2015

Chaos and Learning in Discrete-Time Neural Networks

Jess M. Banks
Oberlin College

CHAOS AND LEARNING IN DISCRETE-TIME NEURAL NETWORKS

JESS BANKS

Abstract. We study a family of discrete-time recurrent neural network models in which the synaptic connectivity changes slowly with respect to the neuronal dynamics. The fast (neuronal) dynamics of these models display a wealth of behaviors ranging from simple convergence and oscillation to chaos, and the addition of slow (synaptic) dynamics which mimic the biological mechanisms of learning and memory induces complex multiscale dynamics which render rigorous analysis quite difficult. Nevertheless, we prove a general result on the interplay of these two dynamical timescales, demarcating a regime of parameter space within which a gradual dampening of chaotic neuronal behavior is induced by a broad class of learning rules.

1. Introduction

Simple models are vital for the study of biological systems. A model which manages to capture some biologically relevant behavior can teach us a great deal about the important structure of a system or phenomenon, even if its simplicity hides much of the true system's apparent complexity. The ways in which such models successfully replicate nature are informative, but perhaps more illuminating are the ways in which they fall short. This paper will study a several related models of the interaction of neurons in the central nervous system which attempt to capture the brain's remarkable capacity for learning and memory, and we will primarily endeavor to understand (i) the mathematical structure of these models which supports this replication of our nervous system's functionality and (ii) the ways in which this structure fails to account for fundamental features of the underlying biology.

Neurons, along with glial cells, are the major cellular constituents of the central nervous system. Although a full account of neural biology is beyond the scope of this paper, a cursory understanding is important both to motivate the models which we will study and to give insight into the biological aspects which are excluded in our mathematization. Cognition emerges from the multitude of minute interactions between neurons; roughly speaking this intercellular communication proceeds via electrochemical impulses called 'action-potentials' which travel a long protrusion from the neural cell body called an *axon*. At the axon's terminal end, located proximate to a 'postsynaptic' neuron, these impulses induce the release of neurotransmitters, chemicals which bind to the receiving neuron and can either promote or inhibit the generation of further action potentials. In general, a neuron will 'fire' as a result

Date: October 26, 2015.

of the net excitation and inhibition from the neurons which synapse onto it.

The models which we study in this paper make a number of simplifying assumptions. Most notably, we will not attempt to model the complicated process by which action potentials are generated, although this is a classic and still-vibrant area of research (see i.e. [15], [19]). Instead we will describe each neuron's behavior in terms of its 'firing rate', i.e. the rate at which it is producing action potentials at any given time. In addition, real neural networks have sparse connectivity which often displays small-world or rich-club structure with its associated network statistics ([28], [14], [18]), but we will assume that our networks are fully connected, with synaptic strengths drawn independently from some ground distribution (usually a Gaussian); we will consequently allow feedback through synaptic cycles of the type $A \to B \to \cdots \to A$. Finally, we will assume that neurons update their firing rates in discrete time-steps according to the firing rates of the other neurons and the strength and type (inhibitory or excitatory) of the synapses between them. In general we may term this type of model a Random Recurrent Neural Network (RRNN).

RRNNs have been used to study numerous features of the nervous system, but our focus will be on their ability to replicate and encode important aspects of learning and memory. The brain has a well-documented capacity for associative memory and pattern recognition, including some tasks at which we are not yet adept in training machines to perform (see [7], [8], [20]). In addition, our neural architecture exhibits a remarkable degree of plasticity, and this malleability of the synaptic couplings between neurons supports episodic, semantic, and procedural memory. We will begin by studying a series of models which endeavor to capture associative memory and pattern recognition. These models rely on some mechanism for 'learning' the patterns to be recognized, and the remainder of the paper will be devoted to extensions of the associative memory models which incorporate explicitly this learning. In particular, Sections 6 and 7 present novel results tying a particular implementation of plasticity to a dampening of chaotic behavior.

2. Associative Learning with the Hopfield Model

A classic and extensively studied neural network model was analyzed most famously by Hopfied in [16], and in this section we will present this model and shadow Hopfield's classic demonstration of its capacity to perform associative pattern recognition. Consider a population of N neurons, each of which can be either 'firing' (1) or 'quiet' (0), and denote the state of the ensemble at time t by a vector $\boldsymbol{x}(t)$. The dynamics of this model are driven by the synaptic couplings between neurons, which we store in an $N \times N$ matrix $\mathcal{W}$. In the original model, Hopfield conceptualized the system as operating in continuous time, each neuron updating its state according to the rule

$$x_i \mapsto \begin{cases} 1 & \sum_j \mathcal{W}_{ij} x_j > d_i \\ 0 & else \end{cases}, \tag{2.1}$$

independently and with some average rate p. The vector $\boldsymbol{d}$ contains each neuron's 'threshold', a parameter which we can consider as tuning a given neuron's proclivity towards either activity or quiescence, and which we will take to be zero throughout our analysis. Because Hopfield's neurons update probabilistically and instantaneously in continuous time, there is a vanishing probability of two neurons updating synchronously, and we can equivalently treat the model as processing in *discrete* time, where at each time-step one neuron is selected uniformly at random to update. Many Hopfield-type models instead allow all neurons to update synchronously, and these variants exhibit the same qualitative features as the original system ([4], [11]). However synchronous systems are arguably *less* biologically consistent, in that they require top-down control by a 'global clock'. We will at some points below have occasion to study such models, though we will always port our analysis over to the asynchronous domain.

Hopfield's primary achievement was to demonstrate that, provided that $\mathcal{W}$ is properly chosen, this network can store binary strings as dynamical attractors; given some initial condition, the Hopfield dynamics drive the system towards a nearby (i.e. with respect to Hamming distance) 'memory' pattern at which the system will be fixed. Hopfield's analysis of this model begins with the observation that its dynamics are, in fact, isomorphic to the classic Ising model for a 'spin-glass' system of the type studied in statistical physics. In such a system, we consider a collection of 'spins' which are permitted to have magnetic moment pointing either 'up' or 'down', and which are subject to pairwise magnetic interactions (in analogy with our neurons, binary firing rates, and synaptic couplings respectively). At positive temperature, it is common to model such systems as evolving stochastically, with spins more likely to flip states in a direction that reduces the overall energy of the system; at zero temperature, however, we model them as only undergoing changes of state which reduce the system's total potential energy. This behavior motivates us to define the following function associating to each state $\boldsymbol{x}$ of the Hopfield model an 'energy' given by

$$\mathcal{H}(\boldsymbol{x}) = -\frac{1}{2}\sum_{i,j}\mathcal{W}_{ij}x_i x_j; \tag{2.2}$$

this function and the following results can be modified for the case where our neuronal thresholds $\boldsymbol{d}$ are nonzero. We will restate Hopfield's classic proof.

Claim 2.3. *(Hopfield* [16]*) For symmetric couplings $\mathcal{W}$, the Hopfield dynamics* (2.1) *induce a monotonic decrease in the energy function* (2.2)*.*

Proof. We begin by computing the change in energy due to the update of the neuron $x_k \mapsto x_k'$; let's denote by $\Delta\boldsymbol{x}$ a vector which has kth component $x_k' - x_k$ and zeros elsewhere. Noticing that $\mathcal{H}(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{x}^T\mathcal{W}\boldsymbol{x}$, we can compute that the change in energy due to this

update is

$$\begin{aligned}\Delta\mathcal{H} &= -\frac{1}{2}\left((\boldsymbol{x}+\Delta\boldsymbol{x})^T\mathcal{W}(\boldsymbol{x}+\Delta\boldsymbol{x}) - \boldsymbol{x}^T\mathcal{W}\boldsymbol{x}\right)\\ &= -\frac{1}{2}\left(\boldsymbol{x}^T\mathcal{W}\boldsymbol{x} + 2\boldsymbol{x}^T\mathcal{W}\Delta\boldsymbol{x}^T + \Delta\boldsymbol{x}^T\mathcal{W}\Delta\boldsymbol{x}^T - \boldsymbol{x}^T\mathcal{W}\boldsymbol{x}^T\right)\\ &= -\sum \mathcal{W}_{ij}x_i\Delta x_j - \frac{1}{2}\mathcal{W}_{kk}\\ &= -\Delta x_k\sum_j \mathcal{W}_{kj}x_j, \qquad (2.4)\end{aligned}$$

where we have repeatedly used the fact that $\mathcal{W}$ is symmetric with zero diagonal. Our update rule (2.1) means that $\Delta x_k = 1$ exactly when the sum in (2.4) is positive, and $\Delta x_k = -1$ when it is nonnegative; thus in either case the update causes a monotonic decrease in $\mathcal{H}$, with this decrease being strictly negative when x_k flips from 0 to 1. □

Claim 2.3 is sufficient to demonstrate that a Hopfield network with symmetric couplings will evolve towards an attracting fixed state, since our energy function (2.2) must achieve a minimum on the system's finite state space. In particular, limit cycles are rendered impossible by the *strict* decrease in $\mathcal{H}$ in the $\Delta x_k = 1$ case. In fact, the system will almost surely (with respect to the stochasticity of the neuronal updating) find a fixed state in finite time. Now, imagine that we are given a modest collection $\boldsymbol{y}_1, ..., \boldsymbol{y}_l$ of binary strings which we would like our network to store as its attractors. If we set the couplings as

$$\mathcal{W}_{ij} = \sum_{p=1}^{l}(2y_i^p - 1)(2y_j^p - 1), \qquad (2.5)$$

where y_i^p denotes the ith entry in $\boldsymbol{y}_p$, these patterns will under appropriate conditions be local minima in the energy landscape of $\mathcal{H}$. We sketch an argument to this effect, again following Hopfield [16].

