Chaos in Random Neural Networks

dated: November 9, 2022

Abstract

Measured electrical signals in the brain are often temporally irregular, seemingly at odds with the orderly behavior that they produce. What is the origin and function of this irregularity? What are the different "phases" of dynamics one can expect in the brain? What characterizes transitions between these phases? We will give a pedagogical overview of the effort to understand the dynamics of coupled neurons in the brain, via the study of toy models. There is a rich phenomenology associated with the high dimensional dynamical systems often used to model large ensembles of neurons, and their analysis draws variously upon areas of physics including spin glasses, dynamical mean field theory, critical phenomena, and random matrix theory. There are many open questions remaining in the field, along with possible connections to the function of artificial neural networks. Relevant references are Sompolinsky, Crisanti, Sommers (PRL 1988), van Vreeswijk & Sompolinsky (*Neural computation* 1998), and Kadmon & Sompolinsky (PRX 2015).

How can theory help us to understand the irregular patterns of electrical activity in the brain via analysis of toy models? Broadly speaking these models fall into two categories: 1) mean field models of firing rate and 2) models of spiking dynamics on a sparse network. We will focus on these two classes of models: rate models and spiking models, and describe general phenomenology and methods of analysis. We will also try to make connections to artificial neural nets and describe open questions.

1 Introduction

1.1 Neuron Physiology

Neurons are comprised of a cell body (soma), dendrites, which receive electrical signals, and an axon, through which signals propagate. The axon terminates in branches which form synapses with dendrites from other neurons, and it can be quite long (up to 1 meter). Synapses are the connections between neurons, e.g. the junctions between axon terminals and dendrites of other cells.

Action potentials: a membrane potential is a difference in electrical potential energy between the inside and outside of a cell. It takes energy to set up this different, and the sodium potassium pump is the mechanism that is largely responsible for setting up this gradient. In neurons, energy spent on this pump can account for 3/4 of the total energy expenditure of the cell. The pump moves 3 sodium ions outside the cell, and 2 potassium ions inside, thereby creating an excess of negative charge inside the cell. Furthermore, the cell membrane has channels which make it more permeable to potassium, so the potassium can diffuse out and further increase the resting potential difference (which is typically around -70 mV). *Most animal cells maintain a potential difference with an excess of negative charge inside the cell.*

Depolarization: a neuron can depolarize when the resting potential flips sign due to an influx of positive charge. This happens if voltage-gated sodium or calcium channels open, triggered by some external stimulus, and positive ions rush into the cell. The channels then close again once the positive charge has entered the cell, and the pump can once again establish resting potential. These voltage-gated channels open at a membrane potential of around -55 mV. Therefore the system is excitable: a small perturbation will simply relax back to resting potential, but a large enough perturbation will cause the membrane potential to make a big excursion before it returns to resting value. There is a refractory period after firing when another depolarization cannot occur: this means that even under a constant current, the response of the cell will be periodic, with the neuron firing (spiking) regularly. The firing rate is a nonlinear (saturating function) of the action potential due to the presence of the refractory period. When the soma depolarizes, a wave of depolarization can travel down an axon: this is an action potential. When an action potential reaches a synapse, it triggers release of neurotransmitters which can open ion channels in the pre-synaptic neuron (the cell sending the signal). Depending on the nature of the neurotransmitter and ion channel, the neuron can inhibit or excite depolarization in the post-synaptic neuron (the cell receiving signal). Inhibitory neurons comprise about 20% of the neurons in the brain but are important; the combined effects of excitatory and inhibitory neurons control the firing of the other neurons in the network. Excitatory synapses tend to transmit glutamate/aspartate, whereas inhibitory synapses use GABA/glycine (among others). The inhibitory synapses result in increases cell permeability to Cl^- which causes hyperpolarization wherein negative ions flow into the cell.

Although the dynamics of a single neuron firing are relatively well understood, the brain is comprised of $\sim 10^{11}$ neurons, each of which receives input from $\sim 10^4$ other neurons. The dynamics of this large nonlinear system are not understood.

