

A Novel Approach for Analyzing Dynamics in Neural Networks with Saturated Characteristics

JIANFENG FENG^{1,2} and DAVID BROWN¹

¹ Laboratory of Biomathematics, The Babraham Institute, Cambridge CB2 4AT UK;

² Mathematisches Institut, Universität München, D-80333, München, Germany

E-mail: jf218@cam.ac.uk

Key words: capacity, Hopfield model, Linsker's model, saturated characteristic

Abstract. Current theories about the dynamics of neural networks with nonlinear characteristics and parameterized by set of parameters are mostly based on approximations in one way or another. In this paper we first introduce a rigorous approach which allows us to check in which parameter region a given saturated state is an attractor of the dynamics: a saturated state $w = (w_i, i = 1, \dots, N) \in \{-1, 1\}^N$ is an attractor of the dynamics if and only if there is a *local field gap* between neurons in $J^+(w) = \{i, w_i = 1\}$ and $J^-(w) = \{i, w_i = -1\}$. Then we apply the result to analyze several models in neural networks. In particular in the Hopfield model we calculate the capacity and give an exact relation between the capacity and the threshold.

1. Introduction

Consider the following N -dimensional dynamics

$$w(t+1) = F(w(t) + G(k_1, k_2, w(t))), \quad t = 1, \dots, \quad (1)$$

where $w(t) = (w_1(t), \dots, w_N(t)) \in \mathbb{R}^N$ and $F = (f, \dots, f)$, $G = (g_1, \dots, g_N)$ with $f, g_i \in C(\mathbb{R})$, $i = 1, \dots, N$, $(k_1, k_2) \in \mathbb{R}^2$ parameters of the dynamics. More specifically we suppose that f is a saturated sigmoidal input-output characteristic which means that f is continuous and $f(x) = 1$ if $x \geq 1$, $f(x) = -1$ if $x \leq -1$, $f(x)$ is increasing inside $[-1, 1]$ and $f(0) = 0$. Many models developed in neural networks to date are special case of the dynamics (1) including

- Linsker's model [6, 7, 12]:

Linsker's model resembles the visual system, with an input feeding onto a number of layers corresponding to the layers of the visual cortex. In this case $w_i(t)$ is explained as the synaptic connection between adjacent layers and as a result of Hebb learning rule we have

$$w_i(t+1) = f(w_i(t) + k_1 + \sum_{j=1}^N (q_{ij} + k_2)r_j w_j(t)) \quad (2)$$

where $Q = (q_{ij}, i, j = 1, \dots, N)$ is the covariance matrix, $\sum_i r_i = 1$ with $r_i \geq 0$, k_1, k_2 are two parameters resulting from the Hebb learning rule.

- The Hopfield model [1, 2, 8]:

The sigmoidal function $\sigma_\beta = \frac{2}{1+\exp(-\beta x)} - 1$ can be approximately saturated for large β and so $f \sim \sigma_\beta$ (β large). Hence, the Hopfield model reads

$$x(t+1) = F(x(t) + G(\theta, h, x(t))) \quad (3)$$

with

$$g_i(\theta, h, x(t)) = \sum_{j=1}^N (T_{ij} + h)x_j(t) + \theta \quad (4)$$

$$T_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu, \quad i, j = 1, \dots, N \quad (5)$$

and θ the threshold, h the external field. $x_i(t)$ is the neural activity at time t of the i -th neuron where $\xi^\mu = (\xi_i^\mu, i = 1, \dots, N)$ the μ -th pattern stored in the network.

- Dynamical link network [9, 11] (DLN):

It is reported that this model has been successfully applied to face recognition. The dynamic link network is essentially a two layer network, say layer X and layer Y with both intra-layer connections and inter-layer connections. After the dynamics of inter-layer connections arrives at its equilibrium state then the intra-layer connections update. Here we are only going to consider the intra-layer dynamics say the X layer which can be written as

$$\begin{cases} \eta_i(t+1) = f(\sum_{j=1}^N k_{ij}\eta_j(t) + I_i(t)) \\ \eta_i(0) = 0 \end{cases} \quad (6)$$

where $i = 1, \dots, N$ and

$$k_{ij} = \gamma p_{ij} - \mu \quad (7)$$

$p_{ij} \geq 0$, $i, j = 1, \dots, N$ is the weight (interaction) function inside the layer X . γ, μ , the intensities of the excitatory and inhibitory connection, are positive parameters. $I_i(t)$ with $\langle I_i(t)I_j(t) \rangle = \delta_{ij}$, $i, j = 1, \dots, N$ is the input signal presented at neuron i of layer X . Note that the interaction kernel k_{ij} consists of short-range excitatory connections and global inhibitory connections of relative strength μ . We assume that p_{ij} depends only on $\|i - j\|$ and is a nonincreasing function of $\|i - j\|$ with $\sum_j p_{ij} = 1$, $i = 1, \dots, N$.