Let's say that the network, with couplings $\mathcal{W}$ defined as in (2.5), is in one of our chosen states $\boldsymbol{y}_\phi$. As above we consider whether or not to update kth neuron by computing

$$\begin{aligned}\sum_j \mathcal{W}_{kj}y_j^\phi &= \sum_j y_j^\phi\left(\sum_p (2y_k^p - 1)(2y_j^p - 1)\right)\\ &= \sum_p (2y_k^p - 1)\left(\sum_j y_j^\phi(2y_j^p - 1)\right).\end{aligned}$$

Notice that if the $\boldsymbol{y}_p$ are chosen uniformly at random from the space of length-N binary strings, the parenthesized inner sum will have zero mean with respect to this randomness with the exception of the $p = \phi$ term. It is in this average sense that our stored patterns are stable, though there are clear caveats here with regard to (i) the specific choices of the $\boldsymbol{y}$ and

(ii) the number of patterns we are attempting to store. A collection $\{\boldsymbol{y}_p\}$ which is too large or contains too many similar 'memories' will yield attractor states which are fragmented combinations of the desired patterns.

3. Modifications of the Hopfield Model

Two assumptions intrinsic to the Hopfield model seem particularly heavy-handed, the symmetry of $\mathcal{W}$ and the restriction of neuronal states to the set $\{0, 1\}$. Synaptic connections in the brain are not symmetric; the fact that one neuron synapses on another does not necessitate that a synapse in the other direction will exist at all, let alone one of comparable type or strength. Moreover, neurons can exhibit a continuum of firing rates between quiescence and maximum rate. The Hopfield model has been modified in numerous ways to relax these assumptions (i.e. [9],[17]), and in this section we will study the effects of these modifications, beginning with synaptic asymmetry and moving on to a model with continuous state space.

In the previous section, we saw that symmetric couplings in the Hopfield model strongly constrain the types of dynamics that these networks can exhibit: the network will invariably converge to a local minimum in the energy landscape. This is in strong contrast with the wealth of ensemble behaviors, including avalanches, long-timescale oscillations, and seizure states ([2], [6]), exhibited by living neurons. When we study Hopfield-type networks with asymmetric coupling matrices, it quickly becomes clear that a vastly wider range of dynamics are accessible, dependent on the model parameters and additional choices we are free to make regarding $\mathcal{W}$. There is a substantial literature on the behavior of various types of biologically motivated coupling matrices, including choices of $\mathcal{W}$ for which each neuron is allowed to make either inhibitory or excitatory synapses, but not both, or for which the weights of all incoming synapses to a given neuron are constrained to have zero sum (i.e. the incoming inhibition and excitation are balanced) [23]. Many such types of asymmetric coupling matrices have enough added structure to admit specialized analytic tools, a study if which is unfortunately outside the scope of this work.

A glance back at our proofs of Claim 2.3 shows that the symmetry of $\mathcal{W}$ is crucial for our argument that the Hopfield dynamics are monotonic in our chosen energy function (2.2) and consequently processes towards a fixed state; these claims in fact fail spectacularly for asymmetric couplings. Figure 1 gives a simple example on three neurons in which an infinite cycle through six states is possible, and a quick computation shows that $\mathcal{H}$ oscillates between 0 and 0.5 along this orbit.

Given that the Hopfield dynamics are stochastic—we choose which neuron to update at random—the difference which we have just illustrated between the symmetric and asymmetric cases is subtle and important. For symmetric weight matrices, Claim 2.3 indicates that despite this stochastic element, Hopfield networks exhibit strongly ordered behavior: the network will always find an attracting fixed state. However, for asymmetric weights,

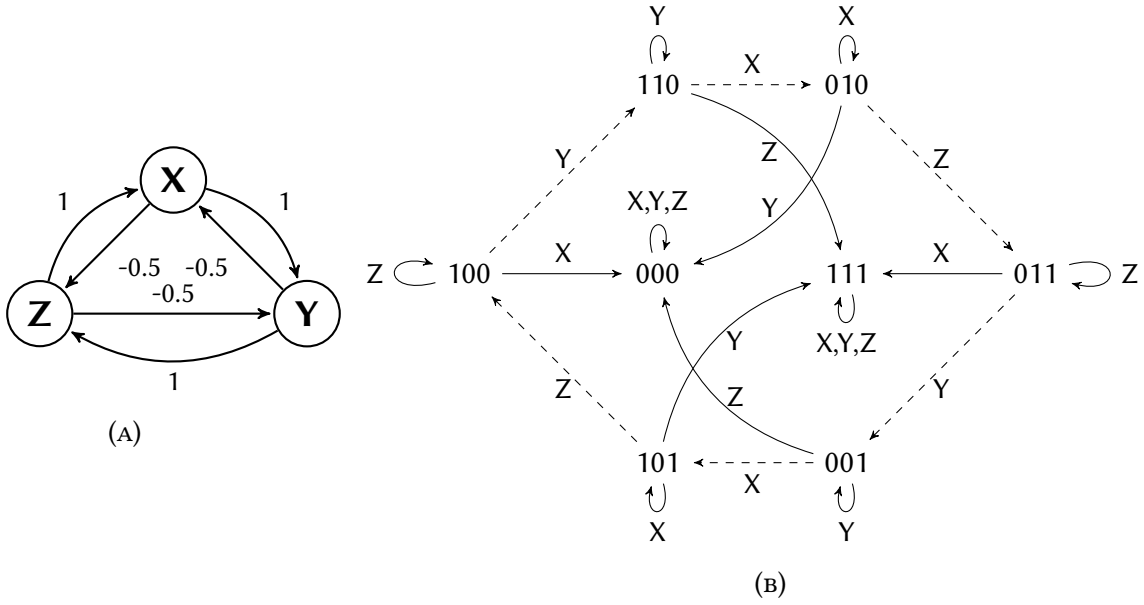

FIGURE 1. A simple asymmetric Hopfield network (1a) on three neurons, **X**,**Y**, and **Z**, and a diagram of the dynamics through its state space (1b), with binary strings corresponding to ordered triples (**X** state, **Y** state, **Z** state). Arrows in (1b) indicate state changes due to the update of a particular neuron. Notice that the network can cycle through the six outer states indefinitely.

the example in Figure 1 shows that convergence to a fixed state is no longer certain. For the discrete state Hopfield model, we can think of the dynamics as a Markov process on the state space of the network, where only a symmetric weight matrix $\mathcal{W}$ guarantees this process to be absorbing.

We can adjust our model still further to address the unreasonable assumption of binary neuronal states. To do so, we consider a population of N neurons, each of which can occupy *any* state in the interval $[0, 1]$, 0 again indicating quiescence and 1 corresponding to firing at maximal rate. We modify the Hopfield dynamics (2.1) so that

$$x_i \mapsto f\left(\sum_j \mathcal{W}_{ij}x_j - d_i\right) := f(u_i), \tag{3.1}$$

where $f : \mathbb{R} \to [0, 1]$ is a sigmoidal function with (maximal) slope σ at the origin and we assume as before that all thresholds are set to 0. We have additionally adopted the notation that $u_i = \sum_j \mathcal{W}_{ij}x_j - d_i$ denotes the 'local field' near the ith neuron.

