Routes to Chaos in Neural Networks with Random Weights

D. J. Albers*†

in collaboration with

J. C. Sprott‡

and

W. D. Dechert§

Chaos and Complexity
Seminar
10/28/97

*Department of Physics, University of Wisconsin, Madison. E-mail: albers@phys-next1.physics.wisc.edu

†D. J. Albers acknowledges special thanks to William Brock, Ian Dobason and Cosma Shalizi for many helpful discussions and Derek Wright for assistance in computing as system adminisrator.

‡Department of Physics, University of Wisconsin, Madison. E-mail: sprott@juno.physics.wisc.edu

§Department of Economics, University of Houston. E-mail: econ71@jetson.uh.edu

Outline

- Motivation
- Brief Definitions
- Neural Networks
  - As Dynamical Systems
  - How Our Networks are Set Up
- Techniques Used
  - Lyapunov Exponent
  - Eigenvalue Calculation
- Results
  - Frequency of Chaos
  - First Bifurcation
  - Qualitive Dynamics
    - The Second Bifurcation
    - The Rest of the Route
- Future/Current Work
- Conclusions and Recap

Pertinent Definitions

- Dynamical System: Systems that Evolve in Time; a Way of Describing the Passage in Time of All Points of a Given Space

- Chaos:
  - Dense Periodic Points
  - Sensitive Dependence on Initial Conditions (Implied by Dense Periodic Points)
  - Topological Transitivity

- Bifurcation: A Qualitative Change in the System Dynamics as a Function of a Control Parameter

  - Saddle-Node
    - Corresponds to the Appearance of 2 Stable Branches of Equilibria
  - Flip
    - Corresponds to the Appearance of a Period 2 Oscillation
  - Hopf (Naimark-Sacker)
    - Corresponds to the Appearance of a Limit Cycle or Torus
- Route to Chaos: The Path of Bifurcations that a System Undergoes from a Steady State to a Chaotic State as a Control Parameter is Varied

General Networks

- Neural Networks as Dynamical Systems
  - Form the Map:

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

- Properties of General Feedforward, Multi-layer Neural Networks
  - Approximation Characteristics

    ".. multilayer feedforward networks with as few as a single hidden layer and fairly arbitary hidden layer activation functions are in fact capable of arbitrarly accurate approximation to an unknown mapping and its derivatives, to as many orders as desired."

– Significance of the Above Statement

> "‥ a net appropriately trained to approximate the transfer function of a (perfectly measured) deterministic chaos could be used to obtain information on the Lyapounov exponents of the underlying chaos. (the Lyapounov exponents are defined in terms of the first derivatives of the transfer function)."

Hornik et al.

