Pergamon

Computers Math. Applic. Vol. 34, No. 10, pp. 71–99, 1997

0898-1221/97 $17.00 + 0.00

PII: S0898-1221(97)00208-3

An Analysis on Neural Dynamics with Saturated Sigmoidal Functions*

J. FENG†
Statistics Group, The Babraham Institute
Cambridge CB2 4AT, United Kingdom
and
Mathematisches Institut, Universität München
D-80333 München, Germany

B. TIROZZI
Mathematical Department, University of Rome "La Sapienza"
P.le A. Moro, 00185 Rome, Italy

(Received October 1996; accepted November 1996)

Abstract—We propose a unified approach to study the relation between the set of saturated attractors and the set of system parameters of the Hopfield model, Linsker's model, and the dynamic link network (DLN), which use saturated sigmoidal functions in its dynamics of the state or weight. The key point for this approach is to rigorously derive a necessary and sufficient condition to test whether a given saturated state (in the Hopfield model) or weight vector (in Linsker's model and the DLN) is stable or not for any given set of system parameters, and used this to determine the complete regime in the parameter space over which the given state or weight is stable. Our approach allows us to give an exact characterization between the parameters and the capacity in the Hopfield model; to generalize our previous results on Linsker's network and the DLN; to have a better understanding of the underlying mechanism among these models. The method reported here could be adopted to analyze a variety of models in the field of the neural networks.

Keywords—Saturated attractor, Saturated sigmoidal function, Hopfield model, Linsker's network, Dynamic link network.

1. INTRODUCTION

The past decade has seen an explosive growth in studies of neural networks, the theory underlying learning and computing in networks has developed into a mature subfield existing somewhere between mathematics, physics, computer science, and neurobiology. In part, this was the result of many deep and interesting theoretical exposition in physics and mathematics, for example, the application of the spin glass theory to the Hopfield model allows us to understand clearly the phase transition from the retrieval to nonretrieval state [1–4]. Another major impulse was provided by the successful explanation of some biological phenomena, at least in a primitive level, for example, Linsker's model mimics the ontogenesis development of the primary visual system [5]. Of course, the most important impulse comes from the learning techniques successfully applied to

One of the authors (F.J.) would like to express his thanks to S. Albeverio, K. D. Hadeler, and H. Pan for their fruitful cooperation on different models.
*Partially supported by the CNR of Italy.
†Partially supported by the A. V. Humboldt Foundation of Germany.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX some practical problems which were traditionally thought of as some of the hardest problems in the AI. One of the recent example of such an application is the face recognition using the dynamic link network (DLN), a model proposed by von der Malsburg first in 1981 [6]. However, at this moment, the theoretical treatment of these models is obviously far away from being satisfactory, mainly due to the lack of theoretical tools to deal with the nonlinearity exploited in most of the models reported today. The dynamic behavior of these models is determined by the underlying nonlinear dynamics that are parameterized by a set of parameters. The difficulties lie in both determining the set of terminal attractors, as well as in characterizing their basins of attraction in the weight space (for learning models) or the state space (for retrieval models).

The purpose of this paper is to gain more insights into the dynamical mechanism of these models by performing a rigorous analysis on their parameter space without approximation which is a further development of our previous work on Linsker's model [7], the DLN [8,9], and a model mimicking the development of the topological map between the tectum and the retinal [10]. We present a unified theoretical framework for studying dynamic properties of the Hopfield model, Linsker's model, and the DLN: to derive a necessary and sufficient condition to test whether a given saturated state (in the Hopfield model) or weight vector (in Linsker's model and the DLN) is stable or not for any given set of system parameters, and used this to determine the complete regime in the parameter space over which the given state is stable.

In particular, our approach allows us to reformulate some problems reported in the literature for the Hopfield model and gives some more exact characterization of them. A concrete criterion to check whether a stored pattern is an attractor of the network is given. The capacity, a quantity which plays a central role in the spin glass approach to the Hopfield model is naturally introduced and calculated here. One advantage of the present approach is that we do not impose the restriction of the symmetry of the connection matrix. Our results also reveal the role of different parameters in the Hopfield model.

We consider Linsker's model with a saturated sigmoidal function in the updating dynamics of its synaptic connections (a definition of a saturated sigmoidal function is in Section 2). All conclusions in [7] are reobtained, where the limiter function, a special case of the saturated sigmoidal function, and so a special case of the present paper is used for the development of the synaptic connections. The present paper tells that in a certain parameter region the potential for an appearance of a structured receptive fields is independent of the specific choice of the limiter function, which is an important, and necessary, aspect of a reasonable biological oriented model. Furthermore, we also take into account on the reason for the appearance of the oriented receptive field in the further layers of Linsker's network.

For the DLN, a principle for choosing all five parameter employed in the model is furnished which is crucial in the application of DLN in the face recognition and confirms our previous claim that all results contained in [8] for the limiter function are true for a more general class of function, i.e., for the sigmoidal function.

Although, here we confine ourselves to the models on which we worked before [2,4,7,11–13], the essential part of our approach here is to analyze the dynamics with the saturated sigmoidal function and it is possible to adopt our method here to analyze other models in the field of neural networks as well. Some further progress is achieved already, see, for example, [10] where we consider a more complex dynamics than here, which marks important new dimensions into which our approach can grow.

A brief report of the present paper is appeared in [14].

2. GENERAL MODELS AND NOTATION

For a given positive integer N, an $N \times N$ matrix $Q = (q_{ij}, i, j = 1, \dots, N)$ and an N-dimensional vector $r = (r_i, i = 1, \dots, N)$, consider the following (synchronous) dynamics:

$$w_i(\tau+1) = f\left(w_i(\tau) + k_1 + \sum_{j=1}^{N} \left[(q_{ij}+k_2)\, r_j w_j(\tau)\right]\right), \tag{1}$$

where $\tau = 1, 2, \dots$ is the discrete time, $w(\tau) = (w_i(\tau), i = 1, \dots, N) \in \mathbb{R}^N$, (k_1, k_2) are two parameters of the dynamics, and f is a continuous function defined on $\mathbb{R}^1$ satisfying

(f1) $f(x) = 1$ if $x \geq 1$, $f(x) = -1$ if $x \leq -1$,

(f2) $f(x)$ is a strictly increasing and continuous function for $x \in [-1, 1]$ and $f(0) = 0$.

We call a function f with Properties (f1) and (f2) a *saturated sigmoidal function*. Furthermore, if

(f3) $f(x) \geq x$ for $x \in (0, 1]$ and $f(x) \leq x$ for $x \in [-1, 0)$.

We call f a dissipative saturated sigmoidal function.

Note that for the sigmoidal function with range between -1 and 1

$$\sigma_\beta(x) = \frac{2}{1+\exp(-\beta x)} - 1, \tag{2}$$

both Conditions (f1) and (f2) are approximately satisfied when β is large. For example, when $\beta = 5$, we have $\sigma_\beta(1) = 0.9866 \sim 1$ and $\sigma_\beta(-1) = -0.9866 \sim -1$. It is reasonable to expect that in numerical simulations both (f1) and (f2) are true for the sigmoidal function given by (2) with large β. Due to this reason, we expect that our results on dynamics (1) with the saturated sigmoidal function below reflect the exact properties of the dynamics (1) mostly observed in numerical simulations with $f = \sigma_\beta$, β large.

The function termed as *limiter function* (or piecewise linear) and utilized in the dynamics of the development of the synaptic connection in Linsker's network is defined by $f_c(x) = x$ if $|x| < 1$, and $f_c(x) = 1$ if $x > 1$, $f_c(x) = -1$ if $x < -1$, which of course fulfills (f1), (f2), and (f3) [5,7].

In the DLN, the fast DLN or the discrete version of it, the function f adopted for their dynamics is either the limiter function or the sigmoidal function [6,8,9].

REMARK 1. The condition on the range of the function f, i.e., (f1) is not an essential restriction. In fact, all results below could be easily generalized to the case $a < f < b$ for $a, b \in \mathbb{R}$.

2.1. Equivalence of Two Dynamics

The dynamics (1) defined on $[-1, 1]^N$ is equivalent to the following commonly used dynamics defined on $\mathbb{R}^N$:

$$v_i(\tau+1) = k_1 + \sum_{j=1}^{N} \left[(\bar{q}_{ij}+k_2)\, r_j f(v_j(\tau))\right], \tag{3}$$

where $\bar{q}_{ij} = q_{ij} + \delta_{ij}/r_i$, $i, j = 1, \dots, N$ if $r_i \neq 0$, $i = 1, \dots, N$. We show this equivalence. Let $w(\tau)$ be given according to dynamics (1), define

$$v_i(\tau) = w_i(\tau) + k_1 + \sum_{j=1}^{N} \left[(q_{ij}+k_2)\, r_j w_j(\tau)\right]. \tag{4}$$

After multiplying the quantity $(\bar{q}_{ij} + k_2) r_j$ on both sides of (1) and taking the summation on j, we easily obtain dynamics (3). To recover from v_i to $w_i, i = 1, \dots, N$. Let $w_i(\tau) = f(v_i(\tau))$, $i = 1, \dots, N$, acting f on both sides of (3), we obtain dynamics (1).

Our arguments above implies that all conclusions below for dynamics (1) are true for dynamics (3).

2.2. Lyapunov Functions

We can associate another dynamics, the asynchronous dynamics to the neural network defined by the parameters Q, f, r, N beside the synchronous dynamics. An asynchronous dynamics is a composition of two dynamics: first we selects a neuron i from $(1, \dots, N)$ with probability $p_i > 0$, $\sum_i p_i = 1$ and the state $x_i(\tau)$ of the i^{th} neuron is updated to the new state according to

$$x_i(\tau + 1) = f\left(\sum_{j=1}^{N} a_{ij} r_j x_j(\tau) + b_i\right),$$

but keep all other states unchanged, i.e., $x_j(\tau + 1) = x_j(\tau)$, $j \neq i$, where $a_{ij} = q_{ij} + k_2 + \delta_{ij}/r_i$ and $b_i = k_1$. So $x(\tau)$ is a stochastic process (a Markov chain).

For the asynchronous and synchronous dynamics, we are able to define a Lyapunov function for them under certain restrictions. Here, we state only the results and for a detailed proof we refer the reader to [13].

THEOREM 1. *Suppose that the matrix $A = \{a_{ij}\}$ is symmetric.*

(1) *Define*

$$L(x(\tau)) = \sum_{j=1}^{N} \int_0^{x_j(\tau)} r_j f^{-1}(y)\, dy - \frac{1}{2} \sum_{j,i=1}^{N} a_{ji} x_j(\tau) x_i(\tau) r_j r_i - \sum_{j=1}^{N} r_j b_j x_j(\tau),$$

then $L(x(\tau))$ is a Lyapunov function (supermartingale) if $a_{ii} \geq 0$, $i = 1, \dots, N$.

(2) *The function*

$$\begin{aligned} V(w(t)) = &- \sum a_{ij} r_i r_j w_i(\tau) w_j(\tau + 1) - \sum r_i b_i \left(w_i(\tau) + w_i(\tau + 1)\right) \\ &+ \sum_i \int_0^{w_i(\tau)} r_i f^{-1}(u)\, du + \sum_i \int_0^{w_i(\tau+1)} r_i f^{-1}(u)\, du \end{aligned}$$

is a Lyapunov function of the synchronous dynamics.

It is worthwhile to point out that the difficulty to prove the conclusions above lies in the fact that the function f is not differentiable, which forces us to apply the Legendre-Fenchel transformation rather than the Taylor expansion in the proof. Theorem 1 indicates that there are differences between the asynchronous and the synchronous dynamics. In the circumstances of Theorem 1, there are only fixed point attractors for the asynchronous dynamics, while there are two-state limit cycle attractors for the synchronous dynamics. In the following, we are going to study the set of saturated fixed point attractors of both dynamics. Since the set of fixed point attractors for both dynamics are common, it is only necessary for us to concentrate on one of the dynamics. We will focus on the synchronous dynamics. Of course we can define more complex dynamics for a given network, for example, the dynamics called distributed dynamics [15,16]. It will be obvious soon that our arguments in the present paper can be applied to the distributed dynamics without essential difficulties.

