Critical Capacity of Hopfield Networks

Kevin Takasaki*
MIT Department of Physics
(Dated: May 17, 2007)

This project investigates the critical storage capacity of Hopfield networks [1] under Hebbian learning, following the thermodynamic treatment given by Amit, *et. al.*, [2] in the noiseless ($T = 0$) limit. Deterministic simulations are coded in MATLAB, and numerical data is collected and compared with predictions from theory.

1. INTRODUCTION

The computational power of the human brain emerges from an intricate network of (on the order of) 10^{11} neurons, each neuron connected by synapses to 10^4 of its fellows on average. This extensive architecture permits the brain to perform a wide range of interesting functions. One such function is that of *associative memory*: the brain is able to store and recall information given partial knowledge of the information's content.

A model describing how associative memory (also known as content-addressable memory), might arise from a structure of neurons joined by synapses was proposed by Hopfield [1]. The question I ask here is how much information can be embedded in a Hopfield network of a given size before retrieval becomes difficult. The critical capacity of Hopfield's model has been extensively studied, and the theory presented in this paper follows the work of Amit, *et. al.* [2].

2. HOPFIELD'S MODEL

A neural network may be modeled as a directed graph, G, composed of N nodes (neurons) indexed by a positive integer, $i = 1, 2, ..., N$, and links (synapses) labeled by a pair of positive integers, ij, denoting the link from node i to node j. For every node i, there is an associated state variable, $s_i = \pm 1$, which indicates the "activity" of the neuron. For every link ij, there is an associated real-valued weighting factor, w_{ij}, which indicates the strength of the synapse. As in [2], we assume symmetric connectivity, i.e. $w_{ij} = w_{ji}$. The influence of the network on a particular neuron can be thought of as interaction with a local field. We will describe the local field interacting with neuron i by the quantity, h_i, which is the linear sum of influences on neuron i by the network. That is,

$$h_i = \sum_{j=1}^{N} w_{ij} s_j. \tag{1}$$

*Electronic address: ktaks07@mit.edu; URL: http://web.mit.edu/physics/

The "energy function", or Lyapunov function, of the system in the state $\mathcal{S}$ is defined as

$$E[\mathcal{S}] = -\frac{1}{2} \sum_{i \neq j} w_{ij} s_i s_j. \tag{2}$$

We consider now the case of a neural network storing p patterns. We describe these patterns by sequences, $\{\sigma_i^\mu\}$, where $\sigma_i^\mu = \pm 1$ denotes the state of node i in the μ^{th} pattern. Ideally, we would like to determine the connectivity matrix w_{ij} in such a way that the network states corresponding with the stored patterns are stable attractors of the network. *Hebb's rule* [3] provides guidelines for how this may be achieved. In the case of large N and small p [12], we may "train" the network to store the patterns σ_i^μ by implementing Hebb's rule as follows:

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \sigma_i^\mu \sigma_j^\mu. \tag{3}$$

The Hopfield model presented above is clearly analogous to a class of Ising spin systems. The usefulness of this analogy will be explored in the next section.

3. MEAN-FIELD THEORY

I now return to the original question concerning the critical capacity of a Hopfield network storing a finite ratio of patterns to elements. Specifically, I am interested in the case where $\alpha \equiv \frac{p}{N}$ remains finite as $N \to \infty$, and the system is in dynamic equilibrium (detailed balance). Work performed by Edwards and Anderson [4], Sherrington and Kirkpatrick [5], and by Amit, *et. al.* [6], on the behavior of "spin-glasses", disordered systems exhibiting magnetic frustration, provide a frame upon which the system described by (2) and (3) can be analyzed. As in the Edwards-Anderson model, I am interested in properties of the network which are valid on the average when the stored patterns are chosen randomly from a large ensemble of possible patterns. Thus, the average denoted by $\langle\langle ... \rangle\rangle$ is the "quenched average" described in [5]. Specifically, I would like to find mean-field equations for the following order parameters:

