Using Hysteresis to Improve Performance in Synchronous Networks

Tirunelveli Anand
Complex Adaptive Systems Laboratory
ECECS Department
University of Cincinnati
Cincinnati, OH 45221-0030
tanand@holmes.ece.uc.edu

Ali A. Minai
Complex Adaptive Systems Laboratory
ECECS Department
University of Cincinnati
Cincinnati, OH 45221-0030
Ali.Minai@uc.edu

Abstract

In this paper, we present a signal-to-noise analysis of synchronous attractor networks of hysteretic threshold elements. The addition of hysteresis is known to enhance the capacity and convergence rate of attractor networks. The aim of this paper is partly to elaborate on results reported previously by other researchers, and to address the issue of whether there is an optimal value for hysteresis. Based on the simplified analysis reported here, we conclude that, for a given network loading (ratio of patterns stored to network size), there is an optimal value of hysteresis, but it changes as recovery proceeds to convergence. We hypothesize that a time-varying "hysteresis schedule" can be used to enhance the performance of attractor networks.

1. Introduction

Attractor networks [5] [6] have been widely investigated as associative memories for storing bipolar patterns, and a number of results about their dynamics, storage capacity, etc., are available [4] [2]. In this paper, we focus on a relatively neglected aspect of attractor networks — the effect of using neurons with hysteresis [9].

Hysteretic neurons add an inertial element to the dynamics of attractor networks, so that small inputs (which are often the result of noise) have little impact on network dynamics. This can significantly reduce recall errors, and bestows a degree of immunity to zero mean noise [9] [10]. However, there has only been limited quantitative characterization of this effect. This paper addresses several important issues in this respect: 1) What is the mechanism by which hysteresis enhances network performance? 2) What is the relationship between performance and the degree of hysteresis? and 3) Is there an optimal value for hysteresis given network size and storage?

2. Network Description

We consider networks of N neurons with ± 1 outputs. The input to neuron i at time t is given by:

$$u_i(t) = \sum_{j \neq i}^{N} w_{ij} z_j(t-1)$$

where z_j is the output of neuron j and w_{ij} is the strength of the connection from j to i. The output of neuron i is then calculated as

$$z_i(t) = sgn\left[u_i(t) + T z_i(t-1)\right]$$

where $sgn(\ )$ is the signum function and T is the hysteresis parameter. Essentially, the effect of the T parameter is to create a "hysteresis region" of width $2T$ in the neuron's switching function, so that output $z_i(t)$ switches from $+1$ to -1 only if $u_i(t) < -T$, and from -1 to $+1$ only if $u_i(t) > T$.

A set of P patterns, $\xi^\mu = \left[\xi_j^\mu\right], j = 1, ..., N$ are stored in the network using the the standard Hebbian rule:

$$w_{ij} = \frac{1}{N} \sum_{\mu} \xi_i^\mu \xi_j^\mu$$

with bits chosen to be $+1$ or -1 independently with probability 0.5. Recall for pattern ξ^ν is then tested by setting the network's initial state to a probe vector, $\bar{\xi}^\nu$, which differs from ξ^ν in K bits. The network is expected to converge to ξ^ν.

The similarity between the network state, $z(t)$, at time t and a pattern, ξ^ν is measured by

$$m^\nu(t) = \frac{1}{N}\sum_j \xi_j^\nu z_j(t)$$

which is between -1 and 1. It is related to the number of incorrect bits by $\#incorrect = N(1-m)/2$, so the number of correct bits is $\#correct = N(1+m)/2$.

3. Analysis

To simplify analysis, we consider the recall of a specific pattern, ξ^ν, with the assumption that, at time 0, $m^\nu(0)$ is the only overlap of magnitude $O(1)$. The method follows the standard "signal-to-noise" approach [7] [4], which is used primarily to analyze one-step stability under the assumption that all m^μ, $\mu \neq \nu$ are small (compared to m^ν) and independent. Analysis of multi-step dynamics in a fully connected network requires the more sophisticated mean-field approximation [2] or more advanced methods of statistical neurodynamics which take temporal correlations into account [1] [8]. Thus, the analysis given below is only a first-cut approximation. One situation where signal-to-noise analysis does become exact is *strong asymmetric dilution* [3]. However, strong dilution typically requires *extremely* large networks. The extent to which the approximation is useful for the fully-connected case is being investigated through simulations.

