Journal of Statistical Physics, Vol. 125, No. 4, November 2006 (© 2006)
DOI: 10.1007/s10955-006-9232-6

Probability of Local Bifurcation Type from a Fixed Point: A Random Matrix Perspective

D. J. Albers[1] and J. C. Sprott[2]

Received October 23, 2005; accepted October 9, 2006
Published Online: December 2, 2006

Results regarding probable bifurcations from fixed points are presented in the context of general dynamical systems (real, random matrices), time-delay dynamical systems (companion matrices), and a set of mappings known for their properties as universal approximators (neural networks). The eigenvalue spectrum is considered both numerically and analytically using previous work of Edelman *et al.* Based upon the numerical evidence, various conjectures are presented. The conclusion is that in many circumstances, most bifurcations from fixed points of large dynamical systems will be due to complex eigenvalues. Nevertheless, surprising situations are presented for which the aforementioned conclusion does not hold, e.g., real random matrices with Gaussian elements with a large positive mean and finite variance.

KEY WORDS: dynamical systems, structural stability, stability conjecture, Lyapunov exponents, complex systems, routes to chaos, random matrices, bifurcation theory

PACS numbers: 05.45.−a, 05.45.Tp, 89.75.−k, 89.75.Fb

1. INTRODUCTION

Determination of the types of random matrices that constitute a physically relevant set depends markedly on the field of study. From the perspective of those interested in quantum mechanical phenomena (e.g., nuclear physics), one might be led to believe that random matrices that are not unitary or Hermitian are of no particular physical interest.[(39)3] However, if one were interested in dynamics a la Poincaré,

[1] Max Planck Institute for Mathematics in the Sciences, Leipzig D-04103, Germany; e-mail: albers@cse.ucdavis.edu
[2] Physics Department, University of Wisconsin, Madison, WI 53706; e-mail: sprott@physics.wisc.edu
[3] In chapter 15 of Mehta's famous book one can find the words "An ensemble of matrices whose elements are complex, quaternion, or real numbers, but with no other restrictions as to their Hermitian

where an understanding of the derivative along an orbit is of utmost importance and where Lyapnov exponents are the key quantities of interest; then matrices of prime concern are real random matrices from the general linear group (the dissipative case) or the special linear group (the non-dissipative case). The interest in real random matrices arises because they form the linear portion of the derivative or the tangent space of a dynamical system at a given point along a trajectory.[1,37,42,45,46] From a dynamical reconstruction perspective, where time-series data are used,[19] it is companion matrices and thus polynomials with random coefficients that are of interest. Because we are concerned with dynamical systems in general, specifically probabilities of bifurcations from fixed points and the factors that determine those probabilities, we will focus on real random matrices with various distributions followed by a practical construction using "universal approximators" with random weights.

To place the current study in context, consider first the following background. Doyon *et al.*[13] argued that the most likely first bifurcation and route to chaos given a particular set of dynamical systems was that of the Ruelle–Takens quasi-periodic route to chaos based upon the random matrix results of Girko[21] (for other useful versions of this random matrix result see Edelman,[15] Bai,[10] and Kanzieper[36]). Likewise, Sompolinsky *et al.*[49] showed that as the dimension of a dynamical system is increased, the location of the bifurcation from a fixed point decreased toward zero; and the routes to chaos region of parameter space subject to parameter variation decreased in length. Doyon *et al.* and Sompolinsky *et al.* both considered neural networks with a single hidden layer (a single layer of neurons) whose input layer was entirely replaced by its output layer at each time-step (i.e., "vector" neural networks). However, Sompolinsky *et al.* considered the continuous-time case where as Doyon *et al.* worked in discrete time—both constructions yielded a similar set of conclusions. Utilizing the local chaos hypothesis of Amari[7] as well as the addition of Gaussian noise to the inputs, Moynot *et al.*[40] rigorously outlined and proved a mean-field theory for the aforementioned vector networks. In previous publications[6] the authors considered time-delay neural networks such as those presented in the work of Hornik *et al.*[31] that have been found to be universal approximators (i.e., they can approximate C^r ($r \geq 0$ mappings and their derivatives on compact, metrizible sets) and also came to some similar conclusions despite the significant differences in the constructions. Differences of note include the fact that the local chaos hypothesis does not apply to the time-delay neural networks and the form of the derivative matrix ($1 -$ jet) is a companion matrix rather than a full matrix which has a significant effect on analytical arguments that

---

or unitary character, is of no immediate physical interest, for their eigenvalues may lie anywhere on the complex plane." He later asserts that these matrices are, nevertheless, of interest in their own right.

can be made. In Ref. 4 the authors claimed that as the dimension of the dynamical system is increased, flip, fold, or any bifurcations due to real eigenvalues, and strong resonance bifurcations will be vanishingly rare and the route to chaos from a fixed point in parameter space will consist of periodic orbits with high-period (>4) and quasi-periodic orbits. The basic idea of the above arguments follows from the matrix theory of Girko,[21] Edelman,[15] Bai,[10] and Kanzieper[36] and is the following: given a square matrix whose elements are real random variables drawn from a distribution with a finite sixth moment,[4] in the limit of infinite dimensions, the normalized spectrum (or eigenvalues) of the matrix will converge to a uniform distribution on the unit disk in the complex plane. For matrices of finite dimension, the measure of the eigenvalues on the unit disc is not absolutely continuous with respect to Lebesgue measure. Nevertheless, if the Jacobian of the map at the "first" bifurcation point (i.e., the bifurcation from the fixed point) is a high-dimensional matrix whose elements have a finite sixth moment, it is reasonable that the bifurcation would be of type Naimark-Sacker (via a complex eigenvalue) instead of a flip or fold (via a real eigenvalue), with probability approaching unity as the dimension goes to infinity. Of course this intuition does not directly apply to the time-delay circumstance because the derivative matrix is not full; thus, for this circumstance the theory of zeros of random polynomials of Edelman *et al.*[16] is more directly applicable.

Each of the studies mentioned above consist of a statistical sampling of a space of mappings via a weight structure imposed upon those mappings. In a sense, it is a statistical sampling of the effects that the weight matrices have on the dynamics. The evidence discussed in the former paragraph leads to at least three important questions: how robust are the results with respect to the measures imposed upon the weight matrices; given that there do exist observable period-doubling bifurcations in high-dimensional dynamical systems, how does this occur given the random matrix style arguments; and how can these results be connected and compared with real world systems—what are the links with the natural world? Addressing the first two questions is a matter of carefully studying how the results from random matrix theory apply specifically to what has been observed computationally. Discussion of the third issue lies with providing a construction that yields comparison with real-world data.

As we will see, the distributions and perturbations of the distributions of the elements of random matrices can have profound and surprising effects on a bifurcation sequence while having negligible effects regarding the generality of the proven theorems. Thus, perturbations of the weight distributions of the

[4] While nearly all random distributions have a finite sixth moment, there are important distributions that do not have a finite sixth moment; a notable example of a distribution without finite sixth moment is the Levy probability distribution.

computationally studied systems can be made in such a way that significantly alters the conclusions of the random matrix arguments as applied to bifurcation theory. This will provide answers to why and how period-doubling sequences can be observed in high-dimensional systems. This of course does not marginalize the results discussed above; it only qualifies them, for despite the effects we observe, the former results remain quite general.

An issue that we will not outrightly address in this work is the effect of the elements of the Jacobians being *iid*, or independent, versus Jacobians whose elements are correlated. In this study, we impose an *iid* weight structure and observe the outcome and consequences of these weight structures. This imposes effects on what is observed. For instance, when identifying Gaussian matrices with the Jacobian, the trace of the Jacobian is mean zero with variance that scales like $d^{1/2}$; thus, there are conservative systems (trace zero), and an equal number of expanding and contracting dynamical systems (positive and negative traces respectively). However, because we are only observing fixed point behavior, this will likely have little effect. But, the point is, the distribution defines what is observed. We could instead impose conditions, such as a negative trace on the Jacobian, which will (likely) induce correlations between the weights, and then observe the implications. One particular implementation of this would be to train neural networks on data, thus inducing correlations between parameters. The most important issue, however, is understanding how, and for what systems, the correlations matter. To achieve an understanding of this, it is likely both of the aforementioned techniques will need to be implemented and compared. Thus, this paper can serve as a first step towards understanding how, when, and for what natural systems, correlations between elements of the Jacobian matter.

We will begin with a general study of linear dynamical systems at fixed points. Generically, all dynamical systems at stable fixed points can be thought of as linear maps via the implicit function theorem. We will identify the Jacobian of the aforementioned dynamical systems with random matrices with various distributions. We will follow this approach, adopting the framework of random polynomials which can be likened to companion matrices. The linear part of the derivative of time-delay dynamical systems is that of a companion matrix. Thus, we will study linear, time-delay dynamical systems subject to weight distributions of the coefficients of the characteristic polynomial. The final framework we will use—which was begun in Ref. 6—is that of scalar, time-delay neural networks that are commonly used to reconstruct a dynamical system from empirical, time-series data. This will begin to forge a connection with natural systems because networks such as the ones we study can be used to fit other data sets, and the weight distributions of the trained networks can be compared with those from more theoretical distributions.

2. CONSTRUCTION

Begin with the space of discrete-time, C^r ($r > 0$, $r \in N$) dynamical systems that map compact subsets (attractors) $U \subset R^q$ (with boxdim$(U) = D$) to themselves, with a single real parameter given by:

$$F : R \times U \to U \tag{1}$$

This is an infinite-dimensional (function) space that admits no sensible measure and thus no direct means of making statements regarding probability. While the notion of prevalence[32,33,44] was invented to address this problem, we will refrain from using this notion and will instead impose measures on function spaces by considering explicit forms of mappings.

Generic dynamical systems with a single parameter have three codimension one[5] bifurcations, $\lambda_{\max} \in R = \pm 1$ (fold and flip bifurcations) and $|\lambda_{\max}| \in C = 1$ (Naimark–Sacker bifurcations).[38,54] Most of the higher codimension bifurcations are combinations of these three codimension one bifurcations occurring simultaneously. In this paper, to determine the bifurcation type we always begin with an initial condition, compute the fixed point, and then calculate the eigenvalues of the Jacobian at the fixed point.

2.1. Discrete-Time Dynamical Systems

Consider the space of mappings defined by

$$x_t = F_T(x_{t-1}) = \epsilon A x_{t-1} + \epsilon G(x_{t-1}) \tag{2}$$

where $A \in R^{n^2}$, $\epsilon \in R$ is non-negative, and $G \in C^r$ is non-linear and represents the higher order terms. For ϵ small, it suffices to consider only $x_t = F_T(x_{t-1}) = \epsilon A x_{t-1}$; we will denote this space as $T(R^n)$. Imposing a measure on the elements of A imposes a measure $T(R^n)$. This construction allows both the ability to put a sensible measure on a space of dynamical systems at fixed points that can undergo bifurcations and the ability to use random matrix theory to make probabilistic statements regarding bifurcation probabilities.

