times. The maximum and minimum number of positive Lyapunov exponents are plotted at the mean of the number of positive Lyapunov exponents for each of the 100 networks. To argue for claim (2), we need to show that as the mean of the number of positive Lyapunov exponents is increased, the degree (i.e. k) of the k-degree LCE stability increases. Considering figure (9.21), the range,

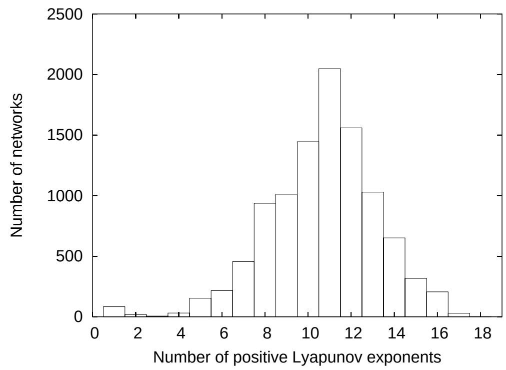

Figure 9.20: Histogram of the number of positive Lyapunov exponents for $d = 64$.

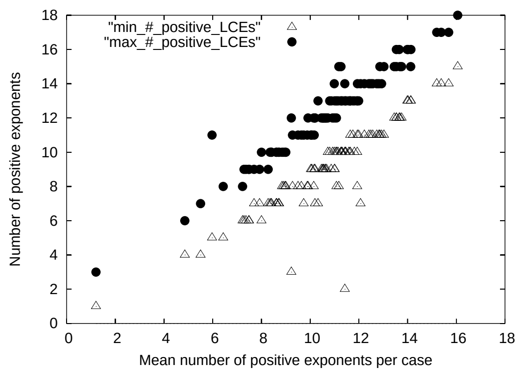

Figure 9.21: The range between the maximum and minimum number of positive Lyapunov exponents versus the mean of the number of positive Lyapunov exponents per case for 100, 64-dimensional networks.

R, between the maximum and minimum number of exponents is, for most attractors, quite narrow. This fact is emphasized in figure (9.22) which depicts the mean of the deviation from the mean number of positive Lyapunov exponents per the mean of the number of positive Lyapunov exponents. The error bars in figure (9.22) are the standard deviation of the range. The important feature to notice is that the deviation from the mean is, in all cases with four or more positive exponents with the exception of five of the 100 networks, less than one. Thus, networks with a higher mean number of positive Lyapunov exponents often have $(\bar{L} - 1)$-degree LCE stability ($\bar{L}$ is the mean number of positive exponents per network). Networks with a higher mean number of exponents are less likely to have periodic windows. Moreover, the lower portion of the distribution of the number of positive Lyapunov exponents comes predominantly from networks that have a low mean number of positive Lyapunov exponents. If we were using our framework to directly test the windows conjecture of Barretto et. al. (16), we would be focusing on topics such as investigating a scaling in the probability of periodic windows with the number of parameters varied. Moreover, we might increase the number of parameters varied per perturbation and attempt to discover what the probability of finding a window might be, given a fixed number of positive exponents. However, we are more concerned with the value of k in our notion of k-degree robust chaos, and thus we must establish, in a general numerical manner, a way of measuring a lower bound on the number of positive Lyapunov exponents for a given network. Nevertheless, figures (9.21) and (9.22) provide strong evidence for claim (2).

When discussing k-degree LCE stability we need to make the distinction between the k-degree LCE stability for an individual network and the k-degree LCE stability for an ensemble of networks. Definitions

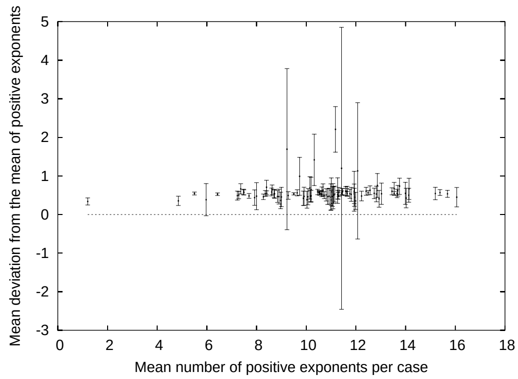

Figure 9.22: The mean deviation from the mean number of positive exponents versus the mean of the number of positive Lyapunov exponents per case for 100, 64-dimensional networks. The error bars represent the standard deviation in the mean deviation.

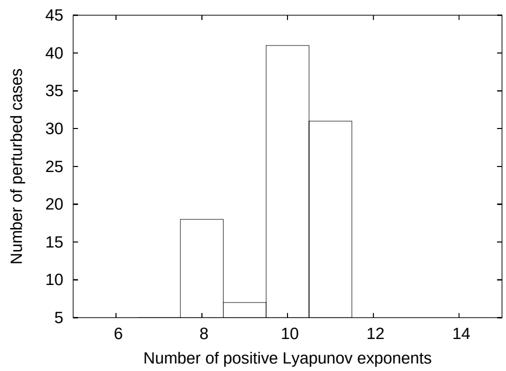

Figure 9.23: The histogram of the number of positive Lyapunov exponents for a typical 64-dimensional network perturbed 100 times.

(6) and (5) pertain to single networks (defined by a specific set of weights). In the case of a single network, one can determine the degree of k-degree LCE stability either from considering figure (9.21) or in figure (9.23) — the histogram of the number of positive Lyapunov exponents for a single network perturbed many times. Considering the k-degree LCE stability of an ensemble, we have many options, of which we list three:

i. the degree of k-degree LCE stability for the ensemble could be defined as the minumum number of positive exponents — this is what is given in claim (4);

ii. the degree of k-degree LCE stability for the ensemble could be defined as the mean number of positive Lyapunov exponents minus the standard deviation about the mean;

iii. the degree of k-degree LCE stability for the ensemble could be defined as the lower bound of the curve under which 99 percent of the area of the distribution of the number of positive Lyapunov exponents is contained;

Besides these three notions, one could devise many others — for our purposes these three will suffice.

The first case, (i), is drawn directly from claim (4) and is useful for establishing the increase in an absolute k-degree LCE stability with dimension. For the dimensions we have considered in this report, the lowest dimension such that R_{min} (R_{min} is the minimum number of positive exponents) was not equal to zero is

$d = 64$ (we considered powers of two, i.e. $d = 4, 8, 16, 32, 64$). Thus, it is not until 64-dimensions that we can claim that for the ensemble of networks, the degree of k-degree robust chaos is at least one for this set of networks.

Case (i) is not particularly enlightening with respect to how the distributions are changing. It will turn out that method (iii) will yield similar issues for the small number of dimensions we are considering. Thus in the interest of a better understanding of the transition to k-degree LCE stability with $k > 0$, we will consider case (ii). The data for case (ii) can be read directly off figure (9.18); the degree of k-degree LCE stability, k, would be the lower bound on the error bars. With respect to (ii), k is increasing linearly with dimension. Method (ii) is a good measure of the k-degree LCE stability for the average network, and is enlightening with respect to how the distribution of the number of positive Lyapunov is changing, but using the lower bound on the standard deviation is not a particularly strong guarantee of robustness of chaos.

The third measure of k-degree robust chaos, (iii), is simply a more restrictive version of (ii). We have chosen to consider the lower bound on the curve under which 99 percent of the distribution of the number of positive Lyapunov exponents lies, which is roughly equivalent to three standard deviations. One could be less restrictive. In our case, it is not until we reach 32 dimensions that 99 percent of the distribution lies at one or above. For 64 dimensions, as we have already mentioned, 99 percent of the distribution lies at five or more positive exponents. Thus, with respect to (iii), once the degree of robust chaos becomes greater than one, it increases reasonably quickly with powers of two of dimension.