$$m^\mu = \left\langle \left\langle \frac{1}{N} \sum_{i=1} N \sigma_i^\mu \langle s_i \rangle \right\rangle \right\rangle \tag{4a}$$

$$q \equiv \left\langle \left\langle \frac{1}{N} \sum_{i=1} N \langle s_i \rangle^2 \right\rangle \right\rangle \tag{4b}$$

$$r \equiv \frac{1}{\alpha} \sum_{\mu > \rho} \left\langle \left\langle \left[\frac{1}{N} \sum_{i=1} N \sigma_i^\mu \langle s_i \rangle \right]^2 \right\rangle \right\rangle \tag{4c}$$

The order parameter m characterizes the mean overlap of a stored pattern and the states of the network visited by the dynamics. In the Ising interpretation, $m \neq 0$ indicates a ferromagnetic phase in which individual spins become frozen into configurations with long-range order. At low temperatures, however, frustration gives rise to frozen-in disorder resulting in zero net magnetization, thus making the spin-glass phase indistinguishable from the paramagnetic phase on the basis of m. The order parameter q discriminates between the paramagnetic and spin-glass phases. In the paramagnetic phase, $\langle s_i \rangle$ vanishes for every i, whereas in the case of the spin-glass, dynamical freezing results in finite q. r is an auxiliary variable that characterizes the noise due to uncondensed patterns, patterns that have vanishing overlap with the network state as $N \to \infty$. Determining the possible values of m, q, and r for a given α will allow characterization of the phases of the network and determination of the critical storage capacity.

To derive the mean-field equations, the free energy must be extremized with respect to the order parameters. The average free energy per element, f, is given by

$$f = \lim_{n \to 0} \frac{-1}{\beta N n} (\langle \langle Z^n \rangle \rangle - 1). \tag{5}$$

where Z is the canonical partition function. Noting that Z^n is the partition function of a system composed of n replicas of the original network, Amit, *et. al.* [2] apply the "replica method" of Sherrington-Kirkpatrick to derive the mean-field equations and the following results for the critical storage capacity of the network.

At $T = 0$, Amit, *et. al.* find the following equation for the variable $y = \frac{m}{\sqrt{2\alpha r}}$:

$$y = \frac{\sqrt{\pi} \Phi(y)}{\sqrt{2\pi\alpha} + 2\exp(-y^2)} \tag{6}$$

where Φ is the error function, $\mathrm{erf}(x)$. First, note that $m = 0$ satisfies (6) for any α, and for $\alpha > 0.138$, $m = 0$ is the *only* solution. Also, at $\alpha = 0.138$, there is an overlap $m = 0.967$ indicating very strong retrieval. Thus, the critical storage capacity $\alpha_c = 0.138$, and over this value, the network undergoes a discontinuous phase transition from a state of nearly perfect retrieval to complete amnesia.

4. NUMERICAL SIMULATION

To further explore the results described in the previous section, I present the results of a program (Appendix A) written in MATLAB to simulate the evolution of a symmetric Hopfield network at zero temperature. Given the number of stored states, p, and the size of the network, N, the program randomly constructs a $p \times N$ matrix of unique, "memorized" patterns, M, with $\langle M_{ij} \rangle = 0$. Subsequently, the $N \times N$ synaptic connectivity matrix, W, is constructed with Hebb's rule (3) from the patterns in M. The state of the network is stored as a vector, S, which at $t = 0$ is initialized to a stored pattern modified with some error. For a given error percentage, a, aN elements in the original pattern are randomly chosen to have their signs flipped. At the end of the simulation, pattern retrieval is determined by the overlap of S with the original pattern. Specifically, I chose to define a successful retrieval as convergence on a state with an overlap, m, above a specified threshold (0.95 or 0.90 in the plots below).