2.3. Notation

Let us now introduce three functions which will play a crucial role in our later development. Let w be a given configuration in $\{-1, 1\}^N$, then

$$J^+(w) = \{i, w_i = 1\}, \qquad J^-(w) = \{i, w_i = -1\} \tag{5}$$

are, respectively, the set of all sites with $w_i = 1$ and all sites with $w_i = -1$.

First, we are going to introduce the *slope function* $c(w)$ on $\{-1,1\}^N$ defined by

$$c(w) = \sum_{i \in J^+(w)} r_j - \sum_{i \in J^-(w)} r_j. \tag{6}$$

Note that if $r_i = 1$, then $c(w) = |J^+(w)| - |J^-(w)|$ is the difference of the number of the sites with $w_i = 1$ and $w_i = -1$.

The second and the third one are the two *intercept functions* $d_1(w)$ and $d_2(w)$ defined on $\{-1,1\}^N$, which are given by

$$d_1(w) = \begin{cases} \max_{i \in J^+(w)} \left[\sum_{j \in J^-(w)} q_{ij} r_j - \sum_{j \in J^+(w)} q_{ij} r_j \right], & \text{if } J^+(w) \neq \phi, \\ -\infty, & \text{otherwise}, \end{cases} \tag{7}$$

and

$$d_2(w) = \begin{cases} \min_{i \in J^-(w)} \left[\sum_{j \in J^-(w)} q_{ij} r_j - \sum_{j \in J^+(w)} q_{ij} r_j \right], & \text{if } J^-(w) \neq \phi, \\ \infty, & \text{otherwise}, \end{cases} \tag{8}$$

respectively.

The reason why we call them as slope function and intercept functions will be clear after Theorem 2.

First, let us have a discussion of the physical meaning for $d_2(w)$ and $d_1(w)$. These two intercept functions d_2 and d_1 were mathematically introduced in [7], however, the physical meaning of them can be understood only after we apply the saturated attractor analysis on the parameter space to the Hopfield model. Considering the local field of each neuron defined by

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

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

$$\min_{i \in J^+(w)} \left[\sum_{j \in J^+(w)} T_{ij} - \sum_{j \in J^-(w)} T_{ij} \right] > \max_{i \in J^-(w)} \left[\sum_{j \in J^+(w)} T_{ij} - \sum_{j \in J^-(w)} T_{ij} \right],$$

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

3. THE SET OF ALL SATURATED ATTRACTORS

The set of all fixed points of the dynamics (1) is

$$FP = \left\{ w; w_i = f\left(w_i + \sum_{j=1}^{N} (q_{ij} + k_2) r_j w_j + k_1 \right), i = 1, \dots, N \right\}. \tag{9}$$

From the compactness of the range of the function f and the continuity of f, we conclude that the set (12) is nonempty by the Brouwer's fixed point theorem which states that if $F = (f, \dots, f)$ is a continuous mapping from a compact convex set (here is $[-1,1]^N$) to itself, then the set of fixed points of the mapping F is nonempty.

A fixed point is called an attractor if it is a stable fixed point. We will confine ourselves to a subset of all attractors in $\{-1,1\}^N$, which is general enough in most of applications (see, Sections 4–6).

DEFINITION 1. *A configuration in the set* $\{-1,1\}^N$ *is called a saturated attractor of dynamics (1) if* $\exists$, *a nonempty neighborhood* $B(w)$ *of* w *in* $[-1,1]^N$ *such that* $\lim_{\tau\to\infty} w(\tau) = w$ *for* $w(0) \in B(w)$ *and* $k_1 + \sum_{j=1}^N (q_{ij} + k_2) r_j w_j \neq 0, \forall i = 1,\dots,N$.

Now we show the general theorem of this paper. The main idea of its proof is fairly direct. Let us consider the dynamics

$$w_i(\tau+1) = \operatorname{sign}\left(\sum_{j=1}^N (q_{ij} + k_2)\, r_j w_j(\tau) + k_1\right), \qquad i = 1,\dots,N \tag{10}$$

for $w \in \{-1,1\}^N$, which is the dynamics of the discrete version of the Hopfield model. The set of all fixed points of the dynamics above is

$$\left\{ w, w_i \cdot \left[\sum_{j=1}^N (q_{ij} + k_2)\, r_j w_j + k_1\right] > 0,\ i = 1,\dots,N \right\}. \tag{11}$$

Namely, if and only if w satisfies the condition

$$w_i \cdot \left[\sum_{j=1}^N (q_{ij} + k_2)\, r_j w_j + k_1\right] > 0, \qquad i = 1,\dots,N, \tag{12}$$

w is a fixed point of dynamics (10). The condition (12) reads

$$\sum_{j=1}^N (q_{ij} + k_2) r_j w_j + k_1 > 0, \tag{13}$$

if $i \in J^+(w)$ and

$$\sum_{j=1}^N (q_{ij} + k_2) r_j w_j + k_1 < 0, \tag{14}$$

if $i \in J^-(w)$. Or equivalently,

$$k_2 c(w) + k_1 > \left[\sum_{j\in J^-(w)} q_{ij} r_j - \sum_{j\in J^+(w)} q_{ij} r_j\right], \tag{15}$$

if $i \in J^+(w)$ and

$$k_2 c(w) + k_1 < \left[\sum_{j\in J^-(w)} q_{ij} r_j - \sum_{j\in J^+(w)} q_{ij} r_j\right], \tag{16}$$

if $i \in J^-(w)$. By noting that the left-hand of the above two inequalities is independent of i, taking the maximum for inequality (19) on the set $J^+(w)$, and taking the minimum for inequality (20) on the set $J^-(w)$, we see that the necessary condition for w to be a fixed point of dynamics (10) is that

$$d_2(w) > k_1 + k_2 c(w) > d_1(w). \tag{17}$$

After reversing the above procedure, we see that this condition is also sufficient.

The following theorem establishes that for dynamics (1), the condition (17) is strong enough to ensure that w is an attractor of the dynamics while this fact does not hold for dynamics (10), where we are only able to assert that it is a fixed point. We call an attractor of a dynamics a dissipative attractor [17] if $\lim w(\tau) = w$ implies there exists a finite time $T > 0$ such that $w(\tau + T) = w$, $\forall \tau > 0$.

THEOREM 2. *If f is a saturated sigmoidal function, then w is a saturated attractor of dynamics (1) if and only if*

$$d_1(w) < k_1 + c(w)k_2 < d_2(w). \tag{18}$$

Furthermore, if f is a dissipative saturated sigmoidal function, then w is a dissipative saturated attractor of dynamics (1).

PROOF. Define a family of functions

$$g_i(x) := x_i \cdot \left[\sum_{j=1}^{N} (q_{ij} + k_2)\, r_j x_j + k_1 \right], \tag{19}$$

for $x \in \mathbb{R}^N$. Then we assert that $g_i(w) > 0$, $i = 1, \ldots, N$. In fact, if $\exists\, i$ with $g_i(w) = 0$, from the definition of the saturated attractor we know that $w_i = 0$. This contradicts our assumption on the function f, i.e.,

$$0 = w_i = f\left(w_i + \sum_{j=1}^{N} (q_{ij} + k_2)\, r_j w_j + k_1 \right) \neq 0. \tag{20}$$

Thus, $g_i(w) \neq 0$ for $i = 1, \ldots, N$.

If $\exists\, i$ such that $g_i(w) < 0$, without loss of generality, we assume that $w_i > 0$ and $\sum_{j=1}^{N}(q_{ij} + k_2)r_j w_j + k_1 < 0$. From the strictly increasing property of the function f, we deduce that

$$w_i = f\left(w_i + \sum_{j=1}^{N} (q_{ij} + k_2)\, r_j w_j + k_1 \right) < f(w_i) \leq 1. \tag{21}$$

Hence, $g_i(w) > 0$ for all $i = 1, \ldots, N$ follows.

"ONLY IF". If w is a saturated attractor of the dynamics, we know from the proof alluded to above that

$$w_i \cdot \left(\sum_{j=1}^{N} (q_{ij} + k_2)\, r_j w_j + k_1 \right) > 0, \qquad \forall i = 1, \ldots, N. \tag{22}$$

So if $i \in J^+(w)$, the above inequality reads

$$\sum_{j=1}^{N} (q_{ij} + k_2) r_j w_j + k_1 > 0, \tag{23}$$

or equivalently,

$$k_1 + c(w)k_2 > \sum_{j \in J^+(w)} q_{ij} r_j - \sum_{j \in J^-(w)} q_{ij} r_j. \tag{24}$$

By noticing that the left-hand side of the inequality above is independent of i, after taking the maximum for $i \in J^+(w)$ on both sides of inequality above, we have that

$$k_1 + c(w)k_2 > d_1(w). \tag{25}$$

After repeating same argument above, we arrive at that

$$k_1 + c(w)k_2 < d_2(w). \tag{26}$$

"IF". After reversing the arguments in the "Only if" part, we conclude that if w satisfies that

$$d_1(w) < k_1 + c(w)d_2 < d_2(w), \tag{27}$$

then w is a fixed point of the dynamics. Since this condition implies that there is a neighborhood δ' of w such that $g_i(x) > 0$ if $x \in \delta'$, $i = 1, \dots, N$. We get, after making the same procedure as above, that w is an attractor.

Under the assumption that f is dissipative from the continuity property of the function $g_i(x)$, we deduce that for each i there is a nonempty neighborhood $\delta_i(w)$ of w such that $g_i(x) > 0$ as $x \in \delta_i(w)$.

Let $\delta = \cap_{i=1}^N \delta_i(w)$, it is again a nonempty open set since $w \in \delta$. From the assumption of the existence of the limit we see that $\exists T_0 > 0$ such that as $\tau > T_0$, $w(\tau) \in \delta$. Since $w_i \neq 0$ for $i = 1, \dots, N$, we could suppose that as $x \in \delta$, x_i is either definite positive or negative. Now without loss of generality, we suppose that $x_i > 0$ for $x \in \delta$, and so

$$b_i = \inf_{x \in \delta} \left[\sum_{j=1}^N (q_{ij} + k_2) r_j x_j + k_1 \right] > 0, \tag{28}$$

which implies that

$$\begin{aligned} w_i(T_0 + T_i) &= f\left(w_i(T_i + T_0 - 1) + \sum_{j=1}^N (q_{ij} + k_2) r_j w_j (T_i + T_0 - 1) + k_1 \right) \\ &\geq f\left(w_i(T_i + T_0 - 1) + b_i \right) \\ &\geq f\left(w_i(T_i + T_0 - 2) + \left(\sum_{j=1}^N (q_{ij} + k_2) r_j w_j\, (T_i + T_0 - 2) + k_1 \right) + b_i \right) \\ &\ \ \vdots \\ &\geq f\left(w_i(T_0) + T_i b_i \right) = 1, \end{aligned} \tag{29}$$

if $T_i \cdot b_i > 1$, the first inequality follows from (f2). Set $T = T_0 + \max_i T_i$, this proves our conclusions of the theorem.

For a given configuration w, Theorem 2 tells that the parameter region in which w is a saturated attractor of dynamics (1) lies between the two parallel lines (see Figure 1)

$$k_1 + k_2 c(w) = d_1(w) \tag{30}$$

and

$$k_1 + k_2 c(w) = d_2(w). \tag{31}$$

Hence, $c(w)$ is the slope function of lines (30) and (31), and d_1, d_2 are the two intercept functions.