- A model mimicking the topological maps between the tectum and the retina [3, 10] (TM)

$$w_{i,j}(t+1) = f(w_{i,j}(t) + \alpha + w_{i,j}(t)(g_{i,j}(w(t)) - B_{i,j}(w(t))g(w(t))) - \beta) \quad (8)$$

where i, j represent a retinal cell and a tectal cell respectively and so $w_{i,j}$ is the synaptic connection between the i -th retinal cell and the j -th tectal cell, $(\alpha, \beta) \in \mathbb{R}^2$,

$$g_{i,j}(w(t)) := \sum_{k,k'} D_{i,k}^T w_{k,k'}(t) \bar{D}_{k',j}^R$$

for the propagation kernel D^T within the tectum, and propagation kernel D^R within the retina

$$\bar{D}_{k',j}^R := \sum_m D_{k',m}^R D_{m,j}^R$$

with D^T and D^R both decreasing function of the distance $\|i - j\|$,

$$B_{i,j}(w(t)g(w(t))) := [\sum_{k,m} w_{k,m}(t)g_{m,j}(w(t)) + \sum_{k,m} w_{i,m}(t)g_{m,k}(w(t))]/2(N+1).$$

For a detailed explanation and the biological background of the model we refer the reader to [10].

In fact many models proposed in neural networks have employed the (saturated) sigmoidal function in their dynamics. We restricted ourselves to these four model because the former three models which are all linear in the function g_i arise from three typical fields in neural networks: a more theoretical, physically oriented model (the Hopfield model), a biologically oriented model (Linsker's model) and a model oriented forwards practical application (the DLN); while the fourth model is nonlinear in function g_i .

2. Saturated Attractor Analysis on the Parameter Space

Our first observation for exposing the relation between the parameters and the set of saturated attractors of the dynamics (1) is quite straightforward:

1. Note that in the Hopfield model the stored memories take the value only in $\{-1, 1\}^N$ and so we are exclusively interested in those attractors of dynamics in the space $\{-1, 1\}^N$; in Linsker's model numerical simulations showed that only attractors in the space $\{-1, 1\}^N$ are observable (the background reason is that the Hebb learning rule ensures the dynamics either to increase or to decrease to the boundary); in DLN we utilize attractors in $\{-1, 1\}^N$ of inter-layer dynamics for the further dynamics in the intra-layer; in TM the topological map is a trivial one, $w_{ij} = 2\delta_{ij} - 1 \in \{-1, 1\}$, $i, j = 1, \dots, N$;
2. Our observations in 1 above allow us to focus on the stability of saturated states, i.e. states in $\{-1, 1\}^N$;
3. Note that for a saturated state to be a fixed point one must have $g_i(k_1, k_2, w) \geq 0$ if $w_i = 1$ and $g_i(k_1, k_2, w) \leq 0$ if $w_i = -1$, $i = 1, \dots, N$;
4. We exclude the (nongeneric) case $g_i(k_1, k_2, w) = 0$. With this exclusion, a fixed point is guaranteed to be stable by the continuity of g_i . Then a state $w \in \{-1, 1\}^N$ is an attractor of the dynamics (1) if and only if

$$w_i g_i(k_1, k_2, w) > 0, \quad i = 1, \dots, N$$

which, after a simple calculation and defining $J^+(w) = \{i, w_i = 1\}$, $J^-(w) = \{i, w_i = -1\}$, implies that