In Hopfield's model [1], dynamical evolution of the state of the network occurs sequentially. In the simplest case, we may write a deterministic law governing network evolution as

$$s_i(t+1) = sgn[h_i(t)] \tag{7}$$

where sgn is the sign function. This simple evolutionary scheme aligns states with the local field. We clearly see that the energy function (2) of a network evolving according to (7) will monotonically decrease in time. Since it is also clear that the energy function is bounded from below, the system converges on a stable state, as shown in Figure 1.

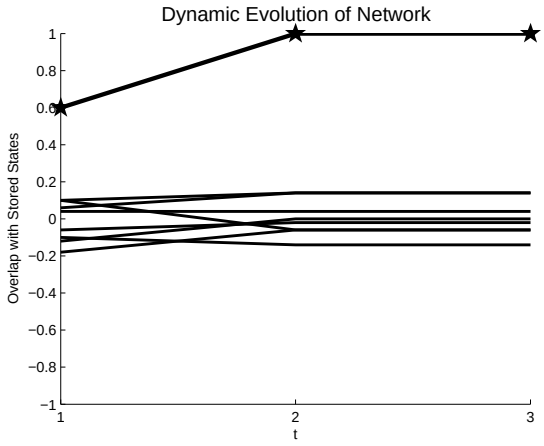

FIG. 1: Dynamic evolution of overlap of the network state with the stored patterns for $p = 10$, $N = 100$. Initial state converges to original pattern in 1 round.

The first simulations I ran were to plot the retrieval rate against the value of α. Specifically, I wanted to observe the collapse of the network over the critical storage capacity, $\alpha_c = 0.138$, predicted by theory. The plot I obtained with an initial error percentage, $a = 20\%$, is shown in Figure 2.

The next set of simulations I ran were to plot the retrieval rate against the initial error percentage for varying

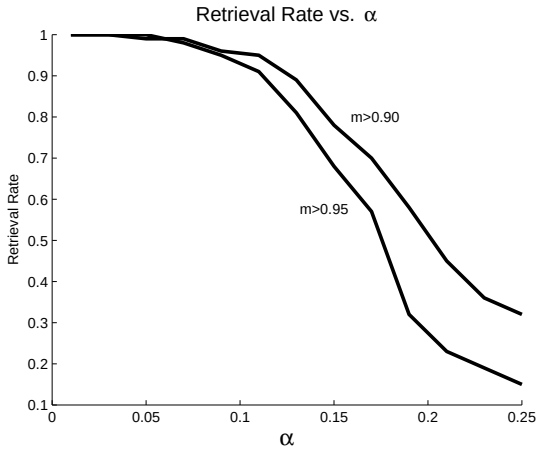

FIG. 2: Retrieval rate plotted against ratio of stored patterns to size of network. Initial states are constructed as stored patterns with an error percentage of 20%. Overlap threshold set to 0.95 and 0.90.

values of α. Specifically, I wanted to observe the dependence of the system's behavior on the aberration of the initial state. The plot I obtained is depicted in Figure 3.

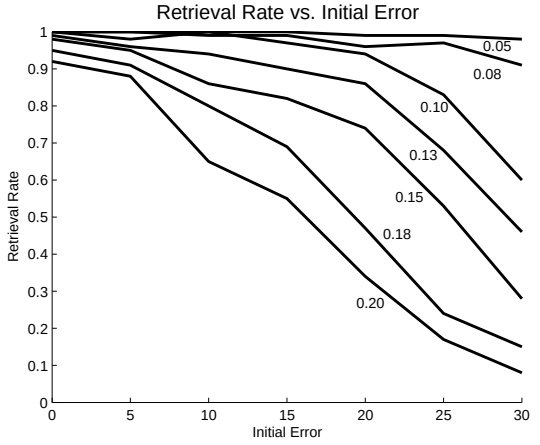

FIG. 3: Retrieval rate plotted against initial error percentages for varying values of α. Overlap threshold set to 0.95.

5. DISCUSSION

Figure 2 exhibits a sharp decline in retrieval efficacy for $\alpha \simeq 0.15$. From his own numerical simulations, Hopfield also estimated the critical storage capacity of his model to be approximately 0.15 [1].