The dynamics of this model are more complicated than that of the original Hopfield model, though again it is fruitful to begin with the $\mathcal{W}$ symmetric case and study the behavior of the energy function (2.2). As above, we can compute the change in energy as we update the kth neuron, leveraging our earlier work to obtain

$$\begin{aligned}\Delta\mathcal{H} &= -\Delta x_k \sum_j \mathcal{W}_{kj} x_j \\ &= -\left(f\left(\sum_j \mathcal{W}_{kj} x_j\right) - x_k\right) \sum_j \mathcal{W}_{kj} x_i \\ &= -\left(f(u_k) - x_k\right) u_k. \end{aligned} \tag{3.2}$$

It is unfortunately not the case that the continuous state Hopfield dynamics with symmetric couplings induce a monotonic decrease in $\Delta\mathcal{H}$. For instance, if $x_k = 1$ and $u_k < 0$, then the fact that $f(\cdot)$ has codomain $(0, 1)$ means that no u_k will cause the $f(u_k) - x_k$ term to be nonnegative. When σ is small, the dynamics of this model are quite noisy. However, when σ is sufficiently large, we can observe heuristically that the neuronal states become concentrated around 0 and 1—i.e. that the Lebesgue measure of $f^{-1}\left[(\epsilon, 1-\epsilon)\right]$ approaches zero for any $\epsilon \in (0, 1)$—meaning that the dynamics begin to approximate the discrete state Hopfield model. We should therefore expect the probability of a positive $\Delta\mathcal{H}$ to vanish. We can relate this model again to a spin glass, in that at nonzero temperature the *probability* that a spin (neuron) flips states is typically given by a sigmoidal function with maximum slope determined by the system's inverse temperature.

For $\mathcal{W}$ asymmetric, the continuous dynamics are quite difficult to study, and many authors elect to study the simpler model where all neurons update synchronously. Dauce et al. in [9] relate chaotic behavior (see Appendix A for technical discussion of chaos and its measurement) to σ in this case as well. Qualitatively we can think of σ as controlling the nonlinearity of the update rule: for small values the model operates predominantly in a near-linear regime where $f(u) = 0.5 + \sigma u + O(u^2)$, and as $\sigma \to \infty$ the transfer function limits on the Heaviside ('on-or-off') dynamics of the Hopfield model. Dauce et al. demonstrate that for coupling matrices $\mathcal{W}$ with I.I.D. entries chosen from a Gaussian distribution with mean 0, the dynamics in fact undergo a phase transition from stability to chaos as σ passes a critical value dependent on this distribution's standard deviation [9]. This is in interesting contrast to the symmetric $\mathcal{W}$ case, where σ small in fact causes the continuous state Hopfield model to deviate substantially from the ordered, energy-minimizing behavior exhibited as $\sigma \to \infty$ and the system approaches the classic Hopfield dynamics.

The remainder of our results—unless explicitly stated—concern the continuous state Hopfield model with synchronous updating and asymmetric couplings, though we modify these findings in Appendix C for the asynchronous case. However, the discussion in the preceding section implies that a study of the energy function (2.2) is no longer sufficient to understand

the dynamics of this model. Instead, we will introduce ideas from the theory of dynamical systems which allow us to quantify 'chaos' and 'stability'; we use these tools to analyze the behavior of our model when it is subjected to learning-based modulation of the synaptic couplings inspired by the behavior of the living brain.

4. Synaptic Plasticity and Hebbian Learning

In our analysis of Hopfield's model, we took the synaptic couplings $\mathcal{W}$ to be permanently fixed. In the brain, however, the strengths of synaptic couplings are not static, but instead are constantly modified by the activity of the pre- and postsynaptic neuron; this plasticity is vital to our ability to learn and refine new tasks and ideas. There is a vast body of experimental literature concerning the mechanisms of learning at the synaptic level, but perhaps the most famous and longstanding model of plasticity is due to Hebb in 1949 [13]; this is the conjecture that correlated activity of two neurons increases the synaptic coupling between them, while uncorrelated activity causes the strength of this coupling to decay.

It is in fact possible to *generate* couplings similar to (2.5) by allowing the couplings in our artificial network to modulate according to a Hebbian learning rule, although we will not go through this result in detail. In the remainder of this work, we will instead study the long-term dynamics of the continuous Hopfield model when it is subjected to a simple implementation of the Hebbian learning rule studied by Dauce et al. and Siri et al. among other ([9],[26], and references therein). Siri et al. demonstrated that in some parameter regimes, Hebbian learning caused the network to transition from chaotic to stable dynamics. We will follow their analysis, ultimately extending this finding to a much broader class of learning rules.

Dauce and Siri incorporate synaptic plasticity into the continuous Hopfield model by updating the coupling matrix $\mathcal{W}$ after each 'learning epoch' of τ time-steps. Within these epochs, the fast (neuronal) dynamics of the Siri model are identical to the continuous Hopfield model with synchronous updating, i.e. where at every discrete time-step, *every* neuron updates its state according to (3.1). In fact their analysis still holds for asynchronous updating, and we will modify their work for this case in Appendix 3. We introduce the notation that $\boldsymbol{x}^{(T)}(t)$ denotes the state of the system at time-step t during the Tth epoch, and similarly for the 'local fields' $\boldsymbol{u}^{(T)}(t)$.

At the end of epoch T, we set

$$\mathcal{W}^{(T+1)} = \lambda \mathcal{W}^{(T)} + \alpha \Gamma^{(T)}, \tag{4.1}$$

where λ and α are parameters which tune the rates of forgetting and learning respectively, and $\Gamma^{(T)}$ is a matrix which is generated by the learning rule; the system's terminal state during this epoch is set as the initial state for the next. Many of the following results concern the $\lambda < 1$ case, in which the decaying contribution of the initial weight matrix allows for a

complete restructuring of the synaptic couplings in finite time, though we present a theorem and some unproven conjecture regarding the $\lambda = 1$ case in Section 7.

In [9] and [26], the synaptic plasticity is a function of neuronal activity throughout the preceding epoch, i.e. $\Gamma_{ij}^{(T)} = h(\tilde{x}_i^{(T)}, \tilde{x}_j^{(T)})$, where $\tilde{x}_i^{(T)}$ is the orbit of neuron i during epoch T. The original model uses a 'Hebbian' function h in the sense that

$$h(\tilde{x}_i^{(T)}, \tilde{x}_j^{(T)}) \begin{cases} > 0 & \text{pre- } (j) \text{ and post- } (i) \text{ synaptic neurons are active} \\ < 0 & \text{postsynaptic } (i) \text{ is inactive, presynaptic } (j) \text{ active} \\ = 0 & \text{presynaptic neuron } (j) \text{ inactive} \end{cases} \tag{4.2}$$

with the additional constraints that (i) the diagonal of $\mathcal{W}^{(T)}$ is always zero, and (ii) synapses are not permitted to change sign. This second stipulation does not impact the mathematical analysis, and we will in fact ignore it here. For simplicity, Siri et al define 'activity' as the average deviation throughout a given epoch of each neuron from some predetermined threshold; the arguments in [26] all hinge on this precise notion of activity. However, much of the analysis (including our first main theorem) holds if we take activity to be any binary quality that a neuron can either have or lack.

5. Reduction of Chaos through Learning

There are many lenses through which we can study the behavior of dynamical systems such as the Hopfield and Siri models. By studying the original Hopfield model in analogy to a physical system—a spin glass—we are able to recognize that for symmetric couplings, the neurons will always make energy-minimizing state changes and therefore that the Hopfield dynamics push the system toward local minima in the energy landscape. However, as we illustrated above, this analytic method is no longer as fruitful in the (more interesting) cases of asymmetric couplings or continuous states. We will instead study the degree of chaos exhibited by our model using Lyapunov exponents, objects which intuitively measure the rate at which nearby points in our system's state space diverge when subjected to the system's dynamics. In general the Lyapunov exponent for a point $\boldsymbol{x}$ in direction $\boldsymbol{v}$ in our state space is given by

$$L(\boldsymbol{x}, \boldsymbol{v}) = \limsup_{t \to \infty} \frac{1}{t} \log \left(\| \mathrm{D}\, \mathrm{F}^t|_{\boldsymbol{x}} \boldsymbol{v} \| \right), \tag{5.1}$$

where $\mathrm{D}\, \mathrm{F}|_{\boldsymbol{x}}$ is the Jacobian of our transfer function evaluated at $\boldsymbol{x}$; $L(\boldsymbol{x}, \boldsymbol{v})$ is independent of $\|\boldsymbol{v}\|$ ([12], [24]). A system operating on a space of dimension N may have at most N Lyapunov exponents at every point, and this *spectrum* of exponents is constant for almost-every $\boldsymbol{x}$ with respect to any ergodic invariant measure for F, and we assume throughout that such a measure exists for our system. The Kolmogorov-Sinai Entropy (KSE), a sophisticated quantification of chaotic behavior, is bounded above by the average sum of a system's positive Lyapunov exponents, and our strategy will be to bound these exponents and therefore understand a rough constraint on the dynamical complexity of our system. We refer the reader

to Appendices A and B for a more substantial discussion of both Lyapunov exponents and Kolmogorov-Sinai Entropy.

