ENERGY MINIMIZATION IN THE NONLINEAR DYNAMIC RECURRENT ASSOCIATIVE MEMORY

Sébastien Hélie*

*Rensselaer Polytechnic Institute, Cognitive science department, 110 Eighth Street, Carnegie 108, Troy, NY 12180-3590 USA. Phone: 1 (518) 276-2692, Fax: 1 (518) 276-3017, E-mail: helies@rpi.edu.

Abstract

Chartier and his colleagues have recently proposed a nonlinear synchronous attractor neural network. In the Nonlinear Dynamic Recurrent Associative Memory (NDRAM), learning has been shown to converge to a set of real-valued attractors in single-layered neural networks and bidirectional associative memories. However, the transmission is highly nonlinear and its global stability has never been proven analytically. In the present, it is shown that NDRAM is an instance of the Cohen-Grossberg class of models and its energy function is defined. Analysis of the energy function shows that the transmission is stable in the entire domain of NDRAM. Numerical simulations further support this analysis.

Keywords: Artificial neural network, dynamical system, energy function, recurrent associative memory, Cohen-Grossberg theorem.

ENERGY MINIMIZATION IN THE NONLINEAR DYNAMIC RECURRENT ASSOCIATIVE MEMORY

1. Introduction

Chartier and his colleagues (Chartier & Boukadoum, 2006_a, 2006_b; Chartier & Proulx, 2005) have recently proposed a nonlinear synchronous attractor neural network (see Fig. 1). In the Nonlinear Dynamic Recurrent Associative Memory (NDRAM), the transmission is highly nonlinear and described by the following:

$$\mathbf{x}_{[t+1]} = f\left(\mathbf{W}\mathbf{x}_{[t]}\right) \tag{1}$$

$$\forall_i,\ 1 \le i \le N:\ f(x_i) = \begin{cases} +1 & ,\ x_i > 1 \\ (\delta+1)x_i - \delta x_i^3 & ,\ -1 \le x_i \le 1 \\ -1 & ,\ x_i < -1 \end{cases} \tag{2}$$

where N is the number of units in the network, $\mathbf{x}_{[t]} = \{x_1, x_2, \ldots, x_N\}$ is the state of the network at time t ($\forall_i$, $-1 \le x_i \le 1$), $\delta > 0$ is a free parameter representing the slope of the transmission function, and $\mathbf{W} = [w_{ij}]$ is the $N \times N$ weight matrix built online using:

$$\mathbf{W}_{[t+1]} = \mathbf{W}_{[t]} + \eta\left(\mathbf{x}_{[0]}\mathbf{x}_{[0]}^{\mathbf{T}} - \mathbf{x}_{[p]}\mathbf{x}_{[p]}^{\mathbf{T}}\right) \tag{3}$$

where $\mathbf{W}_{[t]}$ is the weight matrix at time t ($\mathbf{W}_{[0]} = \mathbf{0}$), $\mathbf{x}_{[0]}$ is the state of the network at time 0 (the input), $\mathbf{x}_{[p]}$ is the state of the network after p applications of Eq. 1 and Eq. 2, and $\eta < \dfrac{1}{2(1-2\delta)N}$ is the learning rate. This learning rule has been shown to converge to a set of real-valued attractors in single-layered neural networks (Chartier & Proulx, 2005) and bidirectional associative memories (Chartier & Boukadoum, 2006_a). It has been used to learn sequences of real-valued stimuli (Chartier & Boukadoum, 2006_b) and to perform density estimation (Hélie, Chartier, & Proulx, 2006).

Insert Fig. 1 about here

While the convergence of Eq. 3 has been proven is several papers (e.g., Chartier & Boukadoum, 2006_a; Chartier & Proulx, 2005), Eq. 2 has not received as much attention. In Chartier and Proulx (2005), the Lyapunov exponent of the transmission function has been numerically estimated in order to constrain the values of some of the free parameters, but no global analysis of the energy landscape of the model has been performed. The aim of this paper is to provide NDRAM with an energy function and prove its global stability.