2.2. Discrete-Time, Time-Delay Dynamical Systems

As previously mentioned, many time-dependent natural systems are studied via time-series data. That time-series data can be used to reconstruct a dynamical system is addressed in Ref. 48; we will discuss this in more depth in Sec. 2.3.1. Note that discrete-time, time-delay dynamical systems, denoted $T_{td}(R^n)$, can be

[5] The number representing the codimension of the bifurcation is the same as the number of parameters for which a bifurcation point occurs simultaneously.

written in the form given in 2 but with special constraints put on A. Specifically, we consider f to be explicitly defined:

$$y_{t+1} = f(y_t, \ldots, y_{t-d}) \tag{3}$$

where f is C^r and $y_i \in R$. For such systems, "A" is a companion matrix given by

$$Df_{T_{td}}(x) = \begin{bmatrix} a_1 & a_2 & a_3 & \ldots & a_{d-2} & a_{d-1} & a_d \\ 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & & & & \\ 0 & 0 & 0 & \ldots & 0 & 1 & 0 \end{bmatrix}$$

where the a_k's are associated with the partial derivatives $a_k = \frac{\partial x_t}{\partial x_{t-k}}$. Imposing a measure on the a_k's imposes a measure on $T_{td}(R^n)$ in the standard way.

2.3. Universal Approximators

To study C^r in a more general setting it is useful to consider a set of functions that can universally approximate a properly chosen time-delay map (f) of F (we will discuss this more in Sec. 2.3.1). This approximation can be done with the space of discrete-time, time-delay, feedforward neural networks given by:

$$x_t = \beta_0 + \sum_{i=1}^{n} \beta_i \tanh\left(s\omega_{i0} + s \sum_{j=1}^{d} \omega_{ij} x_{t-j} \right) \tag{4}$$

which is a map from R^d to R. Here n is the number of hidden units (neurons), d is the number of time lags which determines the system's input (embedding) dimension, and s is a scaling factor on the connection weights w_{ij}. The initial condition is $(x_1, x_2, \ldots, x_d)$, and the state at time t is $(x_t, x_{t+1}, \ldots, x_{t+d-1})$. The approximation theorems of Ref. 31 and well known time-series embedding results Refs. 48, 52 together establish an equivalence via an embedding (given properly chosen weight distributions) between this class of neural networks and the general dynamical systems of interest here.[5]

Each neural network can be associated with a point in the parameter space. Thus, imposing a probability measure on parameter space imposes a measure on the space of neural networks. We sample the $(n(d+1)+1)$-dimensional parameter space taking (i) $\beta_i \in [0, 1]$ uniformly distributed and rescaling them to satisfy $\sum_{i=1}^{n} \beta_i^2 = n$, (ii) w_{ij} as normally distributed with zero mean and unit variance, and (iii) the initial $x_j \in [-1, 1]$ as uniform. We will focus largely on behavior types as a function of the parameter s, which can be interpreted as the standard deviation of the w weight matrix, and the embedding dimension d. Note that for $x \approx 0$, $\tanh(x)$ is nearly linear. Thus, choosing s to be small forces the dynamics

to be mostly linear, yielding fixed-point behavior. Increasing s yields a route to chaos.[4,6] Due to this, s provides a bifurcation parameter that sweeps from linear to highly nonlinear parameter regimes. In this paper we will, however, only consider the networks up to the first bifurcation. Because scalar neural networks are discrete-time, time-delay mappings, the local derivatives are companion matrices. For the space of mappings we consider, the a_k's are given as:

$$a_k = \frac{\partial x_t}{\partial x_{t-k}} = \sum_{i=1}^{n} \beta_i s w_{ik} \operatorname{sech}^2 \left(s w_{i0} + s \sum_{j=1}^{d} w_{ij} x_{t-j} \right) \tag{5}$$

2.3.1. Neural Network Approximation Theory

To understand how the neural networks in our study fit into the dynamical systems framework, we must consider two questions: (i) is there an equivalence between time-delay dynamical systems and non-time-delayed dynamical systems; and (ii) what are the approximation capabilities of neural networks?

The answer to the first issue given by Takens—and more tailored to our construction, by the results of Sauer *et al.*[48]—is yes with certain constraints on the non-time-delayed dynamical system, F. Namely, the box-counting dimension of the set of periodic orbits of period p, (A_p) must be less than $\frac{p}{2}$, and DF_p must have distinct eigenvalues. These constraints are likely satisfied since we do not consider portions of parameter space that easily yield periodic orbits.

The relationship of general time-delay dynamical systems is given in Fig. 1, in which F is a C^r dynamical system, $E : U \to R$ is a "measurement function," (E is a C^k map) embedding $g : U \to R^{2d+1}$ is explicitly given by:

$$g(x_t) = (E(x_t), E(F(x_t)), \ldots, E(F^{2d}(x_t))) \tag{6}$$

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:

$$(y_1, \ldots, y_{2d+1}) \mapsto (y_2, \ldots, y_{2d+1}, g(F(g^{-1}(y_1, \ldots, y_{2d+1})))) \tag{7}$$

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

The response to the question regarding the approximation capabilities of neural networks is significantly more complicated due to two particular issues—the networks have only finitely many parameters, and the induced measure on the weights of the neural networks determines what phenomena will be fit stably. That networks we consider have finitely many parameters implies there will never exist a one-to-one or onto correspondence between the space of C^r diffeomorphisms

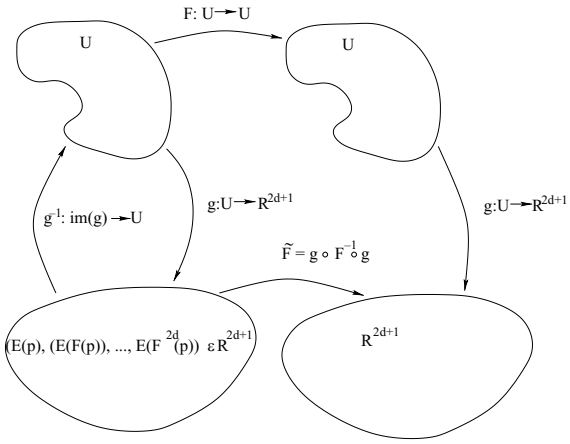

Fig. 1. Schematic diagram of the Takens embedding theorem and how it applies to our construction.

and neural networks. The best that can be achieved is a density in the relevant function space with infinitely many parameters. This is a fundamental functional approximation issue that has no chance of being improved; nevertheless, this issue is clearly not terminal for our study. The more interesting problem is that of the induced measure—specifically, how the space of neural networks with the induced measure on the parameter space represent approximates of the dynamical systems that satisfy the technical restrictions of Sauer *et al.* Further, what subsets of the space of C^r dynamical systems are being selected out by such a measure (or even our measure) is a very interesting question. There surely exists a measure on the space of parameters of neural networks such that (given the dimension is high enough) the selected neural networks will approximate and correspond to a set of the C^r mappings that satisfy the restrictions of Sauer *et al.* In fact our measure often satisfies this criterion given certain restrictions. Such issues hopefully can be addressed by applying information geometry to the parameter space of neural networks following work of Amari.[8] However, a clear understanding of what portion of the C^r mappings we are sampling is currently difficult to discern.

Setting the issues of a finite number of parameters aside and leaving any constraints on the weights behind, the neural networks we utilize can approximate $\tilde{F}$ and its derivatives (to any order) to arbitrary accuracy.[30,31] Precise statements of the neural network approximation theorems require machinery from functional analysis (see Ref. 2) and is covered in detail in the papers by Hornik *et al.* and is discussed in Ref. 5. The neural networks can approximate mappings from Sobolev spaces, $S_p^m(U, \lambda)$, such as the one defined here:

Definition 1. *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:*

$$S_p^m(U, \lambda) = \{f \in C^m(U) | \quad |D^\alpha f||_{p,U,\lambda} < \infty \quad \text{for all } |\alpha| \leq m\} \tag{8}$$

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

Neural networks can also approximate most C^r mappings and their derivatives on compacta. However, if the mapping is piecewise differentiable, the neural networks must be fit independently on both sides of the discontinuity. Nevertheless, neural networks form a set of "universal approximators"—they can approximate nearly any imaginable mapping.

In this work we are not interested in the approximation of a particular mapping, but rather we are interested in approximating the C^r function space via a statistical study. Thus, it is the density in various function space results we are more concerned with; again, recall one of the points of this construction is to be able to put a measure on a diverse function space. It is important to note that with $n < \infty$, density of the space of neural networks in C^r is not achieved. Thus, for finite n, we are instead studying a "grid" of mappings on C^r that approach density as $n \to \infty$.

2.3.2. The Implicit Measure on the Neural Networks

The probability distributions we impose on the space of neural networks form a product measure on $R^{n(d+2)+1}$—this means that all the weights are *independent* and uncorrelated. The weights of ensembles of neural networks that have been *trained* will likely be *correlated*, and thus the probability measures on the parameter space will be *joint* probability distributions instead of *product* measures. This makes little difference to the essence of this work, but hopefully will make a big difference in the results, given the diversity observed in natural systems. The choice of a relatively uniform product measure, which was chosen for its simplicity, 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. Examples of how important this is will be revealed as we consider the numerical results. Unlike actual physical experiments, we could, in principle, prove an invariance of our results to our induced measure; however, this is difficult and from our perspective would be a very unlikely result. It suffices for our purposes to note specifically what our measure is (the weight selection method), and how it might bias our results. The selection method we use 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. It is worth noting that we are, in spirit, attempting to use the prevalence construction of Hunt

et al.[32,33] This would actually allow for rigorous, measure theoretic statements on the full C^r space. However, a full utilization of this framework is beyond the scope of this work.

2.3.3. Why Scalar Neural Networks?

Time-delay scalar neural networks are enticing to study because they allow for a combination of the perspectives of functional analysis, topological dynamics, and the practical, real world. From the functional analysis perspective, the neural networks we study are diverse universal function approximators—they can approximate nearly any mapping one might wish to approximate.[31] Thus, studying the space of neural networks represents a practical study of a common space of mappings used by time-series analysts to reconstruct unknown dynamics from time-series data.[19,35,51] From the perspective of topological dynamics, if one wishes to study general dynamical systems, there is always the issue of how to relate such studies to the natural world that they were originally meant to model. Studying time-delay dynamics as they relate to C^r dynamical systems is a partial connection or link between abstract dynamics and the natural world since the natural world is often studied with time-series data.[43] Yet there is the issue of relating time-delay dynamics to a specific natural system or class of systems. Nevertheless, one of the key sources of the problem of relating the space of neural networks to a space of general dynamical systems is the implied measure on the weight space. This is also where the practical link between the natural world and general time-delay dynamical systems can be found. The weight distributions can form a clear link between the abstract dynamics world via embedding and function approximation of real time-series data from natural phenomena. If the dynamics that arise from the space of neural networks can be understood in terms of their weight distributions and the neural networks can be fit to data from nature—the fit weight distributions can be compared with the dynamics dependent on the weight distributions. Thus, aside from being an extremely malleable class of dynamical systems and universal approximators in their own right, neural network weight distributions are a possible link between the abstract world of mathematical dynamics and the natural world.