A major achievement of Siri et al. is the observation that for $\lambda < 1$ the Hebbian learning rule presented above induces non-chaotic neuronal behavior in the network as $T \to \infty$. They perform this analysis by allowing the length τ of each learning epoch to go to infinity, and studying the largest Lyapunov exponent of the system during each of these epochs, proving that

$$L_1^T \leq \log \left\| \mathcal{W}^{(T)} \right\| + \left\langle \log \max_i f' \left(u_i^{(T)}(t) \right) \right\rangle^{(T)}, \tag{5.2}$$

where $\langle \cdot \rangle^{(T)}$ denotes a time average during the Tth epoch. This upper bound gives a rough understanding of how the sum of the positive Lyapunov exponents, and therefore the KSE, is constrained. Siri et al. prove (5.2) for a synchronously updating network, and we follow their analysis, co-opting it to prove an equivalent result in the asynchronous case in Appendix C; we begin with the derivation of (5.2), building on this method to prove our novel result.

For the synchronous case, we can simplify our notation substantially by noticing that the fast update rule (3.1) becomes

$$\boldsymbol{x}^{(T)}(t+1) = \mathrm{F}\left(\mathcal{W}^{(T)} \boldsymbol{x}^{(T)}(t) \right), \tag{5.3}$$

where $\mathrm{F}_i(\boldsymbol{x}) = f(x_i)$, f being our sigmoidal function from above. We first compute the Jacobian of F as

$$\mathrm{D}\,\mathrm{F}|_{\boldsymbol{x}^{(T)}(t)} = \Lambda^{(T)}(t) \mathcal{W}^{(T)}, \tag{5.4}$$

where Λ is a diagonal matrix with entries

$$\Lambda^{(T)}(t)_{ij} = f' \left(u_i^{(T)}(t) \right) \delta_{ij} \tag{5.5}$$

It's easy to derive (5.4), if we notice that

$$\begin{aligned} \left(\mathrm{D}\,\mathrm{F}|_{\boldsymbol{x}^{(T)}(t)} \right)_{ij} &= \frac{\partial}{\partial x_j^{(T)}(t)} \left[f \left(u_i^{(T)}(t) \right) \right] \\ &= f' \left(u_i^{(T)}(t) \right) \frac{\partial}{\partial x_j^{(T)}(t)} \left[\sum_{k=1}^{N} \mathcal{W}_{ik}^{(T)} x_k^{(T)}(t) - d_i \right] \\ &= f' \left(u_i^{(T)}(t) \right) \mathcal{W}_{ij}^{(T)}. \end{aligned}$$

Finally, we port the above work into the definition of Lyapunov exponents (5.1) to obtain a bound for the magnitude of arbitrary $L(\boldsymbol{x}, \boldsymbol{v})^{(T)}$. Recalling that our precise choice for $\|\boldsymbol{v}\|$

does not impact the exponent, we choose $\boldsymbol{v}$ to have unit norm and compute

$$\begin{aligned}
L^{(T)}\left(\boldsymbol{x}, \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}\right) &= \lim_{t\to\infty} \frac{1}{t} \log\left(\left\|\mathrm{D}\,\mathrm{F}^t|_{\boldsymbol{x}} \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}\right\|\right) \\
&\leq \lim_{t\to\infty} \frac{1}{t} \log\left(\frac{\prod_{k=1}^t \|\,\mathrm{D}\,\mathrm{F}|_{\boldsymbol{x}(k)}\|\|\boldsymbol{v}\|}{\|\boldsymbol{v}\|}\right) \\
&= \lim_{t\to\infty} \frac{1}{t} \log\left(\prod_{k=1}^t \|\,\mathrm{D}\,\mathrm{F}|_{\boldsymbol{x}(k)}\|\right) \\
&= \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^t \log\left(\|\,\mathrm{D}\,\mathrm{F}|_{\boldsymbol{x}(k)}\|\right) \\
&\leq \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^t \log\left(\|\Lambda^{(T)}(t)\|\|\mathcal{W}^{(T)}\|\right) \\
&= \left\langle \log \max_i f'\left(u_i^{(T)}(t)\right)\right\rangle^{(T)} + \log\left\|\mathcal{W}^{(T)}\right\|;
\end{aligned}$$

we have notated $\boldsymbol{x}(k) := F^k(\boldsymbol{x})$. Since the above expression is independent of $\boldsymbol{v}$, it serves as a bound on the maximal Lyapunov exponent $L_1^{(T)}$ as desired. This is a useful expression, because it demonstrates the two major factors that can drive a decrease of $L_1^{(T)}$: contraction of the spectral radius of $\mathcal{W}^{(T)}$ and 'saturation' of the neurons forcing their local fields into a regime far from zero and the maximum slope of f. Neuronal saturation—especially in the average—is hard to analyze, but Siri et al. take a different tact and focus on the effect of learning on $\|\mathcal{W}^{(T)}\|$. Their results are summarized by the following claims, which we have modified slightly from their original statement.

Claim 5.6. *If h is a bounded, Hebbian function as in* (4.2)*, then $\|\Gamma\| \leq N\tilde{h}$, where $\tilde{h}$ is an upper bound for $|h|$.*

Siri et al. prove Claim 5.6 for a particular choice of Hebbian function and conjecture it to be true for all such functions. We do not include their proof, as it is subsumed by our general result in the next section.

Corollary 5.7. $\|\mathcal{W}^{(T+1)}\| \leq \lambda^T\|\mathcal{W}^{(1)}\| + \tilde{h}N\alpha\frac{1-\lambda^T}{1-\lambda}$.

Proof. We begin by expanding (4.1) as

$$\mathcal{W}^{(T+1)} = \lambda^T\mathcal{W}^{(1)} + \alpha\sum_{n=1}^T \lambda^{T-n}\Gamma^{(n)}, \tag{5.8}$$

from which we quickly obtain

$$\|\mathcal{W}^{(T+1)}\| \leq \lambda^T\|\mathcal{W}^{(1)}\| + \alpha\sum_{n=1}^T \lambda^{T-n}\|\Gamma^{(n)}\|. \tag{5.9}$$

Now, invoking Claim 6.3, we can expand (5.9) and observe that

$$\begin{aligned}\|\mathcal{W}^{(T+1)}\| &\leq \lambda^T \|\mathcal{W}^{(1)}\| + \alpha\tilde{h}N\sum_{n=1}^{T}\lambda^{T-n} \\ &= \lambda^T\|\mathcal{W}^{(1)}\| + \alpha\tilde{h}N\frac{1-\lambda^T}{1-\lambda}\end{aligned}$$

as desired. □

Corollary 5.7 indicates that, for $\lambda < 1$, the particular choice of Hebbian learning scheme in [26] induces an exponential contraction of $\|\mathcal{W}^{(T)}\|$ along with an $O(\lambda^{T-1})$ error term. By substituting this into the bound (5.2) on $L_1^{(T)}$, Siri et al. offer a strong heuristic explanation for why their learning scheme should induce a reduction of chaos. In the next section, we show an identical result for a broader collection of learning functions, and discuss parameter regimes in which the relative orders of $\alpha, \sigma, \tilde{h}$ allow us to turn this heuristic argument into a rigorous proof.

6. Alternative Learning Rules

The Hebbian rule studied in [9] and [26] belongs to a much larger class of potential learning rules, any one of which can be defined by the following table

(6.1)

	i active	i inactive
j active	$*$	$*$
j inactive	$*$	$*$

where each $*$ indicates whether h is greater than, less than, or equal to zero, and as above 'activity' is any binary neuronal quality. More formally, this is the class of learning functions $h : [0,1]^\tau \times [0,1]^\tau \to \mathbb{R}$ for which there exist $\eta : [0,1]^\tau \to \{0,1\}$ and $\ell : \{0,1\} \times \{0,1\} \to \{-1,0,1\}$ making the following diagram commute:

(6.2)

$$\begin{array}{ccc} [0,1]^\tau \times [0,1]^\tau & \xrightarrow{\quad h \quad} & \mathbb{R} \\ \Big\downarrow{\scriptstyle \eta\times\eta} & & \Big\downarrow{\scriptstyle \mathrm{sgn}} \\ \{0,1\}\times\{0,1\} & \xrightarrow{\ \ell\ } & \{-1,0,1\} \end{array}$$

We will denote such functions h **admissible**, and our primary achievement is the following extension of Claim 5.6.