1.2 What is a Neural Network?

Once we hook up many neurons with some connections, we have a *neural network*. This might have some channels for input and output, but can also exhibit spontaneous behavior. There are two types of neural networks that are commonly

discussed: feed-forward and recurrent. A feed-forward network has connections which do not form a cycle, whereas a recurrent neural net has cycles which endow it which temporally dynamic behavior. The subject of these notes is recurrent neural nets, but it is worth noting that some systems (i.e. fast aspects of the visual system) are thought to be realized through feed-forward connections.

1.3 Basic Phenomena

Empirical fact: neural assemblies display spontaneous irregular patterns of activity. However, when cortical sections are measured in vitro with applied voltage, spiking patterns are regular. This indicates that internal activity (fluctuating input) is somehow responsible for the observed patterns. However, since a single neuron can have thousands of synaptic connections, it is not expected that the input fluctuates much, unless the inputs are excitatory and inhibitory and largely cancel each other out. Can this balance emerge without fine-tuned network properties? *Why are firing patterns of neurons in the brain so temporally variable, seemingly independent of external stimuli?* Experiments [1] show that distribution of intervals between spike times in vivo is close to exponential, which is what one would expect from a random Poisson process, such as radioactive decay. The firing rate is not constant in time. However if a neuron is placed in vitro an subjected to a constant current stimulus, it fires predictably with a firing rate that is reproducible [2]. Therefore the internal dynamics that lead to irregular patterns of firing are a puzzle.

2 Mean Field Rate Model (Sompolinsky, Crisanti, Sommers 1988)

One of the first papers [3] exploring chaotic dynamics in neural networks models N neurons by variables $-1 \leq S_i \leq 1$ defined as $S_i(t) = \phi(h_i(t))$ for $-\infty < h_i < \infty$ and nonlinear "activation" function ϕ. One can think of S as the firing rate and h as the membrane potential. We take $\phi(s) = \tanh(gx)$ where g is the *gain* parameter, controlling the nonlinearity of the response. Then we investigate the model

$$\dot{h}_i = -h_i + \sum_{j=1}^{N} J_{ij}\phi(h_j), \tag{1}$$

which supposes that the input to neuron i is a random linear combination of the outputs of all the other neurons, where we set $J_{ii} = 0$. We draw the J_{ij} from a gaussian distribution with variance J^2/N and zero mean. If $J_{ij} = J_{ji}$ then the equations have a Lyapunov function which is decreasing along trajectories of the dynamics. Consider the function

$$\mathcal{L} = \sum_{i=1}^{N}[h_i\phi(h_i) - \Phi(h_i)] - \frac{1}{2}\sum_{i\neq j} J_{ij}\phi(h_i)\phi(h_j), \tag{2}$$

where we have summed over $i \neq j$ since the J_{ii} are 0 anyway. Here the notation $\Phi(x)$ denotes the antiderivative of $\phi(x)$. We can show that $\mathcal{L}$ is decreasing along trajectories of our model. In particular, we know that $\frac{d\mathcal{L}}{dt} = \sum_i \dot{h}_i \partial_{h_i}\mathcal{L}$. From the form of $\mathcal{L}$ we can see that

$$-\partial_{h_i}\mathcal{L} = \phi'(h_i)\left[-h_i + \sum_{j=1}^{N} J_{ij}\phi(h_j)\right]. \tag{3}$$

Therefore as long as ϕ is monotonic, $-\partial_{h_i}\mathcal{L}$ has the same sign as $\dot{h}_i$, and so $\frac{d\mathcal{L}}{dt} \leq 0$. Note that this argument only works if $J_{ij} = J_{ji}$, which is what allowed $\partial_{h_i}\mathcal{L}$ to have similar form to $\dot{h}_i$.

