Dynamical Behavior of Artificial Neural Networks with Random Weights

D. J. Albers*†

in collaboration with

J. C. Sprott‡

and

W. D. Dechert§

Chaos and Complexity
Seminar
10/29/96

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

†D. J. Albers acknowledges special thanks to 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
- Introduction to General Properties of Neural Networks
- Our Networks
- Experments and Analysis Performed
  - Largest Lyapunov Exponent
  - Correlation Dimension
- Results
  - Frequency of Chaos
  - Largest Lyapunov Exponent
  - Correlation Dimension

- Future/Current Work

- Conclusions and Recap

Motivation and General Networks

- A General Model of a Deterministic System
- General Feedfoward, Fully Connected, Multi-layer Neural Networks
- What Makes Up a Multi-layer Neural Network
  - Input Layer of Source Nodes
  - Neurons of Hidden Units
  - Outputs
  - Meaning of Multilayer
- 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
- Properties of 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."

– Signifigance 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.

– Note that these Systems are Discrete Time Systems

Our Networks

- Scalar Networks
  - Why Scalar Networks?
  - 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
- Mathematical Form

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

Where β and ω are the matrices of weights, y_j is the matrix of inputs, ϕ is the squashing function, d is the number of inputs, and n is the number of neurons.

- 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
  - Alternative β Distrubution: The RMS Value of the β_i's Equal to 1
  - 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 Rate of Spreading of Nearby Trajectories.

- 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

- Mathematical Representation

– $|\Delta y_0| = \epsilon$

– $\ln \frac{|\Delta y_1|}{|\Delta y_0|} = \lambda$

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

– $\lambda = t^{-1} \Sigma_{s=0}^{t-1} \ln \frac{|\Delta y_{s+1}|}{|\Delta y_s}$
Where t is the total number of iterations, and Δy_s and Δy_{s+1} are as noted above.

- Significance
  - The Lyapunov Exponent is a Measure of the Sensitivity to Initial Conditions
  - We Define A System to be Chaotic if its Largest Lyapunov Exponent is Positive
    * Note: Chaos is Properly Defined as Deterministic, Aperiodic Systems with Sensitive Dependence on Initial Conditions

Dimension

- Embedding Dimension

  - The Embedding Dimension: this is the Number of Delays, or Variables of the System. In our Case the Number of Delays is Exactly the Number of Inputs, or what We Call d, the Dimension of the Network.

- Correlation Dimension; How Dense is the Attractor?

  - A Measure of How Frequently a Typical Trajectory Visits a Region on the Attractor.

- Process of Determining the Correlation Dimension

  - Define a Radius that is Known to Contain all the Points of the System that We Care About.

- Count the Points that Lie within the Given Radius, R, of that Point.

- Repeat this Process for each Point.

- Divide that Sum by the Total Number of Points, N, Multiplied by the Total Number of Points Minus One, $(N-1)$: this Gives a Value for the Correlation Function, $C(R)$, at a Given Radius, R.

- Repeat this Process for Decreasing Radii Until there no Longer are any Points within a Radius R of any Point in the System.

- Take the Slope of the Line Defined by $\ln C(R)$ versus $\ln R$: this is the Correlation Dimension

• How Many Points are Enough? An estimate is given by Tsonis to be:

$$N = 10^{2*d+1} \tag{2}$$

- A Mathematical Discription

  - In One Embedding Dimension:

    $$C(R) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, i \neq j}^{N} \Theta\left(R - |x_i - x_j|\right) \tag{3}$$

    Where R is the radius, N is the number of points, x_i is the point in the time series and Θ is defined as:

    $$\Theta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases} \tag{4}$$

  - Recall the Embedding Dimension, d_e, is the Number of Delays. Again, in our Case the Number of Delays is the Number of Inputs, d, also known as the Dimension of the Network.

  - The Above $C(R)$ can be Generalized for any Embedding Dimension:

$$C(R)=\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1,i\neq j}^{N}\Theta\left(R-\left(\sum_{c=1}^{d}\left(x_{i-c}-x_{j-c}\right)^2\right)^{1/2}\right) \tag{5}$$

- Correlation Dimension's Relation to the Lyapunov Exponent
  - There as Many Lyapunov Exponents as Dimensions
  - If the Largest Lyapunov Exponent is Zero then the Solution is Quasi-Periodic.
    - A n Dimensional Quasi-Periodic Solution Corresponds to a n-Torus.
    - The Correlation Dimension of an n-Torus is n
    - A 1-Dimensional Quasi-Periodic Solution Corresponds to a Discretely Sampled 1-Torus or Limit Cycle

Results

- Frequency of Chaos
  - 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.
    - Note the C Shaped Curve and its Relation to the n Parameter.
    - Changing the β Distribution Changes the Shape of this Contour. It also Allows the

β's to get Bigger than One.

- Largest Lyapunov Exponent
  - Constant s, (8), Variable n and d.
    - $n = 4$ $d = 4$.
    - $n = 64$ $d = 4$.
      - Note the Spike at Zero and the Similarity in Distribution
    - $n = 4$ $d = 64$.
    - $n = 64$ $d = 64$.
      - Note the apparent asymptote to zero. Is this effect real or fake? As d increases so does the time scale that the system evolves on; we believe that this effect is an artifact of the increase in time scale and is not profound.
  - Constant d and n, (64), and Variable s.

- $s = 1$
- $s = 4096$
  - The effect of very small and very large s on the system Note that at very large s the system is forced into the binary region of the tanh function and very small s forces the system into the linear region of the tanh function.

- Correlation Dimension
  - Constant n and s, Variable d.
    * $n = 8$ $d = 2$
    * $n = 8$ $d = 4$
    * $n = 8$ $d = 6$
      · Note that the correlation dimension seems to be much smaller than the embedding dimension of the system.

      · Note the spike at correlation dimension of one, these exist mainly do to the error in calculating the lyapunov exponent and correspond to *limit* cycles or quasi-periodic solutions to maps. (i.e. the ratio of frequencies is an irrational number)

      · Note the lack of the appearance of 2-toruses. A 2-torus is a 2-dimensional quasi-periodic solution. A 2-torus on a map is analgous to a 3 torus in a flow – 3 toruses are unstable, (Piexito's theorem).

- The n parameter seem to have a slight effect on the distibution of the correlation dimension, but this effect is very slight.

- The s parameter seems to have the same effect on the correlation dimension as it does on the Lyapunov exponent; both very small and very large values of s tend to smolder out the dynamics.

* Scatter Plot, $n = 8$, $d = 4$, $s = 8$.

  - Note what happens if the systems with Lyapunov exponent ~ 0 are removed.

Concluding Remarks and Future Work

- Future and Current Work.
  - How the Vector Networks Behave and Relate to the Scalar Networks.
  - How these Systems Tend to Bifurcate.
  - What Routes to Chaos do They Take, and which are Most Prevelant
  - Scaling Laws with Regards to the s, d, and n Parameters with Regards to the Frequency of Chaos, Largest Lyapunov Exponent, and Correlation Dimension.

- 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 Systems Frequency of Chaos of Goes to Unity if a Particular Choice of Weights is Made.
  - In General, as the Complexity of the Systems Increase, the Probability of the Systems being Chaotic Increases to Unity.
  - The Correlation Dimension of the System Tends to be Below 1/2 the Embedding Dimension, or, d of the System.