Claim 6.3. *If h is a bounded, admissible learning function, then $\|\Gamma\| \leq 2N\tilde{h}$, where $\tilde{h}$ is an upper bound for $|h|$.*

Proof. Assuming that m of the N neurons are active, we may (for ease of notation and without loss of generality) assume that the rows (columns) of Γ are ordered such that rows

(columns) $1, ..., m$ correspond to 'active' neurons and rows (columns) $m+1, ..., n$ to 'inactive' ones. Thus we can schematize Γ in the following block form:

$$\Gamma = \begin{pmatrix} \Gamma_{AA} & \Gamma_{IA} \\ \Gamma_{AI} & \Gamma_{II} \end{pmatrix}, \tag{6.4}$$

where (for instance) Γ_{AA} denotes the block of Γ corresponding to couplings between coactive pairs of neurons. From this block decomposition of Γ we obtain

$$\|\Gamma\| \leq \|\Gamma_{AA}\| + \|\Gamma_{II}\| + \|\Gamma_{IA}\| + \|\Gamma_{AI}\|. \tag{6.5}$$

Now, we know that

$$\begin{aligned} \|\Gamma_{AA}\| + \|\Gamma_{II}\| &= \sqrt{\lambda_{\max}\left(\Gamma_{AA}\Gamma_{AA}^{\dagger}\right)} + \sqrt{\lambda_{\max}\left(\Gamma_{II}\Gamma_{II}^{\dagger}\right)} \\ &\leq \sqrt{(N-m)^2\tilde{h}^2} + \sqrt{m^2\tilde{h}^2} \\ &= N\tilde{h}, \end{aligned} \tag{6.6}$$

since the entries of Γ_{AA} and Γ_{II} are (i) nonzero and of uniform sign and (ii) bounded by $(N-m)\tilde{h}^2$ and $m\tilde{h}^2$ respectively, so we can invoke the Perron-Frobenius Theorem. We can perform an identical computation for Γ_{IA} and Γ_{AI}, yielding

$$\|\Gamma\| \leq 2N\tilde{h} \tag{6.7}$$

as desired. □

Theorem 6.8. *For any $\lambda < 1$, $\sigma > 0$, and admissible learning function h, $\lim_{T\to\infty} L_1^{(T)} < 0$ provided that $\alpha \in \left[0, \frac{1-\lambda}{2\tilde{h}N\sigma}\right)$.*

Proof. Similar to above, we can substitute the result of Claim 6.3 into (5.9) to obtain

$$\begin{aligned} \|\mathcal{W}^{(T+1)}\| &\leq \lambda^T\|\mathcal{W}^{(1)}\| + 2\alpha\tilde{h}N\sum_{n=1}^{T}\lambda^{T-n} \\ &= \lambda^T\|\mathcal{W}^{(1)}\| + 2\alpha\tilde{h}N\frac{1-\lambda^T}{1-\lambda}. \end{aligned}$$

Substituting this into (5.2), we obtain

$$\begin{aligned} L_1^{(T)} &\leq \log\left(\lambda^T\|\mathcal{W}^{(1)}\| + 2\alpha\tilde{h}N\frac{1-\lambda^T}{1-\lambda}\right) + \left\langle \log\max_i f'\left(u_i^{(T)}(t)\right)\right\rangle^{(T)} \\ &\leq \log\left(\lambda^T\|\mathcal{W}^{(1)}\| + 2\alpha\tilde{h}N\frac{1-\lambda^T}{1-\lambda}\right) + \log\sigma, \end{aligned}$$

meaning that

$$\lim_{T\to\infty} L_1^{(T)} \leq \log\frac{2\alpha\tilde{h}N\sigma}{1-\lambda}. \tag{6.9}$$

So long as we choose $\alpha < \frac{1-\lambda}{2\tilde{h}N\sigma}$, this limit will be negative. □

Theorem 6.8 tells us that for $\lambda < 1$, the slow dynamics given by (4.1) have a profound effect on the dynamical complexity of the system. As discussed in Appendix 2, the Kolmogorov-Sinai Entropy, a measure of this complexity, is bounded above by the average sum of our system's positive Lyapunov exponents, meaning that *no matter which admissible learning function is used*, a choice of α in our stated regime ensures that that the learning process (4.1) will induce stable dynamics. This result holds even when σ is large enough (as per [9]) to make the initial dynamics chaotic.

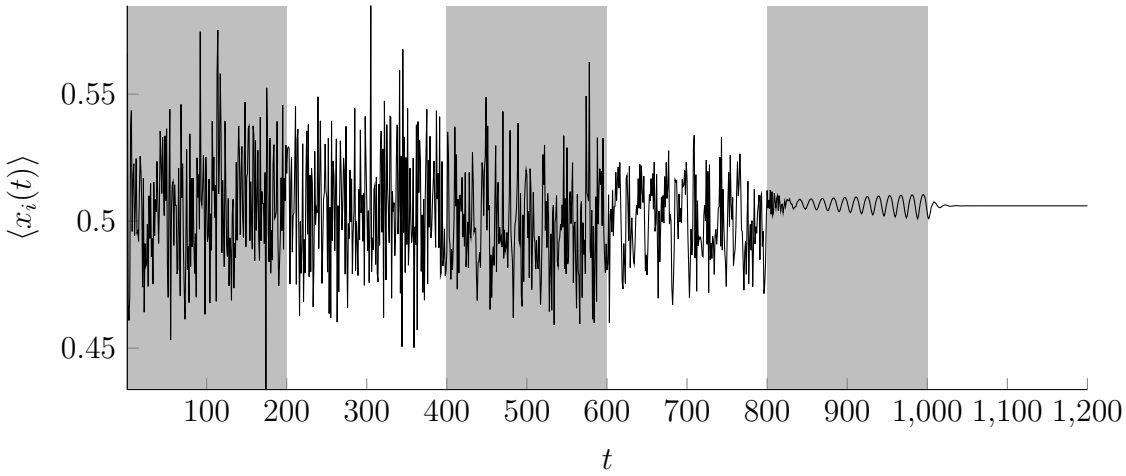

FIGURE 2. Chaotic damping induced by a Hebbian learning rule for large α. Parameter values for this simulation are $N = 100$, $\tau = 200$, $\tilde{h} = 1$, $\lambda = 0.9$, $\sigma = 10$, and $\alpha = 0.1$, substantially larger than the upper bound given in Theorem 6.8. Background shading demarcates different learning epochs. We plot the average neural activity $\langle x_i(t) \rangle = \frac{1}{N} \sum_i x_i(t)$ at each time-step for an example simulation. Entries of $\mathcal{W}$ are chosen I.I.D. from a Gaussian with mean 0 and standard deviation $1/N$.

However, our numerical simulations indicate that even for α larger than the critical value of $\frac{1-\lambda}{2\tilde{h}N\sigma}$ with respect to our upper bound for $L_1^{(T)}$, the long-term dynamics appear stable, as shown in Figure 2. This implies that the relative orders of α, λ, σ, N, and $\tilde{h}$ do not matter so long as $\lambda < 1$. However, this conjecture is difficult to prove, as it requires either a much tighter bound on $L_1^{(T)}$—necessitating a much more sophisticated analysis of the spectral radii of $\mathcal{W}^{(T)}$ and $\Gamma^{(T)}$—or a careful study of neuronal saturation, which involves taking time averages of the system as it evolves.

7. Infinite Memory Systems

Theorem 6.8 demonstrates that the particulars of the learning rule have little bearing on the ultimate fate of the system when $\lambda < 1$. However, it appears from our numerical simulations that the choice of learning rule is in fact *crucial* for determining whether the limiting dynamics will be stable or chaotic in the $\lambda = 1$ case; this corresponds to networks with 'infinite memory', in the sense that the initial synaptic couplings are subject to no decay factor when we update. Our simulations show that there is a large space of possible long-term behavior, all of which is controlled closely by the particulars of the learning rule. An example comparison between two potential learning rules is given in Figure 3. Based on an extensive survey of admissible learning functions we have made the following preliminary observation.

Conjecture 7.1. *When $\lambda = 1$, 'activity' is given by*

$$\eta\left(\tilde{x}_i^{(T)}\right) = \frac{1}{\tau}\sum_{t=1}^{\tau} x_i^{(T)}(t), \tag{7.2}$$

and h is an admissible learning function for which $h(\textit{active}, \textit{active}) > 0$, the continuous state Hopfield model displays non-chaotic dynamics in the large-T limit, for any choices of α, $\tilde{h}$, N, and σ.

Although a full proof of Conjecture 7.1 is as yet beyond our reach, we can prove an analogue for a much simpler case. Let us return to the discrete state Hopfield model with symmetric weights, to which we now apply an admissible learning rule after every learning epoch of τ time-steps, just as we did for the continuous state Hopfield model in the sections above. In order to preserve the symmetry of $\mathcal{W}$ we will have to restrict our attention to *symmetric*, admissible learning rules (i.e. for which $h(\boldsymbol{\theta}, \boldsymbol{\phi}) = h(\boldsymbol{\phi}, \boldsymbol{\theta})$ for all $\boldsymbol{\phi}, \boldsymbol{\theta} \in [0,1]^\tau$ so that learning does not disrupt the symmetry of the weight matrix.