We are more broadly interested in the case where $J_{ij} \neq J_{ji}$ which will give rise to nonrelaxational dynamics. Therefore in the following we choose each J_{ij} as an independent gaussian random variable with variance J^2/N and mean zero. For now we set $J = 1$.

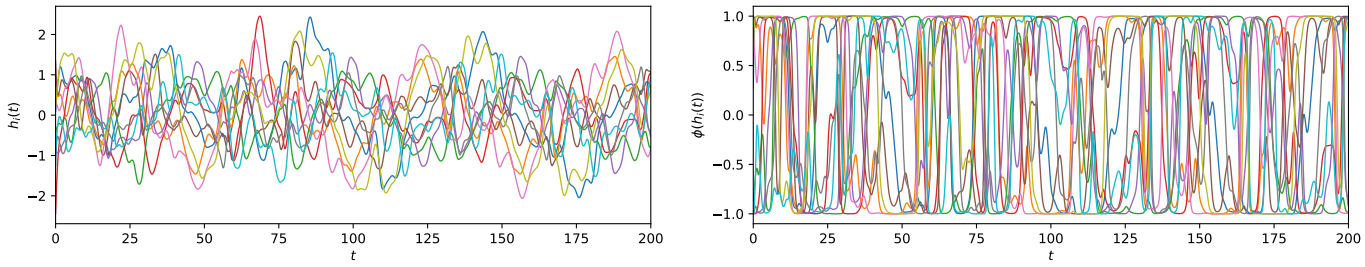

Figure 1: (left) The membrane potentials of a few neurons out of a simulation of $N = 100$ neurons in the chaotic phase of the dynamics. Every neuron interacts with every other neuron, with an independent gaussian-distributed interaction strength. (right) The firing rates of these same neurons, related by the nonlinearity ϕ.

2.1 Dynamical Mean Field Theory

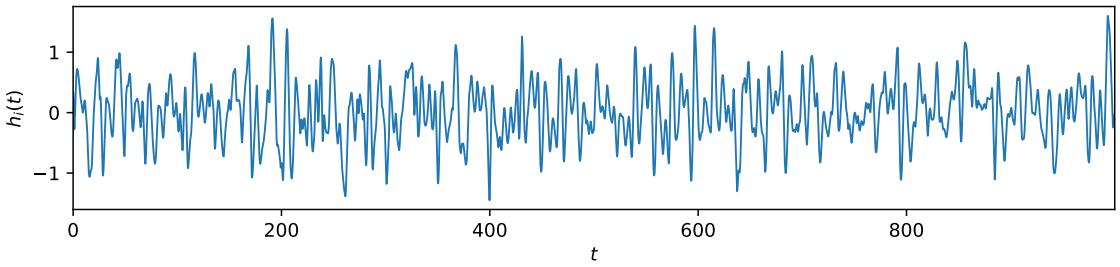

Figure 2: The dynamics of a single neuron in a bath of all the others. The goal of dynamical field theory is to obtain self consistency conditions on these dynamics so that we can describe these dynamics as an effective stochastic process.

In order to obtain a statistical description of the chaos that we observe, we will attempt to recast the many-body disordered deterministic dynamical system above with a single-neuron description that is stochastic: the price we pay is an addition self consistency condition that must be satisfied by the noise driving the stochastic equation $\dot{h}_i = -h_i + \eta(t)$. In other words, we do not know the statistics of $\eta(t)$, which we will assume to be gaussian. From our dynamical equations we can see that the correlations in $\eta(t)$ will be

$$\langle \eta_i(t)\eta_i(t+\tau)\rangle = \left\langle \sum_{j,k} J_{ij}J_{ik}\phi(h_j(t))\phi(h_k(t+\tau)) \right\rangle \approx \langle \phi(h_i(t))\phi(h_i(t+\tau))\rangle, \tag{4}$$

by the self-averaging property of the disorder. We get two self consistent equations for the quantities h and η, namely:

$$\dot{h}_i = -h_i + \eta_i(t) \quad \text{and} \quad \langle \eta_i(t)\eta_i(t+\tau)\rangle = \langle \phi(h_i(t))\phi(h_i(t+\tau)) \equiv C(\tau). \tag{5}$$

Solving these equations is tricky: the authors of this paper formulate an equation for $\Delta(\tau) = \langle h_i(t)h_i(t+\tau)\rangle$ which evolves in a self-consistent potential that depends on $\Delta(0)$.