What Makes Up a Multi-layer Neural Network

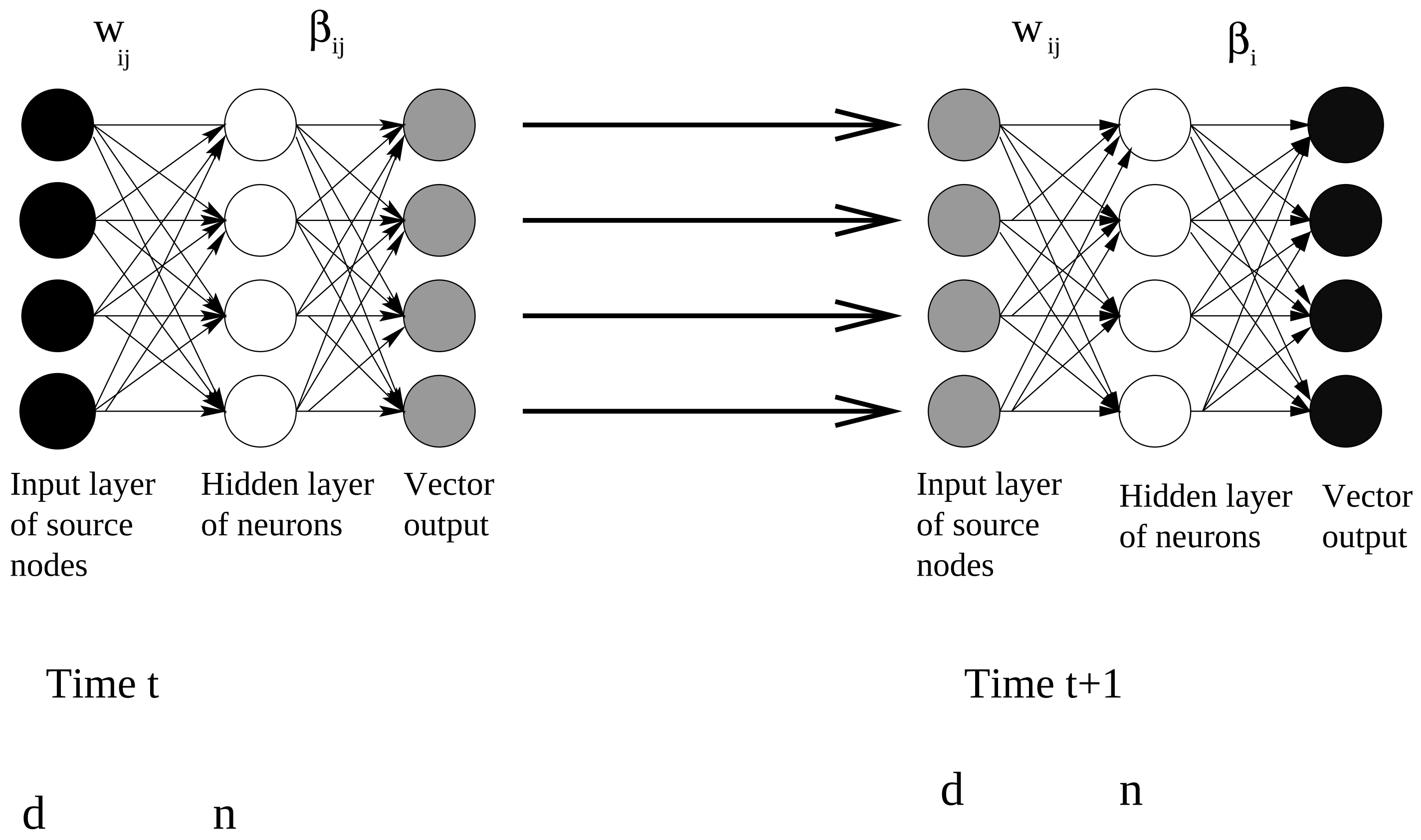
- Input Layer of Source Nodes

- Layers of Neurons or Hidden Units

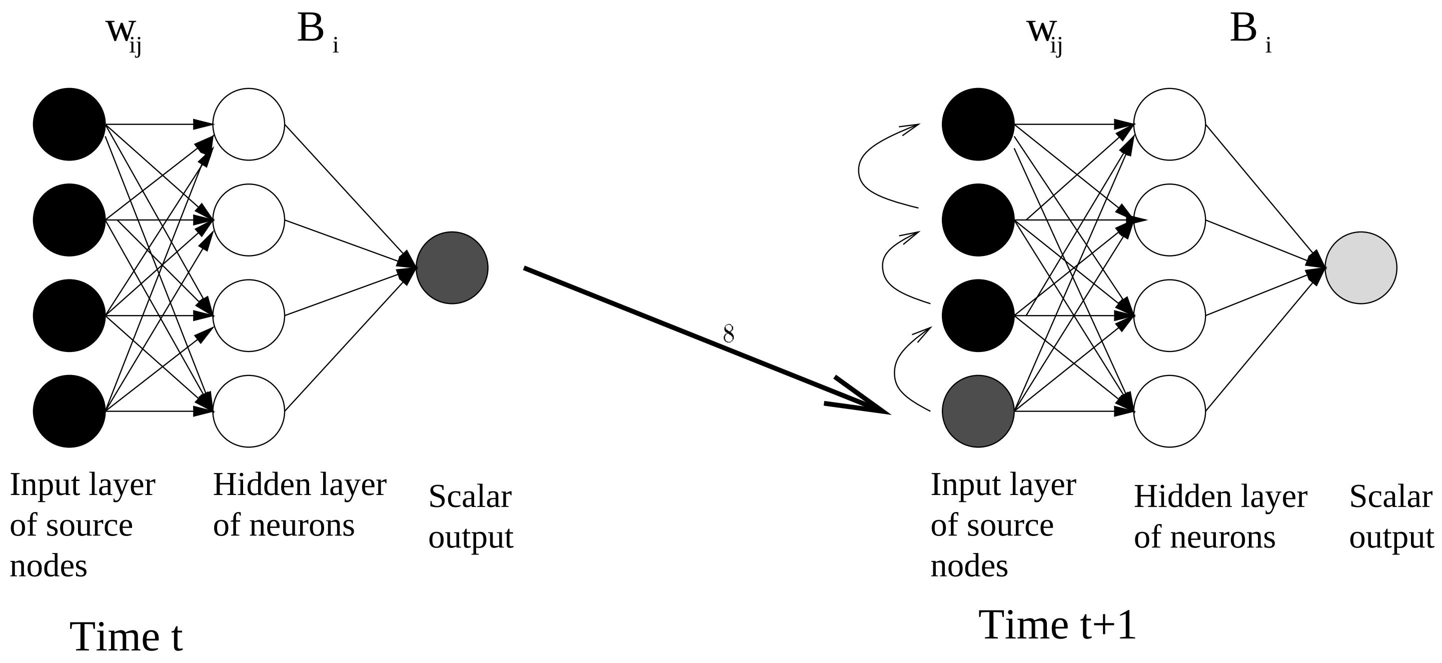
- Outputs

Diagrams of Scalar versus Vector Networks

Vector Networks



Scalar Networks



General Mathematical Form

- $O = \beta * \phi(w * I)$
  Where O is the matrix of outputs, β is a matrix of weights, ϕ is the squashing function, or neuron, w a matrix of weights, and I is the matrix of inputs.

Feedforward

- Feedforward: The Relative Input Layer of Source Nodes Projects onto the Relative Output Layer of Nodes, but Not Vice Versa

Fully Connected

- Every Node in Each Layer is Connected to All Nodes in the Adjacent Layers

Our Networks

- Mathematical Form

$$f(y) = \sum_{i=1}^{n} \beta_i \phi \left(s w_{i0} + s \sum_{j=1}^{d} w_{ij} y_j \right) \tag{2}$$

  – Variables

    * n, the Number of Neurons
    * d, the "Dimension" of the Network, or the Number of Inputs
    * s, the Spread of the Gaussian of the w Matrix of Weights, Used as the Bifurcation Parameter
    * Squashing Function ϕ

- Our Squashing Function
  - The tanh() Function
    - The Hyperbolic Tangent Function is General
  - Note: Region a is Nearly Linear
  - Note: Region b is the Non-Linear Region
  - Note: Region c is the Saturated, (Binary), Region

- Picking the Weight Matrices

  - β_i's Uniform Random from 0 to 1 and then Normalized such that the Sum of the β^2's Equal n

  - w_{ij} Random Gaussian Distribution with a Spread of s

  - Initial Conditions Uniformly Random from -1 to 1

Largest Lyapunov Exponent

- Definition: The Lyapunov Exponent is the Average Rate of Spreading of Nearby Trajectories in the Direction of Maximum Spreading

- The Algorithm

  - Displace the Original Trajectory, $L(y)$ by an Amount ϵ; We Will Call this Displaced Trajectory $L'(y)$