If $d_2(w) > d_1(w)$, which means there exists an local field gap between the neuron in $J^+(w)$ and $J^-(w)$, the parameter region

$$\{\Gamma(w) := (k_1, k_2) \text{ in which } w \text{ is a saturated attractor of dynamics (1)}\}$$

is a nonempty set. If $d_2(w) < d_1(w)$, then $\Gamma(w)$ is an empty set. So in this sense the larger is the difference between $d_2(w)$ and $d_1(w)$, the more stable is the attractor w.

We are in the position to say a few words about Definition 1. One may suggests that the definition of the saturated attractors should include those attractors such that there exists i, $i = 1, \dots, N$ with the equality

$$k_1 + \sum_{j=1}^N (q_{ij} + k_2) r_j w_j = 0, \qquad w \in \{-1, 1\}^N. \tag{32}$$

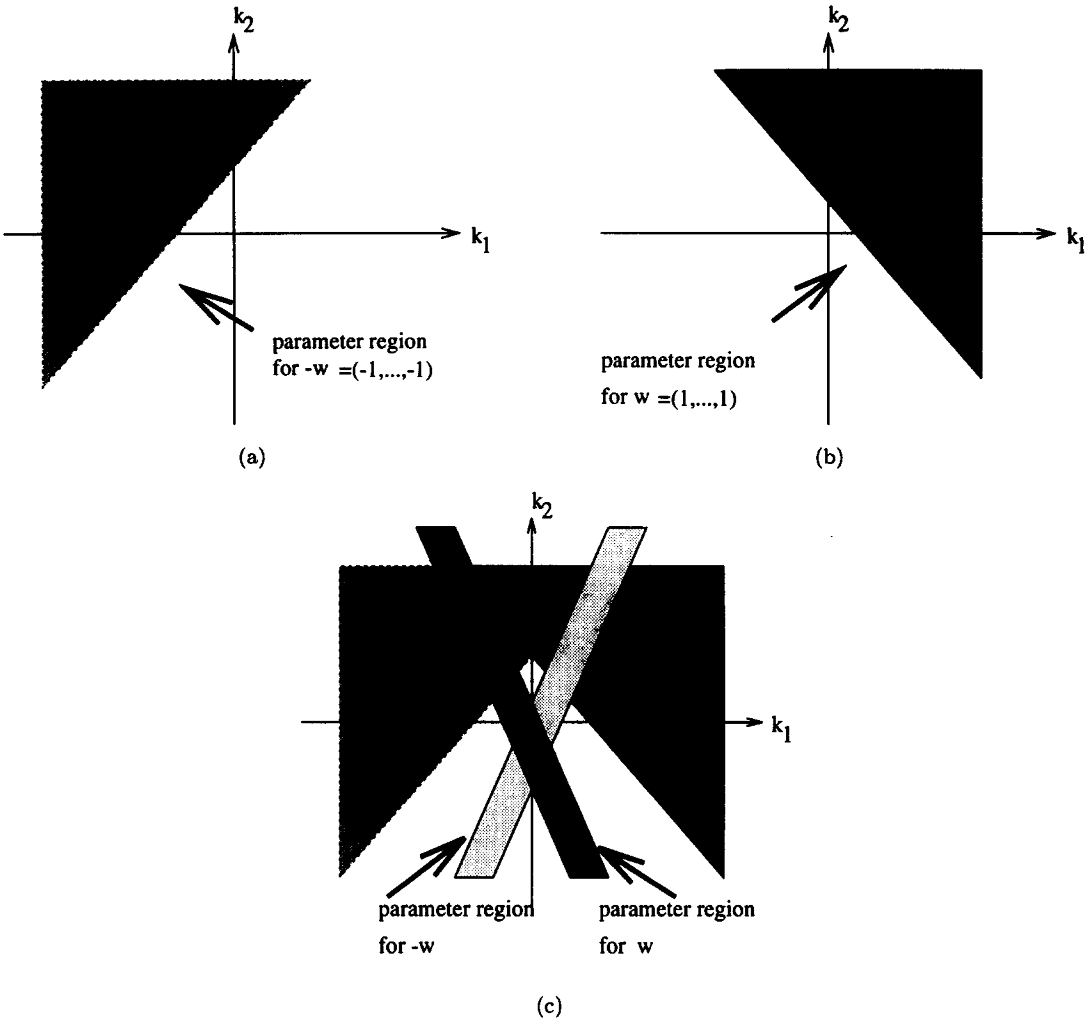

Figure 1. The parameter region of different saturated attractor of dynamics (1).

However, if we look at the parameter space of (k_1, k_2), the Lebesgue measure of the set of parameters (k_1, k_2) satisfying equation (32) is zero (union of finitely many lines). Hence, there is no loss of generality if we consider only the saturated attractors of Definition 1.

COROLLARY 1. *(See Figure 1.)*

(1) *The parameter region of (k_1, k_2) in which $(1, \dots, 1)$ is a saturated attractor of dynamics (1) is*

$$k_1 + \sum_j r_j k_2 > d(+) := - \min_{i=1,\dots,N} \sum_{j=1}^{N} q_{ij} r_j. \tag{33}$$

(2) *The parameter region of (k_1, k_2) in which $(-1, \dots, -1)$ is a saturated attractor of dynamics (1) is*

$$k_1 - \sum_j r_j k_2 < d(-) := \min_{i=1,\dots,N} \sum_{j=1}^{N} q_{ij} r_j. \tag{34}$$

(3) *If q_{ij} depends only on j, then only the configuration $(1, \dots, 1)$ and $(-1, \dots, -1)$ are saturated attractors of dynamics (1).*

(4) *If $q_{ij} = \delta_{ij}$, and $\min\{r_j, j = 1, \dots, N\} > 0$, then any configuration $w \in \{-1, 1\}^N$ is a saturated attractor of dynamics (1).*

PROOF. Conditions 1 and 2 of Corallary 1 are direct consequences of Theorem 2.

Under Condition 3, we see that $d_2(w) = d_1(w)$ if $w \neq (1, \dots, 1), (-1, \dots, -1)$. Condition 4 in this case $d_2(w) = \min\{r_j, j = 1, \dots, N\}$ and $d_1(w) = -\min\{r_j, j = 1, \dots, N\}$, for $w \in \{-1, 1\}^N$.

If $w = (w_i, i = 1, \dots, N)$ is a saturated attractor of dynamics (1), we may ask if $-w = (-w_i, i = 1, \dots, N)$ is also a saturated attractor of dynamics (1). The following proposition gives an answer.

PROPOSITION 1. *w is a saturated attractor of dynamics (1) if and only if $-w$ is a saturated attractor of dynamics (1) and*

$$c(w) = -c(-w), \qquad d_2(w) = -d_1(-w), \qquad d_1(w) = -d_2(-w).$$

PROOF. The relationship between $c(w)$ and $c(-w)$ is an obvious one. For the equality between $d_2(w)$ and $d_1(-w)$, we note

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

Similarly, we have $d_1(w) = -d_2(-w)$. Combining Theorem 2 and relationship above, we yield the conclusion.

The symmetric relation between w and $-w$ is true under our assumption on the symmetry of the function f. Without this symmetry, the theorem above will certainly be violated (see Remark 1).

Finally, we want to point out that all conclusions in this section are a generalization of our previous results on the limiter functions, say Theorem 2 is stated exactly the same way as Theorem 2 in [7].

4. APPLICATIONS TO THE HOPFIELD MODEL

4.1. The Model

The Hopfield model [18], to which most of the theoretical investigations in the field of neural networks has been devoted so far is defined by

$$q_{ij} = T_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu, \qquad i, j = 1, \dots, N, \tag{35}$$

and by setting $k_1 = \theta$ the threshold, $k_2 = h$ the external field and $r_i = 1$, $i = 1, \dots, N$. $w_i(\tau)$ is the neural activity at time τ of the i^{th} neuron where $\xi^\mu = (\xi_i^\mu, i = 1, \dots, N)$ is the μ^{th} pattern stored in the network.

Dynamics (1) now reads

$$w_i(\tau+1) = f\left(w_i(\tau) + \sum_{j=1}^{N} (T_{ij} + h)\, w_j(\tau) + \theta\right), \qquad i = 1, \dots, N. \tag{36}$$

In most of the theoretical investigations, in particular in the statistical physics approach, ξ_i^μ is assume to be i.i.d. and $p(\xi_i^\mu = 1) = p(\xi_i^\mu = -1) = 1/2$, $\forall i, \mu$.

Dynamics (36) is a discrete time version of the continuous Hopfield model, see equation (3.31) in [3]. Next we apply our results of Section 3 to the Hopfield model, which sheds some new light on the dynamics properties of the model.

4.2. Parameter Region

Since the stored patterns take values $+1$ and -1, it justifies our restriction to consider only attractors in $\{-1, 1\}^N$, i.e., in the set of saturated attractors.

For $w \in \{-1, 1\}^N$, now

$$c(w) = \left|J^+(w)\right| - \left|J^-(w)\right|, \tag{37}$$

$d_2(w)$ and $d_1(w)$ turn out to be

$$\begin{aligned} d_1(w) &= \max_{i \in J^+(w)} \left[\sum_{j \in J^-(w)} T_{ij} - \sum_{j \in J^+(w)} T_{ij}\right] \\ &= \max_{i \in J^+(w)} \left[-\sum_{j \in J^-(w)} w_j T_{ij} - \sum_{j \in J^+(w)} w_j T_{ij}\right] \\ &= \max_{i \in J^+(w)} \left[-\sum_{j} w_j T_{ij}\right]. \end{aligned} \tag{38}$$

From the definition of T_{ij}, $i, j = 1, \dots, N$, we see that

$$\begin{aligned} \sum_{j=1}^{N} w_j T_{ij} &= \sum_{j=1}^{N} w_j \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \\ &= \sum_{\mu=1}^{p} \xi_i^\mu \sum_{j=1}^{N} \frac{w_j \xi_j^\mu}{N} \\ &= \sum_{\mu=1}^{p} \xi_i^\mu m(w, \xi^\mu), \end{aligned} \tag{39}$$

where

$$m(w, \xi^\mu) := \frac{1}{N} \sum_{i=1}^{N} w_i \xi_i^\mu \tag{40}$$

is the overlap between the configuration w and the pattern ξ^μ. So now we have that

$$d_1(w) = -\min_{i \in J^+(w)} \sum_{\mu=1}^{p} \xi_i^\mu m(w, \xi^\mu), \tag{41}$$

similarly,

$$d_2(w) = -\max_{i \in J^-(w)} \sum_{\mu=1}^{p} \xi_i^\mu m(w, \xi^\mu). \tag{42}$$

Combining (41), (42), and Theorem 2, we see that the criterion for a saturated attractor of the Hopfield model is the following theorem.

THEOREM 3. *For dynamics (36), a configuration $w \in \{-1,1\}^N$ is a saturated attractor of the Hopfield model if and only if*

$$-\min_{i\in J^+(w)} \sum_{\mu=1}^{p} \xi_i^\mu m(w,\xi^\mu) < \theta + c(w)h < -\max_{i\in J^-(w)} \sum_{\mu=1}^{p} \xi_i^\mu m(w,\xi^\mu). \tag{43}$$

In the practical applications, we are mainly interested in establishing if $w = \xi^\mu$, $\mu = 1,\dots,p$ is a saturated attractor of dynamics (36). Here we furnish a concrete criterion for verifying if a given configuration is an attractor of dynamics (36). Next let us give an example in order to see how to apply the Theorem 2 to a concrete case. Further applications are contained in next section.

EXAMPLE 1.

