Global Convergence Rate of Recurrently Connected Neural Networks

Tianping Chen
tchen@fudan.edu.cn
Wenlian Lu
chiquitita@21cn.com
Laboratory of Nonlinear Mathematics Science, Institute of Mathematics,
Fudan University, Shanghai, China

Shun-ichi Amari
amari@brain.riken.go.jp
RIKEN Brain Science Institute, Wako-shi, Saitama 351-01, Japan

We discuss recurrently connected neural networks, investigating their global exponential stability (GES). Some sufficient conditions for a class of recurrent neural networks belonging to GES are given. Sharp convergence rate is given too.

1 Introduction

Recurrently connected neural networks, sometimes called Grossberg-Hopfield neural networks, have been extensively studied, and many applications in different areas have been found. Such applications heavily depend on the dynamical behavior of the networks. Therefore, analysis of these behaviors is a necessary step for practical design of neural networks.

The dynamical behaviors of recurrently connected neural networks have been studied since the early period of research on neural networks. For example, multistable and oscillatory behaviors were studied by Amari (1971, 1972) and Wilson and Cowan (1972). Chaotic behaviors were studied by Sompolinsky and Crisanti (1988). Hopfield and Tank (1984, 1986) studied the dynamic stability of symmetrically connected networks and showed their practical applicability to optimization problems. Cohen and Grossberg (1983) gave more rigorous results on the global stability of networks.

a neural network should be absolutely stable. They also pointed out that the negative semidefiniteness of the interconnection matrix guarantees the global stability of a Grossberg-Hopfield network and proved the absolute global stability for a class $G$ of activation functions. However, they did not discuss the absolute exponential stability and give the convergence rate. Liang and Wu (1999) discussed the absolute exponential stability; however, they assumed that the connection matrix is a special structure, that is, the off-diagonal entries are nonnegative and the activation functions are bounded. Even with these restrictions, the attraction region is limited. Zhang, Heng, and Fu (1999), Liang and Wang (2000), and Liang and Si (2001) discussed the exponential stability of Grossberg-Hopfield neural networks and estimated the convergence rate. However, these estimates are less sharp. Our goal here is to provide a more accurate estimate of the convergence rate.

2 Some Definitions and Main Results

**Definition 1.** Class $\tilde{G}$ of functions: Let $G = \text{diag}[G_1, \ldots, G_n]$, where $G_i > 0$, $i = 1, \ldots, n$. $g(x) = (g_1(x), \ldots, g_n(x))^T$ is said to belong to $\tilde{G}$, if the functions $g_i(x)$, $i = 1, \ldots, n$, satisfy $0 < \frac{g_i(x+u) - g_i(x)}{u} \leq G_i$.

**Definition 2.** $M^s = \frac{1}{2}(M + M^T)$ is defined as the symmetric part of the matrix $M$.

Here, we investigate following dynamical system

$$\frac{du_i(t)}{dt} = -d_i u_i(t) + \sum_j a_{ij} g_j(u_j(t)) + I_i \quad i = 1, \ldots, n$$

(2.1)

where $d_i > 0$, $g(x) = (g_1(x), \ldots, g_n(x))^T \in \tilde{G}$.

**Definition 3.** Recurrently connected neural network 2.1 is said to be absolutely exponentially stable if any solution of it converges to an equilibrium point exponentially for any activation functions in $\tilde{G}$.

We will prove two theorems on the globally exponential convergence

**Theorem 1.** Let $I_n$ be the identity matrix of order $n$, $D = \text{diag}[d_1, \ldots, d_n]$, $A = (a_{ij})_{i, j=1}^n$, $P = \text{diag}[p_1, \ldots, p_n]$, where $p_i > 0$, $i = 1, \ldots, n$, $g(x) = (g_1(x), \ldots, g_n(x))^T \in \tilde{G}$ with $g_i(x)$ being differentiable. If there is a positive constant $\alpha > 0$ such that

$$\alpha < d_i \quad i = 1, \ldots, n$$

(2.2)
and \( P[(D - \alpha d_n)G^{-1} - A]^s \) is positive definite, then the dynamical system \( 2.1 \) has a unique equilibrium point \( u^* \), such that

\[
|u_i(t) - u_i^*| = O(e^{-\alpha t}) \quad i = 1, \ldots, n. \tag{2.3}
\]

**Proof.** Pick a small \( \epsilon > 0 \) such that \( d_i > \alpha + \epsilon \) and the least eigenvalue \( \lambda_{\text{min}} \) of \( P[(D - \beta I_n)G^{-1} - A]^s \) satisfies

\[
\lambda_{\text{min}} = \lambda_{\text{min}}[P[(D - \beta I_n)G^{-1} - A]^s] > 0, \tag{2.4}
\]

where \( \beta = \alpha + \epsilon \). It is well known that under the conditions given in theorem 1, the dynamical system \( 2.1 \) has an equilibrium point \( u^* = [u_1^*, \ldots, u_n^*]^T \).