The point of all this is to show that as the dimension of a dynamical system in our construction is increased, the degree of the k-degree robust chaos increases. If the dimension is high enough, it becomes difficult to find periodic windows, and k is monotonically increasing.

9.7.4 Integrating data from sections (9.7.2) and (9.7.3)

Since we are using two slightly different data sets (different random seeds) and different numerical methods (computing the largest Lyapunov exponent versus the full spectrum of Lyapunov exponents), it will be useful to discuss apparent differences.

The most striking difference is in the existence of networks with periodic windows. In section (9.7.2) we claim to observe windows in 2 percent of networks perturbed. However, in section (9.7.3), we do not observe any networks for which periodic windows are present. This is due to two different factors. The first and most obvious is that the data set in section (9.7.3) is one-fifth the size of the data set in section (9.7.2). The data in section (9.7.2) is collected in 100 network chunks, and two of the five sets of 64-dimensional networks have no networks with periodic orbits. Moreover, considering the 64-dimensional networks in figure (9.21), this data set has very few networks with less than 5 positive exponents — consideration of another such set might likely have more networks with fewer exponents.

In section (9.7.2), we include results from one higher power of 2 than in section (9.7.3). This is largely due to the fact that the largest Lyapunov exponent is a far less costly calculation. In the periodic window data set for $d = 128$, out of the full 500 networks considered, only one network, for only one perturbation yielded a periodic window. It is highly possible that this is a numerical artifact. The implication is that between $d = 64$ and $d = 128$ there is a significant reduction in the number of networks with periodic windows. This is not surprising given the histograms in figures (9.19) and (9.20). What likely happens between 64 and 128 dimensions is that the range between the minumum and mean number of positive exponents becomes significantly less than the mean number of positive exponents, thus no "normal" 128-dimensional networks, for $s = 3$, will have perturbations that yield nearby periodic networks. Since 4-dimensional networks are contained in the set of 128-dimensional networks, there will clearly exist 128-dimensional networks for which perturbations will yield nearby networks that have periodic orbits — however this set of networks is, in a sense, degenerate. How to quantify this degeneracy remains a somewhat open question.

Aside from these two differences, the data sets from sections (9.7.2) and (9.7.3) are largely equivalent. If we were to show the figure for the data set from section (9.7.3) that is the equivalent of figure (9.14) from section (9.7.2), they would overlay well within numerical accuracy.

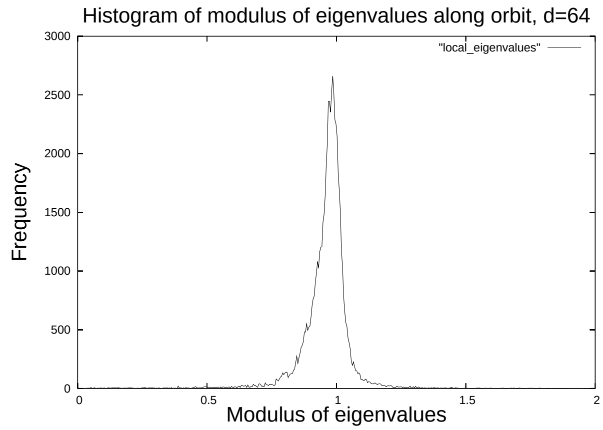

Figure 9.24: Histogram of the modulus of eigenvalues along the orbit for a typical network with 64 neurons and 64 dimensions; mean= 0.964, standard deviation= 0.1135

9.7.5 Existence of Spines: Distribution of the a_k's

We have already outlined a proof of conjecture (12), we will now put forth some numerical evidence for why we believe conjecture (12) is correct. Recall that a given network f will have a spine if and only if the a_i elements are all simultaneously zero for all $t > T$ for some positive T. Instead of presenting a large statistical study of the distribution of a_i's along an orbit for hundreds of networks, we will instead present one typical case, noting that for networks with $d > 32$ and $N > 32$, this is the typical scenario.

Considering figure (9.24), the number of a_k's that are anywhere near zero is extremely low. For a spine to exist, there would be a delta function at zero, which is clearly not the case. Further, it seems, based on the width of this histogram, that any small perturbation in the weights will not likely yield the collapse of all the a_i terms to zero, however, with this data we cannot argue that such a scenario is impossible. Of course, figure (9.24) does not give irrefutable evidence of conjecture (12), or that small perturbations in the parameters will not force all the a_i's to zero. Nevertheless, figure (9.24) contains our intuition and motivation for conjecture (12) since it represents the typical case [12] and because it seems unlikely that small perturbations of the parameters will yield a drastically different distribution of a_k's.

[12]We have observed on the order of 50 such histograms at various values of N and d ranging from 4 to 256.

Chapter 10

Regions III and V

We have so far ignored regions III and V, and we will now briefly discuss what is known about these two regions via a series of examples. Few results will be given; rather we will briefly discuss the dynamics followed by possible future directions for work.

10.1 Region III — Chaos to Bifurcation Chains

Region III, the region along the s-interval between the onset of chaos and the bifurcation chain region exists largely in low-dimensional networks where region IV is non-existent but it also exists in some high-dimensional networks with very few parameters. However, if the dimension of a network with few parameters is increased high enough, region IV appears, and with it, region III disappears.

We know very little of region III since this work is concerned with high-dimensional dynamical systems. A thorough understanding of region III in networks of intermediate dimension is quite interesting, and is a source of future endeavors. It is likely that an analysis using bifurcation tools such as CONTENT (70) will yield considerable insight into the general structure of this region.

One example for a network with no bias term (i.e. $\beta_0 = 0$) and $N = 2$, $d = 2$ is given in figure (10.1). In this figure, there is clearly no region IV, instead region III bleeds directly into region V. Figure (10.2), a close-up or region III in figure (10.1), displays the wonderful structure that exists in this network. One of the main points of our work is that this rich periodic window structure either disappears or becomes extremely rare and unobservable as the dimension of the dynamical systems is increased. Such diversity of periodic and chaotic dynamics, or even the very existence of region III, is very rare for high-dimensional networks.

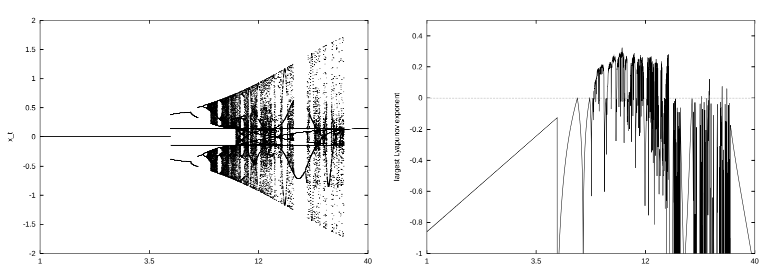

Figure 10.1: Bifurcation diagram with the largest Lyapunov exponent for $N = 2$ and $d = 2$ with $\beta_0 = 0$.