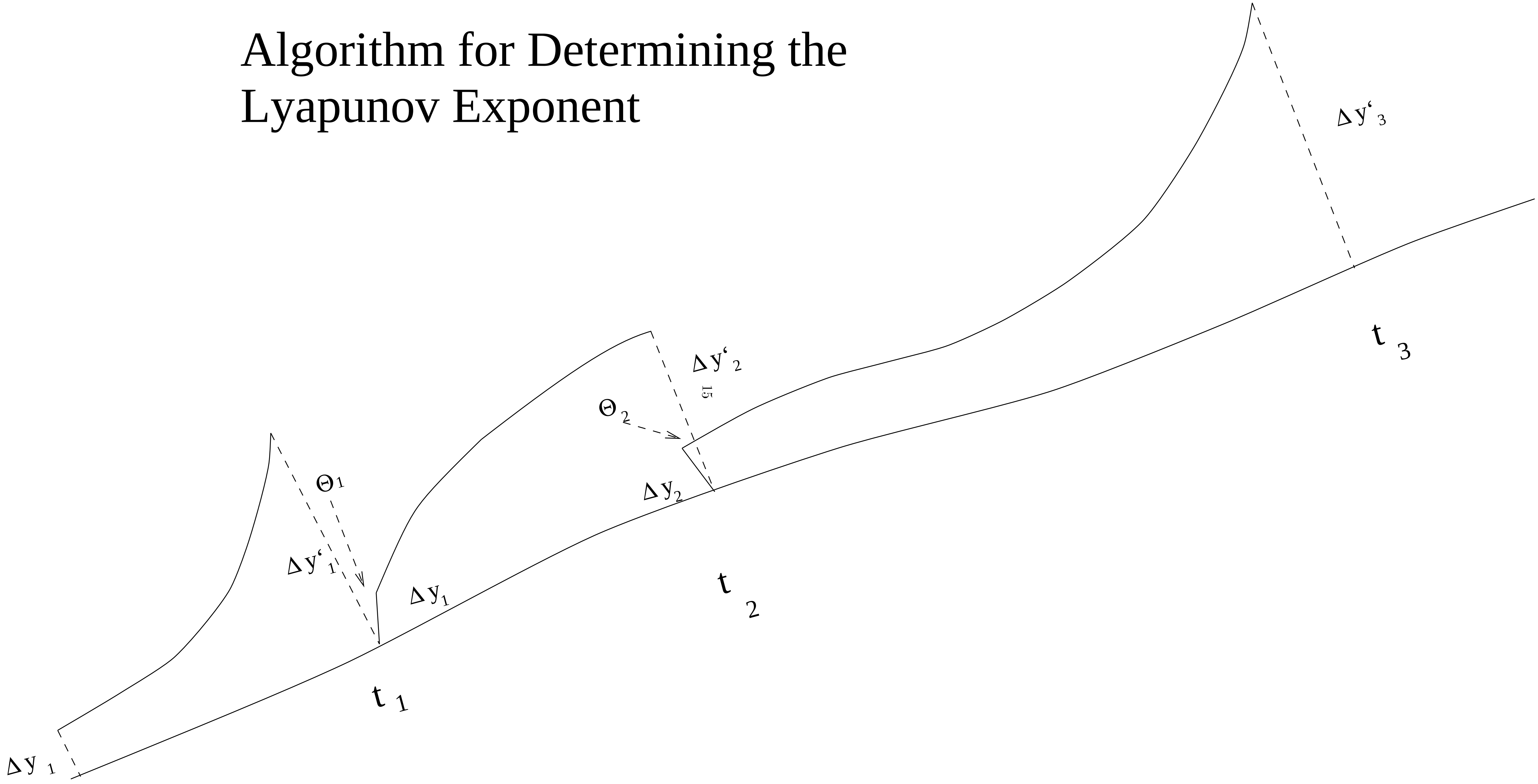
  - Iterate both Points One Time Step

  - Calculate the Trajectory Spreading

  - Calculate the Lyapunov Exponent

  - Renormalize the Trajectory and Continue the Process

Algorithm for Determining the Lyapunov Exponent



- Mathematical Representation

  - $\ln < y_s > = \lambda$

  - Rescale Δy_1 to a Length ϵ From the Trajectory Being Followed

  - $\lambda = t^{-1} \Sigma_{s=0}^{t-1} \ln \frac{|\Delta y_{s+1}|}{|\Delta y_s|}$

- Significance

  - The Lyapunov Exponent is a Measure of the Sensitivity to Initial Conditions

  - We Experimentally Define A System to be Chaotic if its Largest Lyapunov Exponent is Positive

Jacobian and the Eigenvalues

- Jacobian
  - Like the Derivative of a Function of Many Variables with Respect to each Variable
  - The Determinant of a Matrix of Partial Derivatives
  - The Solutions of the Characteristic Equation or Eigenvalues of the Jacobian Determine the Stability of the System
  - The Number of Eigenvalues is Equal to the Dimension of the System

- Eigenvalues
  - The Reason for Stability; the Eigenvalues as Solutions have Roughly the Form

    $$e^{\lambda_{eigen}} \tag{3}$$

    When the Exponent of e is Greater than or Equal to One, the Solution Explodes to Infinity

  - When the Modulus of the Largest Eigenvalue is ≥ 1 the System is Unstable
    * Flip Bifurcation: When the Largest Eigenvalue is Negative and Real Crosses the Unit Circle

    * Saddle Node Bifurcation: When the Largest Eigenvalue is Positive and Real Crosses the Unit Circle