Theorem 7.3. *When subjected to the slow dynamics given in* (4.1)*, with $\lambda = 1$, $\tau \to \infty$, $\boldsymbol{\eta}^{(T)}$ defined as in* (7.2)*, and h an admissible, symmetric learning function with*

$$\ell(0,1) = \ell(1,0) \leq 0 \tag{7.4}$$

$$\ell(1,1) \geq 0, \tag{7.5}$$

a discrete state Hopfield network will remain fixed in whatever attractor it finds during the first epoch.

Proof. Recall first that for an admissible learning function h, ℓ is a function which computes $\operatorname{sgn} \circ h\left(\tilde{x}_i^{(T)}, \tilde{x}_j^{(T)}\right)$ from the 'activies' $\eta\left(\tilde{x}_i^{(T)}\right)$ and $\eta\left(\tilde{x}_j^{(T)}\right)$. Now, let us assume that during some epoch T the network falls into an attracting state $\boldsymbol{y}$, as Claim 2.3 indicates that it (almost surely) must. Allowing τ to grow to infinity, we first observe that

$$\boldsymbol{\eta}^{(T)} = \lim_{\tau\to\infty} \frac{1}{\tau}\sum_{t=1}^{\tau} \boldsymbol{x}^{(T)}(t) = \boldsymbol{y},$$

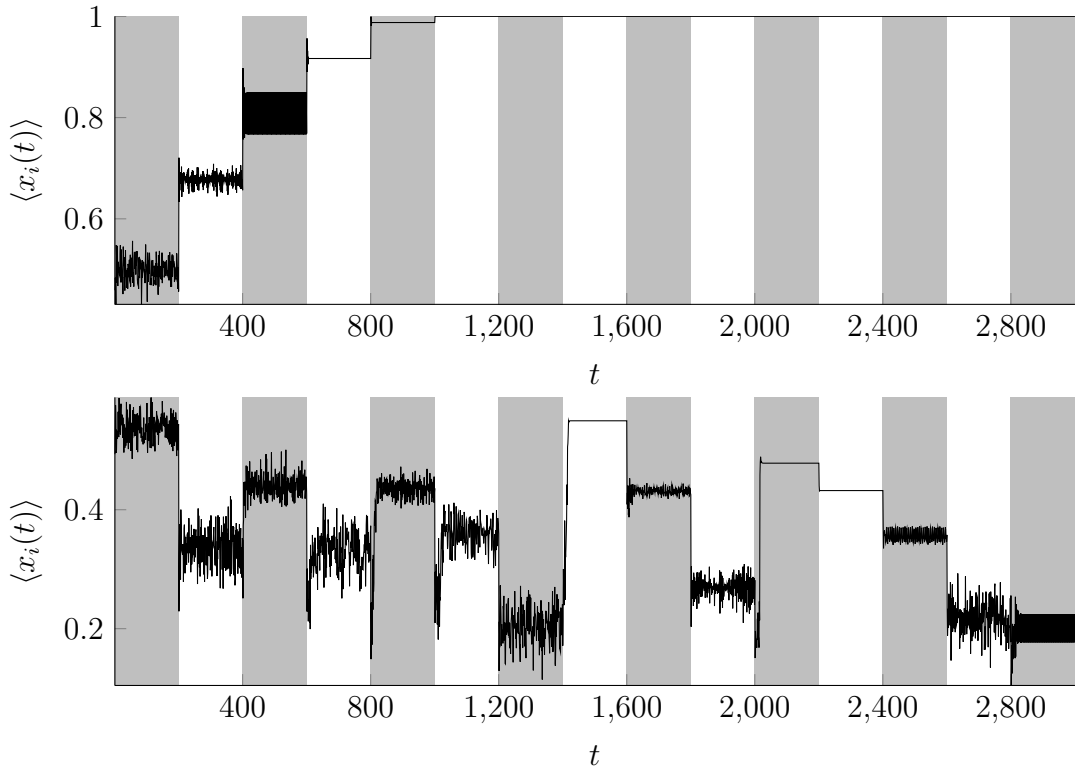

FIGURE 3. The learning rule has a strong impact on dynamical behavior when $\lambda = 1$. We compare two rules. The top system uses a rule which rewards activity and penalizes co-active neurons, wherein $\Gamma_{ij} \geq 0$ if either i or j is active, and negative otherwise. The bottom system uses the exact opposite, where any synapse involving at least one inactive neuron is strengthened. All other simulation parameters are $N = 100$, $\tau = 200$, $\tilde{h} = 1$, $\lambda = 1$, $\sigma = 10$, and $\alpha = 1$. Entries of $\mathcal{W}$ are chosen I.I.D. from a Gaussian with mean 0 and standard deviation $1/N$.

almost surely, since Claim 2.3 indicates our network should (again almost surely) reach $\boldsymbol{y}$ in finite time. Thus, as h is an admissible learning function, $\operatorname{sgn} \circ h$ is a function *only* of the attracting state $\boldsymbol{y}$, since $\boldsymbol{\eta} \to \boldsymbol{y}$ in the $\tau \to \infty$ limit. We can use this fact, as well as our assumptions in (7.5), to understand the behavior of the network once our learning rule has been applied.

As per (4.1), at the termination of our learning epoch the weight matrix will update to $\mathcal{W}^{(T+1)} = \mathcal{W}^{(T)} + \alpha \Gamma^{(T)}$ and the network will retain its final condition as the initial condition for the new epoch. Imagine that our network is in the attracting state $\boldsymbol{y}$ and we wish to

update neuron k with the new matrix $\mathcal{W}^{(T+1)}$. To do so, we compute the sign of

$$\begin{aligned}\sum_j \mathcal{W}_{kj}^{(T+1)} y_j &= \sum_j \left(\mathcal{W}_{kj}^{(T)} + \alpha \Gamma_{kj}^{(T)}\right) y_j \\ &= \sum_j \mathcal{W}_{kj}^{(T)} y_j + \alpha \sum_{y_j=1} \Gamma_{kj}^{(T)} y_j. \end{aligned} \tag{7.6}$$

There are two cases to consider. If $y_k = 0$, $\boldsymbol{y}$ a fixed state means that $\sum_j \mathcal{W}_{kj}^{(T)} y_j \leq 0$. For the second of the two summations in (7.6), we need only analyze the terms where $y_j = 1 = \eta_j$ (the 'active' terms). For these, we can use our assumptions on h, η, and ℓ to see that

$$\operatorname{sgn}\left(\Gamma_{kj}^{(T)}\right) = \operatorname{sgn} \circ h\left(\tilde{x}_k^{(T)}, \tilde{x}_j^{(T)}\right) = \ell\left(\eta\left(\tilde{x}_k^{(T)}\right), \eta\left(\tilde{x}_j^{(T)}\right)\right) = \ell\left(0,1\right) \leq 0. \tag{7.7}$$

Thus $\sum_j \mathcal{W}_{kj}^{(T+1)} y_j \leq 0$ and y_k will remain fixed at zero. Similarly, if $y_k = 1$, $\boldsymbol{y}$ a fixed state means that $\sum_j \mathcal{W}_{kj}^{(T)} y_j > 0$. Once again we need only analyze the terms where $y_j = 1 = \eta_j$, and for these

$$\operatorname{sgn}\left(\Gamma_{kj}^{(T)}\right) = \operatorname{sgn} \circ h\left(\tilde{x}_k^{(T)}, \tilde{x}_j^{(T)}\right) = \ell\left(\eta\left(\tilde{x}_k^{(T)}\right), \eta\left(\tilde{x}_j^{(T)}\right)\right) = \ell(1,1) \geq 0 \tag{7.8}$$

and y_k will remain fixed at 1. We have shown, as desired, that a state $\boldsymbol{y}$ fixed for $\mathcal{W}^{(T)}$ is also fixed for $\mathcal{W}^{(T+1)}$, and consequently that our system will remain in whatever attractor it finds during the first epoch.

□

Notice that the effect of our learning rule on synapses between co-inactive pairs of neurons (i.e. the value of $\ell(0,0)$) *has no impact* on the above result. This co-inactivity response *does* matter if we allow $h(\text{active}, \text{active})$ to be negative, or if we allow $h(\text{active}, \text{inactive}) = h(\text{inactive}, \text{active})$ to be positive, as in these situations the attractor in one learning epoch need not be stable with respect to the updated coupling matrix in the next.