3. RANDOM MATRIX THEORY

Our discussion of random matrix theory will be limited to the circular law of Girko,[21] Bai,[10] Edelman,[15] and Kanzieper.[36] 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 the 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 $d \times d$ matrices with real elements drawn from a random distribution yet to be specified.

Recall that random matrix theory can be applied to dynamical systems at fixed points because dynamical systems at fixed points can be identified with random matrices via the linear derivative map of the dynamical system f where each of the terms in the matrix is given by a number from a random distribution. This is quite general—given that the range and domain of the dynamical system have the same dimensions. Any generic, discrete-time, d-dimensional dynamical system at a fixed point can be recast as a linear map $g = Ax$ where A is a $d \times d$ matrix via the implicit function theorem. Thus, studying the spectrum of random matrices is, in a way, equivalent to studying bifurcations of dynamical systems at non-degenerate fixed points with respect to the variation of a linear scaling parameter; because the spectrum of Df will yield the entire geometric structure of f.

3.1. 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 $d \times d$ matrices with independent and identically distributed elements that were normalized by $\frac{1}{\sqrt{d}}$, converged to a uniform distribution on the unit disk in the complex plane. This is what is refered to as the circular law. In 1965 Ginbre[20] 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). V.I. Girko published, in 1984,[21] 1994,[23] 1997,[24,25] and again in 2004[26–28] papers proving the circular law for real, random matrices. Girko's first circular law roughly states that as $d \to \infty$, the distribution of $\frac{\lambda}{\sqrt{d}}$ tends to uniformity on the unit disk for matrices with Gaussian elements. This result, which implies that as $d \to \infty$, 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 (or near-infinite-dimensional) limit relative to certain measures imposed on the space. 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 when the dimension of the matrix is finite (the infinite-dimensional limit is absolutely continuous). In particular, for finite-dimensional matrices, the probability of an eigenvalue being real with respect to Edelman *et al.* is higher than one

might expect. Thus, convergence in distribution to uniformity on the unit disk is a non-trivial issue. Nevertheless, Edelman[17] derived a formula for the expectation value of real eigenvalues in Girko's measure in finite dimensions; we will discuss that result in Sec. 3.2. In the process of deriving this expectation formula, Edelman also proved Girko's 1984 result. In 1997, Bai[10] provided an alternate proof of the circular law for real random matrices with a significantly weaker hypotheses than either Edelman or Girko in his 1984 work. Bai's result requires that the elements of the matrix *only* be from a distribution with only a *finite sixth moment* (see Ref. 10 (page 496)). We will state a key result that is a corollary of the more powerful results for clarity (for the most relaxed (but complicated) hypothesis see either Refs. 26–28 or 10).

Theorem 1 (**Circular law** (**Girko**)). *Suppose that the entries of an* $d \times d$ *matrix* M *with moments of order* $4 + \delta(\delta > 0)$ *bounded whose entries do not necessarily have existent densities. Then, with probability* 1, *the empirical distribution* $\mu_d(x, y)$ *tends to the uniform distribution over the unit disk in two-dimensional space*.

Understanding the difference between the results of Bai, Girko, Edelman, the convergence of the density of igenvalues on the unit disc, and how this is related to the eigenvalue with the largest magnitude will be the focus of Secs. 6.1 and 6.2. One issue that is of little importance from a random matrix theory perspective, but will prove to be very important from a bifurcation theory perspective is restrictions on the mean of the distribution of the matrix elements.

3.2. Expected Value of Real Eigenvalues and Related Results

The circular law will provide the intuition for the conjecture we will state shortly, but it 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 formulas and theorems 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 $A \in R^{d^2}$ where $a_{ij} \in N(0, 1)$.

We will begin with two definitions, the true density of real eigenvalues of a real random matrix and the probability density of real eigenvalues of the said random matrix. These can be found in Ref. 17.

Assume λ is a real eigenvalue of a fixed, real $d \times d$ matrix A. The true density of real eigenvalues, or the expected number of real eigenvalues per unit length can be defined:

$$\rho_d(\lambda) = \left(\frac{1}{\sqrt{2\pi}} \left[\frac{\Gamma(d-1, \lambda^2)}{\Gamma(d-1)} \right] + \frac{|\lambda^{d-1}| e^{-\frac{\lambda^2}{2}}}{\Gamma\left(\frac{d}{2}\right) 2^{\frac{d}{2}}} \left[\frac{\Gamma\left(\frac{(d-1)}{2}\right), \frac{\lambda^2}{2}}{\Gamma\left(\frac{(d-1)}{2}\right)} \right] \right) \tag{9}$$

or, in a different light:

$$\rho_d(x) = \frac{d}{dx} E_A \#_{(-\infty,x)}(A) \tag{10}$$

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. Moreover, the probability density of $\lambda_d \in R, f_d(\lambda)$ is given by:

$$f_d(\lambda) = \frac{1}{E_d} \left(\frac{1}{\sqrt{2\pi}} \left[\frac{\Gamma(d-1, \lambda^2)}{\Gamma(d-1)} \right] + \frac{|\lambda^{d-1}| e^{-\frac{\lambda^2}{2}}}{\Gamma\left(\frac{d}{2}\right) 2^{\frac{d}{2}}} \left[\frac{\gamma\left(\frac{(d-1)}{2}, \frac{\lambda^2}{2}\right)}{\Gamma\left(\frac{(d-1)}{2}\right)} \right] \right) \tag{11}$$

or more simply:

$$f_d(\lambda) = \frac{1}{E_d} \rho_d(\lambda) \tag{12}$$

where E_d denotes the expected number of real eigenvalues of the $d \times d$ random matrix.

Integrating ρ_d 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 asymptotic series given in Corollary 5.2 of Ref. 17:

$$E_d = \sqrt{\frac{2d}{\pi}} \left(1 - \frac{3}{8d} - \frac{3}{128d^2} + \frac{27}{1024d^3} + \frac{499}{3276d^4} + O\left(\frac{1}{d^5}\right) \right) \tag{13}$$

Again, E_d is the expected number of real eigenvalues for a real, $d \times d$ random matrix. The manner in which the measure is not absolutely continuous (with respect to Lebesgue measure) for finite-dimensional matrices is highly relevant to our results. In particular, the absolute continuity with respect to Lebesgue measure fails along the real line and results in an expected density of real eigenvalues that is higher than other chords of length 2. For a full discussion, see Ref. 17. Nevertheless,

this is not a pathological problem because the distribution of real eigenvalues on the real line is uniform as per the following corollary:

Corollary 1 (**Corollary 4.5 Ref. 17**). *If λ_d denotes a real eigenvalue of an $d \times d$ random matrix, then as $d \to \infty$, the normalized eigenvalue $\frac{\lambda_d}{\sqrt{d}}$ 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 2(**Density of Non-Real Eigenvalues: Theorem 6.2 Ref. 15**). *The density of a random complex eigenvalue of a normally distributed matrix is*:

$$\rho_d(x, y) = \sqrt{\frac{2}{\pi}} y e^{y^2 - x^2} erfc(y\sqrt{2}) e_{d-2}(x^2 + y^2) \tag{14}$$

where $e_d(z) = \sum_{k=0}^{d} \frac{z^k}{k!}$ 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 3 (**Theorem 6.3, Ref. 15**). *The density function $\hat{\rho}$ converges pointwise to a very simple form as $d \to \infty$*:

$$\lim_{d\to\infty} \frac{1}{d} \hat{\rho}(\hat{x}, \hat{y}) = \begin{cases} \frac{1}{\pi} & \hat{x}^2 + \hat{y}^2 < 1 \\ 0 & \hat{x}^2 + \hat{y}^2 > 1 \end{cases} \tag{15}$$

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

Finally, Edelman's version of the circular law can be proved using Theorem 3 and a central limit theorem.

Theorem 4 (**Circular Law: Convergence in Distribution, Ref. 15**). *Let z denote a random eigenvalue of A chosen with probability $\frac{1}{d}$ and normalized by dividing by $\sqrt{d}$. As $d \to \infty$, z converges in distribution to the uniform distribution on the disk $|z| < 1$. Furthermore, as $d \to \infty$, each eigenvalue is almost surely non-real*.

4. RANDOM POLYNOMIALS AND COMPANION MATRICES

Let us recall again why we are concerned with the special case of random polynomials; the linear part of the derivative of time-delay dynamical systems—the

ones often used to fit real time-series data—are companion matrices. In particular, given the companion matrix:

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

the corresponding characteristic polynomial—the equation whose solutions are the eigenvalues of the above matrix—is a polynomial as given in Eq. (16). Thus, the elements of a companion matrix (e.g., the a_k's) can be identified with the coefficients of the characteristic polynomial of the given matrix. The results we will discuss here briefly are from 4.1 while issues that arise due to computation of eigenvalues from companion matrices can be found in Ref. 18.

4.1. Polynomials with Gaussian Coefficients

Let us begin with the polynomial

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d \tag{16}$$

where the a_i coefficients are independent standard normals with mean zero. The expected number of real zeros, E_{real}, as $d \to \infty$ is given by the formula:

$$E_{real}(d) = \frac{2}{\pi} \log(d) + C_1 + \frac{2}{d\pi} + O(1/d^2) \tag{17}$$

where $C_1 = 0.6257358072$ (cf. Theorem 2.1 in Ref. 16 or 34). This formula is calculated by integrating the true density which is given by:

$$\rho_d(x) = \frac{1}{\pi} \sqrt{\frac{1}{(x^2-1)^2} - \frac{(d+1)^2 x^{2d}}{(x^{2d+2}-1)^2}} \tag{18}$$

Careful analysis of Eq. (18) yields the limiting density of real eigenvalues—that as $d \to \infty$, the density of real eigenvalues is concentrated at ± 1. The expected value of real roots is, of course, a crude measurement of interest as we are interested in differentiating between bifurcations due to real and complex eigenvalues. The density provides considerably more insight. Moreover, the convergence of the density will have a significant impact on the probability of the first bifurcation as we will see in Sec. 6 where we compare Eq. (18) with the numerical results.

4.1.1. Polynomials with Gaussian Coefficients with Non-Zero Mean

The random polynomials is the only case we will consider for which there are theoretical results regarding the convergence of the density of real zeros for distributions with non-zero means. The effect of a non-zero mean on the distribution of the coefficients of a random polynomial is neatly summarized by the following theorem:

Theorem 5. *Consider a random polynomial of degree d with coefficients that are independent and identically distributed normal random variables. Define* $m \neq 0$ *to be the mean divided by the standard deviation. Then, as* $d \to \infty$,

$$E_{real}(d) = \frac{1}{x}\log(d) + \frac{C_1}{2} + \frac{1}{2} - \frac{\gamma}{\pi} - \frac{2}{\pi}\log(|m|) + O(1/d) \tag{19}$$

where $C_1 = 0.6257358072\ldots$ *as previously defined, and* $\gamma = 0.5772156649\ldots$ *is Euler's constant. Furthermore, the expected number of positive zeros is asymptotic to*

$$\frac{1}{2} - \frac{1}{2}erf^2\left(\frac{|m|}{\sqrt{2}}\right) + \frac{1}{\pi}\Gamma[0, m^2] \tag{20}$$

The proof can be found in Ref. 16 (Theorem 5.3). Considering Eq. (19) and comparing it with Eq. (17), one arrives at the comparison between $\sim \log(d) + \frac{1}{d}$ for zero mean with $\sim \log(d) - \log(|m|)$ as $d \to \infty$, the effect of the nonzero mean is a shift of the $\log(d)$ curve by $-\log(|m|)$. This dependence can be seen in Fig. 2.