Figure 3 shows a marked decrease in retrieval efficiency as the initial state increasingly deviates from the stored pattern. As the value of α increases, the number of spurious stable attractors also increases, amplifying the network decline.

The work presented here can and has been extended in myriad ways. Behavior at finite T is also investigated by Amit, *et. al.*, and the introduction of noise and stochastic evolution into the simulation would be a logical step forward. As Hopfield points out in [7] and [8], there are significant distinctions between neural models, in particular the model presented here, and real biological systems. Investigations performed on networks with learning rules different from (3) and with asymmetric synaptic connectivities have yielded valuable insights that can be found in the literature [9][10][11][12].

[1] Hopfield, J.J., "Neural networks and physical systems with emergent collective computational abilities", *Proc. Natl. Acad. Sci. USA*, **79**, 2554, [1982].

[2] Amit, D. *et. al.*, "Statistical mechanics of neural networks near saturation", *Ann. of Phys.* A, **173**, 30, [1987].

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

[4] Edwards, S.F. and Anderson, P.W., "Theory of Spin Glasses", *J. Phys.* F, **5**, 965, [1975].

[5] Sherrington, D. and Kirkpatrick, S., "Solvable Model of a Spin-Glass", *Phys. Rev. Lett.*, **35**, 1792, [1975].

[6] Amit, D., *et. al.*, "Spin-glass models of neural networks", *Phys. Rev.* A, **32**, 1007, [1985].

[7] Hopfield, J.J., "Neurons with Graded Response Have Collective Computational Properties like Those of Two-State Neurons", *PNAS*, **81**, 3088, [1984].

[8] Hopfield, J.J., "Neurons, Dynamics and Computation", *Phys. Today*, **47**, Feb., 40, [1994].

[9] Parisi, G., "Asymmetric neural networks and the process of learning", *J. Phys.* A, **19**, 675, [1986].

[10] Feng, J. and Tirozzi, B., "Capacity of the Hopfield Model", *J. Phys. A.: Math. Gen.*, **30**, 3383, [1997].

[11] Amit, D., *Modeling Brain Function*, Cambridge University, Cambridge, [1989].

[12] Müller, B., *et. al.*, *Neural Networks: An Introduction*, Springer, New York, [1995].

APPENDIX A: SIMULATION

```
%%Hopfield Network Simulation
clear;
%p: # of patterns to be embedded
p = 15;
%N: size of network
N = 100;
%a: number of aberrations in initial state from pattern
a = 20;

%%Simulation
%runs: how many times to run simulation
runs = 100;
%suc: # of successful retrievals
suc = 0;

%main loop
for r = 1:runs
%M: random matrix of memories
M = 2*(rand(p,N)¿0.5)-1;
%Ensure states are distinct
for i=1:p-1
for j=i+1:p
if abs(M(i,:)*(M(j,:)')) == N
M(j,:) = 2*(rand(1,N)¿0.5)-1;
i = 1;
end
end
end

%W: Hebbian matrix of synaptic connectivities
W = zeros(N,N);
for i = 1:N
for j = 1:N
W(i,j) = sum(M(:,i).*M(:,j));
end
end
W = W/N;
%S: initial state a elements different from stored pattern
S = M(1,:);
diff = find(randperm(N) ¡= a);
S(diff) = -S(diff);

%tf: runtime limit
tf = 100;
%t: runtime
t = 0;

%evolution of network
while t ¡ tf;
SOld = S;
%random order of updating
seq = randperm(N);
%operations performed each step
for dt = 1:N
k = seq(dt);
h = W(k,:)*(S');
S(k) = sign(h);
end
%increment t
t = t+1;
%check if at stable state
if SOld == S
trun = t;
t = tf;
end
end

%check if retrieval successful
if M(1,:)*(S')/N ¿= 0.90
suc = suc + 1;
end
end
suc/runs
```