- In the Hopfield model

$$\begin{aligned} d_1(w) &:= -\min_{i \in J^+(w)} \sum_{\mu=1}^p \xi_i^\mu m(w, \xi^\mu) < \\ &< \theta + c(w)h < \\ &< d_2(w) := -\max_{i \in J^-(w)} \sum_{\mu=1}^p \xi_i^\mu m(w, \xi^\mu). \end{aligned} \tag{9}$$

where

$$m(w, \xi^\mu) := \frac{1}{N} \sum_{i=1}^N w_i \xi_i^\mu \quad (10)$$

is the overlap between the configuration w and the pattern ξ^μ , $c(w) := |J^+(w)| - |J^-(w)|$, $|\cdot|$ representing the cardinality of a set. Therefore, the parameter region of (θ, h) in which w is an attractor of the dynamics lies between two parallel lines: when $d_2(w) > d_1(w)$ the parameter region $\{(\theta, h) : d_1(w) < \theta + c(w)h < d_2(w)\}$ is a nonempty set; when $d_2(w) < d_1(w)$ it is an empty set. It is easily seen from the Figure 1 that the number of stored patterns, i.e, of saturated attractors, of the Hopfield model depends on the parameters (θ, h) . There is one region in which many saturated attractors coexist (see Figure 1). In this region the network will have the highest capacity, a quantity studied extensively in the literature [1]. Outside this region the capacity will become lower. When h , the external field, is negative there will be only one saturated attractor corresponding to the stored pattern if $c(\xi^\mu) \neq c(\xi^\nu)$, $\mu \neq \nu$, and so the capacity for the network is only $1/N$. However this region is good for retrieving a specific memory w if it is a saturated attractor of the dynamics, since if the dynamics (3) converges to a saturated attractor, it will go to w . This may also suggest an alternative way to recall an item of information avoiding the spurious states [2].

- In Linsker's model w is a saturated attractor of the model if and only if

$$d_1(w) < k_1 + c(w)k_2 < d_2(w), \quad (11)$$

where $c(w) = \sum_{j \in J^+(w)} r_j - \sum_{j \in J^-(w)} r_j$,

$$d_1(w) = \max_{i \in J^+(w)} [\sum_{j \in J^-(w)} q_{ij} r_j - \sum_{j \in J^+(w)} q_{ij} r_j]$$

if $J^+(w) \neq \emptyset$, and $d_1(w) = -\infty$ otherwise;

$$d_2(w) = \min_{j \in J^-(w)} [\sum_{i \in J^-(w)} q_{ij} r_j - \sum_{j \in J^+(w)} q_{ij} r_j]$$

if $J^-(w) \neq \emptyset$, and $d_2(w) = \infty$ otherwise.

- In the DLN, w is an attractor of the dynamics (5) if and only if

$$\gamma e_1(w) + 1 < I(\nu) - c(w)\mu < \gamma e_2(w) - 1$$

where $c(w) = |J^+(w)| - |J^-(w)|$

$$e_2(w) = [1 - 2 \max_{i \in J^-(w)} \sum_{j \in J^+(w)} p_{ij}]$$

$$e_1(w) = [1 - 2 \min_{i \in J^+(w)} \sum_{j \in J^+(w)} p_{ij}] \quad (12)$$

- In the TM, the situation is more complicated than the three models discussed above. In order to get an exact expression to check in which parameter region of (α, β) a given saturated state is an attractor of the model more effort is needed. We refer the reader to [3] for details.

In [6, 7] we have fully reported our results on the problem of determining the critical value of different parameters utilized in Linsker's model which is tractable

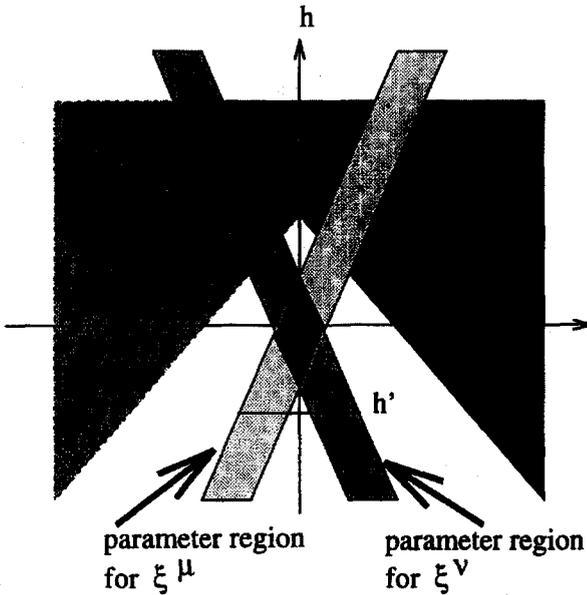


Figure 1. The parameter region of (θ, h) in which w is a saturated attractor of the Hopfield model lies between two parallel lines except for the state $(1, \dots, 1)$ and $(-1, \dots, -1)$ (triangle filled parts). In the darkest region, the Hopfield model has the highest capacity. In this region, for example, ξ^μ, ξ^ν are both attractors of the Hopfield model. When $h = h'$ (horizontal line), the capacity of the model becomes lower.