This equation can be derived by essentially an application of the Wiener Khinchin (WK) theorem, which states the that power spectrum (magnitude squared of the Fourier transform) of a signal is equal to the Fourier transform of the signal's autocorrelation function. We will apply this term to the stochastic signal $\eta(t)$. In Fourier space, our dynamical equation for h reads

$$(1 + i\omega)\tilde{h}(\omega) = \tilde{\eta}(\omega). \tag{6}$$

Therefore the power spectrum of h is given by

$$(1 + \omega^2)|\tilde{h}(\omega)|^2 = |\tilde{\eta}(\omega)|^2 = \tilde{C}(\omega). \tag{7}$$

Further identifying $|\tilde{h}(\omega)|^2$ with $\Delta(\tau)$ via the WK theorem, and inverting the Fourier transform, we obtain

$$\left[1 - \frac{d^2}{d\tau^2}\right]\Delta(\tau) = C(\tau). \tag{8}$$

Now we need an expression for $C(\tau)$, which we can get by representing $h(t)$ and $h(t+\tau)$ as correlated gaussian random variables with

$$h(t) = \sqrt{\Delta(0) - \Delta(\tau)}y_1 + \sqrt{\Delta(\tau)}z \tag{9}$$

$$h(t+\tau) = \sqrt{\Delta(0) - \Delta(\tau)}y_2 + \sqrt{\Delta(\tau)}z \tag{10}$$

where y_1, y_2 and z are i.i.d. standard gaussian random variables. Defining $\alpha = \Delta(0) - \Delta(\tau)$ and $\beta = \Delta(\tau)$, we can write

$$C(\tau) = \int Dy_1 Dy_2 Dz \phi(\sqrt{\alpha}y_1 + \sqrt{\beta}z)\phi(\sqrt{\alpha}y_2 + \sqrt{\beta}z) = \int Dz \left[\int Dy \phi(\sqrt{\alpha}y + \sqrt{\beta}z)\right]^2. \tag{11}$$

Therefore

$$\ddot{\Delta}(\tau) = \Delta(\tau) - \int Dz \left[\int Dy \phi\left(\sqrt{\Delta(0) - \Delta(\tau)}y + \sqrt{\Delta(\tau)}z\right)\right]^2. \tag{12}$$

This equation can be cast in terms of a potential so $\ddot{\Delta} = -\partial V/\partial\Delta$.

2.2 Phase Transition

Different forms of $\Delta(\tau)$ indicate different behaviors. There are three basic regimes for $V(\Delta)$. For $g < 1$ we get an inverted well. Then the only solution is $\Delta(\tau) = 0$ which occurs when all the firing rates go to 0. For $g > 1$ the potential is either a single or double well potential, and the behavior is accordingly a limit cycle or chaos (or fixed point). The chaotic solution for Δ is evident in the double well potential. In the chaotic phase $g > 1$ the only stable solution is that for which Δ is monotonically decreasing to 0, meaning that chaos is the stable phase. In this phase, one can solve the equation for Δ to get

$$\Delta(\tau) = \epsilon \cosh^{-2}\left(\frac{\epsilon t}{\sqrt{3}}\right), \quad \text{where } \epsilon = g - 1 \ll 1. \tag{13}$$