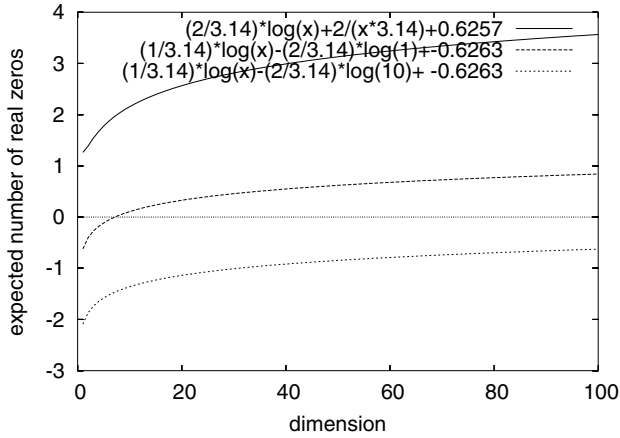

Fig. 2. A plot of Eqs. (17) and (19) for $m = 1, 10$.

4.2. Real Zeros of Random Functions with Random Coefficients

A goal of this paper is to consider the probability of the type of bifurcation from a fixed point in a class of neural networks that can be used to approximate time-series data from actual physical experiments. In Ref. 16 a remarkable theorem was proved regarding sums of differentiable functions with random coefficients linking the distribution to the coefficients and the function type to the distribution of real zeros. This theorem has a high degree of relevance to a study such as ours since it offers the hope of applying to the class of mappings we are considering.

Theorem 6. *Let* $v(x) = (f_0(x), \ldots, f_d(x))^T$ *be any collection of differentiable functions and* $a_0, \ldots, a_d$ *be the elements of a multivariate normal distribution with mean zero and covariance matrix C. The expected number of real zeros on an interval (or measurable set) I of the equation*

$$a_0 f_0(x) + a_1 f_1(x) + \cdots + a_d f_d(x) = 0 \tag{21}$$

is

$$\int_I \frac{1}{\pi} ||\omega'(x)|| dx, \tag{22}$$

where w is given by

$$w(x) = \frac{C^{1/2} v(x)}{||C^{1/2} v(x)||} \tag{23}$$

In logarithmic derivative notation this is:

$$\frac{1}{\pi} \int_I \left(\frac{\partial^2}{\partial x \partial y} (\log(v(x)^T C v(y))|_{y=x=t}) \right)^{1/2} dt \tag{24}$$

There are many applications of this profound theorem presented in Ref. 16. One of particular interest is an application to a trigonometric series such as

$$\sum_{k=0}^{\infty} a_k \cos(\nu_k \theta) + b_k \sin(\nu_k \theta) \tag{25}$$

where a_k and b_k are independent normal random variables with mean zero and variance σ_k^2. The density and, thus, the expected values of real zeros is quite easily computed given Theorem 6. The density is constant and, therefore, the real zeros of

the above random trigonometric sum are uniformly distributed on the real line. An important point to note is the sharp difference between the trigonometric case and the random polynomial case. Thus, different functional forms and distributions of a_k's give very different distributions of real zeros of polynomials. These two results show a sharp contrast regarding bifurcations from fixed points—the trigonometric series above is likely to yield all real bifurcations; whereas, the standard random polynomial case yields significantly more bifurcations due to complex eigenvalues. Providing a similar analysis for neural networks like those given in Eq. (4) is complicated by the sum inside the activation function and, therefore, is beyond the scope of this paper. However, as we will discuss later, because neural networks can approximate nearly any C^r mapping which include both the random polynomials and the trigonometric series, altering the distributions of the β_i's and the w_{ij} will clearly yield very different densities of real zeros. A relation between the distribution of the β_i's and w_{ij}'s will be the focus of Sec. 6.4.

5. 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.[11,38] 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 Naimark-Sacker,[41,47] corresponding to a complex conjugate pair of eigenvalues with 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 Naimark-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 Naimark-Sacker bifurcation. A conclusion along these lines will prove incorrect as we will show in Sec. 6. The restriction of a finite sixth moment will turn out to be too weak because lower-order moments can affect quantities like the spectral radius or the eigenvalue with the largest modulus. Limiting ourselves to the case where the matrix has real Gaussian elements with mean zero and unit variance for which we have more restricted and detailed results will be fruitful. For example, if the real eigenvalues concentrate near 1 and -1, we will run into problems, but Edelman (Corollary (1)) has shown that this circumstance will not occur. Instead, the real eigenvalues will be distributed uniformly on the real axis. Results akin to the aforementioned result will prove necessary for arguments involving bifurcations. In the standard and general 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 where the parameterized curve is linear and forms an interval in say, R^1, the difficulties are greatly reduced. Thus, we can make the following statement:

Corollary 2 (**First bifurcation probability**). *Given the dynamical system F*

$$F(x_{t-1}) = x_t = \varepsilon A x_{t-1} + \varepsilon G(x_{t-1}) \tag{26}$$

where $x_t \in R^d$, $\varepsilon \in R$, $A \in R^{d^2}$, $a_{ij} \in N(0, 1)$, *and where* $G(x_{t-1})$ *is a nonlinear* C^r $(r > 0)$ *mapping of* x_{t-1} *which is of order 2 or higher. Thus* $F(x_{t-1}) \cong \varepsilon A x_{t-1}$ *for* ε *small. Assume F has a fixed point at* $\varepsilon = 0$ *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* $A \in R^{d^2}$, $d \to \infty$), *the probability that the first bifurcation will be of type Naimark-Sacker will converge to one*.

Proof: This result follows trivially from the results of Edelman[15,17] and Girko.[21–25]

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,[6] 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 was originally hoped that we could 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.[10] However, considering only random matrices with elements chosen from a Gaussian distribution, it can be shown numerically that the mean of the distribution has a significant effect on the eigenvalue with the largest modulus.

Corollary (2) falls far short of satisfying our desires. First, Corollary (2) does not speak to the probability of its hypothesis being satisfied in C^r function space relative to a measure. It is, however, hoped that such a result can eventually be formulated using the notion of prevalence.[32,33,44] Another shortcoming of Corollary (2) is it is not cast in the general parameterized curves of the bifurcation theory framework we desire—linear "curves" are very limiting (see Refs. 50, 53 for a construction of the general bifurcation framework to which we are referring.) Moreover, Corollary (2) does not provide any information regarding how large, but finite-dimensional, dynamical systems that satisfy all the hypotheses behave. Lastly, Corollary (2) does not provide any insight into how the convergence to

[6] In theory we can impose a measure on the weights of the neural networks that mimics the results of random matrices, however we have not found this said measure. It is likely that it could be determined via training of neural networks on data from dynamical systems with random ($N(0, 1)$) linear parts.

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 want.

Conjecture 1 (**Genericity of Naimark–Sacker bifurcations in high-dimensional dynamical systems**). *Begin with a subset of the space of dynamical systems, $\mathcal{V} \subset C^r (r > 0)$, whose elements have at least one fixed point on a measurable interval of the parameter space, and at least one local bifurcation upon a continuous variation of a parameter. Further, assume these dynamical systems have first derivatives whose elements form a distribution with mean zero. Then, there exists a probability measure on the parameters for such dynamical systems such that, as the dimension d of the dynamical system is increased, the probability of the bifurcation from fixed points via the Naimark-Sacker bifurcation will increase and approach probability one as $d \to \infty$.*

One noticeable and important omission in the hypothesis is the requirement that the elements of the Jacobian be *iid*, thus we are conjecturing this result will *not* depend heavily on correlations between the elements of the Jacobian but will depend on the *mean* of the distribution. This conjecture makes the most sense in a framework such as that provided by neural networks for which the parameter space is large enough that it can have some approximation of C^r function space. We will not discuss this conjecture further here, but upon the presentation of the numerical results, we will provide a discussion of what this conjecture might mean and where and how it is known to fail.

Finally, the above construction does not apply explicitly to the time-delay dynamics in general—whose linear derivative matrices are companion matrices—or to the neural networks we are focusing on in particular. Such a circumstance would require a slightly different conjecture:

Conjecture 2 (**Genericity of Naimark–Sacker bifurcations in high-dimensional time-delay dynamical systems**). *Begin with a subset of the space of time-delay dynamical systems, $\mathcal{V} \subset C^r (r > 0)$, whose elements have at least one fixed point on a measurable interval of the parameter space, and at least one local bifurcation upon a continuous variation of a parameter. Further, assume these dynamical systems have bounded first derivatives whose elements form a distribution with mean zero and compact support. Then, there exists a probability measure on the parameters for such dynamical systems such that, as the dimension d of the dynamical system is increased, the probability of the bifurcation from fixed points via the Naimark–Sacker bifurcation will increase and approach probability one as $d \to \infty$.*

Analysis of this conjecture is akin to a study of random polynomials or a sum of random functions with coefficients drawn from a random distribution. This case is of interest because many experimental results come from time-series data—data

that are reconstructed with a "universal approximator" that forms a time-delay dynamical system (neural networks of the type we consider fall into this class). We will compare and contrast the various results and implications of the constructions of the two conjectures. Because neural networks are universal approximators, we will, at the end, suggest how to find the measure on the a_k's that will link them to the general random matrix framework.

6. NUMERICAL CASES

We will begin our numerical investigations with a careful analysis of Gaussian random matrices with mean zero and variance 1—one of the cases for which Edelman has provided analytical results. We will then begin to investigate the difference between, and hence generality of, what we can imply regarding Conjecture 1 with the results of Edelman versus those of Bai by perturbing the first and second moments of the Gaussian matrices and observing differences. Following this, we will study the case of uniform random matrices—a case for which Bai's results still apply—to begin to understand the invariance of various results to different random distributions. We will again study the perturbation of the first and second moments of the uniform distribution on the relevant quantities. Since the Jacobian of a time-delay dynamical system at a fixed point is a companion matrix, we will then briefly study companion matrices and finally move on to the case of time-delay neural networks.

6.1. M_n with Gaussian Elements

As a base case, consider matrices G that have elements chosen from a Gaussian distribution with mean zero and variance one ($N(0, 1)$). This is the case that Girko, Edelman and Kanzieper investigated, and thus there exist several analytical results and theorems which we can apply to the current framework. We will then perturb various moments of this distribution to better understand the different implications and limitations of the work of Edelman, Girko, and Bai as applied to bifurcation theory.

6.1.1. Gaussian Matrices with Zero Mean

Begin with the mean-zero, variance-one case, a comparison between Edelman's expected number of real eigenvalues given in Eq. (13) and those empirically calculated from random matrices is depicted in Fig. 3. The lines overlay nearly perfectly even at low dimensions yielding a power-law dependence of $E[\lambda_{real}] = 0.978d^{-0.529}$ as expected.

