MIT Department of Brain and Cognitive Sciences
9.641J, Spring 2005 - Introduction to Neural Networks
Instructor: Professor Sebastian Seung

Problem Set 7 (due Thursday, Apr. 14)
Models of Associative Memory

April 8, 2005

1. Rivalry model.

   This neural circuit has two neurons x_1 and x_2 interacting via mutual inhibition, and adaptation variables z_1 and z_2.

$$\dot{x}_1 + x_1 = [b_1 - \beta x_2 - \gamma z_1]^+ \quad (1)$$
$$\dot{x}_2 + x_2 = [b_2 - \beta x_1 - \gamma z_2]^+ \quad (2)$$
$$\tau \dot{z}_1 + z_1 = x_1 \quad (3)$$
$$\tau \dot{z}_2 + z_2 = x_2 \quad (4)$$

   Consider the case of very large τ, and $\beta > 1$. Suppose that $b_1 > b_2 > 0$, and the circuit is initialized with all variables equal to zero. Initially x_1 will win, and x_2 will be suppressed. Prove that this is the steady state for

$$\frac{b_1}{b_2} > \frac{1+\gamma}{\beta}$$

   Prove that if b_1/b_2 is less than this value, the circuit will switch back and forth.

   Show numerical simulations supporting the statements above.

2. Lyapunov analysis.

   (a) In class we presented the function

$$E = -\frac{1}{2}\sum_{ij} W_{ij} s_i s_j$$

   as a Lyapuov function (we use E here for "energy function", which is a synonym for Lyapunov function) for the Hopfield model using a sequential update:

$$s_i = \text{sign}(\sum_j W_{ij} s_j)$$

$$W_{ij} = \begin{cases} \frac{1}{N}\sum_\mu \xi_i^\mu \xi_j^\mu & i \neq j \\ 0 & i = j \end{cases}$$

   where $s \in \{-1, 1\}^N$. Prove this by showing:

   i. E is lower bounded.

   ii. $\Delta E \leq 0$, where ΔE is the change in energy after performing a sequential update.

   (b) Generalize the Lyapunov function to work for units with a bias term:

$$s_i = \text{sign}(\sum_j W_{ij} s_j + b_i).$$

   Prove your result as necessary.

(c) Find a Lyapunov function for a network which uses units $x_i \in \{0,1\}$ instead of $s_i \in \{-1,1\}$:

$$x_i = H(\sum_j W_{ij} x_j - \theta_i),$$

where H is the Heaviside function: $H(u) = 1$ if $u > 0$, otherwise $H(u) = 0$. Prove your result as necessary.

3. Capacity simulations.

One definition of the capacity of the Hopfield model is the number of patterns that can be stored where some small fraction ($P_{err} \leq 0.01$) of the bits are corrupted. Using this definition, the capacity of the original Hopfield model is approximately $0.14N$ for large N, where N is the number of units in the network. In this problem, we will validate this capacity using a simple MATLAB simulation, and then use our simulation to compare the capacities of original Hopfield model with the capacities an network storing sparse patterns using $\{0,1\}$ units.

(a) The original Hopfield model.
Construct P random $\{-1,1\}$ patterns, $\xi^1, ..., \xi^P$, each of size N. Find W using the prescription given in problem 2a.

As described in class, we investigate the capacity by checking if each of the stored patterns are actually steady states of the system. In class we showed the weight update from a stored pattern ξ^ν can be written as:

$$s_i = \text{sign}(\xi_i^\nu + \frac{1}{N} \sum_{\mu \neq \nu} \sum_{j \neq i} \xi_i^\mu \xi_j^\mu \xi_j^\nu).$$

We would like s_i to equal ξ_i^ν, but our steady state could be corrupted by the zero-mean crosstalk term. To visualize this in MATLAB, collect the terms $\sum_j W_{ij} \xi_j^u$ for all i and all μ and make a histogram of the results. To get a nice plot, use $N = 1000$ and 50 bins instead of MATLAB's default of 10.

Submit your matlab code and plots for $P = 100, 200$, and 140 (the known capacity for $N \to \infty$). Describe in words how the shape of the histogram changes as we change P, and how this impacts the capacity.

(b) Storing sparse patterns.
The notion of sparsity doesn't make sense for the $\{-1,1\}$ network, so we will use $\{0,1\}$ units for this part. The patterns that we wish to store are random with each bit having probability f of being a 1. We are interested in the case where f is small.
Consider the network using units $x_i \in \{0,1\}$ and the covariance rule:

$$W_{ij} = \begin{cases} \frac{1}{Nf(1-f)} \sum_\mu (\xi_i^\mu - f)(\xi_j^\mu - f) & i \neq j \\ 0 & i = j \end{cases}$$

with the discrete dynamics:

$$x_i = H(\sum_j W_{ij} x_j - \theta_i).$$

i. Show that for large N and small f the sum $\sum_j W_{ij} \xi_j^\nu$ can be seperated into ξ_i^ν and a crosstalk term.
ii. Show that this crosstalk term has zero mean.
iii. Construct P random $\{0,1\}$ patterns, each of size N, using f as the probability of a 1 bit. Plot the histogram of $\sum_j W_{ij} \xi_j^\mu$ as in part a. Experiment with P to estimate the capacity for $N = 1000$ and $f = 0.05$.
iv. According to your simulations, what value of the threshold θ_i maximizes the capacity?
v. One published result estimates the capacity of the sparse network as $P = \frac{N}{2f|\log(f)|}$. How well does this quantity compare to your results (test this by varying N and f)?

(c) How does the capacity of the sparse network compare to the original model?