2. Global stability analyses and the Cohen-Grossberg model

Cohen and Grossberg (1983) have defined a general class of neural networks that have been shown to be globally stable. The general form of the model is:

$$\frac{d}{dt}x_i = a_i(x_i)\left[b_i(x_i) - \sum_{j=1}^{N} c_{ij} d_j(x_j)\right] \tag{4}$$

where N is the number of units in the network, $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ is the state of the network, $\mathbf{C} = [c_{ij}]$ is a coefficient matrix, $a_i(\bullet)$ is the amplification function, $b_i(\bullet)$ is the self-signal function, and $d_j(\bullet)$ is the 'other signal' function (Grossberg, 1988). Networks that can be written in this general form (Eq. 4) and obey the following conditions:

Symmetry: $c_{ij} = c_{ji}$ (C1)

Positivity: $a_i(x_i) \geq 0$ (C2)

Monotonicity: $d'_j(x_j) \geq 0$ (C3)

allows the following Lyapunov function:

$$L(\mathbf{x}) = -\sum_{i=1}^{N} \int_0^{x_i} b_i(\xi_i) d_i'(\xi_i) d\xi_i + \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} c_{jk} d_j(x_j) d_k(x_k) \tag{5}$$

where the symbols are the same as in Eq. 4. Grossberg (1988) has shown that many neural networks fall in this class of models: e.g., the Brain-State-in-a-Box (Anderson, Silverstein, Ritz, & Jones, 1977), the McCulloch-Pitts model (McCulloch & Pitts, 1943), and the Boltzmann machine (Ackley, Hinton, & Sejnowski, 1985). In the following section, it is shown that NDRAM is also an instance of the Cohen-Grossberg model.

3. A global analysis of the energy landscape in NDRAM

This section is organized as follow. First, it is shown that the transmission in NDRAM (Eq. 1 and Eq. 2; Chartier & Boukadoum, 2006_a, 2006_b; Chartier & Proulx, 2005) can be written in the general form of the Cohen-Grossberg model (Eq. 4; Cohen & Grossberg, 1983; Grossberg, 1988). Second, the transmission in NDRAM is shown to obey conditions C1 – C3. Finally, an energy function is derived and comments on the stability of NDRAM are presented.

The first step in showing that NDRAM is an instance of the Cohen-Grossberg class of models is to approximate its transmission in continuous form. Eq. 1 becomes:

$$\frac{d}{dt}x_i = -x_i + f\left(\sum_{j=1}^{N} w_{ij}x_j\right) \tag{6}$$

where the symbols are the same as in Eq. 1. To facilitate the demonstration that Eq. 6 has the same general form as Eq. 4, the following change of coordinates is made (Grossberg, 1988): $y_i = \sum_{j=1}^{N} w_{ij}x_j$. Hence, Eq. 6 becomes:

$$\frac{d}{dt}y_i = -y_i + \sum_{j=1}^{N} w_{ij}f\left(y_j\right) \tag{7}$$

where $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$ is the state of the network in the new coordinates, and the remaining symbols are the same as in Eq. 6. By making the following substitutions, Eq. 7 has the same general form as Eq. 4: $a_i(y_i) = 1$, $b_i(y_i) = - y_i$, $c_{ij} = -w_{ij}$, and $d_j(y_j) = f(y_j)$. ■

Next, it is shown that NDRAM obeys conditions C1 – C3. Showing that conditions C1 and C2 are satisfied by NDRAM is trivial. First, the weight matrix in NDRAM is built online using Eq. 3. As seen, this learning rule is the difference between two covariance matrices, which are by definition symmetric: their difference is also symmetric. Hence, C1 is satisfied provided that the weight matrix was initiated as a symmetric matrix. This is the case, as the $\mathbf{W}_{[0]} = \mathbf{0}$. Second, $a_i(y_i)$ is constant and equals unity, which satisfies C2 (positivity). As for monotonicity (C3), the saturation limits in Eq. 2 are constant and thus satisfy C3. However, showing that the nonlinear part of the transmission function obeys this last condition requires constraining the slope of the transmission function (δ). First, the derivative of the nonlinear part of Eq. 2 is defined:

$$\frac{d}{dx_i} f(x_i) = \delta + 1 - 3\delta x_i^2 \tag{8}$$

where the symbols are the same as in Eq. 2. To constrain the value of δ and satisfy C3 (Monotonicity), Eq. 8 can be more conveniently rewritten:

$$\delta\left(1 - 3x_i^2\right) \geq -1 \tag{9}$$

where again the symbols are the same as in Eq. 2. By setting $x_i = \pm 1$, which are the limit cases, and solving for δ, it is shown that C3 is satisfied when $\delta \leq 1/2$. ■