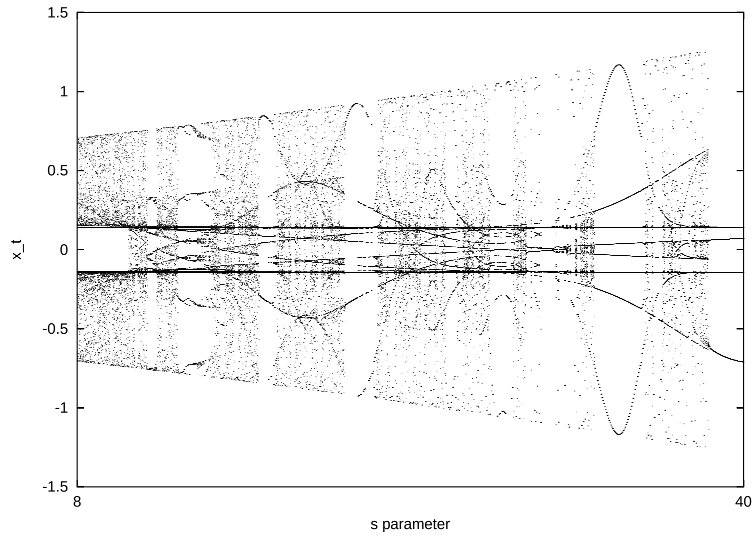

Figure 10.2: Close-up of the bifurcation diagram of region III of the example given in figure (10.1).

10.2 Region V — Bifurcation Chains to Binary Dynamics

Much like region III, region V is not well understood. For low-dimensional networks with few parameters, region V is most often trivial, however, for low-dimensional networks with many neurons, region V can be very rich. Nevertheless, low dimensional networks with few parameters can be shown to have behavior indicative of region V. To demonstrate the types of behavior in region V, we will consider two examples, the case given in figure (6.3) with $N = 4$ and $d = 64$ and the case given in figure (10.1) with $N = 2$ and $d = 2$.

Figure (10.3) shows the state diagram for the network given in figure (10.1) for $s = 12.2301$. When this network is truly binary, it can take any one of eight possible states which are combinations of $\pm(\beta_1 + \beta_2) = \pm 1.9949$, $\pm(\beta_1 - \beta_2) = \pm 0.1423$. For this s value, the network is chaotic, but the orbit dances around, and frequently returns very near the four points, $(0.1423, 0.1423)$, $(0.1423, -0.1423)$, $(-0.1423, 0.1423)$, $(-0.1423, -0.1423)$.

Figure (10.4), which is also chaotic, displays the same effect to a much more exaggerated degree. Here the network dances around on 64 different states. If one were to examine the time series, the orbit stays very near the various 64 states, eventually drifts away from the 64 states very briefly before returning to a different, nearly periodic orbit. Such behavior is clearly some type of intermittancy. With such behavior, it is difficult to interpret the meaning of the Lyapunov exponents since upon an increase in s, the deviations from the nearly periodic orbits become less frequent and a trajectory that is chaotic (or forever transient) could easily numerically produce a largest Lyapunov exponent average negative. It is of course even possible that these dynamical systems never really converge to any time average. Further, it is in this region there are very severe Milnor attractors and riddled basin issues are likely to exist (78) (66) (122). These are all important and yet unanswered questions.

One possible solution lies in a calculation using the anti-integrable limit akin to that used in (124) to show both a method of computing periodic orbits of the Hénon map and a method of showing the existence of a horseshoe akin to the result in (35). Such an application of the anti-integrable limit and the ensuing symbolic dynamics analysis to networks such as those in the aforementioned figures is a subject of future work.

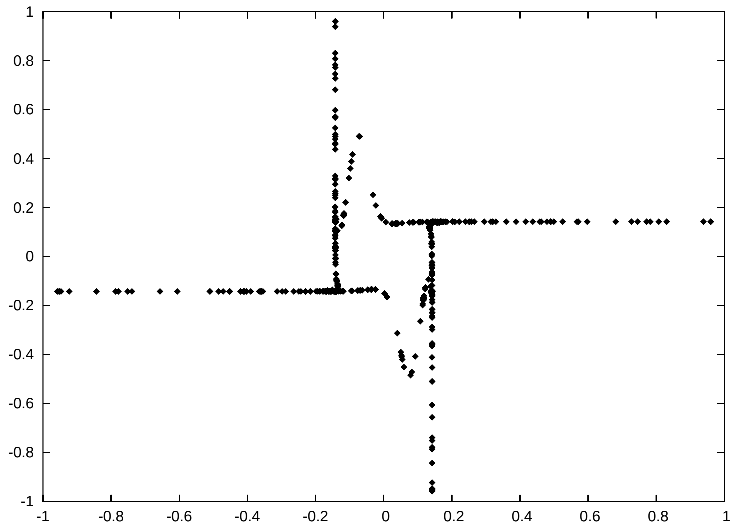

Figure 10.3: Plot of the state diagram for the network given in figure (10.1) at $s = 12.2301$

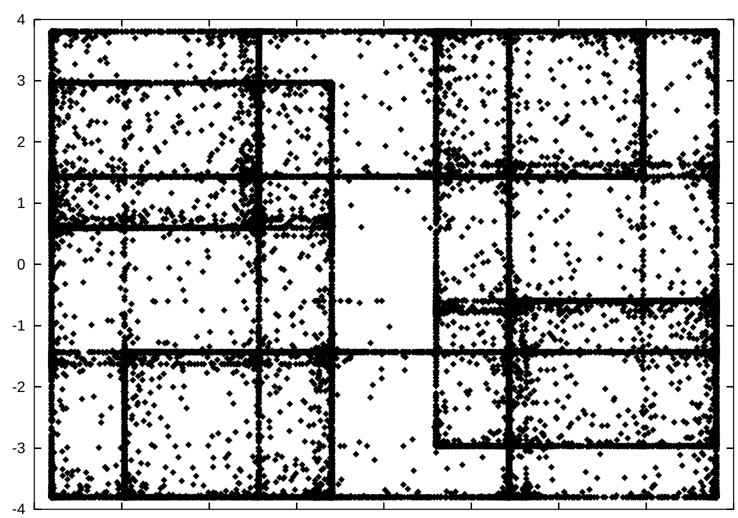

Figure 10.4: Plot of the state diagram for the network given in figure (6.3) at $s = 1.10079$

Chapter 11

Summary

Let us finally summarize by mentioning the major points of this thesis in two ways. First, we will state, in colloquial language, the major punch lines from this work. We will conclude with a list of "rules of thumb" that form our current intuition regarding high-dimensional dynamical systems.

11.1 Summary of major points

The major results are accumulated in chapters (7), (8), and (9). A summary of the general dynamical features of our study will be given in the next section.

First bifurcation: There are three generic, codimension one bifurcations from fixed points in maps of dimension two or higher. Two of these bifurcations are due to real eigenvalues, and one is due to complex eigenvalues. As the dimension of the dynamical system is increased, the real bifurcations tend to occur with equal probability and the bifurcation due to the complex conjugate pair of eigenvalues becomes the dominant bifurcation. In the limit of an infinite number of dimensions, the probability that the first bifurcation is due to the complex conjugate pair of eigenvalues, and thus a bifurcation to a quasi-periodic orbit or high-period periodic orbit, is above 90 percent. This work is contained in (8).

Route to chaos: As the dimension of a dynamical system is increased, the probability that the route to chaos from a fixed point is a cascade of bifurcations of and between quasi-periodic orbits, high-period periodic orbits, and tori of dimension as high as two increases to become the dominate route. As the dimension of the dynamical system increases, the length of the interval in parameter space between the first bifurcation and a chaotic orbit decreases. This work is contained in (5).

Hyperbolicity and chaos along an interval in parameter space: We find an interval in parameter space such that, as the dimension of the dynamical system is increased, the length of sub-intervals on which strict hyperbolicity is preserved decreases, each interval being interrupted by a point in the interval for which there exists a zero Lyapunov exponent. However, these points of "structural instability" are countable, and non-catastrophic. In general, in the region of parameter space where the dynamical system is chaotic, the chaos, the dynamic stability, or the persistence of chaos with perturbations increases with increasing dimension (in the dynamical system). This work is contained in (4).