(1) Storage of one pattern $\xi \neq (1,\dots,1)$, $(-1,\dots,-1)$.
In this case,

$$T_{ij} = \frac{\xi_i \xi_j}{N}, \qquad i,j = 1,\dots,N, \tag{44}$$

and so

$$d_2(\xi) = 1, \qquad d_1(\xi) = -1. \tag{45}$$

Hence, the parameter region in which ξ is a saturated attractor of dynamics (36) is

$$1 > \theta + c(\xi)h > -1. \tag{46}$$

Furthermore, we should note here that if

$$w \neq \xi,\ -\xi,\ (1,\dots,1),\ (-1,\dots,-1),$$

then

$$d_2(w) - d_1(w) < d_2(\xi) - d_1(\xi), \tag{47}$$

namely, ξ is the most stable attractor in the sense that the larger is the difference between d_2 and d_1, the more stable is the attractor.

(2) Storage of two patterns $\xi \neq (1,\dots,1)$, $(-1,\dots,-1)$, and $-\xi$.
In this case, after making an easy calculation, we obtain that

$$d_2(\xi) = 2, \qquad d_1(\xi) = -2. \tag{48}$$

Therefore, the parameter region of (θ, h) in which ξ is a saturated attractor of dynamics (36) is

$$2 > \theta + c(\xi)h > -2. \tag{49}$$

In spite of the extensive investigation of the Hopfield model, little attention was paid to the parameters (θ, h). Our theorem allows us to have a clear understanding of the role played by the two parameters in dynamics (36) as explained below. The Hopfield model is described by a picture of the type of Figure 1 which is redrawn in Figure 2. It is easily seen from Figure 2 that the number of stored patterns, i.e., of saturated attractors, of the Hopfield depends on the parameters (θ, h). There is one region in which many saturated attractors coexist (see Figure 2). In this region, the network will have the highest capacity, a quantity studied extensively in the literature [1,3,19]. Outside this region, the capacity will become lower and 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

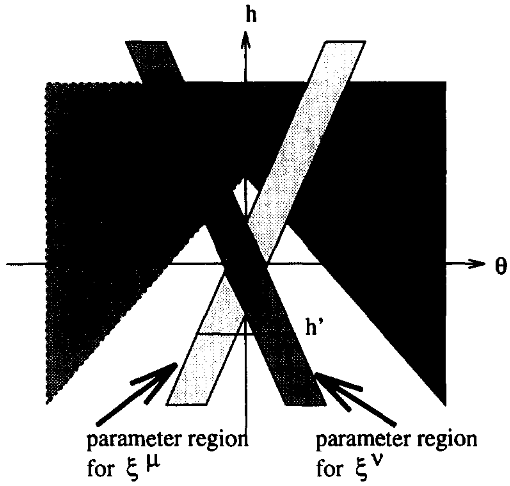

Figure 2. The parameter region of (θ, h) in which w is a saturated attractor of the Hopfield model (see Figure 1, also). In the dark region, the Hopfield model have the highest capacity. In this region, for example, ξ^μ, ξ^ν are both attractors of the Hopfield model. When $h = h'$ (horizontal line), the capacity of the model becomes lower.

since if dynamics (36) converges to a saturated attractor, it will go to w. This may also suggest a way to recall an information avoiding the spurious states [11].

4.3. Capacity

As we already discussed before the difference $d_2(w) - d_1(w)$ 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 dynamics (36). Or in other words, the existence of an energy gap for a state w between the neurons in $J^+(w)$ and $J^-(w)$ is a 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.

DEFINITION 2. *The critical capacity α_c for perfect retrieval of dynamics (36) is*

$$\alpha_c := \inf \left\{ \alpha = \frac{p}{N}, \langle d_2(\xi^\mu) \rangle - \langle d_1(\xi^\mu) \rangle = 0, \text{ for any } \mu = 1, \dots, p \right\}, \tag{50}$$

where $\langle \cdot \rangle$ represents the expectation with respect to the distribution P of ξ^μ.

It is reasonable to expect that the capacity defined above will be larger than that of dynamics (10) [1,20]. For dynamics (10), we are only able to assert that a configuration $w \in \{-1, 1\}^N$ satisfying

$$d_2(w) > k_1 + k_2 c(w) > d_1(w) \tag{51}$$

is a fixed point of (10), while for dynamics (36), any configuration $w \in \{-1, 1\}^N$ with the property (51) is already an attractor of dynamics (36).

Since $\langle d_1(\xi^\mu) \rangle$ and $\langle d_2(\xi^\mu) \rangle$ are symmetric with respect to μ under the condition, the matrix T is given by (35), we only need to compute

$$\langle d_2(\xi^1) \rangle - \langle d_1(\xi^1) \rangle, \tag{52}$$

for estimating the capacity of the network. Furthermore, in terms of the symmetry 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 $\langle d_1(\xi^1) \rangle > 0$ or $\langle d_2(\xi^1) \rangle < 0$.

THEOREM 4. *Iff* $p < N/(2\ln N)$, *we have* $\langle d_1(\xi^1)\rangle < 0$.

PROOF. By the central limit theorem, we obtain that

$$\begin{aligned}\max_{i\in J^-(\xi^1)} \frac{\sum_{\mu=1}^p \sum_{j=1}^N \xi_i^\mu \xi_j^\mu \xi_j^1}{N} &= -1 + \max_{i\in J^-(\xi^1)} \frac{\sum_{\mu=2}^p \sum_{j=1}^N \xi_i^\mu \xi_j^\mu \xi_j^1}{N} \\ &= -1 + \max_{i\in J^-(\xi^1)} \frac{\sum_{\mu=2}^p \xi_i^\mu \zeta^\mu}{\sqrt{N}} \\ &= -1 + \frac{\sqrt{p}}{\sqrt{N}} \max_{i\in J^-(\xi^1)} \eta_i,\end{aligned} \tag{53}$$

where ζ^μ and η_i are both random variables of standard normal distribution. Hence, to ensure $\langle d_1(\xi^1)\rangle < 0$ iff $\langle(\sqrt{p}/\sqrt{N})\max_{i\in J^-(\xi^1)}\eta_i\rangle < 1$. Next we are going to estimate the distribution of the random variable $\max_{i\in J^-(\xi^1)}\eta_i$,

$$\begin{aligned}P\left(\max_{i\in J^-(\xi^1)} \eta_i < x\right) &= P\left(\left\langle\left(\max_{i\in J^-(\xi^1)} \eta_i < x \mid \xi^1\right)\right\rangle\right) \\ &= \sum_k C_N^k \frac{1}{2^N} P\left(\max_{1\le i\le k} \eta_i \le x\right) - \frac{1}{2^N} \\ &= \left(\frac{1+P(\eta_i\le x)}{2}\right)^N - \frac{1}{2^N}.\end{aligned} \tag{54}$$

Define

$$a_N = (2\log N)^{1/2}$$

and

$$b_N = (2\log N)^{1/2} - \frac{1}{2}(2\log N)^{-1/2}(\log\log N + \log 4\pi),$$

we have

$$P\left(\max_{i\in J^-(\xi^1)} \eta_i \le \frac{x}{a_N} + b_N + o(a_N)\right) = \left(\frac{1+P(\eta_i \le (x/a_N) + b_N + o(a_N))}{2}\right)^N - \frac{1}{2^N}. \tag{55}$$

Following the arguments in [21, p. 15], we know that

$$\left(\frac{1+P(\eta_i \le (x/a_N) + b_N + o(a_n))}{2}\right)^N \to e^{-(\exp(-x)/2)},$$

as N tends to infinity which implies that

$$\left\langle \frac{\max_{i\in J^-(\xi^1)} \eta_i}{a_N} \right\rangle = 1.$$

So the conclusions of the present theorem follow.

Now we go a step further to consider the parameter region in which the Hopfield model has the capacity as in Theorem 4. In order to make sense for inequality (43) as $N \to \infty$, we consider the parameter region of θ only[1]. Theorem 3 tells that when $-\langle d_1(\xi^1)\rangle < \theta < \langle d_1(\xi^1)\rangle$, 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 $p(N)/N$, but when θ is not in the region $[-\langle d_1(\xi^1)\rangle, \langle d_1(\xi^1)\rangle]$ the capacity is zero.

[1]By the law of iterated logarithm, we know that $\limsup_{N\to\infty} c(\xi^1)/(\sqrt{N\log\log N}) = +1$ and $\liminf_{N\to\infty} c(\xi^1)/(\sqrt{N\log\log N}) = -1$.

Let us now have a comparison of conclusions in Theorem 3 with the existing results. For the Hopfield model in space $\{-1,1\}^N$ in [1], it is proven that if $p < N/(2\log N)$, a *given* pattern is a fixed point and if $p < N/(4\log N)$ *all* patterns are fixed point of the dynamics. In [19], the authors rigorously proved that if $p < N/(4\log N)$, then all patterns are attractors of the dynamics. Here we conclude that as $p < N/(2\log N)$ all patterns are stable fixed points.

REMARK 3. If a small fraction of errors is tolerated in retrieved patterns, it is possible for us to find a positive critical capacity if we define

$$\alpha_c' = \inf \left\{ \frac{p}{N}, \left\langle \max_{w \in S(\xi^{(1)},\delta)} d_2(w) - d_1(w) \right\rangle > 0 \right\},$$

for $S(x,\delta)$ representing the ball with the radius of δ and the center at x (see [22,23]).

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 the capacity in a more general context [24,25].

5. APPLICATION TO LINSKER'S MODEL

5.1. The Model

Linsker's model [5,7,26] resembles the visual system, with an input feeding onto a number of layers corresponding to the layers of the visual cortex. The units of the network are linear and are organized into two-dimensional layers indexed L_0 (input), $L_1, \dots,$ and so on. For the simplicity of the notation, suppose that each layer has N neurons and has periodic boundary conditions (wrapped up). There are feed-forward connections between adjacent layers, with each unit receiving inputs decreasing monotonically with the distance from the neurons belonging to the underlying layer. Figure 3 shows the arrangement.

More specifically, let $x_i^{(n)}(\tau)$ be the activity of the i^{th} neuron at time τ in the n^{th} layer,

$$x_k^{(n)}(\tau) = \sum_{i=1}^{N} x_i^{(n-1)}(\tau) w_{ki}^{(n)}(\tau) r_{ki}^{(n)} + a_1, \qquad n = 1, \dots, \tag{56}$$

where $w_{ki}^{(n)}(\tau)$ is the synaptic connection between the $(n-1)^{\text{th}}$ layer and the n^{th} layer, $r_{ki}^{(n)}$ is the synaptic density function between the $(n-1)^{\text{th}}$ layer and the n^{th} layer,

$$\sum_{i=1}^{N} r_{ki}^{(n)} = 1, \qquad \forall\, n, k, \tag{57}$$

a_1 is a parameter. For $n = 0$ (input layer), let $x_i^{(0)}(\tau)$ be the i.i.d. noise, i.e.,

$$\left\langle x_i^{(0)}(\tau) x_j^{(0)}(\tau) \right\rangle = \delta_{ij}. \tag{58}$$

For the development process of the synaptic connections $w_{ki}^{(n)}$ in Linsker's network, the Hebb-type learning rule is used, namely,

$$w_{ki}^{(n)}(\tau+1) = w_{ki}^{(n)}(\tau) + a_2 + a_3 \left(x_k^{(n)}(\tau) - a_4 \right) \left(x_i^{(n-1)}(\tau) - a_5 \right), \tag{59}$$

where a_2, $a_3 > 0$, a_4, a_5 are all constants again.