The effect of hysteresis on network dynamics can be understood by dividing neurons into four classes at any time step $t+1$. These are:

1. *Lost Neurons (L)*: All neurons i which have incorrect output at step $t+1$ and input of the incorrect sign, i.e $\xi_i^\nu z_i(t+1) = -1$ and $\xi_i^\nu u_i(t+1) = -1$. These neurons would have been incorrect regardless of hysteresis.

2. *Blocked Neurons (B)*: All neurons i which have incorrect output at step $t+1$ but input of the correct sign, i.e $\xi_i^\nu z_i(t+1) = -1$ and $\xi_i^\nu u_i(t+1) = +1$. These neurons are incorrect only because hysteresis has locked them into their state at step t.

3. *Protected Neurons (P)*: All neurons i which have correct output at step $t+1$ but input of the incorrect sign, i.e $\xi_i^\nu z_i(t+1) = +1$ and $\xi_i^\nu u_i(t+1) = -1$. These neurons are correct only because hysteresis has locked them into their state at step t.

4. *Safe Neurons (S)*: All the remaining neurons i, with $\xi_i^\nu z_i(t+1) = +1$ and $\xi_i^\nu u_i(t+1) = +1$. These neurons would have been correct regardless of hysteresis.

Using signal-to-noise analysis (e.g., [4]), it can be shown that, when $m^\nu(t)$ is the only significant ($O(1)$) overlap and N and P are large, the probabilities of a neuron i belonging to each of the four classes at step $t+1$, with $m^\nu(t) = m$, are:

$$\begin{aligned}
P(i \in L) &\equiv P_L(m,T) \\
&= Q\left(\frac{m+T}{\sqrt{\alpha}}\right) \\
&\quad + \frac{1-m}{2}\left[Q\left(\frac{m}{\sqrt{\alpha}}\right) - Q\left(\frac{m+T}{\sqrt{\alpha}}\right)\right] \\
P(i \in B) &\equiv P_B(m,T) \\
&= \frac{1-m}{2}\left[Q\left(\frac{m-T}{\sqrt{\alpha}}\right) - Q\left(\frac{m}{\sqrt{\alpha}}\right)\right] \\
P(i \in P) &\equiv P_P(m,T) \\
&= \frac{1+m}{2}\left[Q\left(\frac{m}{\sqrt{\alpha}}\right) - Q\left(\frac{m+T}{\sqrt{\alpha}}\right)\right] \\
P(i \in S) &\equiv P_S(m,T) \\
&= 1 - (P_L + P_B + P_P) \\
&= 1 - Q\left(\frac{m-T}{\sqrt{\alpha}}\right) \\
&\quad + \frac{1+m}{2}\left[Q\left(\frac{m-T}{\sqrt{\alpha}}\right) - Q\left(\frac{m}{\sqrt{\alpha}}\right)\right]
\end{aligned}$$

where $\alpha \equiv P/N$, and

$$Q(x) \equiv \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-x^2/2}dx$$

From this, it is simple to calculate the mean-field value of $m^\nu(t+1)$ as

$$\begin{aligned}
\langle m(t+1)\rangle &= 1 - \\
&\quad \left[(1+m)Q\left(\frac{m+T}{\sqrt{\alpha}}\right)\right. \\
&\quad \left. +(1-m)Q\left(\frac{m-T}{\sqrt{\alpha}}\right)\right]
\end{aligned}$$

This is the same as the expression found by Wang and Ross [9]. The effect of hysteresis (compared to the no hysteresis case) can be measured by the difference of cardinality between sets P and B. For large N, this reduces to $D(m,T) \equiv N(P_P - P_B)$. The optimal value of hysteresis, $T_{opt}(m)$, defined as that T which gives the greatest $D(m,T)$ at the next step (or the greatest $m(t+1)$), is then given by:

$$T_{opt} = \frac{P}{2mN} ln \left[\frac{1+m}{1-m} \right]$$

To ascertain the long term behavior of the system, the mean-field dynamics of $m^{\nu}(t)$ can be calculated iteratively from Eqn. (5). The final overlap, $\langle m(\infty) \rangle$, is a function of $m(0)$, T, and α, as is the rate of convergence.