Given corollary 1—that the distribution of real eigenvalues converges to a uniform distribution on the real line—a disproportionate fraction of the

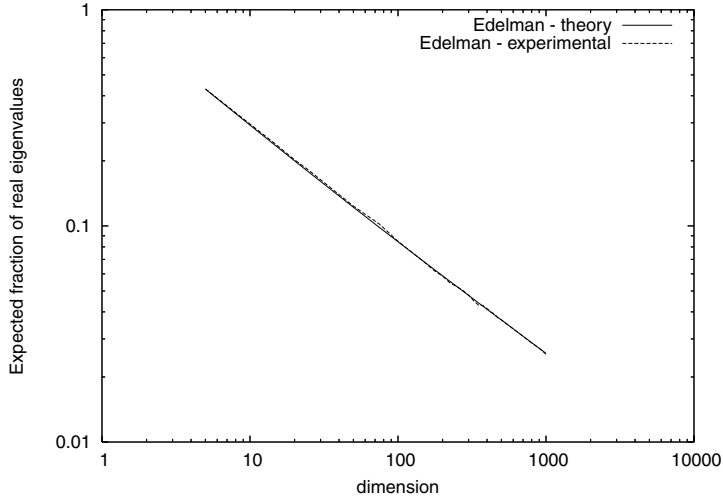

Fig. 3. Edelman's prediction for the expected fraction of real eigenvalues and the empirically calculated expected number of real eigenvalues. Both quantities were calculated in increments of 5 dimensions up to $d = 50$ and then increments of 25 thereafter until $d = 1000$. The line is that of a power law with $E_{\lambda_{real}}[d] \sim 0.9784d^{-0.5291}$ as expected.

bifurcations should be due to complex eigenvalues, especially as $d \to \infty$. Figure 4 depicts the fraction of bifurcations that correspond to Naimark–Sacker, flip, and fold bifurcations as well as the fraction of eigenvalues that are real. For systems we are considering, these distinctions are easily made by simply calculating the largest eigenvalue and determining whether it is complex or real, combined with the sign of the eigenvalue. The results are as one might expect in the sense that the number of real bifurcations due to positive and negative eigenvalues are nearly identical. Nevertheless, the fraction of real eigenvalues converges to zero considerably faster than the fraction of bifurcations due to real eigenvalues. Considering the right plot of Fig. 4, the spectrum of eigenvalues of a 1024×1024 matrix, the real line is clearly highly and evenly populated with $\sim$30 real eigenvalues. Hence the convergence to uniformity seems to be well-behaved by $d = 1024$. Nevertheless, the fraction of bifurcations due to real eigenvalues is clearly decreasing with d in a power law. Moreover, the decrease of bifurcations due to 1 and -1 are nearly identical as expected.

6.1.2. Perturbing the First Moment—Gaussian Matrices with a Non-Zero Mean

The results of Girko, Edelman and Kanzieper pertain to real random matrices with Gaussian matrices with mean zero and finite variance, while the results

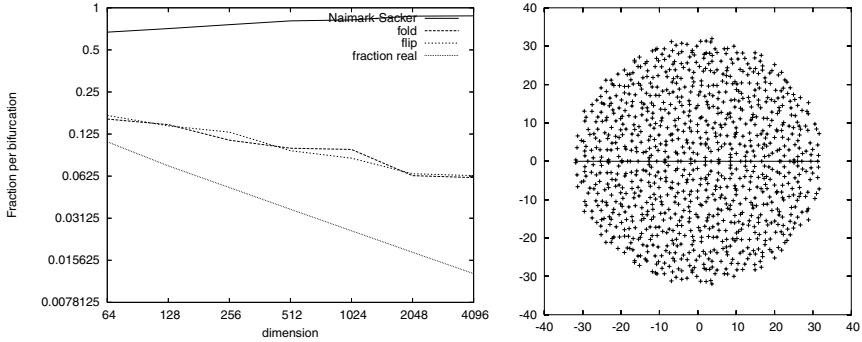

Fig. 4. On the left, the observed probability of each bifurcation was recorded for 1000 matrices with *iid*, mean zero, variance one, Gaussian elements for each d (in powers of 2) along with the fraction of eigenvalues that are real. On the right is the spectrum of eigenvalues in the complex plane that corresponds to a single 1024×1024 matrix ($d = 1024$).

of Bai apply to any real random matrix with a distribution with a finite sixth moment. There is a great difference between these two formulations. Beginning with a standard Gaussian random matrix and perturbing the first moment (the mean), yields a result that has little relevance for the random matrix theory but has significant implications from the perspective of bifurcation theory. Summing a $d \times d$ matrix G as a matrix with elements g_{ij} drawn from a Gaussian distribution with mean zero and variance one ($N(0, 1)$) and a constant $d \times d$ matrix P_m with elements $p_{ij} = m$ yields

$$\Lambda_m = G + P_m \tag{27}$$

Figure 5 has the modulus of the largest and second largest eigenvalues plotted for Λ_m with $d = 64$. For $|m| > 0.1289$, the largest eigenvalue is always real and increases with m according to:

$$\lambda_d(m) = dm \tag{28}$$

The next largest eigenvalue is most often complex as would be expected if the distribution of eigenvalues on the unit disk is uniform. The modulus of the second largest eigenvalue(s), however, is independent of m ($\lambda_{d-1} = \text{constant}$) is approximately $\sqrt{d}$ as expected from Edelman's normalization formulas. Thus, aside from the largest real eigenvalue, the distribution of eigenvalues behaves like the spectrum of G. Therefore, a dynamical system with a DF matrix

$$\frac{\epsilon}{\sqrt{d} + a} \Lambda_{m=\frac{1}{d}} \tag{29}$$

for $a \ll 1$, upon increasing ϵ, will always undergo a flip bifurcation from a fixed point. The point of this is that, given a random matrix with elements drawn from

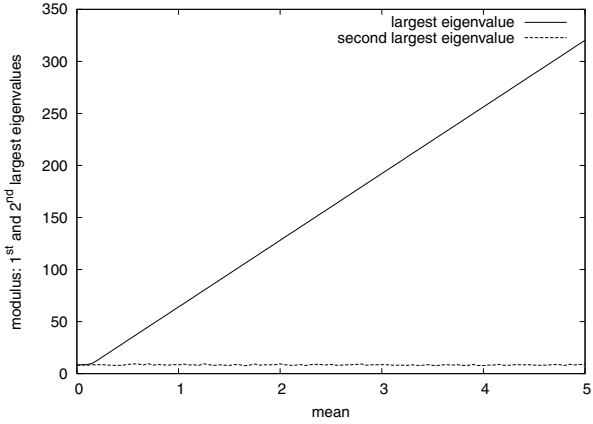

Fig. 5. This figure represents an ensemble of 1000 $d \times d$ matrices with $d = 64$. Depicted are the modulus of largest and second largest eigenvalues. The line representing the modulus of the largest eigenvalue is given by 64 m while the line for the modulus of the second largest eigenvalue is given by $\sim\sqrt{d}$.

a distribution with finite sixth moment, convergence of the spectrum to a uniform distribution on the unit disk in the complex plane is not sufficient to guarantee that a dynamical system with a Jacobian matrix (at a fixed point) with elements that converge to a distribution with a finite sixth moment, will undergo a Naimark–Sacker bifurcation with probability increasing with d. In this case, a measure-zero set (a single eigenvalue), happens to be of utmost importance when considering the most probable bifurcation from a fixed point. Moreover, this single eigenvalue is not a counter-example to any of the circular laws because, aside from the one single eigenvalue, the rest of the spectrum converges uniformly on the unit disk in the complex plane as required by the various theorems.

A complete explanation of why the largest eigenvalue scales with the mean is unknown. However, it is likely that the explanation will rely on the Perron–Frobenius theorem for non-negative matrices that says if A is a non-negative matrix, then there will exist an eigenvalue that equals the spectral radius, $\rho(A)$.[29] That the largest eigenvalue scales with d is not difficult to see, but Gaussian random matrices with positive mean have a positive probability of having negative elements, and thus are not non-negative matrices. However, as the mean increases, the probability of having negative elements decreases considerably. Thus, it is likely we are observing a "convergence" to the Perron–Frobenius result for non-negative matrices. While the Perron–Frobenius result will never strictly apply to this context, it is likely that it can be shown that the Perron–Frobenius theorem will apply with probability approaching one as the mean approaches infinity.

6.1.3. Perturbing Higher Moments

Perturbation of the second moment of the distribution of G, which amounts to perturbing the variance, has a simple effect on the spectrum—the spectral radius increases linearly with the variance. This will have very little effect on the bifurcation structure aside from decreasing the ϵ value for which the first bifurcation from a fixed point will occur. Perturbing higher moments of the distribution and the subsequent effects remains an open problem.

6.2. M_n with Uniform Elements

Above we considered matrices with entries drawn from Gaussian distributions. As a result, we could bring to bear a large body of analytical machinery. Now, however, we turn to matrices with entries drawn from *iid* uniform random variables on the interval (a, b). Unfortunately, in this circumstance, very little analytical machinery is available.

6.2.1. Uniform Matrices with Zero Mean

For simplicity, we will set $a = -1$ and $b = 1$ such that the u_{ij}'s are *iid* uniform random variables on $(-1, 1)$, with mean zero. Concentrating first on the fraction of real eigenvalues, Fig. 6 yields a scaling law of $E[\lambda_{real}] = 0.919d^{-0.517}$. This is similar to the theoretical and empirically calculated $E[\lambda_{real}]$ for a random matrix with Gaussian elements with zero mean. However, $E[\lambda_{real}]$ approaches zero faster in the case of Gaussian random matrices than with uniform random matrices. The difference is, nevertheless, quite small (about 3 percent). Accordingly, the fraction

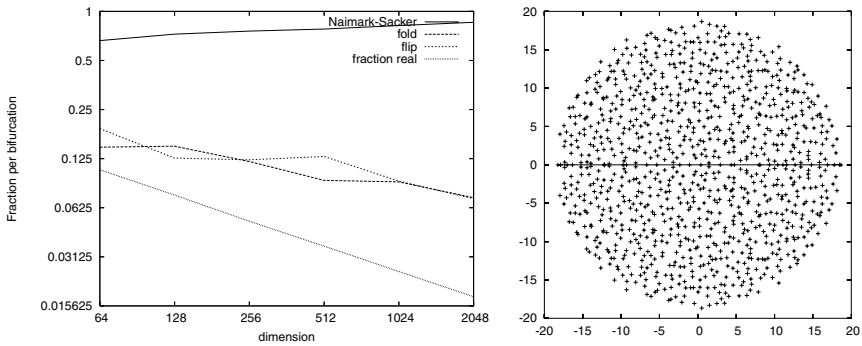

Fig. 6. On the left, the observed probability of each bifurcation was recorded for 1000 matrices with i.i.d. uniform elements for each d (in powers of 2) along with the fraction of eigenvalues that are real. On the right is the spectrum of eigenvalues in the complex plane that corresponds to a single 1024×1024 matrix ($d = 1024$).

of real bifurcations decreases slightly slower for uniform random matrices than for the Gaussian case, while the fraction of flip and fold bifurcations are identical up to standard error. The right plot in Fig. 6 is indistinguishable from the analog plot for Gaussian random matrices given in Fig. 4. The real line is highly and evenly populated with ∼25 real eigenvalues.