Taking the expectation on both sides of the equality (59) above, and supposing that the expectation of $x_i^{(n-1)}(\tau)$ is independent of τ and i, which is the real situation in Linsker's simulation since he trained the network layer by layer, we obtain finally that

$$w_{ki}^{(n)}(\tau+1) = w_{ki}^{(n)}(\tau) + k_1 + \sum_{j=1}^{N} \left(q_{ij}^{(n)} + k_2 \right) r_{kj}^{(n)} w_{kj}^{(n)}(\tau), \tag{60}$$

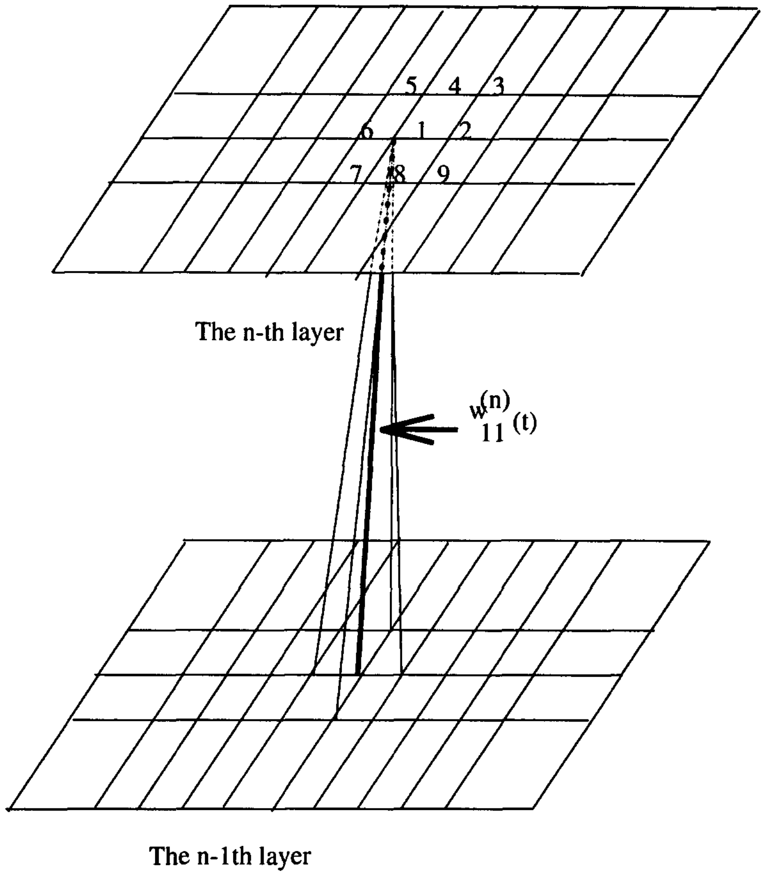

Figure 3. A schematic representation of Linsker's network between the $n-1^{\text{th}}$ layer and the n^{th} layer.

where k_1, k_2 are combinations of the parameter a_1, a_2, a_3 (we take $a_3 = 1$), a_4, and a_5, $Q^{(n)} = (q_{ij}^{(n)},\ i,j = 1,\dots,N)$ is the matrix given by

$$q_{ij}^{(n)} = \sum_{m,k} r_{ik}^{(n-1)} w_{ik}^{(n-1)} r_{jm}^{(n-1)} w_{jm}^{(n-1)} q_{km}^{(n-1)}, \qquad n = 1,2,\dots, \tag{61}$$

with $q_{km}^{(0)} = \delta_{km}$ in accordingly to (58),

$$w_{ik}^{(n-1)} = \lim_{\tau\to\infty} w_{ik}^{(n-1)}(\tau),$$

the equilibrium value of the synaptic connection between the $(n-2)^{\text{th}}$ layer and the $(n-1)^{\text{th}}$ layer, $r_{ik}^{(0)} = 1$, $w_{ik}^{(0)} = 1$, $i,k = 1,\dots,N$.

In order to avoid the unboundedness of $w_{ki}^{(n)}(\tau)$, Linsker used the limiter function f_c defined in the Section 2 to restrict the range of $w_{ki}^{(n)}(\tau)$ to $[-1,1]$. So the dynamics used in the simulation of Linsker's network is

$$w_{ki}^{(n)}(\tau+1) = f_c\left(w_{ki}^{(n)}(\tau) + k_1 + \sum_{j=1}^{N}\left(q_{ij}^{(n)} + k_2\right) r_{kj}^{(n)} w_{kj}^{(n)}(\tau)\right).$$

A lot of facts on Linsker's model with the limiter function are given in [7], in particular, in the first three layers. Here, we consider the more general dynamics defined by

$$w_{ki}^{(n)}(\tau+1) = f\left(w_{ki}^{(n)}(\tau) + k_1 + \sum_{j=1}^{N}\left(q_{ij}^{(n)} + k_2\right) r_{kj}^{(n)} w_{kj}^{(n)}(\tau)\right). \tag{62}$$

Unlike the Hopfield model Linsker's model is a feed forward multilayer network. Dynamics (62) is the updating process of the synaptic connections rather than the neuron activities. In the sequel, we refer to the model with dynamics (62) as (generalized) Linsker's network. For fixed n, the appearance of a structured receptive field is independent of the index k thus, we can rewrite (62) as

$$w_i^{(n)}(\tau+1) = f\left(w_i^{(n)}(\tau) + k_1 + \sum_{j=1}^{N}\left(q_{ij}^{(n)} + k_2\right) r_j^{(n)} w_j^{(n)}(\tau)\right). \tag{63}$$

5.2. Parameter Region

We change our notation a little bit in order to apply Theorem 2 in Section 3 to dynamics (63). Let

$$d_1(w,n) = \begin{cases} \max_{i\in J^+(w)}\left[\sum_{j\in J^-(w)} q_{ij}^{(n)} r_j^{(n)} - \sum_{j\in J^+(w)} q_{ij}^{(n)} r_j^{(n)}\right], & \text{if } J^+(w) \neq \phi, \\ -\infty, & \text{otherwise,} \end{cases} \tag{64}$$

and

$$d_2(w,n) = \begin{cases} \min_{j\in J^-(w)}\left[\sum_{i\in J^-(w)} q_{ij}^{(n)} r_j^{(n)} - \sum_{j\in J^+(w)} q_{ij}^{(n)} r_j^{(n)}\right], & \text{if } J^-(w) \neq \phi, \\ \infty, & \text{otherwise.} \end{cases} \tag{65}$$

THEOREM 5. *w is a saturated attractor of Linsker's model if and only if*

$$d_1(w,n) < k_1 + c(w)k_2 < d_2(w,n), \qquad n = 1, \dots, \tag{66}$$

furthermore, if f is a dissipative saturated sigmoidal function then there exists $T > 0$ such that $w = w(T+\tau)$, $\tau \geq 0$.

In Linsker's model a structured (an on-center or an oriented) receptive field is of particular interest. By keeping all the synaptic connections between the L_0 layer and the L_1 layer positive an on-center (off-center) receptive field appears between the L_1 and L_2 layer. This kind of structured receptive field is also recently founded important in the application of similar network to image recognition. Let us make a comparison between what has been discovered in Linsker's numerical simulation for the third layer (L_2) (Figure 4) and the more exactly discovery in [7].

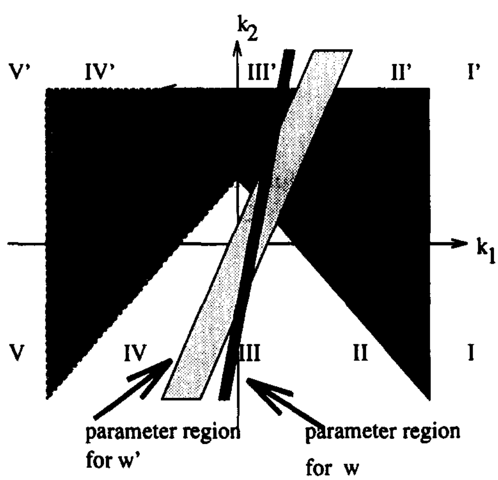

Figure 4. The smaller the size of the on-center of an on-center receptive field, the narrower the band in which that the on-center receptive field is an attractor [7]. The same conclusion is true for off-center receptive field (Proposition 1). When k_1 decreases ($k_2 < 0$), we go from the Region I (all-excitatory), II (on-center), III (several attractors coexist), IV off-center), V (all-inhibitory). If k_1 decreases ($k_2 > 0$), we go from the Region I′ (all-excitatory), II′ (off-center), III′ (several attractors coexist), IV′ (on-center), V′ (all-inhibitory).

For $k_2 < 0$, we pass through a series of regimes as k_1 is decreased.

(1) Each cell is all-excitatory (Region I in Figure 4). This happens when

$$k_1 + k_2 > d(+,2) := - \min_{i=1,\dots,N} \sum_{j=1}^{N} q_{ij}^{(2)} r_j^{(2)}.$$

(2) Each cell is an on-center circularly symmetric opponent cell (Region II in Figure 4). This happens when

$$d_1(w,2) < k_1 + c(w)k_2 < d_2(w,2),$$

where $c(w) > 0$. The width of the band is $d_2(w,2) - d_1(w,2)$.

(3) As we continue to lower k_1, a more complex situation appeared (Region III in Figure 4). For example, in Figure 2 of [7], it is shown that the oriented receptive field is also an attractor of dynamics (63).

(4) As k_1 is made more negative, we reach an off-center circularly symmetric opponent cell (Region IV in Figure 4). This happens when

$$d_1(w,2) < k_1 + c(w)k_2 < d_2(w,2),$$

where $c(w) < 0$. The width of the band is $d_2(w,2) - d_1(w,2)$ (see Proposition 1 for the symmetry between the on-center and off-center).

(5) Finally, an all inhibitory regime (Region V in Figure 4). This is the region

$$k_1 - k_2 < d(-,2) := \min_{i=1,\dots,N} \sum_{j=1}^{N} q_{ij}^{(2)} r_j^{(2)}.$$

The above phenomena is observed in the numerical simulation of [5]. As $k_2 > 0$, the similar phenomena is observable if k_1 decreases (in another order, from (1) $\to$ (4) $\to$ (3) $\to$ (2) $\to$ (5), in Figure 4 from I$'$ $\to$ II$'$ $\to$ III$'$ $\to$ IV$'$ $\to$ V$'$).

This discovery becomes more important as we encounter the necessity in the practical applications to control the size of the on-center receptive field by selecting the parameter of dynamics (63) as in the next section.

5.3. The n^{th} Layer

In fact, all the above descriptions are true for any layer in Linsker's model. The only difference is to replace $Q^{(2)} = (q_{ij}^{(2)}, i,j = 1,\dots,N)$ by $Q^n = (q_{ij}^{(n)}, i,j = 1,\dots,N)$. Let w be a given structured receptive field, say the on-center receptive field w. The problem is how to choose a certain layer n so that $d_2(w,n) - d_1(w,n)$ is as large as possible. In the following, we will always assume that $r_{ki}^{(n)}, k,i = 1,\dots,N$ is independent of n. This is done only for the convenience of the theoretical treatment of our consideration for the case $n \to \infty$. The case in which r is dependent on n is fully discussed in [7] for the first three layers. We can ask whether the difference $d_2(w,n) - d_1(w,n)$ will become larger and larger if we keep all the connections positive between adjacent layers as those between the L_0 layer and the L_1 layer. The answer is negative as indicated by the theorem below.

THEOREM 6. *If $w_{ij}^{(n)} = 1$, $i,j = 1,\dots,N$, $n = 1,\dots,$ then as $n \to \infty$, the only attractor of dynamics (63) will be $(1,\dots,1)$ and $(-1,\dots,-1)$.*

PROOF. First, note that

$$Q^{(n)} = A^{2n},$$

where

$$A = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ r_{21} & r_{22} & \cdots & r_{2N} \\ & & \cdots & \\ r_{N1} & r_{N2} & \cdots & r_{NN} \end{pmatrix},$$

with

$$\sum_{j=1}^{N} r_{ij} = 1, \qquad i = 1, \dots, N.$$

So A now is a probability matrix. From the general theory of the Markov chain, we know that

$$\lim_{n\to\infty} Q^{(n)} = \begin{pmatrix} q_1 & q_2 & \cdots & q_N \\ q_1 & q_2 & \cdots & q_N \\ & & \cdots & \\ q_1 & q_2 & \cdots & q_N \end{pmatrix},$$

since the Markov chain defined by matrix A is irreducible, where

$$\sum_i q_i = 1.$$

By Corollary 2 of Section 3, we obtain the conclusion of the theorem.