The questions of interest for recall are whether $m(\infty) > m(0)$, and how close it is to 1. This can be seen in terms of the 1-step dynamics of $m(t)$. The map from $m(t)$ to $m(t+1)$ depends on the loading, α, and the hysteresis, T (Figure 1).

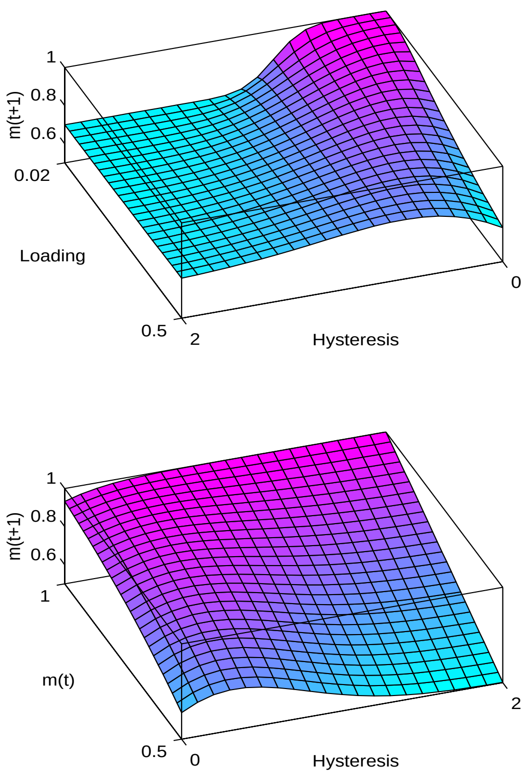

Figure 1. The dependence of $m(t+1)$ on loading and $m(t)$. In the plot on the left, $m(t) = 0.7$, while in the other, $\alpha = 0.3$.

There exists a point $m(t+1) = m(t) < 1 \equiv m^*(\alpha, T)$, which is the fixed point of the dynamics and the best mean retrieval achievable at that loading and hysteresis (Figures 2 & 3).

Probe patterns with $m(0) > m^*$ will degrade down to m^*. The recall quality desired, m_d, therefore, implies a critical loading, α_c, which gives the capacity of the network. It is that α beyond which $m(t+1) < m(t)$ for $m(t) = m_d$, and is a decreasing function of m_d. (It must be emphasized, however, that this analysis ignores temporal correlations produced by the recurrent dynamics. More accurate — but more complicated — approximations can be obtained through statistical neurodynamics techniques [1] [10] [8].)

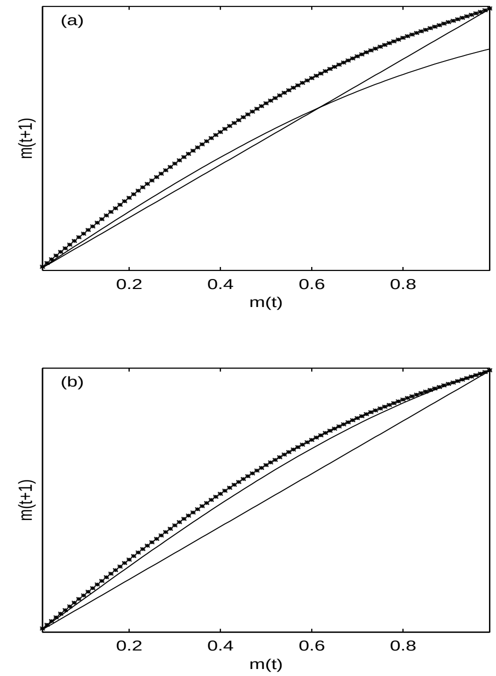

Figure 2. The dependence of the $m(t) \to m(t+1)$ map on hysteresis. $T = 0$ for (a) and 1.0 for (b), with $\alpha = 0.5$. The bold curve is the map obtained when T is set to its optimal value, T_{opt}, for each $m(t)$. The intersection with the diagonal is the fixed point, m^*.