only after we have the results above. [9] contained our findings on choosing five parameters which are useful in practical application of the DLN. In [3] the existence of a parameter region such that only the topological map is an attractor of the TM model was proved after a modification of the original model. Although here we consider more general models than those in [3, 5–7, 9] where the function $f(x) = x$ for $-1 < x < 1$ was utilized in their dynamics all analyses in [3, 6, 7, 9] are essentially valid, and we thus refer the reader to these papers for details.

3. Capacity of the Hopfield Model

As a further demonstration of the power of our approach in the previous section we apply it to the Hopfield model and calculate the capacity of the Hopfield model (see also [1]).

First let us discuss the physical meaning of $d_2(w)$ and $d_1(w)$ defined in Equation (9). These two intercept functions d_2 and d_1 were mathematically introduced early in [5]. However, their physical meaning can be understood only after we apply

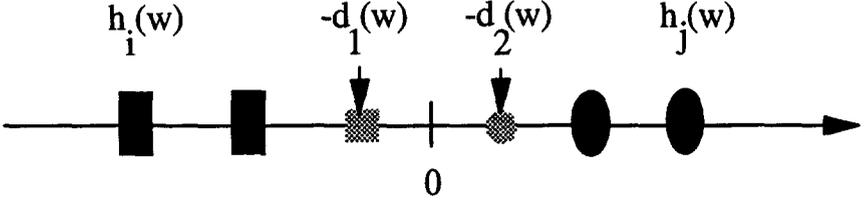


Figure 2. An explanation of the existence of the local field gap corresponding to w . w is an attractor of the Hopfield model if and only if there exists a local field gap between the neurons in $i \in J^+(w)$ and $j \in J^-(w)$.

the saturated attractor analysis on the parameter space to the Hopfield model. Considering the local field of each neuron defined by

$$\begin{aligned} h_i(w) &:= \sum_{j=1}^N T_{ij} w_j \\ &= \sum_{j \in J^+(w)} T_{ij} - \sum_{j \in J^-(w)} T_{ij} \end{aligned}$$

we see that $d_2(w) > d_1(w)$ if and only if

$$\begin{aligned} -d_1(w) &= \min_{i \in J^+(w)} [\sum_{j \in J^+(w)} T_{ij} - \sum_{j \in J^-(w)} T_{ij}] > \\ -d_2(w) &= \max_{i \in J^-(w)} [\sum_{j \in J^+(w)} T_{ij} - \sum_{j \in J^-(w)} T_{ij}]. \end{aligned}$$

or equivalently if and only if there exists a *local field gap* between the neurons in $J^+(w)$ and $J^-(w)$ (see Figure 2).

As we already discussed before, the difference (local field gap) $d_2(w) - d_1(w)$ reflects the stability of a saturated attractor w . If it is negative or equal to zero, w will no longer be a saturated attractor of the dynamics (3). In other words, the existence of a local field gap between the neurons in $J^+(w)$ and $J^-(w)$ is necessary and sufficient for w to be an attractor of the Hopfield network. From this point of view here we are also able to give a definition of the critical capacity of the Hopfield model in terms of the intercept functions d_1 and d_2 . Assume that $\xi_i^\mu, i = 1, \dots, N, \mu = 1, \dots, p$ are i.i.d. random variables with $P(\xi_i^\mu = \pm 1) = 1/2$.

Definition 1 The critical number of patterns p_c for perfect retrieval of the dynamics (3) is $p_c := \inf\{p, \langle d_2(\xi^\mu) \rangle - \langle d_1(\xi^\mu) \rangle = 0 \text{ for any } \mu = 1, \dots, p\}$ where $\langle \cdot \rangle$ represents the expectation with respect to the distribution P of ξ^μ .

Since $\langle d_1(\xi^\mu) \rangle$ and $\langle d_2(\xi^\mu) \rangle$ are symmetric with respect to μ under the condition that matrix T is given by (5), we only need to compute

$$\langle d_2(\xi^1) \rangle - \langle d_1(\xi^1) \rangle \tag{13}$$

for estimating the critical number of patterns of the network. Furthermore in terms of the symmetries between d_2 and d_1 we see that $\langle d_2(\xi^1) \rangle > \langle d_1(\xi^1) \rangle$ if and only

if $\langle d_1(\xi^1) \rangle < 0$.