Theorem 6 tells that the further the layer, the smaller the difference of $d_2(w,n) - d_1(w,n)$ will be if w is an on-center receptive field. A confirm of this statement is the fact that in Linsker's network an on-center receptive field switches on between the second layer (L_1) and the third layer (L_2).

Now we would like to ask ourselves that what kind of matrix $Q = \lim_{n\to\infty} Q^{(n)}$ defined by equation (61) favors the appearance of an oriented receptive field. Returning to the trivial Example 1 in the Section 3, we see that if we store one pattern ξ of the oriented receptive field then

$$\xi = (1, \dots, 1, -1, \dots, -1, 1, \dots, 1, \dots\dots, -1, \dots, \dots, -1),$$

and we get the following synaptic matrix:

$$T = \left\{ T_{ij}, T_{ij} = \frac{\xi_i \xi_j}{N}, \; i,j = 1, \dots, N \right\}.$$

We observe that the feature of the matrix T is that each row of $(T_{ij}, j = 1, \dots, N)$ oscillates many times between $+1/N$ and $-1/N$. As we already know, this matrix makes the difference between $d_2(\xi)$ and $d_1(\xi)$ the biggest. It is founded in the numerical simulations of Linker ([5, Figure 1, p. 8786]) that the further the layer, the deeper the oscillations of $q_{ij}^{(n)}$, $j = 1, \dots, N$ between the positive and the negative values. This implies that the further the layer, the more the matrices $Q^{(n)}$ and T are similar. And thus, the quantity $d_2(w,n) - d_1(w,n)$ should become larger and larger if w is an oriented receptive field. A rigorous proof of the above conclusion relies on the limit behavior of the matrix $Q^{(n)}$. We believe that it is possible to get some results on it.

6. APPLICATION TO THE DLN

The power of the dynamic link network(DLN), a model proposed by von der Malsburg first in 1981 is demonstrated and developed in recent years in different applications, see, for example, [6]. In [8], a discrete version of the DLN is proposed and a principle for choosing the parameters used in the DLN is given for the limiter function defined as in Section 2. Here by our results of Section 3, we are able to reobtain all the results in [8]. We first briefly review the DLN.

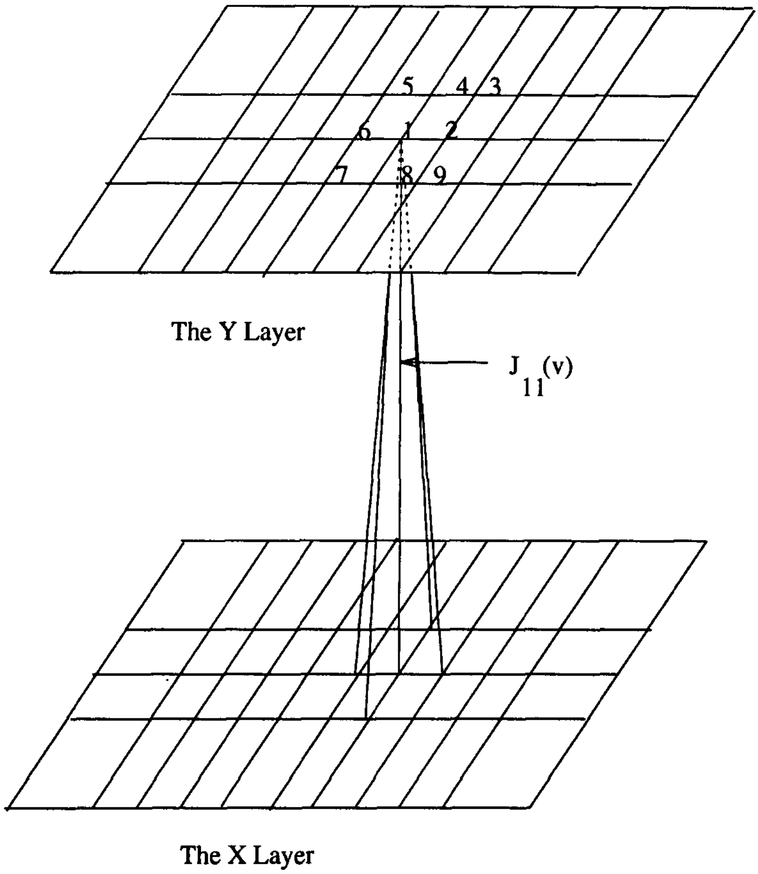

Figure 5. A schematic representation of the DLN.

6.1. The Model

The dynamic link network is essentially a two layer network, say layer X and layer Y with both inter-layer connections and intra-layer connections. Suppose that there are N neurons both in the layer X and Y, and all neurons in the layer $X(Y)$ are arranged in a two-dimensional torus (i.e., with periodic boundary conditions) as shown in Figure 5.

The periodic boundary conditions are adopted here only for avoiding the boundary effects. Choose a coordinate system so that the first neuron sits at the origin. For $i = (r_1, r_2)$, $j = (r_3, r_4)$, $i, j = 1, \dots, N$, or $i, j = 1, \dots N$, the distance between i, j is given by

$$\|i - j\| = \sqrt{|r_1 - r_3|^2 + |r_2 - r_4|^2}.$$

One main feature of the DLN is that there are two time scales, a slow varying one $\nu = 1, 2, \dots,$ and a fast varying one $\tau \in \mathbb{R}^+$. For fixed ν, let $X_i(\tau, \nu)$ denote the activity of the i^{th} neuron at time τ in the layer X and $Y_i(\tau, \nu)$ be the activity of the i^{th} neuron in the layer Y at time τ. $X_i(\tau, \nu)$, $i = 1, \dots, N$ is obtained by a weighted linear combination of the activities of the other neurons in the same layer and then by an application of the sigmoid transformation σ_β. More precisely,

$$\begin{aligned} X_i(\tau, \nu) &= \sigma_\beta\left(x_i(\tau, \nu)\right), \\ \dot{x}_i(\tau, \nu) &= -\alpha x_i(\tau, \nu) + \sum_{j=1}^{N} k_{ij} X_j(\tau, \nu) + I_i^X(\tau, \nu), \qquad (67) \\ X_i(0, \nu) &= 0, \end{aligned}$$

where $i = 1, \dots, N$, $\alpha > 0$ is a parameter of the dynamics,

$$k_{ij} = \gamma p_{ij} - \mu, \tag{68}$$

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

For the activities in the Y layer, we have the same dynamics as the X layer except for the different input signal $I_i^Y(\tau, \nu)$, i.e., for $\nu = 1, 2 \dots$, $i = 1, \dots, N$

$$\begin{aligned} Y_i(\tau, \nu) &= \sigma_\beta \left(y_i(\tau, \nu) \right), \\ \dot{y}_i(\tau, \nu) &= -\alpha y_i(\tau, \nu) + \sum_{j=1}^{N} k_{ij} Y_j(\tau, \nu) + I_i^Y(\tau, \nu), \\ Y_i(0, \nu) &= 0, \end{aligned} \tag{69}$$

here $I_i^Y(\tau, \nu)$, the input signal in the layer Y will be specified in equation (72).

Let a function $T(i, j)$ be defined according to the matching algorithm, i.e., it is equal to one if the feature presented at the neuron i in the X layer, and that at the neuron j in the Y layer is similar and equals to zero otherwise. Then the inter-layer connections $J_{ij}(\nu)$, $\nu = 1, 2, \dots$, $i = 1, \dots, N$, $j = 1, \dots, N$ between the i^{th} neuron in the layer X and j^{th} neuron in the layer Y evolve according to a version of the normalized Hebb learning rule

$$J_{ij}(\nu + 1) = \frac{J_{ij}(\nu) + \epsilon J_{ij}(\nu) T(i, j) Y_j(\nu) X_i(\nu)}{\sum_{i=1}^{N} \left[J_{ij}(\nu) + \epsilon J_{ij}(\nu) T(i, j) Y_j(\nu) X_i(\nu) \right]}, \tag{70}$$

where

$$Y_i(\nu) = \frac{(\lim_{\tau \to \infty} Y_i(\tau, \nu) + 1)}{2}, \qquad X_i(\nu) = \frac{(\lim_{\tau \to \infty} X_i(\tau, \nu) + 1)}{2} \tag{71}$$

are the equilibrium state of the i^{th} neuron activity in the layer Y and the equilibrium state of the i^{th} neuron activity in the layer X, respectively. The matrix $T(i, j)$, with a suitable normalization defines the initial condition for the synaptic matrix J_{ij}:

$$J_{ij}(0) = \frac{T(i, j)}{\sum_{i=1}^{N} T(i, j)}.$$

Note that the existence of the limit in equation (71) is ensured by the existence of the Lyapunov function corresponding to dynamics (67) and dynamics (69) [9].

Now we can give the definition of the input signal $I_j^Y(\tau, \nu)$, $j = 1, \dots, N$, $\nu = 1, 2, \dots$, in the layer Y,

$$I_j^Y(\tau, \nu) = \sum_{i=1}^{N} J_{ij}(\nu) X_i(\tau, \nu) T(i, j). \tag{72}$$

Hence, in the dynamic link network, the time scaling τ is explained as a kind of 'short term memory' and ν is a kind of 'long term memory'. Neurons in the layer X and layer Y are grouped according to the self-organizations mechanism (67) and (69) first. And then the learning procedure for $J_{ij}(\nu)$ is evolved in accordance with the self-organization (67) and (69) through dynamic (70).

Since all the conclusions below are true for both X and Y layer, let us agree to use ξ, I to represent either X, I^X or Y, I^Y. For the sake of simplicity, we take the parameter $\alpha = 1$ and note that this parameter also does not appear in the fast DLN proposed in [9]. So now dynamics (67) and (69) read

$$\begin{aligned} \xi_i(\tau,\nu) &= f\left(x_i(\tau,\nu)\right), \\ \dot{x}_i(\tau,\nu) &= -x_i(\tau,\nu) + \sum_{j=1}^{N} k_{ij}\xi_j(\tau,\nu) + I_i(\tau,\nu), \\ \xi_i(0,\nu) &= 0, \end{aligned} \tag{73}$$

where f is the saturated sigmoidal function.

Discretising (73) with time step h, without loss of generality, we set $h = 1$, we have

$$\begin{aligned} x_i(\tau+1,\nu) &= \sum_{j=1}^{N} k_{ij} f\left(x_j(\tau,\nu)\right) + I_i(\tau,\nu), \\ x_i(0,\nu) &= 0, \end{aligned} \tag{74}$$

for $\tau \in N$. As pointed out in Section 2, we can transform the dynamic system generated by the solutions $x_i(\tau,\nu)$ of (74) in a more suitable system by making the transformation $\eta_i(\tau,\nu) = f(x_i(\tau,\nu))$

$$\begin{aligned} \eta_i(\tau+1,\nu) &= f\left(\sum_{j=1}^{N} k_{ij}\eta_j(\tau,\nu) + I_i(\tau,\nu)\right), \\ \eta_i(0,\nu) &= 0, \end{aligned} \tag{75}$$

which is the dynamics we will focus on.

We suppose now that $I_\tau(i,\nu)$ is independent of i and τ denoting it as $I(\nu)$. The case of the dependence of I on i and τ is considered in (5) of Theorem 7 below.