It is obvious from Eqn. (5) that, for a given m^*, the critical loading is an increasing function of T (Figure 4), but setting T too high has two negative consequences: 1) It slows convergence drastically; and 2) In finite size networks, it can preclude any completion whatsoever.

For any given loading, However, if T is updated every step to the T_{opt} corresponding to the current overlap, convergence is much faster (since T_{opt} is defined as giving the maximal increment of m), and, in theory, has $m^* = 1$ for any loading (Figure 3), since, by definition, we choose T to maximize the current $m(t+1) - m(t)$.

Of course, as loading increases, the convergence becomes very slow (since T_{opt} becomes very large), and the assumption of small, independent overlaps m^μ is violated. Nevertheless, we can conclude that: 1) Hysteresis always increases network capacity (at least in infinite size networks); and 2) For any reasonable loading, α, and sizeable $m(0)$, there exists an update schedule for T which would assure full completion. Of course, this schedule will not be available in a practical situation, since its calculation requires knowledge of $m(t)$, but it might be possible to define near-optimal default schedules based on an estimated $m(0)$ and the loading. This is especially so since T_{opt} is a strictly increasing function of m, and, in a completion situation, must increase monotonically in time.

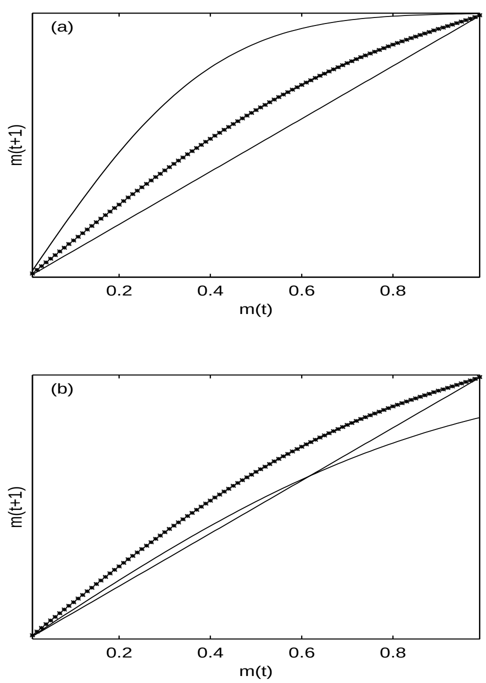

Figure 3. The dependence of the $m(t) \to m(t+1)$ map on α with $T = 0$. The settings are: $\alpha = 0.1$ for (a) and 0.5 for (b).

4. Simulation Results

Figure 5 shows some preliminary simulation results for a 500 neuron network storing 200 patterns. The loading of 0.4 is clearly too high for recovery at no hysteresis. However, the introduction of a little hysteresis begins to produce recall, which peaks near $T = 1$ before succumbing to finite-size effects. For very large hysteresis, there is no dynamics because all neurons have u_i in the hysteresis band. The shape of the overlap curve is of particular interest, reflecting the dependence of $m(t+1)$ on T (Figure 4(b)).

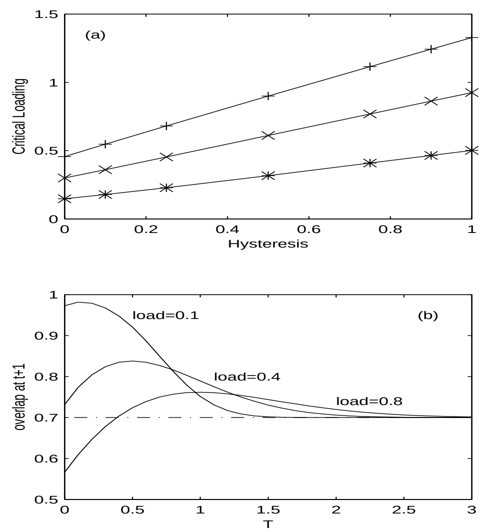

Figure 4. Graph (a) shows the dependence of the critical loading, α_c on hysteresis for $m(t) = 0.99$ (*), $m(t) = 0.9$ (x), and $m(t) = 0.7$ (+). Graph (b) shows the value of $m(t+1)$ vs. hysteresis, T, for various loadings. In all cases, $m(t) = 0.7$. The critical loading is that α at which $m(t+1)$ falls below $m(t)$, implying failure of recall. However, $m(t+1) > m(t)$ does not imply convergence to $m(\infty) = 1$, only that $m(\infty) > m(0)$.