Define \( v_i(t) = u_i(t) - u_i^* \), and

\[
g_i^*(s) = g_i(s + u_i^*) - g_i(u_i^*). \tag{2.5}
\]

Then we have

\[
\frac{dv_i(t)}{dt} = -d_iv_i(t) + \sum_j a_{ij}g_j^*(t) \quad i = 1, \ldots, n. \tag{2.6}
\]

The new variables \( v_i \) put the unique equilibrium at zero.

Let \( v(t) = [v_1(t), \ldots, v_n(t)]^T \), \( g^*(t) = \text{diag}[g_1^*(u_1(t)), \ldots, g_n^*(u_n(t))] \), and define the Lyapunov function, which was used in Forti and Tesi (1995),

\[
L(k, t) = \frac{1}{2} \sum_{i=1}^n |v_i(t)|^2 + k \sum_{i=1}^n P_i \int_0^{v_i(t)} g_i(\xi) \, d\xi. \tag{2.7}
\]

Direct differentiating leads to

\[
\dot{L}(k, t) = -Dv^T(t)v(t) + v^T(t)Ag^*(t) \\
- k[g^*(t)]^TPDv(t) + k[g^*(t)]^T P A[g^*(t)] \tag{2.8}
\]

\[
= -\beta v^T(t)v(t) - v^T(t)(D - \beta I_n)v(t) + v^T(t)Ag^*(t) \\
- k\beta[g^*(t)]^TPv(t) - k[g^*(t)]^T P(D - \beta I_n)v(t) \\
+ k[g^*(t)]^T P A[g^*(t)] \tag{2.9}
\]

\[
\leq -\beta v^T(t)v(t) - v^T(t)(D - \beta I_n)v(t) + v^T(t)Ag^*(t) \\
- k\beta[g^*(t)]^TPv(t) - k[g^*(t)]^T \\
\times [P[(D - \beta I_n)G^{-1} - A]^s]g^*(t) \tag{2.10}
\]
\[
\begin{align*}
&= - \left\{ v(t)(D - \beta I_n)^{1/2} - \frac{1}{2}(D - \beta I_n)^{-1/2} A g^*(t) \right\}^T \\
&\times \left\{ v(t)(D - \beta I_n)^{1/2} - \frac{1}{2}(D - \beta I_n)^{-1/2} A g^*(t) \right\} \\
&+ \frac{1}{4} [g^*(t)]^T A^T (D - \beta I_n)^{-1} A [g^*(t)] \\
&- k [g^*(t)]^T [P[(D - \beta I_n)G^{-1} - A]]^s [g^*(t)] \\
&- \beta I_n v^T(t)v(t) - k\beta [g^*(t)]^T P v(t) \\
&\leq \frac{1}{4} [g^*(t)]^T A^T (D - \beta I_n)^{-1} A [g^*(t)] \\
&- k [g^*(t)]^T [P[(D - \beta I_n)G^{-1} - A]]^s [g^*(t)] \\
&- \beta v^T(t)v(t) - k\beta [g^*(t)]^T P v(t) \\
&= -[g^*(t)]^T \left\{ k[P[(D - \beta I_n)G^{-1} - A]]^s \\
&\quad - \frac{1}{4} A^T (D - \beta I_n)^{-1} A \right\} [g^*(t)] \\
&- \beta v^T(t)v(t) - k\beta [g^*(t)]^T P v(t).
\end{align*}
\]

Pick a sufficiently large positive \( k \) such that the least eigenvalue

\[
\lambda_{\text{min}} = \lambda_{\text{min}} \left\{ k[P[(D - \beta I_n)G^{-1} - A]]^s \\
- \frac{1}{4} A^T (D - \beta I_n)^{-1} A \right\} > 0.
\]