Now that NDRAM has been shown to be an instance of the Cohen-Grossberg class of models, its energy function can be computed. By inserting the previously defined substitutions into Eq. 5:

$$E(\mathbf{y}) = \sum_{i=1}^{N} \int_0^{y_i} \xi_i f'(\xi_i) d\xi_i - \frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} w_{jk} f(y_j) f(y_k) \tag{10}$$

where the symbols are the same as in Eq. 7. The next step is to insert the derivative of $f(\bullet)$ (Eq. 8) into Eq. 10 and solve the integral:

$$\begin{aligned} E(\mathbf{y}) &= \sum_{i=1}^{N} \int_0^{y_i} \xi_i + \delta\xi_i - 3\delta\xi_i^3 d\xi_i - \frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} w_{jk} f(y_j) f(y_k) \\ &= \frac{1}{2} \left[\sum_{i=1}^{N} y_i^2 \left(1 + \delta - \frac{3}{2} \delta y_i^2 \right) - \sum_{j=1}^{N} \sum_{k=1}^{N} w_{jk} f(y_j) f(y_k) \right] \end{aligned} \tag{11}$$

And by returning to the original coordinates:

$$E(\mathbf{x}) = \frac{1}{2} \left[\sum_{i=1}^{N} \left(\sum_{l=1}^{N} w_{il} x_l \right)^2 \left[1 + \delta - \frac{3}{2} \delta \left(\sum_{l=1}^{N} w_{il} x_l \right)^2 \right] - \sum_{j=1}^{N} \sum_{k=1}^{N} w_{jk} f\left(\sum_{l=1}^{N} w_{jl} x_l \right) f\left(\sum_{l=1}^{N} w_{kl} x_l \right) \right] \tag{12}$$

■

Because this energy function can be interpreted as a Lyapunov function, its slope is informative about the stability of the transmission in NDRAM. Cohen and Grossberg (1983) have shown that the derivative of Eq. 5 is:

$$\frac{d}{dt} L(\mathbf{x}) = -\sum_{i=1}^{N} a_i(x_i) d_i'(x_i) \left[b_i(x_i) - \sum_{j=1}^{N} c_{ij} d_j(x_j) \right]^2 \tag{13}$$

where the symbols are the same as in Eq. 4. Hence, using the previous substitutions:

$$\frac{d}{dt} E(\mathbf{y}) = -\sum_{i=1}^{N} f'(y_i) \left[-y_i + \sum_{j=1}^{N} w_{ij} f(y_j) \right]^2 \tag{14}$$

where the symbols are the same as in Eq. 7. Going back to the original coordinates, the following function is found to illustrate the slope of Eq. 12:

$$\frac{d}{dt} E(\mathbf{x}) = -\sum_{i=1}^{N} f'\left(\sum_{j=1}^{N} w_{ij} x_j \right) \left[-\sum_{j=1}^{N} w_{ij} x_j + \sum_{j=1}^{N} w_{ij} f\left(\sum_{k=1}^{N} w_{jk} x_k \right) \right]^2 \tag{15}$$

where the symbols are the same as in Eq. 6. Because the transmission in NDRAM has been shown to obey C3, Eq. 15 is negative throughout its domain, and the transmission is stable everywhere in NDRAM: every trajectory converges on one of the learned attractors (i.e., a learned pattern or a spurious attractor). ■

4. Simulation

This section present some numerical simulations to support the derivations detailed in Section 3. Three simulations were run, varying the slope of the energy function (δ). All other parameters and methodology were fixed for all three simulations.

For graphical purposes, and without lost of generality, all the simulations were conducted in two dimensions ($N = 2$). There were two training stimulus, $\{-1, -1\}^T$ and $\{-1, 1\}^T$. In each simulation, NDRAM was trained for 1000 epochs with the following parameter values: $\eta = 0.01$, $\zeta = 0.9999$, and $p = 1$. After training, 100 random stimuli $[-1, 1]^N$ were input to the trained networks and allowed to settle.

The slope of the transmission function was the only difference between the simulations. Precisely, three values were used: $\delta = 0.1$, $\delta = 0.4$, and $\delta = 0.5$. These values were chosen because the first one is rather small, the second is the usual value given to this parameter (e.g., Chartier & Proulx, 2005; Hélie et al., 2006), while the third value represents the limit case. The simulation results are shown in Fig. 2.