    * Naimark-Sacker Bifurcation: When the Largest Eigenvalue is Complex Crosses the Unit Circle
  - Bifurcations Occur when the Modulus of the Largest Eigenvalue is Equal to 1

- The Relationship between the Largest Eigenvalue and the Largest Lyapunov Exponent:

$$\lambda_{Lyap} = log|\lambda_{eigen}| \tag{4}$$

- Problems and Computational Woes
  - We Calculate the Lyapunov Exponent Numerically, and Thus It Tends to be Rough on Fine Scale; It is Numerically Stable Throughout the Entire Route to Chaos
  - The Jacobian for Our Systems Gets Very Complicated For High Periodic Orbits, Limit Cycles, and Chaos. Numerical Stability Problems Arise, and Thus can Only Really be Used to Determine the First Bifrucation

The Quasi-Periodic Route

- Quasi-Periodic Route: From the Initial State of the System (Usually a Fixed Point), All the Bifurcations Up to Chaos Consist of Hopf Bifurcations

- Theorem by Girko: The Measure of the Set of Real Eigenvalues is Zero Under Girko's Limit Distribution Which is Uniform Over a Circle in the Complex Plane

- Apply Girko to the Jacobian which is a $d \times d$ Matrix

- Based on this We are Inclined to Think that Hopf Bifurcations are Most Probable at High Dimension, thus the Quasi-Periodic Route would be the Dominate Route to Chaos at High Dimension

Probability of Chaos

- The Experiment:
  - Monte Carlo Study of Several Hundred Thousand Cases
  - Select s, d, and n at Random
  - Select and Fix the β and w Matrices
  - Iterate the System and Calculate the Largest Lyapunov Exponent
- Constant n, Variable s and d.
  - Note the Increasing Frequency of Chaos as the Complexity of the Networks Increase.
  - Note the C Shaped Curve and its Relation to the s Parameter.
- Constant s, Variable d and n.
  - Again Note the Increase in Frequency of Chaos as the Complexity Rises.

- Dependence on Means of Weight Selection

Probability of First Bifurcation

- Based on the Calculation of the Eigenvalues from the Jacobian

- n Dependence
  - Artifact of How We Choose the β Matrix
    * For Naimark-Sacker Bifurcation to Occur, Discriminant Must be Negative

    * Increasing n Increases the Variance of the Coefficients of the Matrix; This Pushes the Expected Value of the Discriminant Postive
- d Dependence
  - Increase in d Gives Rise to a Increase in the Probability of a Naimark-Sacker Bifurcation

Qualitative Results After the First Bifurcation

- Low d Low n
  - High Dynamic Diversity
  - Not Very Chaotic
  - "Jumpy" Dynamics
- Low d High n
  - High Dynamic Diversity
  - More Chaotic Than At Low n
  - "Jumpy" Dynamics
- High d Low n
  - Less Dynamic Diversity
  - Dominated by Limit Cycles and Tori
  - Magnitude of the Largest Lyapunov Exponent Decreased Due to the Increase in Time Lags

- High d High n
  - Less Dynamic Diversity
  - Chaos Over a much Larger s Range
  - Dominated by Limit Cycles, Tori, and Chaos
  - Magnitude of the Largest Lyapunov Exponent Decreased Due to the Increase in Time Lags
- Basin Jumping
- The Nearly Binary Region, Point Intermittant Chaos
- Saturated Network Highest Period

$$P_{highest} = d2^n \tag{5}$$

Concluding Remarks

- Conclusions
  - As the Dimension of the System Increases the Systems Frequency of Chaos of Goes to Unity Depending on the Choice of s.
  - As the Number of Neurons in the System Increases, the System's Probability of Chaos of Goes to Unity Given Our Particular Method of Choosing Weights.
  - In General, as the Complexity of the Systems Increase, the Probability of the Systems being Chaotic Increases to Unity.
  - The Probabilty of the First Bifurcation Being Naimark-Sacker Increases to Near Unity as the Dimension is Increased
  - Qualitatively the Quasi-Periodic Route Dominates at High d

Future and Current Work.

- Understand Vector Networks Behavior and Relationship to Scalar Networks; the Extent of the Space of Neural Networks Sampled

- Closer Investigation Into the Nearly Binary Region

- Construct a More Rigorous Result After the First Bifurcation