6.2.2. Perturbing the First Moment—Uniform Matrices with a Non-Zero Mean

The uniform random variable case is fundamentally different from the Gaussian case in several ways. First, the uniform case has finite support on R. Second, the mean is directly related to the endpoints of the support of the distribution, i.e., $\bar{x} = \frac{b+a}{2}$. Thus perturbing the mean from zero amounts to upsetting the symmetry of the end points of the support about zero. To investigate the effects of perturbing the mean we fix $b = 1$ and vary a from 0 to -1. Or, more explicitly, given

$$\Lambda_a = U + P_a \tag{30}$$

with u_{ij} uniform random variables on (0, 1) and with $p_{ij} = a\,\forall\, i, j$, a is varied on $(-1, 0)$. Figure 7 depicts the dependence of the magnitude of the modulus of the two largest eigenvalues versus a for a collection of 64×64 matrices. For $a = 0$ to $a \sim -0.78$ the largest eigenvalue is real and its magnitude is given by

$$\lambda_d(a) \sim -\frac{a}{4} \tag{31}$$

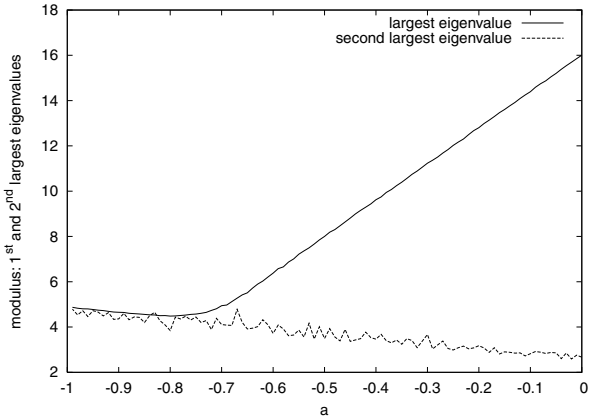

Fig. 7. This figure represents an ensemble of 1000 $d \times d$ matrices with $d = 64$ per a increment where $\delta a = 0.01$. Depicted are the modulus of largest and second largest eigenvalues. The line representing the modulus of the largest eigenvalue is given by $\frac{a}{4}$.

The knee in the curve which occurs at ~ -0.78 for $d = 64$ is dependent upon b—the upper bound on the support of the distribution. Moreover, λ_{d-1} does not remain constant with variation of the mean of the distribution but rather increases. However, aside from the largest eigenvalue, for $a \in (\sim -0.78, 0)$, the distribution of eigenvalues appears uniform on a disk of radius $|\lambda_{d-1}|$. This effect is symmetric about zero. Fixing $a = -1$ and varying b from 0 to 1 will net the same effects. Thus, for $a \in (\sim -0.78, 0)$, the corresponding dynamical system will have a 100 percent probability of undergoing a fold bifurcation. This result does not violate the circular law because the convergence to uniformity of the eigenvalues on the unit disk is only violated by a single eigenvalue which is not relevant for random matrix results regarding distributions of eigenvalues. Nevertheless, it is very relevant from our bifurcation theory perspective.

In the situation where $a = 0$, the Perron–Frobenius theorem for non-negative matrices will always apply, and thus there will exist an eigenvalue equal to $\rho(A)$ with probability one. For $a < 0$, however, the elements of the matrices will have a positive probability of being negative, and thus the matrices will have a positive probability of not satisfying the hypothesis for the Perron–Frobenius theorem. Thus, again what we are observing when a is moved from -1 to 0 is a "convergence" of sorts to the situation where the Perron–Frobenius theorem applies with probability one.

6.2.3. Perturbing Higher Moments

The variance of a uniform distribution on an interval (a, b) is given by $\text{var} = \frac{(b-a)^2}{12}$. Thus, increasing the variance by a factor of c while leaving the mean at zero is identical to multiplying a and b by $\sqrt{c}$, which is in turn identical to multiplying u_{ij} by $\sqrt{c}$. Therefore, increases in the variance of a mean-zero uniform distribution of the elements of a random matrix U has the effect of increasing the spectral radius by a factor of $\sqrt{c}$, which is simply a normalization factor and makes little difference to either the bifurcation perspective or the random matrix perspective. Again, perturbations of moments >2 are beyond the scope of this paper.

6.3. M_n Companion with Gaussian Elements

Beginning with Fig. 8, there are two important features. First, the eigenvalues appear to be distributed uniformly on the real line and on the unit circle, not on the entire disk. Second, the probability of a bifurcation due to a complex eigenvalue hovers near 55 percent over a range of $d = 128$ to 1024. This result, which seems surprising considering the spectrum depicted in Fig. 8, can be explained considering the convergence of the density of real zeros given by Edelman. This

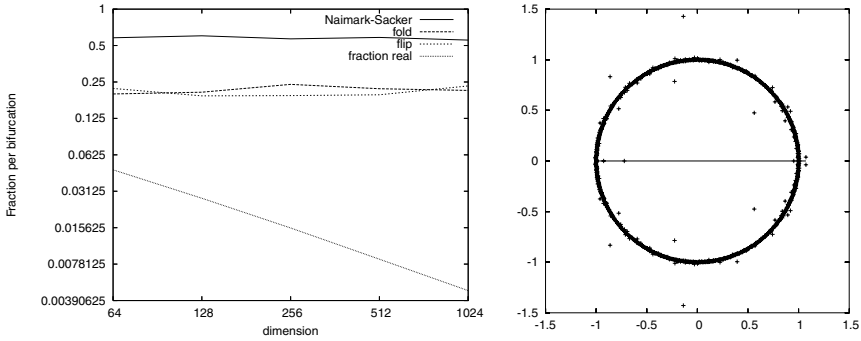

Fig. 8. On the left, the observed probability of each bifurcation was recorded for 1000 matrices with i.i.d., mean zero, variance one, Gaussian a_k's for each d (in powers of 2) along with the fraction of eigenvalues that are real. On the right is the spectrum of eigenvalues in the complex plane that corresponds to a single 1024×1024 matrix ($d = 1024$).

demonstrates how very important it is to consider the convergence of the densities when making claims regarding probable bifurcations.

Before we consider the densities, let us begin by considering the fraction of zeros that are due to real roots. Figure 9 portrays the numerically calculated fraction of real eigenvalues for a set of companion matrices with a_k's drawn from a normal

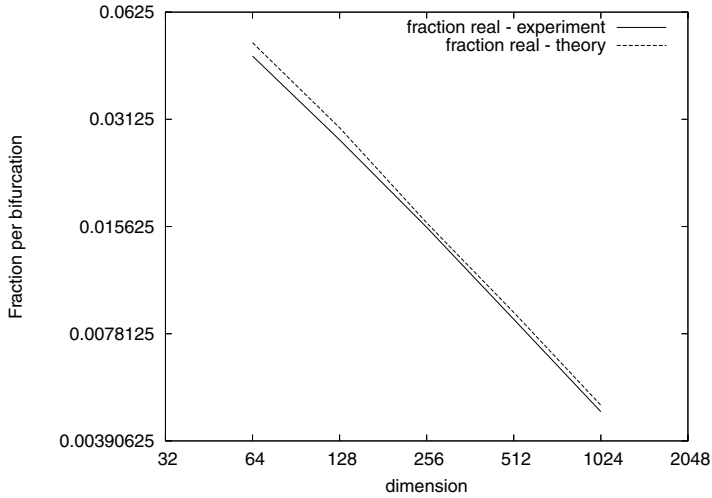

Fig. 9. Edelman's prediction for the expected number of real eigenvalues and the empirically calculated expected number of real eigenvalues. Both quantities were calculated for $d = 64, 128, 256, 512, 1024$ and have a clear power-law dependence with $\frac{E[\lambda_{real}]}{d} = 1.5d^{-0.831}$ for the numerically generated set and $E[\lambda_{real}] = 1.78d^{-0.849}$ as calculated from Edelman's formula (Eq. (17)).

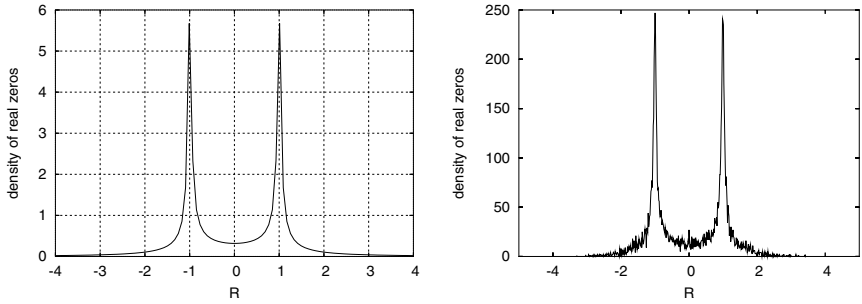

Fig. 10. On the left is the theoretical real zero density for a 64-degree polynomial with random coefficients drawn from normals with mean zero and unit variance. On the right is the real zero density for a set of 3000 companion matrices with a_k's drawn from standard normals with mean zero and unit variance.

distribution with mean zero and unit variance along with the predicted fraction of real zeros of Edelman. Clearly the two lines are in considerable agreement and have a power-law dependence with $\frac{E[\lambda_{real}]}{d} = 1.52d^{-0.831}$ for the numerically generated set and $E[\lambda_{real}] = 1.78d^{-0.849}$ as calculated from Edelman's formula given in Eq. (17). Note that the falloff of the fraction of zeros that are real is considerably faster than the other cases considered. This adds to the surprise found in Fig. 8. The fraction of eigenvalues that are real is decreasing like $d^{-0.85}$, yet the fraction of bifurcations due to real eigenvalues remains roughly constant for $d > 128$.

This seeming contradiction, a decreasing fraction of real zeros with a constant fraction of bifurcations due to real eigenvalues, is rooted in the convergence of the density of real eigenvalues. Figure 10 shows both the empirical distribution (of 1000 polynomials) and the theoretical density of real zeros for polynomials of degree 64 with coefficients drawn from a Gaussian distribution with unit variance. The obvious spikes at ± 1 are of interest, of course; however, it is the tails that extend above and beyond ± 1 that have the most impact. This feature can be highlighted at $d = 1024$ if one considers Fig. 11. The point is that the complex eigenvalues exist largely on the unit circle while the real zeros have a non-zero measure set of zeros that have magnitude greater than one. The fraction of real bifurcations is largely determined by the convergence of the density of real eigenvalues—and because the convergence has a non-zero density with magnitudes greater than one, bifurcations due to real eigenvalues will persist to very high-dimensional matrices or polynomials. The fraction of real bifurcations is constant because the ratio of the measure of *complex* eigenvalues with magnitude greater than one to the measure of real eigenvalues with magnitude greater than one is relatively constant with increasing dimension.