Dynamic stability in high-dimensional dynamical systems: There exist several notions of dynamic equivalence on function spaces, with respect to parameters, with respect to initial conditions, and with respect to perturbations (parameter and function) to different initial conditions. We primarily are concerned with parameter perturbations. With respect to perturbations of all parameters, we have two key results. The probability of periodic windows existing for networks with dimensions $d = 2 - 64$

falls with dimension like $\sim \frac{1}{d}$. Moreover, the mean number of positive exponents is increasing linearly with dimension, and the variance in the mean number of the number of positive exponents remains relatively constant. Thus there exists a dimension, likely between $d = 64$ and $d = 128$, such that the lower tail of the distribution of positive exponents of an ensemble of perturbed networks no longer reaches zero. For dimensions higher than 64, the fall-off of the periodic window probability is likely much faster than $\frac{1}{d}$. In general, the number of degree of the k-degree LCE stability depends on the specific value of s, however, there exist regions of parameter space such that $\frac{k}{d}$ is significant. This work is contained in (6) and is a high-dimensional generalization of the results in (4).

11.2 Relationship between results from region IV and mathematical dynamics results

We will briefly discuss relationships with our ideas and results with respect to region IV with respect to those of others.

11.2.1 Windows conjecture

There are really two questions to address with respect to the windows conjecture. The first such question is in respect to the commonality of fragile attractors, i.e. given a high-dimensional dynamical system, what is the likelihood it will be fragile. The second question addresses the windows conjecture directly. Of the attractors that are fragile, when more parameters are varied than there exist positive Lyapunov exponents, is the windows set W dense. Or, if a dynamical system is fragile, what is the scaling between the existence of windows and the number of parameters varied.

With respect to the first question, our results show that the networks that have observed (stable) periodic windows likely have many more periodic windows than average and these networks become more rare as the dimension is increased. Thus, we would claim that fewer networks are fragile as the dimension is increased.

In our experiments, we varied all the parameters at once, thus, for our networks, the ratio of the number of parameters varied versus the maximum number of positive Lyapunov exponents is $N + (2N)/d + 1/d$. However, given figure (9.18), the former estimate is much more likely to be $4N + (8N)/d + 4/d$. In other words, at each perturbation, we are varying many more parameters than positive exponents. For the networks which are fragile, the window set seems not to be dense. Thus, for the s interval, $[0.5, 30]$, it is unlikely that the windows conjecture will hold. However, for much larger values of s, when the networks begin to become binary, the windows conjecture is much more likely to be reasonable.

From the perspective of the mathematical dynamics conjectures and theorems, the number of parameters is irrelevant; it is the geometry of how the stable, unstable, and center manifolds are pieced together that determines the resistance to a change in the dynamics. How parameter variation, as opposed to general C^r $(r > 0)$ perturbations, affect and constrain dynamic stability is an very important open question.

11.2.2 Structural stability

The structural stability theorem of Robbin, Robinson and Mañé requires two properties, strong transversality and axiom A. It is nearly impossible to verify the strong transversality condition numerically, thus we will refrain from such a discussion. Considering (4), it is likely that for $s = 3$, the high dimensional networks are very near to a parameter value such that there exists a zero exponent. However, this set of parameter values is likely measure zero with respect to Lebesgue measure in R^p. If our chaotic networks satisfied the hypotheses of the structural stability theorem, they would certainly not be fragile, and thus any C^r $(r > 0)$ perturbation, which included every parameter perturbation, would make the given network periodic. However, the neighborhood of the network such that the structural stability would hold could be vanishingly small.

Understanding the specific nature of the high-dimensional networks which yield periodic orbits upon parameter perturbations, with respect to their relationship to axiom A is an important subject of future work. Such an understanding would aid greatly in a practical interpretation of the structural stability theorem. We find networks which likely satisfy axiom A, yet there do exist parameter perturbations for some of the networks that yield many periodic windows. Understanding how axiom A and the dependence on dimension affect the overall observed dynamic stability would be interesting and useful.

What we do know is that many of our dynamical systems seem not to require large parameter perturbations to destroy topological equivalence since the distribution in the number of positive Lyapunov exponents is not a zero variance spike. Along these lines, how the variance scales with perturbation size would serve as a useful study. That our dynamical systems do not obey the stability theorem for reasonably small perturbations is not fatal because the disobedience is not severe, especially as the dimension is increased.

11.2.3 Closing lemma

Speaking to the closing lemma numerically is not particularly feasible in our current framework; testing for the existence of a single periodic point that yields a C^2 approximation is difficult. However, testing for the observable effects of the closing lemma is something that lies in the scope of this work. One of our key conclusions which is missing from the mathematical dynamics community is that the number of dimensions of the dynamical systems matters. There is a huge difference between high and low-dimensional systems. Of course, for a physicist, this seems obvious, but this point is lacking from most of the conjectures in the mathematical dynamics community. Of course, why this is the case is due to difficulties with respect to defining dimension, and devising a notion that captures the difference between dimensions effectively. Nevertheless, we would claim that the effects of a closing type lemma are significantly more difficult to detect as the dimension of the dynamical system is increased.

One useful direction which we have neglected, but which our neural network framework would be able to investigate is a construction or a numerical implementation of Pugh and Robinson's C^1 approximation technique. Such a study would not only aid in an understanding of our high-dimensional networks with windows, but it would also aid in a better understanding of high-dimensional dynamical systems in general.

11.2.4 Stable ergodicity

The theory of stable ergodicity reflects the beliefs and observations of the authors' of this report most closely of all the mathematical dynamics work. Non-trivial partial hyperbolicity is likely not particularly uncommon; strict hyperbolicity is likely the most common behavior, however it might be perturbed away with relative ease. There are two issues we would like to specifically address. The first is that stable ergodicity, and specifically the conjecture of Pugh and Shub (conjecture (1). The second is with respect to convergence in dynamical and physical systems.

There is little we are likely to be able to say about strict ergodic behavior, dynamic coherence, and accessibility. Partial hyperbolicity, uniform hyperbolicity, and non-uniform hyperbolicity are however, quantities we can and will measure. The Pugh-Shub conjecture (and some of Palis's conjectures (91)) are aligned with our results both in spirit and in stylistic features. With respect to our results, it is apparent that there exists some type of dynamic stability, persistent chaos, with various neutral directions around for good measure. As the dimension is increased, the dynamic stability increases, and take on traits that resemble stable ergodicity.

The second issue, which is noted in (113), is the general problem with convergence times of quantities like Lyapunov exponents in dynamical and physical systems. There are many systems for which there will not exist any real notion of convergence; many dynamical systems that reflect nature may exist in a continuous state of transients. This will likely be a problem for stable ergodicity, since if there is no convergence, many issues arise with some of the current proofs, e.g. the existence almost everywhere (with respect to Lebesgue measure) of ergodic limits. Strict Lyapunov exponents will also not likely exist in the current sense. The neural networks of course can approximate such systems, all that needs be done is training on data, at which point neural networks can produce the localized Lyapunov exponents (Lyapunov exponents might not converge in the strict sense) for the system they are modeling (42) on that data set. Thus, one line of investigation might follow training neural networks on data from a complex system and monitoring the weight distribution. From our experiments, and experiments like ours, given a weight distribution, it is likely that we will be able to predict behavior types. In other words, there exists the possibility for a much more weak notion of equivalence between dynamical systems than has been so far discussed in this report. One could speak of distributions of weight distributions in neural networks. Or, the evolution of the mean, variance and any other statistical notions with respect to the weight distributions. If networks can be understood in a statistical manner with respect to weight distributions, then insight might be gained with respect to dynamical and physical systems for which long-term convergence of quantities such as Lyapunov exponents is troublesome. In this report we present a partial understanding of neural networks with our weight distribution; in preliminary work, we claim that our results are not particularly sensitive to many of the properties of the weight distributions aside from the mean. If the mean is non-zero, the neurons quickly saturate, and the networks become periodic.