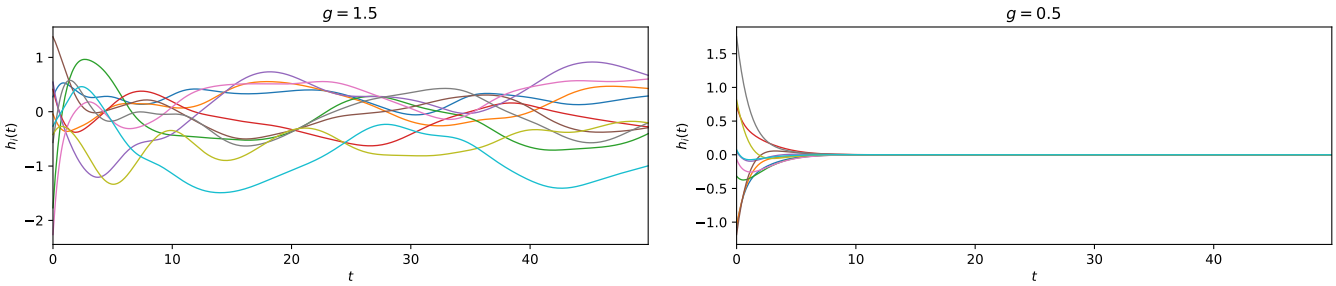

Figure 3: The dynamics of the firing rates on either side of the phase transition, for $J = 1$ and $N = 100$.

2.3 Lyapunov Exponents

The authors also derive some results on the Lyapunov exponents, through the off-diagonal average susceptibility defined as $\chi_{ij}(t, t') = \frac{\delta h_i(t)}{\eta_j(t')}$. Based on the long-time behavior of this, information about the leading Lyapunov exponent can be derived. In particular the leading Lyapunov exponent dominates the exponential growth of $\chi^2(t, 0) = \sum_{ij} \chi_{ij}(t, 0)$ at late t. By taking the derivative of the original dynamical equations we have that

$$\left(1 + \frac{d}{dt}\right) \chi_{ij}(t, t') = \sum_k J_{ik} \phi'(h_k) \chi_{kj}(t, t'). \tag{14}$$

Based on this, one can find that $\chi^2(t) = \sum_{n=0}^{\infty} \chi_n e^{2\omega_n t}$ where the ω_n are solutions to an eigenvalue problem. One can find that the larges λ is positive for the constant Δ solutions from the chaotic phase, implying that these solutions are unstable. The only consistent solution is positive λ for the decaying Δ. In the chaotic phase one finds that $\lambda = \epsilon^2/2$, indicating that the phase transition is continuous. However, in this paper the full distribution of Lyapunov exponents is not calculated.

3 Sparse Networks & Neuron Firing (van Vreeswijk & Sompolinsky, 1998)

The previous assumption of all-to-all coupling of neurons is hardly justifiable. How might we come up with a more realistic model? Here we consult reference [4], which explicitly models a population of excitatory and inhibitory neurons, along with some other number outside the network, which provide constant input to the network. The authors find that the network structure results in dynamical balance between excitatory and inhibitory inputs for each neuron close to the firing threshold, meaning that small fluctuations around this threshold cause chaotic spiking patterns.

The model tracks the state of a number of binary neurons σ_k^i belonging to population k. A neuron is active if its input is above a threshold, and its activity is updated according to

$$\sigma_k^i(t) = \Theta\left[u_k^i(t)\right] \quad \text{where } u_k^i(t) = \sum_{j,l} J_{kl}^{ij} \sigma_l^i(t) + u_k^0 - \theta_k. \tag{15}$$

Here θ_k is the firing threshold, and u_k^0 is the constant input to each neuron due to external input neurons which are not part of the network being studied. Connectivity is random and sparse, meaning that the number of inputs to a neuron does not scale with the total number of neurons. We say that each neuron receives inputs from an average of K excitatory and K inhibitory neurons, while the total number of neurons is N. Nevertheless one can make a mean field theory for this model by first taking $N \to \infty$ and then taking the number of connections $K \to \infty$.