Theorem 1 *If and only if $p < N/(2 \ln N)$ we have $\langle d_1(\xi^1) \rangle < 0$*

For a detailed proof of Theorem 1 we refer the reader to our full paper [4]. The idea is to use the properties of the N -th order statistics developed in [13].

Now we go a step further to consider the parameter region in which the Hopfield model has the capacity as in Theorem 1. In order to make sense of the inequality in Equation (9) as $N \rightarrow \infty$ we only consider the parameter region of θ since by the law of iterated logarithm we know that $\limsup_{N \rightarrow \infty} c(\xi^1)/(\sqrt{N \log \log N}) = +1$ and $\liminf_{N \rightarrow \infty} c(\xi^1)/(\sqrt{N \log \log N}) = -1$. Theorem 1 tells that when $\langle d_1(\xi^1) \rangle < \theta < \langle d_2(\xi^1) \rangle$ the capacity for the network is $p = N/(2 \log N)$. For a given $p(N)$ we could easily decide the exact parameter region of θ in which the network has a capacity $p(N)/N$, while when θ is not in the region $[\langle d_1(\xi^1) \rangle, \langle d_2(\xi^1) \rangle]$ the capacity is zero. For example let $p(N) = N/(4 \log N)$. When $-1/2 < \theta < 1/2$ all patterns are stable, but when $\theta \notin [-1/2, 1/2]$ all patterns are no longer stable.

Finally, we want to point out that our approach to the Hopfield model is independent of the symmetry of the matrix Q and so we could calculate its capacity in a more general context [4].

Acknowledgement

This paper was partially supported by the A. von Humboldt Foundation of Germany.

References

1. D. Amit, *Modeling Brain Function*, Cambridge University Press: Cambridge, UK, 1989.
2. S. Albeverio, J. Feng and M. Qian, "Role of noises in neural networks", *Phys. Rev. E.*, Vol. 52, pp. 6593–6606, 1995.
3. J. Feng, "Establishment of topological maps – a model study", *Neural Processing Letters*, Vol. 2, pp. 1–4, 1995.
4. J. Feng and B. Tirozzi, "An analysis on neural dynamics with saturated sigmoidal functions", *Computers and Mathematics with Applications*, 1997 (in press).
5. J. Feng and H. Pan, "Analysis of Linsker-type Hebbian learning: rigorous results", in *Proc. 1993 IEEE Int. Conf. on Neural Networks – San Francisco*, Vol. III, pp. 1516–1521, IEEE: Piscataway, NJ, 1993.
6. J. Feng, H. Pan and V.P. Roychowdhury, "A rigorous analysis of Linsker-type Hebbian learning", in G. Tesauro, D.S. Touretzky and T.K. Leen (eds) *Advances in Neural Information Processing Systems 7*, pp. 319–326, MIT Press: Cambridge, MA, 1995.
7. J. Feng, H. Pan and V.P. Roychowdhury, "On neurodynamics with limiter function and Linsker's developmental model", *Neural Computation*, Vol. 8, pp. 1003–1019, 1996.
8. J. Feng and B. Tirozzi, "The SLLN for the free-energy of the Hopfield and spin glass model", *Helvetica Physica Acta*, Vol. 68, pp. 365–379, 1995.
9. J. Feng and B. Tirozzi, "On choosing the parameters in the dynamic link network", in M. Marinaro and R. Tagliaferri (eds) *Neural Nets, WIRN VIETRI-95*, pp. 245–250, World Scientific: Singapore, 1995.
10. A.F. Häussler and C. von der Malsburg, "Development of retinotopic projections. An analytical treatment", *Journal of Theoretical Neurobiology*, Vol. 2, pp. 47–73, 1983.
11. W. Konen, T. Maurer and C. von der Malsburg, "A fast dynamic link matching algorithm for invariant pattern recognition", *Neural Network*, Vol. 7, Nos. 6–7, pp. 1019–1030, 1994.

12. R. Linsker, "From basic network principle to neural architecture (series)". Proc. Natl. Acad. Sci. USA, Vol. 83, pp. 7508–7512, 8390–8394, 8779–8783, 1986.
13. M.R. Leadbetter, G. Lindgren and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag: New York, Heidelberg, Berlin, 1983.