11.3 Lore

The overall dynamics of our set of neural networks covers an enormous range of behaviors and qualitative types of dynamical systems. From a general study of these dynamical systems via the analysis given in this work as well as studying hundreds of individual networks, we have developed several rules of thumb that comprise much of our intuition for high-dimensional dynamical systems. These rules are not strict, and there clearly exist counterexamples. However, these are the stylized facts from our study. Many of these results will be contained in a future report.

Rules of thumb

1. As the dimension is increased, the probability of a network being chaotic over some portion of its parameter space goes to unity.

2. For the networks considered, at $d = 50$ there exists an s value such that the maximum probability of chaos is about fifty percent.

3. As the dimension is increased, the (maximum) Kaplan-Yorke dimension of chaotic attractors scales like $\sim \frac{d}{2}$.

4. As the dimension is increased, the probability of the first bifurcation being Neimark-Sacker approaches unity (or is at least greater than 90 percent)

5. As the dimension is increased, the dominant route to chaos from a fixed point along a one-dimensional interval is the quasi-periodic route involving tori (T^2 for $d = 64$) and quasi-periodic orbits;

6. As the dimension is increased, the probability (and length) of periodic windows in parameter space scales like $\frac{1}{d}$.

7. As the dimension is increased, the number of positive exponents scales like $\sim \frac{d}{4}$.

8. As the dimension is increased, in the chaotic region of parameter space, hyperbolicity is violated on a countable, increasingly countable, "dense" interval in parameter space.

9. As the dimension is increased chaos becomes more robust with respect to parameter changes.

11.4 Future work

We will conclude with a list of some future and current projects and suggested extensions of this work.

1. Considering figures (9.4) and (9.9), there is likely a scaling between the positive Lyapunov exponents. If there exists a single scaling such that all the Lyapunov exponents can be collapsed onto a single line, there is much we can study. In other words, it is possible that the Lyapunov exponent spectrum is scale invariant (see (31) for more information regarding scale invariance in dynamical systems). Armed with a better understanding of such systems, high-dimensional asymptotic claims will be much easier to make.

2. Study the universality of the largest Lyapunov exponent curve.

3. Study the existence and prevalence of uniform and non-uniform partial hyperbolicity along the orbit of high-dimensional neural networks. Compare this with results of both Pesin and Bonatti and Viana (19) regarding SRB measures and Lyapunov exponents.

4. A study regarding the basins of attraction and multiple attractors for our neural networks is badly needed. This would include a study of the existence of finitely many SRB measures as has been suggested by Palis (91), and the prevalence of Milnor type attractors as suggested by Kaneko (66).

5. Perform a full statistical study of the routes to chaos supplementing chapter (8), pending a better orthonormalization routine.

6. A symbolic dynamics, anti-integrable limit study of region V is desired. In other words, a detailed study of the transition from chaos to finite state orbits.

7. Training the networks to high-dimensional experimental and numerical data sets. With this information, an analysis of common weight distributions can be uncovered, and with that information comes a general picture of how our results are representative in nature.

8. A generalized notion of dynamic stability in systems never allowed to time to converge to an ergodic-type limit. An example of such a system in our framework would involve a slow evolution of the weights or the weight distribution. Such an analysis would aid in an understanding of the nature of dynamics that are, in a sense, forever transient.

9. A study of the robustness of these results with respect to the weight distribution would provide insight into how general our results are with respect to the neural network framework. Such a study would also likely prove useful upon considering item 7, the analysis of these results with respect to networks trained on data sets.

Bibliography

[1] R. Adams and J. J. F. Fournier. *Sobolev Spaces*. Elsevier, 2nd edition, 2003.

[2] D. J. Albers and J. C. Sprott. Dynamic stability of high-dimensional dynamical systems. Submitted.

[3] D. J. Albers and J. C. Sprott. Point intermittant chaos. contact albers@santafe.edu for notes.

[4] D. J. Albers and J. C. Sprott. Structural stability and hyperbolicity violation in large dynamical systems. in preparation.

[5] D. J. Albers and J. C. Sprott. Routes to chaos in high-dimensional dynamical systems: A qualitative numerical study. *submmitted*, 2004.

[6] D. J. Albers, J. C. Sprott, and J. P. Crutchfield. A dynamics stability conjecture for high-dimensional dynamical systems. in preparation.

[7] D. J. Albers, J. C. Sprott, and W. D. Dechert. Dynamical behavior of artifical neural networks with random weights. In *Intelligent Engineering Systems Through Artifical Neural Networks*, volume 6, pages 17–22. ASME, 1996.

[8] D. J. Albers, J. C. Sprott, and W. D. Dechert. Routes to chaos in neural networks with random weights. *Int. J. Bif. Chaos*, 8:1463–1478, 1998.

[9] D. V. Anosov. Geodesic flows on closed Riemannian manifolds with negative curvature. *Proc. Steklov Inst. Math.*, 90:1–235, 1967.

[10] V. Arnold. *Geometric methods in the theory of ordinary differential equations.* Grundlehrn de mathematischen Wissenschaften. Springer-Verlag, 2^{nd} edition, 1983.

[11] V. I. Arnold. *Ordinary differential equations*. Springer-Verlag, 1992.

[12] V. I. Arnold, editor. *Dynamical systems V.* Springer-Verlag, 1994.

[13] D. K. Arrowsmith and C. M. Place. *An introduction to dynamical systems*. Cambridge University Press, 1990.

[14] Z. D. Bai. Circular law. *Ann. Probab.*, 25:494–529, 1997.

[15] Luis Barreira and Ya. Pesin. *Lyapunov exponents and smooth ergodic theory.* AMS, 2002.

[16] E. Barreto, B. Hunt, C. Grebogi, and J. A. Yorke. From high dimensional chaos to stable periodic orbits. *Phys. Rev. Lett.*, 78:4561–4564, 1997.

[17] Giancarlo Benettin, Luigi Galgani, Antonio Giorgilli, and Jean-Marie Strelcyn. Lyapunov characteristic exponents from smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Meccanica*, 15:9–20, 1979.

[18] Giancarlo Benettin, Luigi Galgani, Antonio Giorgilli, and Jean-Marie Strelcyn. Lyapunov characteristic exponents from smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 2: Numerical application. *Meccanica*, 15:21–30, 1979.

[19] C. Bonatti and M. Viana. SRB measures for partially hyperbolic systems whose central direction is most contracting. *Israel J. Math.*, 115:157–194, 2000.

[20] B. Braaksma, H. Broer, and G Huitema. Toward a quasi-periodic bifurcation theory. *Mem. Amer. Math. Soc.*, 83(421):83–175, 1990.

[21] M. I. Brin and Ja. B. Pesin. Partially hyperbolic dynamical systems. *English transl. Math. Ussr Izv*, 8:177–218, 1974.

[22] H. Broer. Coupled hopf-bifurcations: persistent examples of n-quasiperiodicity determined by families of 3-jets. In *Geometric methods in dynamics (I)*, volume 286 of *Astérisque*, pages 223–229, 2003.