In this model the scaling of various quantities with K is very important. In particular, the synaptic efficacies scale as $1/\sqrt{K}$ with K excitatory and inhibitory neurons projecting onto each other neuron. The external excitatory neurons

have input $\sim \sqrt{K}$ and the threshold is ~ 1. Therefore things must cancel out to order $\sqrt{K}$ in order for fluctuations in the input of a given neuron to be close to the threshold. Note that the additivity of the activities inside the threshold function is responsible for a global linear response to input despite the nonlinearity in the system.

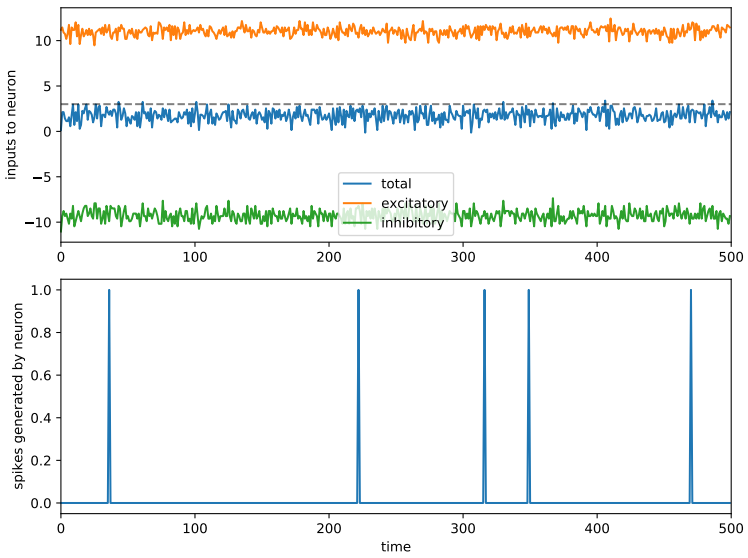

Figure 4: The inputs and behavior of a single excitatory neuron in a simulation of a spiking model with 300 excitatory neurons and 100 inhibitory neurons, and an average connectivity of $K = 50$. The state of each neuron is updated according to some predetermined schedule which updates each neuron in the same order each time step. The neural network finds a state where the inhibitory and excitatory inputs to the neuron almost cancel out, causing the neuron to spike intermittently due to fluctuations in its input around the spiking threshold. In this simulation we take all the J_{kl}^{ij} to be the same $(1/\sqrt{K})$ and so the random connectivity is what gives rise to fluctuations.

3.1 Mean Field Theory

We denote the input to each neuron in population k (either excitatory or inhibitory) as u_k^i, and the state of each neuron (active or inactive) as σ_k^i. Then the activity level of a given neuron as a function of time is $m_k^i(t) = \langle \sigma_k^i(t) \rangle$, where the average is over initial conditions and update times.

The continuous time dynamics of the activity level will obey

$$\tau_k \frac{d}{dt} m_k^i(t) = -m_k^i(t) + \Theta(u_k^i(t)), \tag{16}$$

where the decay term comes from flipping the active spins to inactive and the activation term comes from the flipping of inactive spins to active. Mean field theory assumes that the activities of different input cells are uncorrelated. We can then make the approximation that

$$\tau_k \frac{d}{dt} m_k(t) = -m_k(t) + \sum_{n_1, n_2} p(n_1, n_2) \Theta(u_k^i(t)), \tag{17}$$

where we have averaged over the joint distribution of the independent random variables n_1 and n_2, which are the number of excitatory and inhibitory inputs respectively. Assuming independence gives a binomial distribution of the number of active inputs, and since the number of connections is Poisson distributed,

$$p_l(n) = \sum_{s=n}^{\infty} \frac{K^s}{s!} e^{-K} \binom{s}{n} m_l^n (1 - m_l)^{s-n}. \tag{18}$$

If we take the $K \to \infty$ limit we can use a gaussian approximation for the Poisson and binomial distributions, and reduce our dynamics for $m_k(t)$ to something in terms of error functions. In a similar fashion one can derive an equation for the autocorrelation function.