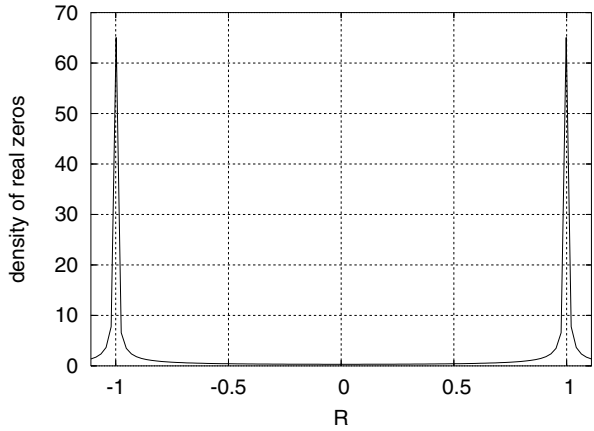

Fig. 11. Theoretical density of real zeros for polynomials of degree 1024 with coefficients drawn from a Gaussian distribution with mean zero and unit variance. For $|x| > 1.12$ the density is zero.

6.4. Neural Networks

In the case of neural networks, we do not have the luxury of having a formula to guide our understanding of the empirical distribution of the real eigenvalues or the a_k values. Compare Figs. 8 and 12. In a preliminary comparison between companion matrices with Gaussian a_k's and the neural networks with the weight structure (defined in Sec. 2), the distribution of eigenvalues appears nearly the same. However, the fraction of bifurcations of a fixed point that correspond to

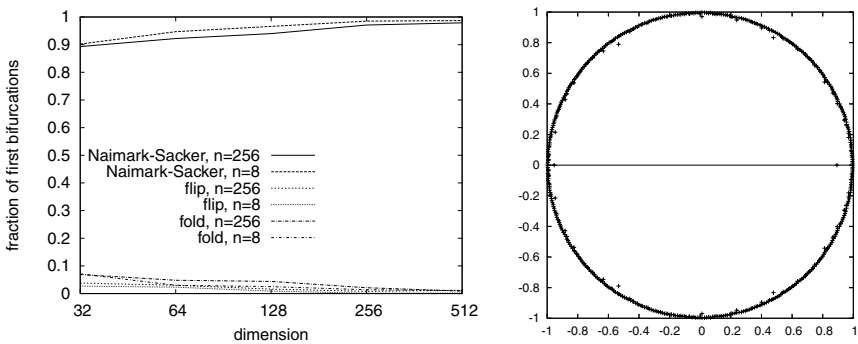

Fig. 12. On the left, the observed probability of each bifurcation was recorded for 1000 neural networks for each d along with the fraction of eigenvalues that are real. On the right is the spectrum of eigenvalues in the complex plane that corresponds to a single neural network with $n = 256$ and $d = 512$.

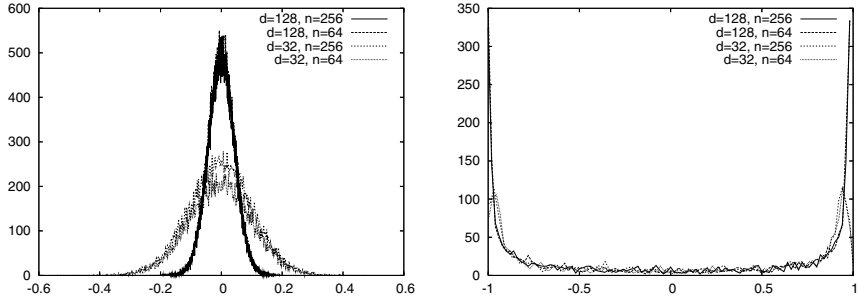

Fig. 13. The plot on the left is of the distribution of a_k's for 1000 neural networks with $n = 256$, 64 and $d = 128$, 32. The plot on the right is of the distribution of real eigenvalues along the real axis for the same set of neural networks.

Naimark–Sacker type in the neural networks tends towards unity as the dimension tends toward infinity while the fraction of Naimark–Sacker bifurcations in the companion matrices with Gaussian a_k's tends toward a constant value ($\sim$0.58) as the dimension is increased. This difference can be rectified considering Fig. 13 which depicts the distribution of real zeros and the a_k's for the neural networks. With respect to the real zeros, adding dimensions has very little effect on the interval $(-0.9, 0.9)$. However, near ± 1, the real eigenvalues are considerably more dense. Nevertheless, there do not exist the tails above and below ± 1 that are present in the companion matrices with Gaussian a_k's, which is enough to allow the fraction of bifurcations due to real zeros to tend to zero as the dimension is increased. At first glance, increasing d has a significant effect on the variance a_k's at the first bifurcation. Because s controls the variance of the a_k's, this effect is due to the s dependence of the first bifurcation. The s dependence can be understood by considering Fig. 14 which characterizes the decrease in the mean s at the first bifurcation. The decrease in the mean s-location of the first bifurcation obeys the power law $\sim a d^{0.63}$ where a depends on n. Thus, with increasing d, the variance of the a_k's at the first bifurcation point will decrease. In general, the number of neurons has a negligible effect on the distribution of real eigenvalues and only a relatively minor effect on the peak of the a_k's. Nevertheless, this power law scaling does suggest that linear coordinates are likely not the best coordinates for making asymptotic, $d \to \infty$ arguments. Moreover, this s-decrease of the first bifurcation from a fixed point is mostly an artifact of the activation function tanh (due to increasing the number and size of the terms inside tanh) and can, in a sense, be scaled away. Finally, it is worth noting that the neural networks we considered can have more than one fixed point; however, as this will have little effect on our conclusions, we will leave such an analysis for another study.

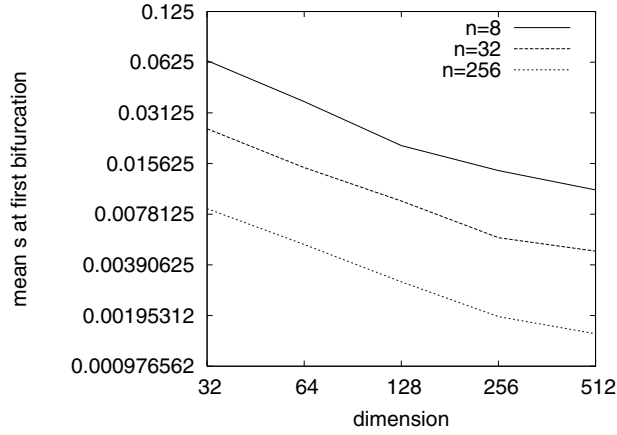

Fig. 14. The mean value of s at the first bifurcation for $n = 8$, $n = 32$, and $n = 256$ for networks of dimensions varying from 32 to 512.

7. REVISITING THE CONJECTURE

Measures such as the one provided by Bai on random matrices—any distribution with a finite sixth moment—is quite inclusive; it was originally hoped that such a measure would be enough to qualify probable bifurcations from fixed points in a large set of dynamical systems. As is now clear, simply perturbing the first moment of the distribution, while having no effects upon Bai's results, completely alters the probability of a bifurcation. In an intuitive sense, the difference corresponds to dynamical systems whose most slowly contracting directions are rotations versus dynamical systems for which a single, non-rotational contraction dominates the slowly contracting dimension. What we are left with is a much more complicated picture. It is likely, given our numerical results, that Corollary 2 can be generalized to distributions with a finite sixth moment and zero mean; however, we have not attempted to do so. Conjecture 1 is in part justified by the random matrix results presented in Sec. 6, but clearly there is much room for a more complete numerical and analytical study. The apparent distribution independence in the standard random matrix case is not present in the companion matrix case. Simply considering the examples given by Edelman or the ones presented in this paper are enough to demonstrate the existence of the diversity in the eigenvalue spectra with changes in the measure imposed on the a_k's. Analysis of the Gaussian a_k's yields evidence for the necessity of the first derivatives to be bounded in Conjecture 2 (or for the distribution of the elements of the Jacobian to have compact support). The neural network analysis yields an example of a distribution of a_k's that both satisfies the hypotheses of Conjecture 2 and provides evidence in support of Conjecture 2. Nevertheless, the constraints on the distribution of a_k's that will

yield adherence to Conjecture 2 are clearly not fully known nor understood. It is likely that an insightful application of Theorem 6 will yield a rigorous solution to this issue.

Comparison of the full random matrix and companion matrix cases is striking; the eigenvalue distributions for full random matrices are distributed on the unit disk as opposed to the unit circle in the case of companion matrices. Nevertheless, there is, via the neural network approximation scheme, a very strong connection between these two situations that has yet to be found. Recalling the discussion in Sec. 2.3.1, begin with a dynamical system

$$F(x_{t-1}) = x_t = \epsilon A x_{t-1} \tag{32}$$

where $\epsilon \in R$ is small enough such that F is a fixed point, A is a $d \times d$ real random matrix with Gaussian elements with mean zero and unit variance. There exists a neural network f of dimension $2d + 1$ that can be trained on the time-series generated by F such that f will have a spectrum for which d of its eigenvalues will be identical (within the desired numerical accuracy) as well as $d + 1$ free eigenvalues that will have magnitude less than one. Despite how the $d + 1$ "free" eigenvalues are distributed, the remaining d eigenvalues will have a distribution that is not on the unit circle and thus significantly different than any situations we have presented or know about. This connection yields insight into the distributions of a_k's of companion matrices (and thus the characteristic polynomial), as well as forging a connection between standard dynamical systems and general time-delay dynamics. Finally, the neural networks provide both an opportunity for a connection between real-world systems and the abstract dynamical models that many in the field in dynamical systems study via training and an understanding how their weight distributions affect their spectra.

We have refrained from making any rigorous asymptotic limit statements in this paper; while the companion matrix and full matrix have only $d \to \infty$ to consider, the neural networks or other function approximation spaces have both to contend with limits in the number of parameters ($n \to \infty$) as well as the number of dimensions. While this work is primarily concerned with the convergence to $d = \infty$ but not the actual limit, there are several issues we have not addressed. Focusing on the neural network case: (i) there are issues regarding the order with which asymptotic limits are taken (i.e., whether to let n or d go to infinity first); (ii) there are issues with the specific nature of the behavior of the activation function; (iii) there are issues with the existence of an infinite number of parameters and dimensions which bring about issues regarding bifurcations with infinite codimension; and finally, (iv) there is an issue regarding the appropriate bifurcation parameter coordinates for characterizing the first bifurcation. For instance, Sompolinsky *et al.*[49] and others have demonstrated that as $d \to \infty$ in their networks, the first bifurcation point goes to zero, the "routes to chaos" region disappears, and there is an abrupt bifurcation into chaos. However in Refs. 3 and 4, we have

found that both the first bifurcation point and the bifurcation to chaos follow a power-law scaling dependent on both n and d. That the s location of the first bifurcation decreases in magnitude with n, and d is mostly a function of the behavior of the tanh activation function, which is more easily saturated or activated when the number of terms in the internal sum is increased. Moreover, because of the power-law scaling, in the right log-coordinates, the "routes to chaos" region does not disappear but rather occupies the same number of s-decades. Thus the tactic we would employ to study the $n \to \infty, d \to \infty$ limits would be to find a universal rescaling and take limits with respect to a rescaled mapping. This topic is beyond the scope of this work.