6.2. Choosing the Parameters

Define

$$\begin{aligned} e_1(w) &:= \max_{i\in J^+(w)} \left[\sum_{j\in J^-(w)} p_{ij} - \sum_{j\in J^+(w)} p_{ij}\right], \\ e_2(w) &:= \min_{i\in J^-(w)} \left[\sum_{j\in J^-(w)} p_{ij} - \sum_{j\in J^+(w)} p_{ij}\right]. \end{aligned} \tag{76}$$

By Theorem 1 of Section 3, we know that if we look at the parameter space of $(I(\nu),\mu)$, the region of them ensuring that ω is an attractor of dynamic (75) is a band between two parallel lines

$$d_1(w) = \gamma e_1(w) + 1 = I(\nu) - c(w)\mu \tag{77}$$

and

$$I(\nu) - c(w)\mu = \gamma e_2(w) - 1 = d_2(w). \tag{78}$$

Assume that $c = \sum_{j=1}^{N} p_{ij}$. Differently from the previous two sections, here we could easily calculate the function $e_2(w)$ and $e_1(w)$ if w is an on-center activity pattern with radius r. Without loss of generality, we suppose that $w_i = 1$ if $\|i\| \le r = \sqrt{m^2+n^2}$, and $w_i = -1$ if $\|i\| > r$, where r is a positive number, the radius of the excitatory neuron activities. In this setting, from the nonincreasing property of p_{ij}, we know that

$$\begin{aligned} e_2(w) &= \left[c - 2\max_{i\in J^-(w)} \sum_{j\in J^+(w)} p_{ij}\right] \\ &= \left[c - 2\sum_{j\in J^+(w)} p_{i^*j}\right], \end{aligned} \tag{79}$$

where $i^* = (m^*, n^*) := (m, n+1)$ if $\sqrt{m^2 + (n+1)^2} < m + 1$, $m \geq n > 0$, $m^2 + n^2 = r^2$, and $i^* = (m^*, n^*) := (m+1, 0)$ if $r = m+1$, $m \geq 0$, i.e., the point i^* lies on the nearest circle passing through the integer lattice outside the circle $m^2 + n^2 = r^2$, and similarly,

$$\begin{aligned} e_1(w) &= \left[c - 2 \min_{i \in J^+(w)} \sum_{j \in J^+(w)} p_{ij} \right] \\ &= \left[c - 2 \sum_{j \in J^+(w)} p_{i_* j} \right], \end{aligned} \tag{80}$$

where $i_* = (m_*, n_*)$ with $m_*^2 + n_*^2 = r^2$.

THEOREM 7.

(1) *For* $\forall w \in \{-1, 1\}^N$, $w \neq (1 \dots, 1)$, $(-1, \dots, -1)$, *a necessary and sufficient condition ensuring that there exists a nonempty set of* (μ, γ, s, I) *in which* w *is a saturated attractor of dynamic (75) is*

$$e_2(w) > e_1(w) \tag{81}$$

and

$$\gamma > \gamma_0 := \frac{2}{e_2(w) - e_1(w)}. \tag{82}$$

Furthermore, the larger the γ, *the bigger the parameter region ensuring that* w *is a saturated attractor of dynamics (2).*

(2) *In the circumstances of (1), there exists a positive number* μ_0 *such that when* μ *is in the set*

$$\{\mu, \mu \geq \mu_0\} \cap \{\mu, \gamma e_1(w) + 1 < I(\nu) - c(w)\mu < \gamma e_2(w) - 1\}, \tag{83}$$

then w *is a saturated attractor of dynamics (75) and* $(1, \dots, 1)$, $(-1, \dots, -1)$ *will no longer be attractors of dynamics (75).*

(3) *If* s *is large enough so that* p_{ij}, $i, j = 1, \dots, N$ *are constants independent of* i, j, *then only* $(1, \dots, 1)$ *and* $(-1, \dots, -1)$ *are the possible saturated attractors of dynamics (75).*

(4) *If* s *is small enough so that* $p_{ij} = \delta_{ij}$ *with* $\gamma > 1$, *then any state* $w \in \{-1, 1\}^N$ *is an attractor of dynamics (75).*

(5) w *is a saturated attractor of dynamics (75) if and only if*

$$I(\nu) \in [\gamma e_1(w) + c(w)\mu + 1, \gamma e_2(w) + c(w)\mu - 1].$$

PROOF.

(1) We know from Theorem 1 that there exists a set of the parameters $(I(\nu), \mu)$ such that w is a saturated attractor of dynamic (75) if and only if

$$\gamma e_2(w) - 1 > \gamma e_1(w) + 1,$$

which implies the first conclusion of the theorem. Furthermore, since the bigger the γ, the wider the band between the lines

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

and

$$I(\nu) - c(w)\mu = \gamma e_2(w) - 1,$$

we arrive at the second conclusion.

(2) For $w_1 = (1, \dots, 1)$, we have that $c(w_1) = N$ and $e_1(w_1) = -c$. So in terms of Theorem 1, we deduce that in the region

$$I(\nu) - N\mu > -\gamma c + 1,$$

the configuration $w_2 = (1, \dots, 1)$ is an attractor of dynamic (75). Similarly, we also have that in the region

$$I(\nu) + N\mu < \gamma c - 1,$$

the configuration $w_2 = (-1, \dots, -1)$ is an attractor of dynamic (75). For fixed γ, w, denote μ_1, μ_2, μ_3, and μ_4 as the solution of the following four equations:

$$\begin{aligned} I(\nu) - N\mu &= -\gamma c + 1, & I(\nu) + N\mu &= \gamma c - 1, \\ I(\nu) - c(w)\mu &= \gamma e_1(w) + 1; & I(\nu) - c(w)\mu &= \gamma e_1(w) + 1; \\ I(\nu) - N\mu &= -\gamma c + 1, & I(\nu) + N\mu &= \gamma c - 1, \\ I(\nu) - c(w)\mu &= \gamma e_2(w) - 1; & I(\nu) - c(w)\mu &= \gamma e_2(w) - 1; \end{aligned}$$

respectively. Then we could choose $\mu_0 = \max(\mu_1, \mu_2, \mu_3, \mu_4)$, and one obtains the conclusion. In fact, we have

$$\begin{aligned} \mu_0 &= \max\left(\frac{\gamma c - \gamma e_1(w) - 2}{N + c(w)}, \frac{\gamma c + \gamma e_2(w) - 2}{N - c(w)} \right) \\ &= \max\left(\frac{\gamma(\sum_{j=1}^N p_{ij} - e_1(w)) - 2}{N + c(w)}, \frac{\gamma\left(\sum_{j=1}^N p_{ij} + e_2(w)\right) - 2}{N - c(w)} \right). \end{aligned}$$

(3) In this case, $e_2(w) = e_1(w)$, from Theorem 2, we know that any $w \in \{-1, 1\}^N$ will not be a saturated attractor of dynamics (75) if $w \neq (1, \dots, 1), (-1, \dots, -1)$. And $(1, \dots, 1)$ is a saturated attractor of dynamics (75) if

$$I(\nu) - N\mu > -\gamma c + 1,$$

$(-1, \dots, -1)$ is a saturated attractor of dynamics (75) if

$$I(\nu) + N\mu < \gamma c - 1.$$

(4) In this setting, $\gamma e_2(w) - 1 = \gamma - 1 > \gamma e_1(w) + 1 = -\gamma + 1$ for any $w \in \{-1, 1\}^N$, so we prove the conclusion by Corollary 1.

(5) It is an easy consequence of Theorem 2 of Section 3.

For an explanation of Theorem 7, we refer the reader to Figure 6.

For dynamics (75), we could define a Lyapunov function as in Section 2,

$$V(w) = -\sum_{i,j}^{N} \bar{k}_{ij} w_i w_j + 2\sum_i \int_0^{w_i} f^{-1}(z)\,dz - 2\sum_i I(\nu) w_i, \tag{84}$$

if $\bar{k}_{ii} \geq 0$, $i = 1, \dots, n$, $w \in [-1, 1]^N$, where $\bar{k}_{ij} = k_{ij} - \delta_{ij} - \mu$, $i, j = 1, \dots, N$. It is readily seen that $(1, \dots, 1)$ $((-1, \dots, -1))$ is the global minima of the dynamics if the input $I(\nu) > 0 (I(\nu) < 0)$ and $\mu \leq 0$. So $(1, \dots, 1)$ $((-1, \dots, -1))$ will dominate the behavior of the dynamics in the sense that if the input is contaminated by the noise, the neural configuration will converge to the global minima with large probability as in the simulated annealing. However, the lateral inhibition $\mu > \mu_0$ ($|\mu_0|$ is small usually) guarantees that some nontrivial activity patterns (not all the neuron activities are excitatory or inhibitory) will be reached by the system in the evolution

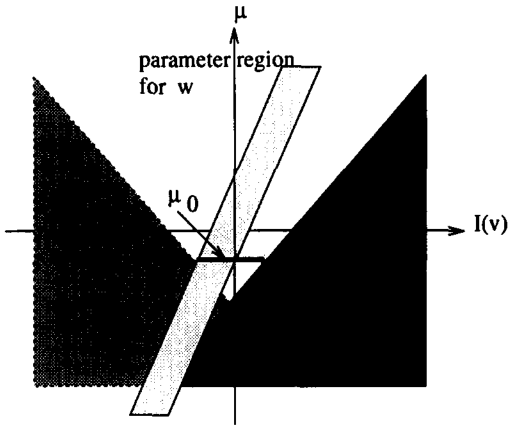

Figure 6. The parameter region of $(I(\nu), \mu)$ in the DLN. Note that the slope of the dark lines is $c(w) = 2|(m,n) \in Z^2, m^2+n^2 \le r^2| - N$ (see equation (88)) if w is a pattern with an on-center field of radius r. So for fixed $\mu = \mu_0$, the smaller the $I(\nu)$, the smaller the size of the on-center (r) of w, which is an attractor of dynamics (75). If r is small enough (Table 1), the pattern with an on-center field of radius r will no long be an attractor of the dynamics (75). For fixed $I(\nu) > 0$, as $\mu > 0$ becomes smaller and smaller, the size of the on-center of w will become larger and larger (see Figure 4, also).

of the neuron activities since now the parameters are outside of the region, where $(1,\dots,1)$ or $(-1,\dots,-1)$ is the attractor of dynamics (75) (Figure 6).

Now we could explain *the role of the lateral inhibition plays in the neural model.* The lateral inhibition pulls the dynamics outside of the region dominated by the attractors $(1,\dots,1)$ or $(-1,\dots,-1)$. If these models are a good approximation of the biological systems, then here we supply an argument that explains why the lateral inhibition is necessary in the natural biological network and it is surprising to us that the biological system is so cleverly devised.

Theorem 7, Conditions 3 and 4 tell that the range of the parameter s of the excitatory connection controls the correlation length of the activity patterns. As s is small, the activity of each neuron could change independently, so any activity pattern could be a saturated attractor of dynamic. As $s \to \infty$, the activity of each neuron is highly correlated, only all excitatory and all inhibitory attractors are saturated attractors of the dynamic.

Theorem 7, Conditions 5 gives an exact fluctuation region of the fluctuation of the input signal. If $I(\nu)$ is in the region of $[\gamma e_1(w) + c(w)\mu + 1, \gamma e_2(w) + c(w)\mu - 1]$, w will remain as an attractor of dynamic (75) (Figure 6). The interval in which $I(\nu)$ changes can be taken as an estimate for an effective interval in the case when the input signal is not translation invariant and depends on the time τ of the neural dynamic.

From all the above arguments we now can give a useful way for choosing the four parameters μ, γ, s, I^X. For a given neuron activity pattern w with an on-center of radius r, which is used in the simulation of fast DLN and numerically founded in the simulation of dynamics (67) and (69), the slope function

$$c(w) = 2\left|(m,n) \in Z^2, m^2+n^2 \le r^2\right| - N, \tag{85}$$

where $|\cdot|$ represent the cardinality of a set. Here, we consider the case

$$p_{ij} = \begin{cases} 1, & \text{if } \|i-j\| \le s, \\ 0, & \text{otherwise.} \end{cases}$$

The case when p_{ij} has the Gaussian distribution is contained in [8]. In this setting, the two intercept functions $e_2(w)$, $e_1(w)$ are obtained by (79) and (80), and we have the following proposition.