There is also a decrease in convergence time with hysteresis. In a large enough network, this speedup would eventually reverse as increments in $m(t)$ diminished, but for a finite size network, convergence at high hysteresis in instantaneous because there is no dynamics!

5. Discussion

The foregoing analysis illustrates how hysteresis works in an attractor network. It has two opposing effects: protection of already correct bits that might have been corrupted, and blocking of incorrect bits from switching to their correct states. The interaction between these two effects determines the optimal hysteresis.

Heuristically, the argument can be understood as follows. Starting from a reasonably high overlap with the correct pattern, it can be expected that neurons which get input of large magnitude are already well-aligned with the desired pattern.

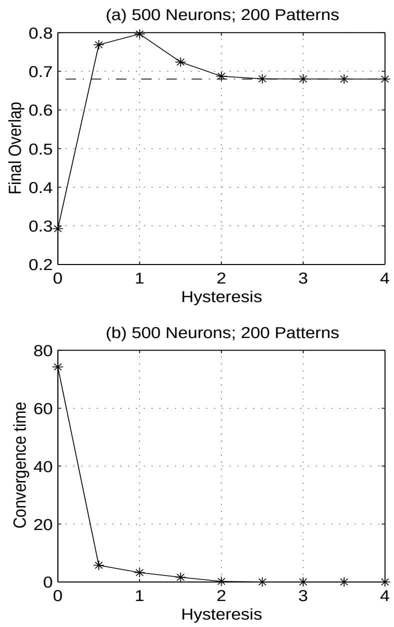

Figure 5. The graph on top shows the final average overlap as a function of hysteresis for recall in a 500 neuron network with 200 patterns. The dashed horizontal line indicates the initial overlap (0.68). The graph on the bottom shows the mean convergence time for the same simulations. All data points are averaged over 10 runs with independent data sets.

However, neurons which get a small recurrent input are in a "risky" situation, and their input might well have an incorrect polarity. By clamping these neurons, we essentially try to "preserve" the existing ratio of correct bits in this risky population.

If the overlap with the desired pattern is high enough, probability will generally ensure that most of the clamped bits are correct (protected) rather than incorrect (blocked). However, as T becomes too large, the clamping begins to encroach upon neurons that are not at risk, preventing them from (correctly) following their input. The larger the existing overlap, the larger hysteresis can be made, since preserving the status quo becomes increasingly optimal. When overlap is perfect ($m^\nu = 1$), hysteresis can be made infinite because no further switching is needed.

References

[1] S.-I. Amari and K. Maginu. Statistical neurodynamics of associative memory. *Neural Networks*, 1:63–73, 1988.

[2] D. Amit. *Modeling Brain Function*. Cambridge University Press, Cambridge, UK, 1989.

[3] B. Derrida, E. Gardner, , and A. Zippelius. An exactly solvable asymmetric neural network model. *Europhys. Lett.*, 4:167–173, 1987.

[4] J. Hertz, A. Krogh, and R. Palmer. *An introduction to the theory of neural computation*. Addison-Wesley, Redwood City, CA, 1991.

[5] J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci. USA*, 79:2554–2558, 1982.

[6] J. Hopfield. Neurons with graded response have collective computational properties like those of two state neurons. *Proc. Nat. Acad. Sci. USA*, 81:3088–3092, 1984.

[7] W. Kinzel. Learning and pattern recognition in spin glass models. *Z. Phys. B.*, 60:205–213, 1985.

[8] M. Okada. A hierarchy of macrodynamical equations for associative memory. *Neural Networks*, 8:833–838, 1995.

[9] L. Wang and J. Ross. Synchronous neural networks of nonlinear threshold elements with hysteresis. *Proc. Nat. Acad. Sci. USA*, 87:988–992, 1990.

[10] H. Yanai and Y. Sawada. Associative memory network composed of neurons with hysteretic property. *Neural Networks*, 3:223–228, 1990.