8. BEYOND FIXED POINTS

Extending this construction beyond bifurcations of fixed points to the routes to chaos offers considerable problems. There are two basic approaches. One involves reduction of bifurcations for periodic orbits to bifurcations of fixed points in appropriately chosen coordinates. The other involves studying products of matrices of derivatives of periodic orbits.

Regarding the reduction of bifurcations of periodic orbits, one major problem arises because bifurcations of periodic orbits must be understood well enough to be approximated and reduced to analysis of fixed points. To see some of the currently open problems see Ref. 38. Likewise, see Ref. 9 for nice explanations of the various approximation schemes. Then the measures on the random matrices must be carried through the various approximations.

Taking products of random matrices might be a fruitful approach (see Ref. 1 for results and techniques along these lines). However, linking them to periodic orbits might be difficult.

Various authors have studied the routes to chaos computationally. Utilizing the neural network framework discussed here[4] concludes that the most likely route to chaos is a quasi-periodic one—however these results are subject to the measures imposed upon the weight matrices. Likewise, Doyon *et al.*,[13,14] Cessac *et al.*,[12] and Sompolinsky *et al.*,[49] have arrived at similar conclusions in a variety of circumstances. Clearly any results such as these will be subject to the same dependences on the measures imposed on the weight matrices as were present in the case of the bifurcation of a fixed point mentioned in the previous section.

9. FINAL REMARKS

In the context of a general dynamical system if (i) the Jacobian of the fixed point can be identified with a full random matrix and (ii) if the distribution of elements has zero mean, then, for the cases we considered (uniform and Gaussian), as the dimension approaches infinity, the probability of a

Naimark–Sacker bifurcation approaches unity. However, the largest *real* eigenvalue in the aforementioned circumstances scales linearly with the mean of the distribution of the elements of the matrix. Therefore, if the mean is large, the most probable bifurcation will be due to a real eigenvalue (e.g., a flip or a fold bifurcation), regardless of the dimension. An analytical understanding of linear scaling of the largest real eigenvalue with the mean of the distribution is unknown. Aside from the effect of the mean on the largest real eigenvalue, the result of the probability of a Naimark–Sacker bifurcation increasing with dimension is quite independent of the distribution. This is because, for nearly all distributions of elements, real random matrices have a spectrum that converges to uniformity on the unit disk as the dimension of the matrix goes to infinity.

In the context of time-delay dynamical systems, the story is different. In this case, the Jacobian forms a companion matrix. If the companion matrix has elements drawn from a Gaussian distribution, the probability of a Naimark–Sacker bifurcation saturates at $\sim$58 percent as the dimension goes to infinity. This is due to tails in the distribution of real zeros. Numerical results with time-delay, feed-forward neural networks however, behave very differently. For neural networks, as the dimension goes to infinity, the probably of a Naimark–Sacker bifurcation goes to unity. For companion matrices, the first bifurcation probability is highly dependent on the distribution of the elements of the matrix. In general, the spectrum of a time-delay dynamical system lies on the unit circle and the real line, not the entire unit disk as in the case with full matrices. However, the distriution of real zeros can vary significantly from distribution to distribution. Providing a link between the time-delay case and the standard dynamical systems framework via neural network training is suggested as future work.

ACKNOWLEDGMENTS

We thank J. Albers, R. A. Bayliss, W. Brock, J. P. Crutchfield, W. D. Dechert, I. Dobson, J. Jost, C. McTague, and J. Thompson for helpful discussions and advise. This work was partially supported at the Santa Fe Institute under the Networks Dynamics Program funded by the Intel Corporation and under the Computation, Dynamics, and Inference Program via SFI's core grants from the National Science and MacArthur Foundations. Computing was done primarily on the Max Planck Institute for Mathematics in the Sciences Beowulf cluster, however computation was carried out on the U.C. Davis Computational Science Beowulf cluster and the Santa Fe Institute Beowulf cluster.

REFERENCES

1. *Random matrices and their applications*, volume 50 of *Contemporary mathematics*. AMS (1986).
2. R. Adams and J. J. F. Fournier, *Sobolev Spaces*, 2nd edition (Elsevier, 2003).

3. D. J. Albers, J. P. Crutchfield and J. C. Sprott, Phenomenological scaling in the organization of high-dimensional dynamics. In preparation.
4. D. J. Albers and J. C. Sprott, Routes to chaos in high-dimensional dynamical systems: A qualitative numerical study. submitted, see http://www.santafe.edu/∼albers/research/publications.html (2005).
5. D. J. Albers and J. C. Sprott, Structural stability and hyperbolicity violation in high-dimensional dynamical systems. *Nonlinearity* **19**: 1801–1847 (2006).
6. 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).
7. S. Amari. Characteristics of random nets of analog neuron-like elements. *IEEE Trans. Syst. Man. Cyb.*, SMC-2 N **5**: 643–657 (1972).
8. S. Amari, H. Nagaoka and S.-I. Amari, *Methods of Information Geometry*. Translations of Mathematical Monographs. AMS (2001).
9. D. K. Arrowsmith and C. M. Place, *An Introduction to Dynamical Systems* (Cambridge University Press, 1990).
10. Z. D. Bai, Circular law. *Ann. Probab.* **25**: 494–529 (1997).
11. P. Brunovsky, On one parameter families of diffeomorphisms. *Commentationes mathematicae Universitatis Carolinae* **11**: 559–582 (1970).
12. B. Cessac, B. Doyon, M. Quoy and M. Samuelides, Mean-field equations, bifurcation map and route to chaos in discrete time neural networks. *Physica D* **74**: 24–44 (1994).
13. 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).
14. B. Doyon, B. Cessac, M. Quoy and M. Samuelides. On bifurcations and chaos in random neural networks. *Acta Biotheo.* **42**: 215–225 (1994).
15. 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).
16. A. Edelman and E. Kostlan, How many zeros of a random polynomial are real? *Bull. Amer. Math. Soc.* **32**: 1–37 (1995).
17. A. Edelman, E. Kostlan and M. Shub, How many eigenvalues of a random matrix are real? *J. Am. Math. Soc.* **7**: 247–267 (1994).
18. A. Edelman and H. Murakami, Polynomial roots from companion matrix eigenvalues. *Math. Comput.* **64**: 763–776 (1995).
19. R. Gencay and W. D. Dechert, An algorithm for the *n* Lyapunov exponents of an *n*-dimensional unknown dynamical system. *Physica D* **59**: 142–157 (1992).
20. J. Ginbre, Statistical ensembles of complex, quaternion and real matrices. *J. Math. Phys.* **6**: 440–449 (1965).
21. V. Girko, Circular law. *Theory Probab. Appl.* **29**: 694–706 (1984).
22. V. Girko, *Theory of Random Determinants* (Kluwer Academic, 1990).
23. V. Girko, The circular law: Ten years later. *Random Oper. and Stoch. Eqns.* **2**: 235–276 (1994).
24. V. Girko. Strong circular law. *Random Oper. and Stoch. Eqns.* **5**: 173–196 (1997).
25. 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).
26. V. Girko, The strong circular law. Twenty years later. Part I. *Random Oper. and Stoch. Equ.* **12**: 49–104 (2004).
27. V. Girko, The strong circular law. Twenty years later. Part II. *Random Oper. and Stoch. Equ.* **12**: 255–312 (2004).
28. V. Girko, The strong circular law. Twenty years later. Part III. *Random Oper. and Stoch. Equ.* **13**: 53–109 (2005).
29. R. Horn and C. Johnson, *Matrix Analysis* (Cambridge Univerisity Press, 1985).

30. K. Hornik, M. Stinchocombe and H. White, Mulitlayer feedforward networks are universal approximators. *Neural Networks* **2**: 359–366 (1989).
31. K. Hornik, M. Stinchocombe and H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Netw.* **3**: 551 (1990).
32. B. R. Hunt, T. Sauer and J. A. Yorke, Prevalence: a translation invariant "almost everywhere" on infinite-dimensional spaces. *Bull. Am. Math. Soc.* **27**: 217–238 (1992).
33. B. R. Hunt, T. Sauer and J. A. Yorke, Prevalence: an addendum. *Bull. Am. Math. Soc.* **28**: 2306–307 (1993).
34. M. Kac, On the average number of real roots of a random algebraic equation. *Bull. Am. Math. Soc.* **49**: 314–320 and 938 (1943).
35. H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd edition (Cambridge University Press, 2003).
36. E. Kanzieper and G. Akemann, Towards spectral theory of Ginibre's ensemble of real random matrices. *Phys. Rev. Lett.* **95**: 230201 (2005).
37. A. Katok, Lyapunov exponents, entropy, and periodic orbits for diffeomorphisms. *Publ. Math. I.H.E.S.* **51**: 137–174 (1980).
38. Y. Kuznetzov, *Bifurcation Theory*, 2nd edition (Springer-Verlag, 1998).
39. M. L. Mehta, *Random Matrices*, chapter 15 (Elsevier, 2004).
40. O. Moynot and M. Samuelides, Large deviations and mean-field threory for asymmetric random recurrent neural networks. *Probab. Theory Relat. Fields* **123**: 41–75 (2002).
41. J. Neimark, On some cases of periodic motions depending on parameters. *Dokl. Acad. Nauk SSSR* **129**: 736–739 (1959).
42. V. I. Oseledec, A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Tras. Moscow Math. Soc.* **19**: 197–221 (1968).
43. W. Ott and J. Yorke, Learning about reality from observations. *SIAM J. Appl. Dyn. Syst.*, **3**: 297–322 (2003).
44. W. Ott and J. A. Yorke, Prevalence. *Bull. Am. Math. Soc.* **42**: 263–290 (2005).
45. Y. B. Pesin, Lyapunov characteristic exponents and smooth ergodic theory. *English Transl. Russian Math. Surv.* **32**: 55–114 (1977).
46. D. Ruelle, Characteristic exponents and invariant manifolds in Hilbert space. *Ann. Math.* **115**: 243–290 (1982).
47. R. Sacker, On invariant surfaces and bifurcation of periodic solutions of ordinary differential equations. Technical Report 333 (New York State University, 1964).
48. T. Sauer, J. Yorke and M. Casdagli, Embedology. *J. Stat. Phys.* **65**: 579–616 (1991).
49. H. Sompolinsky, A. Crisanti and H. J. Sommers, Chaos in random neural networks. *Phys. Rev. Lett.* (1988).
50. J. Sotomayor, Structural stability and bifurcation theory. *Dyn. Syst.* 549–560 (1973).
51. J. C. Sprott, *Chaos and Time-series Analysis* (Oxford University Press, 2003).
52. F. Takens, Detecting atrange attractors in turbulence. In D. Rand and L. Young (Eds.), *Lecture Notes in Mathematics*, vol. 898, pp. 366–381, Dynamical Systems and Turbulence, Warwick (Springer-Verlag, Berlin, 1981).
53. S. Wiggins, *Global Bifurcations and Chaos: Analytical Methods* (AMS. Springer-Verlag, 1988).
54. S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, vol. 2 of TAM. (Springer-Verlag, 1994).