Define

$$\Delta e(r,s) := \left|\{(m,n) \in Z^2, m^2+n^2 \le r^2\} \cap \{(m,n) \in Z^2, (m-m_*)^2(n-n_*)^2 \le s^2\}\right| - \left|\{(m,m) \in Z^2, m^2+n^2 \le r^2\} \cap \left\{(m,n) \in Z^2, (m-m^*)^2 + (n-n^*)^2 \le s^2\right\}\right|.$$

PROPOSITION 2. *The parameter range of $(I(\nu), \mu)$ in which an on-center pattern w with the on-center of radius r is a saturated attractor of dynamics (75) is not an empty set if and only if*

$$\Delta e(r,s) > 0, \tag{86}$$

where (m_*, n_*) and (m^*, n^*) are defined by (79) and (80), respectively. μ_0 defined by (84) is given by

$$\mu_0 = \frac{1}{\Delta e(r,s)}. \tag{87}$$

PROOF. We see that

$$e_1(w) = N - 2\left|\{(m,n) \in Z^2, m^2+n^2 \le r^2\} \cap \left\{(m,n) \in Z^2, (m-m_*)^2 + (n-n_*)^2 \le s^2\right\}\right|,$$
$$e_2(w) = N - 2\left|\{(m,m) \in Z^2, m^2+n^2 \le r^2\} \cap \left\{(m,n) \in Z^2, (m-m^*)^2 + (n-n^*)^2 \le s^2\right\}\right|,$$

hence, the conclusions follow from Theorem 7.

Proposition 2 tells that for fixed s, if r is too small with respect to s, then condition (89) will be violated since (upper right corner of Table 1)

$$\{(m,n) \in Z^2, m^2+n^2 \le r^2\} \subset \left\{(m,n) \in Z^2, (m-m_*)^2 + (n-n_*)^2 \le s^2\right\} \cap \left\{(m,n) \in Z^2, (m-m^*)^2 + (n-n^*)^2 \le s^2\right\}.$$

If r is too large with respect to s, condition (90) is also violated (lower left corner of Table 1).

Let

$$\gamma > \gamma_0 = \frac{2}{e_2(w) - e_1(w)} = \frac{1}{\Delta e(r,s)}, \tag{88}$$

be fixed which is independent of the size N if $\Delta e(r,s) > 0$. In Table 1, the value of $\Delta e(r,s)$ is given for $r = 1, \ldots, 10$, $s = 1, \ldots, 10$.

Table 1. $\Delta e(r,s)$ is shown in the table for $r, s = 1, \ldots, 10$.

s	1	2	3	4	5	6	7	8	9	10
$\Delta e(1,\cdot)$	0	2	0	0	0	0	0	0	0	0
$\Delta e(2,\cdot)$	0	1	1	1	0	0	0	0	0	0
$\Delta e(3,\cdot)$	0	0	2	0	0	1	0	0	0	0
$\Delta e(4,\cdot)$	0	0	1	1	0	2	1	1	0	0
$\Delta e(5,\cdot)$	0	0	0	1	2	1	0	0	1	1
$\Delta e(6,\cdot)$	0	0	0	1	1	1	1	0	2	0
$\Delta e(7,\cdot)$	0	0	0	1	1	1	0	2	1	2
$\Delta e(8,\cdot)$	0	0	0	1	0	1	1	1	1	1
$\Delta e(9,\cdot)$	0	0	0	0	1	1	1	0	0	2
$\Delta e(10,\cdot)$	0	0	0	0	1	0	1	1	1	1

Then we easily find μ_0 as in the proof of Theorem 7, i.e.,

$$\mu_0 = \max\left(\frac{\gamma\left(\left|\{(m,n) \in Z^2, m^2+n^2 \le r^2\}\right| - e_1(w)\right) - 2}{N + c(w)}, \frac{\gamma\left(\left|\{(m,n) \in Z^2, m^2+n^2 \le r^2\}\right| + e_2(w)\right) - 2}{N - c(w)}\right). \quad (89)$$

Without loss of generality, set $\mu = \mu_0$. Now we only leave one parameter I^X free, which is determined by the relation

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

The effect of the size of the on-center pattern in the DLN is studied in [9]. As we pointed in Section 5, the input $I(\tau)$ increases, the size of the on-center field will also increase (see Figure 6). This gives a way to control the size of the on-center of an on-center pattern in the DLN by selecting the parameter of dynamics.

Another important fact from our analysis here is that for the DLN, as in many networks proposed today, the problem on how to ensure the convergence of the algorithm is not clear. A simple way to achieve it, as in the case of the Kohonen network [27,28], is that to shrink the size of the on-center field of an on-center configuration. This can be done by decreasing the input $I(\tau)$ as well (Figure 6). However, our Proposition 2 claims that, in general, it is impossible to shrink to any small value the size of the on-center field if s is fixed. If $s > 0$, $r > 0$, then $J_{ij}(\nu)$ is distributed over several neurons in general on the X layer rather than concentrated on a single neuron only. This is a main difference between the Kohonen network and the DLN, as it is noted heuristically in [9].

7. CONCLUSIONS

This paper unifies an approach to study the dynamics properties of the Hopfield model, Linsker's model, and the DLN, three typical networks arising from three typical areas in the study of neural networks. Since most of models proposed to date in the field of neural networks use the sigmoidal function in their dynamics of learning or retrieving procedure, as discussed in Section 2, we are able to analyze the attractors of these models in terms of the present method. So the power of the present analysis is not restricted to these three models.

In the Hopfield model, we give a sufficient and necessary condition to check if a given pattern is an attractor of the network. The capacity of the network is considered from a different point of view of the statistical physics approach. It is also obvious that we could apply our method here to analyze other versions of the Hopfield model.

The present approach becomes more efficient if we are mainly interested in one or a few kind of patterns. This is the case in Linsker's network and the DLN. For the former network the appearance of the on-center and oriented receptive field are the core of its dynamics. For the latter the on-center structured pattern is an important one in its dynamics. This paper asserts that the potential for the appearance of a structured receptive field in Linsker's model is universal in the sense that the appearance of such a field is independent of the specific choice of the limiter function used in the numerical simulations. For the DLN we propose a principle for the selection of these parameters employed in the model.

The significance of this unified approach is obvious: it helps us to understand the mechanism underlying each model more deeply. Besides these findings reported here, there is still a lot of work to be done further, as we have already pointed out from Sections 3–5.

REFERENCES

1. D. Amit, *Modeling Brain Function*, Cambridge University Press, Cambridge, United Kingdom, (1989).
2. J. Feng and B. Tirozzi, The SLLN for the free-energy of the Hopfield and spin glass model, *Helvetica Physica Acta* **68**, 365–379, (1995).
3. J. Hertz, A. Krogh and R. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, (1991).
4. L. Pastur, M. Shcherbin and B. Tirozzi, The replica symmetric solution of the Hopfield model without replica trick, *Jour. of Stat. Phys.* **74** (5/6), 1161–1183, (1994).
5. R. Linsker, From basic network principle to neural architecture (series), *Proc. Natl. Acad. Sci. USA* **83**, 7508–7512, 8390–8394, 8779–8783, (1986).
6. M. Lades, J.C. Vorbrüggen, J. Buhrmann, J. Lange, C. von der Malsburg, R.P. Würtz and W. Konen, Distortion invariant object recognition in the dynamic link architecture, *IEEE Transaction on Computers* **42**, 300–311, (1993).
7. J. Feng, H. Pan and V.P. Roychowdhury, On neurodynamics with limiter function and Linsker's developmental model, *Neural Computation* **8** (5), 1003–1019, (1996).
8. J. Feng and B. Tirozzi, On choosing the parameters in the dynamic link network, In *Neural Nets, WIRN VIETRI-95*, (Edited by M. Marinaro and R.Tagliaferri), pp. 245–250, World Scientific, Singapore, (1996).
9. W. Konen, T. Maurer and C. von der Malsburg, A fast dynamic link matching algorithm for invariant pattern recognition, *Neural Network* **7** (6/7), 1019–1030, (1994).
10. J. Feng, Establishment of topological maps—A model study, *Neural Processing Letters* **2** (6), 1–4, (1995).
11. S. Albeverio, J. Feng and M. Qian, The role of noises in neural networks, *Phys. Rev. E.* **52** (6), 6593–6606, (1995).
12. M. Antonucci, B. Tirozzi, N.D. Yarunin and Dotsenko, Numerical simulation of neural networks with translation and rotation invariant pattern recognition, *Inter. Jour. of Modern Physics B* **8** (11/12), 1529–1541, (1994).
13. J. Feng, Lyapunov functions for neural nets with nondifferentiable input-output characteristics, *Neural Computation* **9** (1), (1997) (to appear).
14. J. Feng and B. Tirozzi, An application of the saturated attractor aanalysis to three typical models, *Lecture Notes in Computer Science* **930**, 353–360, (1995).
15. A.V.M. Herz and C.M. Marcus, Distributed dynamics in neural networks, *Phys. Rev. E* **47**, 2155–2161, (1993).
16. C.M. Marcus and R.M. Westervelt, Dynamics of iterated map networks, *Phys. Rev. A* **40**, 501–504, (1989).
17. J. Feng and K.P. Hadeler, Qualitative behavior of some simple networks, *Jour. of Phys. A: Gener. and Math.* **29**, 5019–5033, (1996).
18. J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554–2558, (1982).
19. S.I. Amari, Mathematical foundation of neurocomputing, *Proc. of the IEEE* **78** (9), 1443–1457, (1990).
20. R.J. MacEliece, E.C. Posner, E.R. Rodemich and S.S. Venkatesh, The capacity of the Hopfield associative memory, *IEEE Trans. Inform. Theory* **33**, 461–482, (1987).
21. M.R. Leadbetter, G. Lindgren and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York, (1983).
22. D. Loukianova, Two rigorous Bounds in the Hopfield model of associative memory, URA CNRS 1321, Université Paris 7 (Preprint), (November 1994).
23. C. Newman, Memory capacity and neural network models: Rigorous lower bounds, *Neural Networks* **1**, 223–238, (1988).
24. J.L. van Hemmen, D. Grensing, A. Huber and R. Kuhn, Nonlinear neural networks, I. 231–258; II. 259–272, *Jour. of Stat. Phys.* **50** (1/2), (1988).
25. J. Feng, B. Tirozzi and R. Zucchi, Rigorous results and critical capacity for a short-term model, In *Markov Processes and Related Fields*, Volume 2, (in press), (1996).
26. D. MacKay and K. Miller, Analysis of Linsker's application of Hebbian rules to linear networks, *Network* **1**, 257–297, (1990).
27. T. Kohonen, *Self-Organization and Associative Memory*, Third edition, Spring-Verlag, Berlin, (1989).
28. J. Feng and B. Tirozzi, Convergence theorems for the Kohonen feature mapping algorithm with VLEPs, *Computers Mathl. Applic.* **33** (3), 45–63, (1997).
29. M.A. Cohen and S. Grossberg, 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, (1983).
30. S. Grossberg, Nonlinear neural networks: Principles, mechanisms, and architectures, *Neural Networks* **1**, 17–61, (1988).
31. C. von der Malsburg, Network self-organization in the ontogenesis of the mammalian visual system, In *An Introduction to Neural and Electronic Networks*, Second edition, (Edited by S.F. Zornetzer, J. Davis and C. Lau), Academic Press, (1995).

32. K.D. Miller, A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between on- and off-center inputs, *Jour. of Neurosciences* **14** (1), 409–441, (1994).
33. K.D. Miller, J.B. Keller and M.P. Stryker, Ocular dominance column development: Analysis and simulation, *Science* **245**, 605–615, (1989).