The upshot is that we can show the existence and stability of a balanced state, wherein the excitatory and inhibitory inputs cancel each other out to first order in $\sqrt{K}$, and fluctuations around this cause chaotic spiking behavior. Note that in this kind of model the transition to chaos has been less established than in the mean field model mentioned earlier. Some more recent work has explored this transition in sparse networks [5]. This work has also explored the distinction between rate dynamics and spiking dynamics, and shown that in a spiking model like the one considered here, there can be a transition to chaos provided the synapses are sufficiently slow, which allows a coarse-graining to smooth rate dynamics.

4 Artificial Neural Networks (Poole et al 2016)

There is an ongoing effort to understand signal processing in artificial neural nets, which are unexpectedly good at many tasks for reasons that we do not understand. Does this have anything to do with the brain? We do not know. However, recent work [6] has shown that there are manifestations of chaos as signals propagate through large random feed-forward artificial neural nets.

These feed-forward nets are defined as

$$\mathbf{x}^l = \phi(\mathbf{h}^l), \quad \mathbf{h}^l = \mathbf{W}^l \mathbf{x}^{l-1} + \mathbf{b}^l, \tag{19}$$

where l indexes the layer in the deep neural network. We will consider a random set of weights and biases such that the elements of $\mathbf{W}$ and $\mathbf{b}$ are gaussian distributed, with the entries of $\mathbf{W}$ normalized by the network width N (which could vary in each layer) with variance σ_w^2/N_{l-1} and biases with variance σ_b^2. It turns out that by looking at the divergence of two similar inputs to the neural net, whose dynamics can be derived by a similar mean field approach, we can find a transition in the σ_w–σ_b plane where the differences between two inputs become large (i.e. the signals become uncorrelated) on one side of the transition, and the signals converge on the other side. Note that this is somewhat similar to the transition observed in the chaotic mean field neural nets, as far as tuning the strength of disorder induces an order-chaos transition (since the dimensionless control parameter in that model is gJ).

In addition to applying this notion to a pair of points, we can apply it to a whole collection of points, i.e. a manifold, and explore how the geometry of the manifold changes as it is propagated through the neural net. One finds that in the ordered phase, the manifold remains smooth, while in the chaotic phase it gets folded over and over again and its curvature increases exponentially as it propagates through the network. This behavior is not found in a shallow neural network with one hidden layer. This implies that the ability to compute highly nonlinear/curved functions is strongly enhanced by neural network depth. For example, if I have to locate a decision boundary that is highly curved, this will be much easier if I can increase the depth of my network. Although here we have focuses on chaos as the input propagates through the network, there may be ties to *weight chaos*, wherein the neural net function is chaotic in the weights.

5 Ongoing Work & Open Questions

There is continued work trying to relate these neural phenomena to machine learning. In addition, the understanding of the chaotic phase in a sparsely connected rate model does not seem to have been fully explored. How does the topology of the connection network influence the phase of the dynamics? Another interesting set of questions perhaps has to do with the expressivity of recurrent neural nets, and how/whether chaotic dynamics in these systems could result in increased expressivity. How does chaos help (if at all) with computation? This seems to be a big unanswered question.

References

[1] William R Softky and Christof Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of neuroscience*, 13(1):334–350, 1993.

[2] Zachary F Mainen and Terrence J Sejnowski. Reliability of spike timing in neocortical neurons. *Science*, 268(5216):1503–1506, 1995.

[3] Haim Sompolinsky, Andrea Crisanti, and Hans-Jurgen Sommers. Chaos in random neural networks. *Physical review letters*, 61(3):259, 1988.

[4] Carl Van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.

[5] Jonathan Kadmon and Haim Sompolinsky. Transition to chaos in random neuronal networks. *Physical Review X*, 5(4):041030, 2015.

[6] Ben Poole, Subhaneil Lahiri, Maithra Raghu, Jascha Sohl-Dickstein, and Surya Ganguli. Exponential expressivity in deep neural networks through transient chaos. *Advances in neural information processing systems*, 29, 2016.