8. Discussion

It is surprising that *any* admissible learning-rule causes the continuous-state Hopfield model to gravitate toward a stable regime. Certainly the choice of learning rule is important for what we are hoping that the network will *learn*; the Hebbian rule is provably capable of generating Hopfield-type couplings (2.5) that allow for the storage of input patterns for later retrieval, and it is unclear which other admissible functions are capable of doing this. However, from the standpoint of chaotic damping, the particulars of the learning rule *do not matter* in the $\lambda < 1$ case. In fact, Theorem 6.8 was even agnostic as to how we define 'activity' in the first place, and still holds if we denote each neuron 'active' or 'inactive' at random each time the synaptic weights are to be updated.

How much does this result tell us about the role of synaptic plasticity in the brain? The simple (and naive) answer to this question is that it tells us very little. The nervous system

evinces architecture and mechanics with complexity many orders of magnitude beyond the capacity of our simple system to capture. Living neurons are varied in form and function, their interactions governed by intricate dynamics at several temporal and spatial scales. The learning that occurs *in vivo* is not the storage of simple patterns, but the slow refinement of tasks, concepts, and memories which are complicated and multimodal. Moreover, measuring chaos in noisy, living systems is quite difficult in practice and interpretation, but certainly most living nervous systems are not locked in the fixed states or stable oscillations that our learning rules produce; there has been much academic focus on the ways in systems at the 'edge of chaos' may exhibit optimal computational and information processing capacity ([1], [3], [21]).

However, the continuous state Hopfield dynamics with incorporated synaptic plasticity are designed to model a particular aspect of cognition—learning—and each simplifying assumption that enters into this model, together with the ways in which it fails to conform to the ground truth, tells us about an aspect of the living brain that plays an important role in the phenomena which we hope to understand. At first blush it is tempting to view results like that of Siri as indicating the importance of Hebbian learning in regulating the brain's chaotic behavior. However, Theorem 6.8 tells us that it is in fact the *forgetting*, and not the learning, which is at the root of our model's chaotic damping. Perhaps this is further evidence that our model of forgetting (uniform decay of synaptic weights in time) is far too simple, or possibly it tells us that we should be seeking out biological mechanisms for non-Hebbian synaptic plasticity. In addition, the fact that Theorem 6.8 is valid for any 'activity' function should hint to us that our results pertain to a much broader class of systems than the brain, as learning is known to have a biological mechanism related to a notion of activity subsumed by our general case. All of these questions are best explored in conjunction with *in vivo* study (of which there is already a formidable body of literature concerning learning and memory) but we present them to illustrate the importance of modeling, and the ways that a thorough mathematical understanding of our models can feed back to biological insight.

Beyond their biological usefulness, the models we present here evince rich and complex mathematical structure that continues to stymie our best analytic efforts. Their strong nonlinearity and high dimensionality require a mathematical treatment which blends traditional dynamical systems theory with 'large-n' analyses at the thermodynamic limit ported from statistical physics; this paper has given us the opportunity only to scratch the surface of these techniques. However, a highly idealized model which nevertheless reveals deep and beautiful mathematics is certainly the best kind of all. Mathematical structure flows from the intricacies of the universe and our scientific interface with it, and the structure which we discover within a simple model is the first, tantalizing intimation of the beauty just beyond our reach.

Appendix A. Chaos and Lyapunov Exponents

We now give a brief survey of chaotic systems and Lyapunov Exponents for the unfamiliar reader. Such systems are defined by a common phenotype: although deterministic, their behavior appears highly stochastic and disordered. 'Chaos' has a range of definitions, but most involve some notion of **sensitivity**: nearby points in the system's state space will eventually be moved far apart. For the precise definition of this phenomenon, given first by Devaney in [10], we let (X, d) be a metric space and consider some function $F : X \to X$. We term a dynamical system as sensitive provided that there exists some $\delta > 0$ such that for any $x \in X$ and $\epsilon > 0$ there exist $y \in X$ with $d(x, y) < \delta$ and some n for which $d\left(F^n(x), F^n(y)\right) > \delta$. It's important to note sensitivity is a necessary but not sufficient ingredient for chaotic behavior, but as our results concern the *reduction* of chaos, it will suffice for us to demonstrate that our systems are *not* sensitive in the manner described above.

Our results primarily utilize the measurement of sensitivity with *Lyapunov exponents*; as discussed in Section 5, these exponents intuitively measure the rate at which nearby points in a system's state space deviate from (or converge toward) one another when acted on by F. We may think of this divergence (or convergence) as being exponential, i.e. for $x, y \in \mathbb{R}$ we approximate

$$\frac{|F^t(x) - F^t(y)|}{|x - y|} \approx e^{Lt},$$

with L being the eponymous Lyapunov exponent; the sign of L indicates whether our system is expanding, contracting, or fixed. We can compute L by allowing (i) the distance between x and y to become small and (ii) the system to evolve indefinitely, i.e.

$$\begin{aligned} L(x) &= \lim_{t\to\infty} \lim_{|x-y|\to 0} \frac{1}{t} \log\left(\frac{|F^t(x) - F^t(y)|}{|x - y|}\right) \\ &= \lim_{t\to\infty} \frac{1}{t} \log\left|\frac{\mathrm{d}}{\mathrm{d}t}\left[F^t(x)\right]\right|, \end{aligned} \tag{A.1}$$

where L may depend on our choice of x.

In maximal generality (following the presentations in [12] and [24]), we let M be a manifold of dimension m, and $F : M \to M$ a diffeomorphism. For any $\boldsymbol{x} \in M$ and $\boldsymbol{v}$ in the tangent space $T_{\boldsymbol{x}}X$, the Jacobian of F gives a linearization of F which acts on $T_{\boldsymbol{x}}X$. We compute the Lyapunov exponent $L(\boldsymbol{x}, \boldsymbol{v})$ by using this linearization to estimate the expansion or contraction in the direction of $\boldsymbol{v}$, calculating

$$L(\boldsymbol{x}, \boldsymbol{v}) = \limsup_{t\to\infty} \frac{1}{t} \log\left(\| \mathrm{D}\, F^t|_{\boldsymbol{x}}\boldsymbol{v}\|\right);$$

this value depends on the direction of $\boldsymbol{v}$ but not its norm.

Appendix B. Kolmogorov-Sinai Entropy

The Kolmogorov-Sinai Entropy (KSE) is a measurement of chaotic behavior defined in analogy to the information theoretic entropy of a random variable. If $X : \Omega \to A$ is a measurable function from a space Ω endowed with a probability measure μ to a finite set A, we compute the *entropy* of X by

$$H(X) = -\sum_{\alpha \in A} \mu\left(X^{-1}(\alpha)\right) \log\left(\mu\left(X^{-1}\right)\right); \tag{B.1}$$

H measures the degree of 'surprise' we experience when we observe the state of X. Entropy is maximized when all outcomes are equally likely, and minimized when one is certain and the others have measure zero. The intuition behind KSE is to port this idea into the realm of dynamical systems, treating the trajectories of points in our state space with respect to the system's dynamics as if they were random variables.

To do this, shadowing [22] and [27], we consider a measure space $(M, \mathcal{M}, \mu)$, and a measurable function $F : M \to M$ under which μ is invariant. We can think of an arbitrary partition $\mathcal{P}$ of M as a 'coarse-graining' of our system's dynamics, giving us a rough description of an orbit in terms of the sequence of partition sets which it visits. In this way, orbits of length n induce a *new* partition $\mathcal{P}_n$, the elements of which correspond to possible trajectories through the elements of $\mathcal{P}$. To be precise, let $P_{i_1}, ..., P_{i_n} \in \mathcal{P}$; the set of all $x \in M$ for which $F^k(x) \in P_{i_k}$ for $k = 1, .., n$ is the intersection $\bigcap_{k=0}^{n-1} F^{-k} P_{i_n}$. Our new partition $\mathcal{P}_n$ is therefore

$$\mathcal{P}_n = \bigvee_{t=1}^{n} F^{-k}(\mathcal{P}), \tag{B.2}$$

where $F^{-k}(\mathcal{P}) = \{F^{-k}(P) : P \in \mathcal{P}\}$, and $\mathcal{Q} \vee \mathcal{R} = \{Q \cap R : Q \in \mathcal{Q}, R \in \mathcal{R}\}$ is the *common refinement* of partitions $\mathcal{Q}$ and $\mathcal{R}$. We can think of $\mathcal{P}_n$ as inducing a random variable $X_{\mathcal{P}_n} : M \to \mathcal{P}_n$, and we can compute its entropy

$$H(X_{\mathcal{P}_n}) = \frac{1}{n} \sum_{P' \in \mathcal{P}_n} \mu(P') \log \mu(P'). \tag{B.3}$$

Finally, then, the Kolmogorov-Sinai (metric) entropy is given by

$$h(F) = \sup_{\mathcal{P}} \lim_{n \to \infty} H(X_{\mathcal{P}_n}). \tag{B.4}$$

This quantity gives us insight into the diversity of infinite F-orbits, in that if all points in M converge to a single attracting point all orbits are eventually indistinguishable (no matter which partition of M we choose). The following important result (due to Ruelle) allows us to compute an upper bound for $h(F)$ using Lyapunov exponents.