Insert Fig. 2 about here

As can be seen, all the random stimuli have converged, notwithstanding the value given to the slope parameter. The main difference between the simulations is the minimum of the energy function: when $\delta = 0.1$, the energy function is relatively flat and the random stimuli converged to an energy minimum of -0.04. When $\delta = 0.4$, the energy function is much more 'bumpy' and the random stimuli converged to an energy minimum of -0.18. When $\delta = 0.5$, the energy function is similar to the preceding and the random stimuli converged to an energy minimum of -0.19. Hence, higher δ values produce lower energy minimum, and accelerate convergence (see the scale of the *x*-axis in the simulation graphs). Nevertheless, convergence occurs in all cases, which support the previous analysis of global stability.

5. Summary and conclusion

In the present, it was shown that NDRAM is an instance of the Cohen-Grossberg class of models (Cohen & Grossberg, 1983) when $\delta \leq 1/2$. Using this property, an energy function was defined and the global stability of the transmission in NDRAM was proven and supported by numerical simulations. It is interesting to note that the constraint on the slope of the transmission function is the same that was found to prevent the model from being chaotic (Chartier & Proulx, 2005). However, links between the monotonicity of the transmission function and the behavior of the model outside of the attractor fields have not been explored prior to the present.

Global energy landscape analyses of attractor neural networks are very important in understanding their stability and proving the convergence of their trajectories. Given its generality, the stability theorem presented by Cohen & Grossberg (1983) is very useful and makes stability analyses for a large class of dynamical systems accessible to non-experts. It is our hope that other theorems in nonlinear dynamics can be made available to provide the mathematical rigor needed to understand cognitive models in general and artificial neural networks in particular.

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, *9*, 147-169.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413-451.

Chartier, S., & Boukadoum, M. (2006_a). A bidirectional heteroassociative memory for binary and grey-level patterns. *IEEE Transactions on Neural Networks*, *17*, 385-396.

Chartier, S., & Boukadoum, M. (2006_b). A sequential dynamic heteroassociative memory for multistep pattern recognition and one-to-many association. *IEEE Transactions on Neural Networks*, *17*, 59-68.

Chartier, S., & Proulx, R. (2005). NDRAM: Nonlinear Dynamic Recurrent Associative Memory for learning bipolar and non-bipolar correlated patterns. *IEEE Transactions on Neural Networks*, *16*, 1393-1400.

Cohen, M. A., & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-13*, 815-826.

Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, *1*, 17-61.

Hélie, S., Chartier, S., & Proulx, R. (2006). Are unsupervised neural networks ignorant? Sizing the effect of environmental distributions on unsupervised learning. *Cognitive Systems Research*, *7*, 357-371.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115-133.