[23] H. W. Broer, G. B. Huitema, and M. B. Sevryuk. *Quasi-periodic motions in families of dynamical systems*, volume 1645 of *Lecture Notes in Mathematics*. Springer-Verlag, 1996.

[24] H. W. Broer, G. B. Huitema, and F. Takens. Unfoldings and bifurcations of quasi-periodic tori. *Mem. Amer. Math. Soc.*, 83(421):1–81, 1990.

[25] R. Brown, P. Bryant, and H. Abarbanel. Computing the Lyapunov spectrum of a dynamical system from an observed time-series. *Phys. Rev. A.*, 43:2787–2806, 1991.

[26] P. Brunovsky. On one parameter families of diffeomorphisms. *Commentationes mathematicae Universitatis Carolinae*, 11:559–582, 1970.

[27] K. Burns, D. Dolgopyat, and Y. Pesin. Partial hyperbolicity, Lyapunov exponents and stable ergodicity. *J. Stat. Phys.*, 108:927–942, 2002.

[28] K. Burns and A. Wilkinson. Stable ergodicity of skew products. *Ann. Sci. École Norm. Sup.*, 32:859–889, 1999.

[29] A. Chenciner and G. Iooss. Bifurcaions de tores invariants. *Arch. Rat. Mech. Anal.*, 69:109–98, 1979.

[30] S. Chern and S. Smale, editors. *Global Analysis*, volume 14 of *Proc. Sympos. Pure Math.*, Bekeley, Ca, July 1968 1970. A.M.S.

[31] James P. Crutchfield. *Noisy Chaos*. PhD thesis, University of California, Santa Cruz, 1983.

[32] W. D. Dechert, J. C. Sprott, and D. J. Albers. On the probability of chaos in large dynamical systems: A monte carlo study. *J. Econ. Dynamics and Control*, 23:1197–1206, 1999.

[33] W. Davis Dechert and Ramazan Gencay. The topological invariance of Lyapunov exponents in embedded dynamics. *Physica D*, 90:40–55, 1996.

[34] R. L. Devaney. *An Introduction to Chaotic Dynamical Systems*. Addison-Wesley, Redwood City, Calif., 2nd edition, 1989.

[35] R. L. Devaney and Z. Nitecki. Shift automorphisms in the henon mapping. *Commun. Math. Phys.*, 67:137–146, 1979.

[36] B. Doyon, B. Cessac, M. Quoy, and M. Samuelides. "Control of the Transition to Chaos in Neural Networks with Random Connectivity". *IJBC*, 3:279–291, 1993.

[37] J.-P. Eckmann. Roads to turbulence in dissapative dynamical systems. *Rev. Mod. Phys.*, 53:643–654, 1981.

[38] J. P. Eckmann, S. Kamphorst, D. Ruelle, and S. Ciliberto. Liapunov exponents from time-series. *Phys. Rev. A.*, 34:4971–4979, 1986.

[39] A. Edelman. The probability that a random gaussian matrix has k real eigenvalues, related distributions, and the circular law. *J. Multivariate Anal.*, 60:203–232, 1997.

[40] A. Edelman, E. Kostlan, and M. Shub. How many eigenvalues of a random matrix are real? *J. Amer. Math. Soc.*, 7:247–267, 1994.

[41] K. Geist, U. Parlitz, and W. Lauterborn. Comparison of different methods for computing Lyapunov exponents. *Prog. Theor. Phys.*, 83(5):875–893, 1990.

[42] Ramazan Gencay and W. Davis Dechert. An algorithm for the n Lyapunov exponents of an $n-$dimensional unknown dynamical system. *Physica D*, 59:142–157, 1992.

[43] J. Ginbre. Statistical ensembles of complex, quaternion and real matrices. *J. Math. Phys.*, 6:440–449, 1965.

[44] V. Girko. Circular law. *Theory Probab. Appl.*, 29:694–706, 1984.

[45] V. Girko. *Theory of random determinants*. Kluer Academic, 1990.

[46] V. Girko. The circular law: ten years later. *Random Oper. and Stoch. Eqns.*, 2:235–276, 1994.

[47] V. Girko. Strong circular law. *Random Oper. and Stoch. Eqns.*, 5:173–196, 1997.

[48] V. Girko. The V-density of eigenvalues of non-symmetric random matrices and rigorous proof of the strong circular law. *Random Oper. and Stoch. Eqns.*, 5:371–406, 1997.

[49] G. Golub and C. Van Lorn. *Matrix computations*. The Johns Hopkins University Press, 3^{rd} edition, 1996.

[50] M. Golubitsky and V. Guillemin. *Stable mappings and their singularities*. Springer-Verlag, 1973.

[51] J. Graxzyk and G. Światek. *The real Fatou conjecture*. Princeton univeristy press, 1998.

[52] E. Hebey. *Nonlinear analysis on manifolds: Sobolev spaces and inequalities*. American Mathematical Society, 1999.

[53] M. Hénon. A two-dimensional mapping with a strange attractor. *Comm. Math. Phys.*, 27:291–312, 1976.

[54] N. J. Higham. *Accuracy and stability of numerical algorithms*. SIAM, 1996.

[55] M. Hirsch. *Differential Topology*. Graduate studies in mathematics. Springer-Verlag, 1976.

[56] M. W. Hirsch, C. C. Pugh, and M. Shub. *Invariant Manifolds*, volume 583 of *Springer Lecture Notes in Mathematics*. Springer-Verlag, 1977.

[57] E. Hopf. A mathematical example displaying the features of turbulence. *Comm. Pure App. Math*, 1:303–322, 1948.

[58] R. Horn and C. Johnson. *Matrix analysis*. Cambridge univerisity press, 1985.

[59] K. Hornik, M. Stinchocombe, and H. White. "Mulitlayer Feedforward Networks are Universal Approximators". *Neural Networks*, 2:359–366, 1989.

[60] K. Hornik, M. Stinchocombe, and H. White. "Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks". *Neural Networks*, 3:535–549, 1990.

[61] G. Iooss and M. Adelmeyer. *Topics in bifurcation theory and applications*. World Scientific, 1998.

[62] G. Iooss and W. F. Langford. *Conjectures on the routes to turbulence via bifurcation*. New York Academy of Sciences, 1980.

[63] G. Iooss and J. E. Los. Quasi-genericity of bifurcations to high dimensional invariant tori for maps. *Commum. Math. Phys.*, 119:543–500, 1988.

[64] M. Jakobson. Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. *Commun. Math. Phys.*, 81:39–88, 1981.

[65] K. Burns and C. Pugh and M. Shub and A. Wilkinson. Recent results about stable ergodicity. In *Smooth ergodic theory and its applications*, volume 69 of *Proc. Sympos. Pure Math.*, pages 327–366. Amer. Math. Soc., 2001.

[66] Kunihiko Kaneko. Dominance of milnor attractors in globally coupled dynamical systems with more than 7 ± 2 degrees of freedom. *Phys. Rev. E*, 66:055201, 2002.

[67] J. Kaplan and J. Yorke. *Chaotic behavior of multidimensional difference equations*, volume 730 of *Lecture notes in mathemtics*, pages 228–37. Springer-Verlag, 1979.

[68] A. Katok. Lyapunov exponents, entropy, and periodic orbits for diffeomorphisms. *Publ. Math. I.H.E.S.*, 51:137–174, 1980.

[69] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, 1995.

[70] Y. Kuznetzov. *Bifurcation Theory*. Springer-Verlag, 2^{nd} edition, 1998.