Theorem B.5. *(Ruelle* [25]*) Let M be a C^∞ compact manifold and $F : M \to M$ a C^1 map. For each measure μ invariant under F there is some Borel subset Ω of full measure where at*

each $\boldsymbol{x} \in \Omega$ the tangent space splits into an increasing sequence of subspaces $0 = V_{\boldsymbol{x}}^0 \subset \cdots \subset V_{\boldsymbol{x}}^{s(x)} = T_{\boldsymbol{x}}M$, so that $L(\boldsymbol{x}, \boldsymbol{v})$ is constant for each $\boldsymbol{v} \in V_{\boldsymbol{x}}^r \setminus V_{\boldsymbol{x}}^{r-1}$. Moreover, if we let $L_+ : M \to \mathbb{R}$ map $\boldsymbol{x} \in M$ to the sum of all positive Lyapunov exponents at $\boldsymbol{x}$,

$$h(\mathrm{F}, \mu) \leq \int_M \mu(d\boldsymbol{x}) L_+(\boldsymbol{x}).$$

Finally, the Krylov-Bogolyubov Theorem [22] tels us that for any continuous map from a continuous, metrizable topological space to itself admits an invariant Borel probability measure. As our transfer function F satisfies these criteria, we can be sure that our work in the main text demonstrating an α-regime in which all Lyapunov exponents become negative is sufficient to show that our system will be non-chaotic everywhere in this regime.

Appendix C. Learning in Asynchronous Networks

It is simple to extend our proof of Theorem 6.8 to the case where neurons update asynchronously as opposed to in unison. To do so, it suffices to return to our derivation of equation (5.2). If we think of the neurons in our system updating in a prescribed order which is an infinite sequence of draws $\{\theta_n\}$ from the uniform distribution on $\mathbb{Z}_N$, then we may think of our system as processing according to a function $\mathrm{G} : [0,1]^N \to [0,1]^N$ where

$$\mathrm{G}_i(\boldsymbol{x}^{(T)}(t)) = \begin{cases} x_i^{(T)}(t) & i \neq \theta_t \\ f(u_i^{(T)}(t)) & i = \theta_t \end{cases}, \tag{C.1}$$

f being our friendly sigmoidal function with maximal slope σ. The Jacobian of this function is given by

$$\mathrm{D}\,\mathrm{G}|_{\boldsymbol{x}^{(T)}(t)} = \Upsilon(t)\mathcal{W}^{(T)}, \tag{C.2}$$

where $\Upsilon(t) = f'(u_{\theta t})\delta_{i,\theta_t}\delta_{\theta_t,j}$ is a matrix with (θ_t, θ_t) entry equal to $f'(u_{\theta_t})$ and zeros elsewhere. This follows from the derivation of (5.4) and the fact that F fixes all but the θ_tth component of it's input. Finally, then, we can prove an analogous expression to (5.2) by noticing that

$$\begin{aligned} L_1^{(T)} \leq L^{(T)}(\boldsymbol{x}, \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} &= \lim_{t\to\infty} \frac{1}{t} \log\left(\frac{\| \left(\prod_{k=1}^t \mathrm{D}\,\mathrm{G}|_{\boldsymbol{x}(k)}\right) \boldsymbol{v}\|}{\|\boldsymbol{v}\|} \right) \\ &\leq \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^t \log\left(\| \mathrm{D}\,\mathrm{G}|_{\boldsymbol{x}(k)} \| \right) \\ &\leq \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^t \log\left(\|\Upsilon\| \|\mathcal{W}^{(T)}\| \right) \\ &\leq \log \left\|\mathcal{W}^{(T)}\right\| + \log \sigma. \end{aligned}$$

We then invoke Claim 6.3—which is not affected by asynchronous updating—to prove Theorem 6.8 for the asynchronous case.

References

[1] J. Beggs, *The criticality hypothesis: how local cortical networks might optimize information processing*, Phil. Trans. R. Soc. A **366** (2008), no. 1864, 329–343

[2] J. Beggs and D. Plenz, *Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures*, J Neurosci **24** (2004), no. 22, 5216–5229

[3] N. Bertschinger and T. Natschläger, *Real-time computation at the edge of chaos in recurrent neural networks*, Neural Computation **16** (2004), no. 7, 1413–1436

[4] J. Bruck, *On the convergence properties of the Hopfield mdoel*, Proceedings of the IEEE **78** (1990), no. 10, 1579–1585

[5] E. Bullmore and O. Sporns, *The economy of brain network organization*, Nat. Rev. Neurosci. **13** (2012), no. 5, 336–349

[6] G. Buzsàkil and A. Draguhn, *Neuronal oscillations in cortical networks*, science **304** (2004), no. 5679, 1926–1929

[7] S. Cooper et al., *Predicting protein structures with a multiplayer online game*, Nature **466** (2010), no. 7307, 756–760

[8] R. Coulom, *Efficient selectivity and backup operators in Monte-Carlo tree search*, Computers and Games, Springer Berlin Heidelberg 2007. 72–83

[9] E. Dauce et al., *Self-organization and dynamics reduction in recurrent networks: stimulus presentation and learning*, Neural Networks **11** (1998), no. 3, 521–533

[10] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Bejamin/Cummings, Menlo Park, 1986

[11] E. Goles-Chacc et al., *Decreasing energy functions as a tool for studying threshold networks*, Discrete Appl. Math. **12** (1985), no. 3, 261–277

[12] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, Springer-Verlag, New York, 1983

[13] D. O. Hebb ,*The Organization of Behavior*, Wiley & Sons, New York, 1949

[14] M. P. van den Heuvel and O. Sporns, *Rich-club organization of the human connectome*, J. Neurosci., **44** (2011) no. 35, 15775–15786

[15] A. L. Hodgkin and A. F. Huxley, *A quantitative description of membrane current and its application to conduction and excitation in nerve*, J. Physiol. **117** (1952), no. 4, 500–544

[16] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. U.S.A. **79** (1982), no. 8, 2554–2558

[17] J. J. Hopfield, *Neurons with graded response have collective computational properties like those of two-state neurons* Proc. Natl. Acad. Sci. U.S.A. **81** (1984), no. 10, 3088–3092

[18] Humphries, Mark D., Kevin Gurney, and Tony J. Prescott. *The brainstem reticular formation is a small-world, not scale-free, network* Proc. R. Soc. Lond. (Biol) **273** (2006), no. 1585, 503–511.

[19] E. G. Izhikevich, *Dynamical systems in neuroscience*, MIT Press (2007).

[20] M. Krishnan et al., *Segregating complex sound sources through temporal coherence*, PLoS Comp. Biol. **10** (2014), no. 12, e1003985

[21] C. G. Langton, *Computation at the edge of chaos: phase transitions and emergent computation*, Physica D **42** (1990), 12–37

[22] R. Mañè, *Ergodic Theory and Differentiable Dynamics*, Springer-Verlag, Berlin, 1987

[23] K. Rajan and L. F. Abbott, *Eigenvalue spectra of random matrices for neural networks* Phys. Rev. Lett. **97** (2006), no. 18, 188104

[24] C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos*, CRC Press, Boca Raton, 1995

[25] D. Ruelle, *An inequality for the entropy of differentiable maps*, Bol. Soc. Bras. Mat. **9** (1978), no. 1, 83–87

[26] B. Siri et al., *A Mathematical Analysis of the Effects of Hebbian Learning Rules on the Dynamics and Structure of Discrete-Time Random Recurrent Neural Networks*, Neural Computation **20** (2008), no. 12, 2937–2966

[27] P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, New York, 1982
[28] Watts, Duncan J., and Steven H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature **393** (1998), no. 6684, 440–442.

Acknowledgements

Thanks are first and foremost due to my advisor Jim Walsh for his incalculable support in the completion of this project. I would in addition like to acknowledge Matthew Banks for his gentle, guiding hand through the world of neuroscience, and as well to thank Jessica Flack, David Krakauer, Bryan Daniels, Christopher Ellison, and Phillip Poon for invaluable discussion and insight on the nature of science and quantitative modeling. Finally, the deepest of gratitude to my friends and family not already mentioned, in particular Judi Bartfeld, Maya Banks, Andrea Allen, Cricket Epstein, Shelby Lorman, Jesse Wiener, and Miles Schulman, for their love and allyship throughout these months.

I affirm that I have adhered to the Honor Code.

1107 Canyon Rd., Apartment D, Santa Fe, NM 87501
E-mail address: Banks.Jess.M@gmail.com