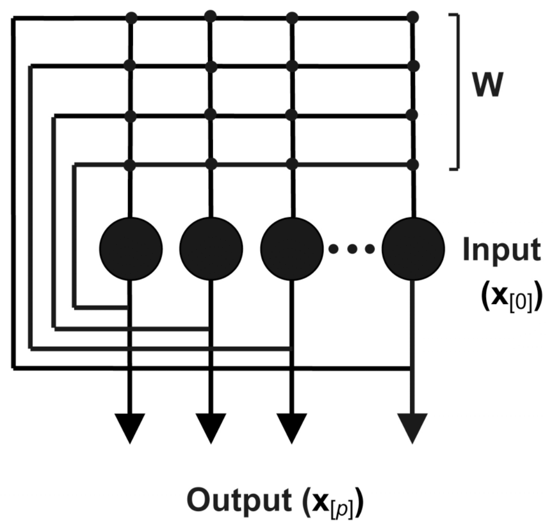
Acknowledgments

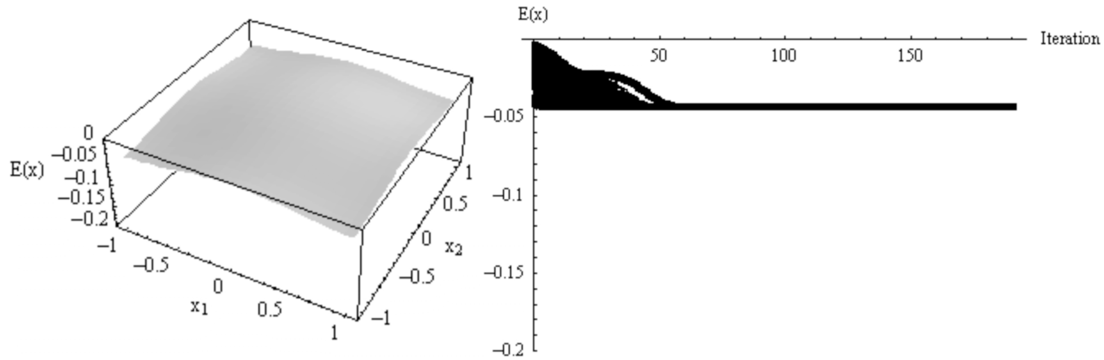
This research was supported by a postdoctoral research fellowship from *Le Fonds Québecois de la Recherche sur la Nature et les Technologies*. The author would like to thank Dr. Sylvain Chartier, Dr. Denis Cousineau, Gyslain Giguère, and two anonymous reviewers for their useful comments on an earlier version. Requests for reprints should be addressed to Sébastien Hélie, Cognitive science department, Rensselaer Polytechnic Institute, 110 Eighth Street, Carnegie 108, Troy, NY 12180-3590 USA, or using e-mail at helies@rpi.edu.

Figure captions

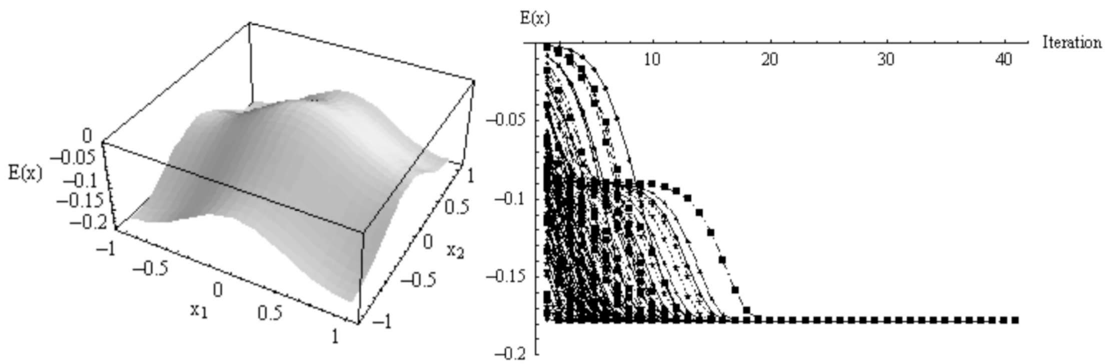
Fig. 1. Architecture of NDRAM.

Fig. 2. Simulation results. For each parameter value, the left graph shows the energy function and the right graph shows the energy of the random stimuli at each iteration (each line represents a different random stimulus).

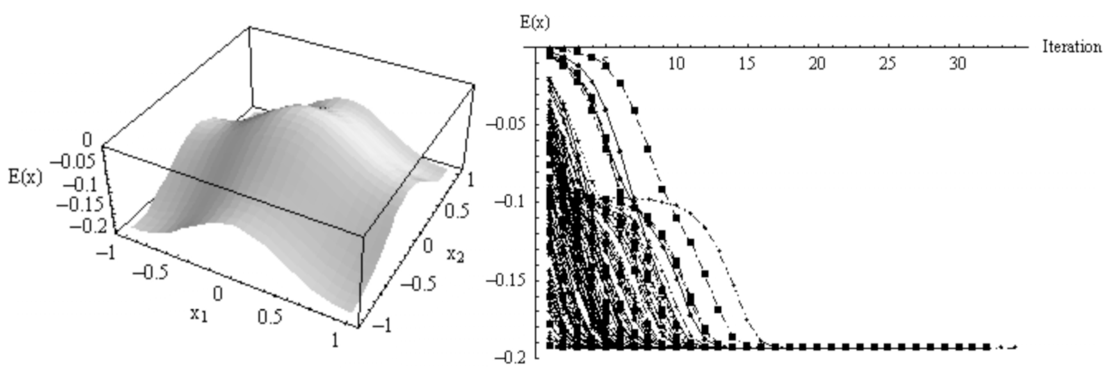

W
Input
($\mathbf{x}_{[0]}$)
Output ($\mathbf{x}_{[p]}$)



$\delta = 0.1$



$\delta = 0.4$



$\delta = 0.5$