[71] D. Landau and L. Lifshitz. *Fluid Mechanics*. Pergamon Press, 1^{st} edition, 1959.

[72] P. L'Ecuyer. *Communications of the ACM*, 31:742–774, 1988.

[73] Loéve. *Probability Theory I*. Springer-Verlag, 1977.

[74] E. N. Lorenz. Deterministic nonperiodic flow. *J. Atmosph. Sci.*, 20:130–141, 1963.

[75] R. Mañé. A proof of the C^1 stability conjecture. *Publ. Math. IHES*, 66:161–210, 1988.

[76] J. Milnor. *Topology from the differential viewpoint*. Princeton University Press, 1997.

[77] J. Milnor and W. Thurston. *On iterated maps of the interval*, volume 1342 of *Lecture notes in mathematics*. Springer-Verlag, 1988.

[78] John Milnor. On the concept of attractor. *Commun. Math. Phys.*, 1985.

[79] R. Ma né. An ergodic closing lemma. *Ann. of Math.*, 1982.

[80] Ricardo Ma né. *Ergodic theory and differentiable dynamics*. Springer-Verlag, 1987.

[81] J. Neimark. On some cases of periodic motions depending on parameters. *Dokl. Acad. Nauk SSSR*, 129:736–739, 1959.

[82] V. V. Nemytskii and V. V. Stepanov. *Qualitative theory of differential equations*. Dover, 1989.

[83] S. Newhouse. The abundance of wild hyperbolic sets and nonsmooth stable set for diffeomorphisms. *Publ. Math. IHES*, 50:101–151, 1979.

[84] S. Newhouse, J. Palis, and F. Takens. Bifurcation and stability of families of diffeomorphisms. *IHES*, 57:5–71, 1983.

[85] S. Newhouse, D. Ruelle, and F. Takens. Occurrence of strange Axiom A attractors near quasiperiodic flows on T^m, $m \leq 3$. *Comm. Math. Phys.*, 64:35–40, 1978.

[86] L. Noakes. The Takens embedding theorem. *IJBC*, 1991.

[87] V. I. Oseledec. A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Tras. Moscow Math. Soc.*, 19:197–221, 1968.

[88] E. Ott. *Chaos in dynamical systems*. Cambridge University Press, 1993.

[89] J. Palis. Vector fields generate few diffeomorphisms. *Bull. Amer. Math. Soc.*, 80:503–505, 1974.

[90] J. Palis and S. Smale. Structural stability theorems. In *Global Analysis*, volume 14 of *Proc. Sympos. Pure Math.* A.M.S., 1970.

[91] Jacob Palis. A global view of dynamics and a conjecture on the denseness of finitude of attractors. *Asterisque*, 261:339–351, 2000.

[92] M. Peixoto. Structural stability on two-dimensional manifolds. *Topology*, 1:101–120, 1962.

[93] M. M Peixoto and C. C. Pugh. Structurally stable systems on open manifolds are never dense. *Ann. of Math.*, 87:423–430, 1968.

[94] L. Perko. *Differential equations and dynamical systems*. Springer-Verlag, 2000.

[95] Y. Pesin. *Lectures on partial hyperbolicity and stable ergodicity*. European Mathematical Society, 2004.

[96] Ya. B. Pesin. Invariant manifold families which correspond to non-vanishing characteristic exponents. *English transl. Math. USSR Izv.*, 10:1261–1305, 1976.

[97] Ya. B. Pesin. Lyapunov characteristic exponents and ergodic properties of smooth dynamical systems with an invariant measure. *English transl. Soviet math. Dokl*, 17:196–199, 1976.

[98] Ya. B. Pesin. Lyapunov characteristic exponents and smooth ergodic theory. *English transl. Russian Math. Surveys.*, 32:55–114, 1977.

[99] K. Peterson. *Ergodic theory.* Cambridge University Press, 1983.

[100] W. A. Press and S. A. Teukolsky. Portable random number generators. *Computers in Physics*, 6:522–524, 1992.

[101] C. Pugh. The closing lemma. *Amer. J. Math.*, 89:956–1009, 1967.

[102] C. Pugh. Against the C^2 closing lemma. *J. Diff. Eqns.*, 17:435–443, 1975.

[103] C. Pugh and C. R. Robinson. The C^1 closing lemma, including Hamiltonians. *Ergod. Th. Dynam. Sys.*, 3:261–313, 1983.

[104] C. Pugh and M. Shub. Stably ergodic dynamical systems and partial hyperbolicity. *J. of Complexity*, 13:125–179, 1997.

[105] C. Pugh and M. Shub. Stable ergodicity and Julienne quasiconformality. *J. Eur. Math. Soc.*, 2:1–52, 2000.

[106] C. Pugh and M. Shub. Stable ergodicity. *Bull. Amer. Math. Soc.*, 41:1–41, 2003.

[107] J. Robbin. A structural stability theorem. *Annals math.*, 94:447–493, 1971.

[108] C. Robinson. C^r structural stability implies Kupka-Smale. In *Dynamical Systems*, pages 443–449, Salvador 1971, 1973. Academic Press.

[109] C. Robinson. Structural stability of C^1 diffeomorphisms. *J. Diff. Eq.*, 22:28–73, 1976.

[110] C. R. Robinson. *Introduction to the closing lemma, structure of attractors in dynamical systems*, volume 668 of *Lecture notes in mathematics*, pages 223–230. Springer-Verlag, 1968.

[111] D. Ruelle. Ergodic theory of differentiable dynamical systems. *Publ. Math. I.H.E.S.*, 50:27–58, 1979.

[112] D. Ruelle. Characteristic exponents and invariant manifolds in Hilbert space. *Ann. of math.*, 115:243–290, 1982.

[113] D. Ruelle. Some ill-formulated problems on regular and messy behavior in statistical mechanics and smooth dynamics for which I would like the advise of Yasha Sinai. *J. Stat. Phys.*, 108:723–728, 2002.

[114] D. Ruelle and F. Takens. A route to turbulence. *Comm. Math. Phys.*, 20:167–192, 1970.

[115] R. Sacker. On invariant surfaces and bifurcation of periodic solutions of ordinary differential equations. Technical Report 333, New York State University, 1964.

[116] M. Sano and Y. Sawada. Measurement of the Lyapunov spectrum from a chaotic time-series. *Phys. Rev. Lett.*, 55:1082–1085, 1985.

[117] T. Sauer and J. Yorke. Reconstructing the Jacobian from data with observational noise. *Phys. Rev. Lett.*, 83:1331–1334, 1999.

[118] I. Shimada and T. Nagashima. A numerical approach to ergodic problems of dissapative dynamical systems. *Prog. Theor. Phys.*, 61:1605–1635, 1979.

[119] Ya. Sinai. *Topics in ergodic theory.* Princeton University press, 1994.

[120] S. Smale. Structurally stable systems are not dense. *American Journal of Mathematics*, 88:491–496, 1966.

[121] S. Smale. Differentiable dynamical systems. *Bull. A.M.S.*, 73:747–817, 1967.

[122] J.C. Sommerer and E. Ott. A qualitatively nondeterministic physical system. *Nature*, 365:135, 1993.

[123] Jorge Sotomayor. Structural stability and bifurcation theory. In *Dynamical Systems*, pages 549–560, 1973.

[124] D. Sterling and J. D. Meiss. Computing periodic orbist using the anti-integrable limit. *Phys. Lett. A*, 241:46–52, 1998.

[125] F. Takens. Detecting atrange attractors in turbulence. In D. Rand and L. Young, editors, *Lecture Notes in Mathematics*, volume 898, pages 366–381, Dynamical Systems and Turbulence, Warwick, 1981. Springer-Verlag, Berlin.

[126] H. R. von Bremen, F. E. Udwadia, and W. Proskurowski. An efficient QR based method for the computation of Lyapunov exponents. *Physica D*, 101:1–16, 1997.

[127] P. Walters. *An introduction to ergodic theory.* Graduate texts in mathematics. Springer-Verlag, 1982.

[128] S. Wiggins. *Global bifurcations and chaos: Analytical methods.* AMS. Springer-Verlag, 1988.

[129] R. F. Williams. The DA maps of smale and structural stability. In *Global Analysis*, volume 14 of *Proc. Sympos. Pure Math.* A.M.S., 1970.

[130] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano. Determining Lyapunov exponents from a time-series. *Physica D*, 16:285–317, 1984.

[131] L-S Young. What are SRB measures, and which dynamical systems have them? *J. Stat. Phys.*, 108:733–754, 2002.

Appendix A

Acknowledgments

I am deeply indebted to both of my advisors, Clint Sprott and Jim Crutchfield. I have been studying under Clint since 1994 and he has influenced my work in every way; he was my first teacher with regards to dynamical systems and his experimental physics mindset has had a major impact on my thinking. There are not enough words for what Professor Sprott has given me. I have also been studying under Jim Crutchfield since 2001. Again, there are not words available to describe the level of help I have received from Jim. Professor Crutchfield's intuitive, geometric, global view of dynamical systems helped form much of my thinking on the subject and much of this work. His guidance has been invaluable. I would like to thank both of these wonderful people for their endless patience and kindness. Without the understanding and skill of these two people, I would have gone nowhere.

Other professors who have had a major influence on this work and my thinking include Joel Robbin, Sergey Bolitin, Ian Dobson, Dave Feldman, Dee Dechert, Buz Brock, Moe Hirsch and Keith Burns. Joel has spent many hours explaining many wonderful results from dynamical systems as well as giving me a historical context from which to work. By doing this Professor Robbin opened many doors for me, including an introduction to Moe. Moe Hirsch also helped considerably in pointing out important results that I had overlooked. More importantly (and generously) Moe spent many afternoons patiently explaining important examples of structural instability (counter-examples to the stability dream) which had a large impact on the present work. Dave Feldman always helped reign in my abstractions. Several times I would confront him with a complicated abstract mess and would leave two hours later with an intuitive, useful explanation of my results. Dee Dechert was one of my first teachers with respect to nonlinear dynamics. He aided much in my understanding, patiently corrected my misconceptions, and generously read and commented on my (and our) work. Buz Brock spent many hours sketching results from topology, ergodic theory, dynamical systems, game theory, mathematical finance, and mathematical economics in general. His enthusiasm was boundless and his unique insight into dynamics problems had a big impact on my thinking. Ian Dobson taught my first formal, mathematical dynamics course. Ian's interdisciplinary approach, integrating abstract mathematics with the world is deeply embedded in this work. There were many times when I was initially attempting to formulate some of the results in this thesis that he gave me guidance and helped me begin to understand many things I was, until then, completely unconscious of. Furthermore, his careful reading of my thesis has made it considerably better both in terms of general content and with respect to my errors. Keith Burns gave generously of his valuable time, he read my manuscripts, he patiently corrected my poor mathematical prose, and he spent many hours explaining current dynamics results to me. Without his help, my thinking would both be much less clear and much less complete. His willingness to thanklessly help a physics graduate student try to understand his mathematics and the mathematics of others was wonderful. Lastly, I'd like to thank the entire Dynlearn group at the Santa Fe Institute for the patience sitting though my talks and all the helpful insight and conversation.

I also received generous help from many fellow students (some of which are postdocs now). The endless hours spent discussing physics, mathematics, turbulence, and everything else with Adam Bayliss have been wonderful and have greatly influenced this work. Adam taught me so much and always pulled me out of an abstract world and back to a more physical one. The little I know about turbulence can be attributed to his expertise. I'd like to thank Jodi Supanich for the endless discussions about my mathematical problems and constructs, much of my thinking was made more clear through her insight. Battling through differential topology, analysis and ordinary differential equations with her aided much in my limited understanding of mathematics. The hours discussing topology, geometry, computation theory, and various aspects of both our various projects with Carl McTague helped more than I can say. His desire for clarity, his patience, and his comments on my various results have been invaluable. Cosma Shalizi has been around since I began thinking about mathematics and physics, and his influence and guidance has been far-reaching. Cosma has taught me about everything from quantum mechanics, to probability theory, to how to compile my code using gcc. He patiently answered my most naive questions with grace. In my early years as a graduate student, Magda Gonzales helped a great deal in clarifying how to think and make physical arguments. When I was bogged down in mathematical formalism, she always pulled me back into reality (as did Adam later in my graduate career). Finally, my brother John read much of my work and provided many useful comments. Further, he always demanded intuitive explanations of my arguments which caught many mistakes and aided in many clarifications.

Marcus Daniels, Derek Wright, and Carl McTague gave me endless help writing code and in general with computing. All three of these people saved me years of pain with their wonderful skill.

I would like to thank my two sets of parents, Mom and Wayne and Dad and Linda for all the love and support. My surrogate parents, Bonnie and Ronnie Fingerhut always provided me with a home away from home out on the farm when I needed space to think. I'd like to thank my brother John for always supporting me, putting up with my impatience, editing my work, and for all the support over the years. I'd like to thank my cousin Andy who gave me a talk in eighth grade that made all the difference. Of course, I'd like to thank my external family, John Dopp, Jon Zeisser, Marin Dobson, Ross Spitz, Rachel Senseman, Coco Ballantyne, and Rob Lauer for 20 years of friendship and always being there when life looked bleak (and when it looked great). Terry Chrudinsky spent hours driving us all over Colorado to climb rock and ice at a time when I had little time for anything but work. Several chapters were composed in the passenger seat while he drove through blizzards and over snow bound passes. These trips kept me sane while writing this thesis. Without the generious help from the Durands (Bernice and Loyal), I would never have passed my qualifying exams and thus would not be writing this thesis. I learned a lot of physics from them, including my first global vision of how a field was built up. Other folks around the Madison physics department (and the U. W. in general) who have graciously given their support include (in no particular order): Ugo Camerini, Aimee Lefkow, Jean Buehlman, Fatima Ebrahimi, Chris Odell, Ben Hudson, Steve Peterson, Tyler Rooker, Travis Warwick, Bill Grogan, Professor Richard Davis, Professor Charles Sih, Professor George Zografi, and the plasma physics group (for putting up with me in their space). There are many people at the Santa Fe Institute who made my time there very happy, and how helped me keep my head up (in no particular order): Chris Lantz, Cate Griebel, Sarah Knutson, Ginger Richardson, Andi Surtherland, Ronda Butler-Villa, Olivia Posner, Tim Carlson, Alex Herman, Van Savage, Todd Kaplan, Erik van Nimwegen, Bae Smith, Harry Owen, Carolyn Resnicke, Lara Ware, and John Miller are but a few of the wonderful people who made my life better at SFI. Lastly, I'd like to thank my grandparents for their love, support, and inspiration. Words cannot describe what they have given me. Without these people, I would not have written this thesis.

Portions of the computing for this project was done on the Condor pools at the University of Wisconsin at Madison. The majority of the computing for this 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 D. J. Albers was provided by NSF grants DMR-9820816 and PHY-9910217 and DARPA Agreement F30602-00-2-0583. Further support for D. J. Albers was provided by the University of Wisconsin Graduate School.