Then we have

\[
\dot{L}(k, t) \leq \beta v^T(t)v(t) - k\beta [g^*(t)]^T P v(t) - \lambda_{\text{min}} [g^*(t)]^T [g^*(t)].
\]

To obtain the desired convergence rate, we use following equalities:

\[
g_i^*(\xi) = g'_i(u_i^*)\xi_i + o(|\xi_i|) \quad (2.16)
\]

\[
g_i^*(v_i(t))v_i(t) = g'_i(u_i^*)v_i^2(t) + o(|v_i(t)|^2) \quad (2.17)
\]

\[
2 \int_0^{v_i(t)} g_i^*(\xi) d\xi = g'(u_i^*)v_i^2(t) + o(|v_i(t)|^2). \quad (2.18)
\]
Combining previous equalities, we have
\[ g_i^*(v_i(t))v_i(t) = 2 \int_0^{v_i(t)} g_i^*(\xi) \, d\xi + o(|v_i(t)|^2). \tag{2.19} \]

Substituting into equation 2.15, we see that
\[
\dot{L}(k, t) \leq \beta v^T(t)v(t) - k\beta \{g_i^*(t)\}^T P v(t) - \lambda_{\min} \{g_i^*(t)\}^T \{g_i^*(t)\}
\]
\[
\leq -\beta \sum_{i=1}^n |v_i(t)|^2 - 2k\beta \sum_{i=1}^n P_i \int_0^{v_i(t)} g_i^*(\xi) \, d\xi
\]
\[
- \lambda_{\min} \sum_{i=1}^n (g_i'(u_i^*))^2 |v_i(t)|^2 + o \left( \sum_{i=1}^n |v_i(t)|^2 \right)
\]
\[
\leq -\alpha \sum_{i=1}^n |v_i(t)|^2 - 2k\alpha \sum_{i=1}^n P_i \int_0^{v_i(t)} g_i^*(\xi) \, d\xi = -2\alpha L(k, t) \tag{2.20}
\]
holds for sufficiently large \( t \). Therefore,
\[
\frac{1}{2} \sum_{i=1}^n |v_i(t)|^2 \leq L(k, t) \leq L(k, 0)e^{-2\alpha t} \tag{2.21}
\]
and
\[
v_i(t) = O(e^{-\alpha t}) \quad i = 1, \ldots, n. \tag{2.22}
\]

Theorem 1 is proved.

**Remark 1.** The conditions that the derivatives \( g_i(x) \) exist imposed on the activations \( g_i(x) \) in theorem 1 are a little stronger than in the existing literature. However, in practice, they are not very restrictive because the monotonically increasing function is differentiable almost everywhere.

**Remark 2.** The assumption \( g_i(x) \) \( i = 1, \ldots, N \) is differentiable can be relaxed to the right and left derivatives of \( g_i(x) \) \( i = 1, \ldots, N \). This relaxation is significant. For example, monotonically increasing piecewise linear functions satisfy this condition.

**Remark 3.** Let \( g_i'(x) \) denote the right and left derivatives, respectively. Under the following assumption that for any \( \epsilon > 0 \), there is \( \delta > 0 \), such that
\[
\left| \frac{g_i(x + h) - g_i(x)}{h} - g_i'(x) \right| < \epsilon \tag{2.23}
\]
holds for all $i = 1, \ldots, N$ and $x$ whenever $0 < h < \epsilon$, then
\[
|u_i(t) - u_i^*| = O(e^{-\alpha t}) \quad i = 1, \ldots, n
\]  
hold uniformly for all $g(x) = (g_1(x), \ldots, g_n(x))^T \in G$.

**Corollary 1.** Let the activation functions satisfy the assumptions given in theorem 1, and let $\xi_i > 0$, $i = 1, \ldots, N$ be some constants. Then under any of the following three inequalities,
\[
d_{ij} \xi_j - G_j \left\{ \sum_{i=1, i \neq j}^{N} \xi_i |A_{ij}| \right\} > \alpha \xi_j, \quad j = 1, \ldots, N
\]  
\[
\xi_i [d_i - G_i A_{ii}] - \sum_{j=1, j \neq i}^{N} 1/2 \{ |G_i A_{ij}| + |G_j A_{ji}| \} > \alpha \xi_i,
\]  
\[
i = 1, \ldots, N,
\]  
there is a unique equilibrium $u^* = [u_1^*, \ldots, u_n^*]^T \in \mathbb{R}^N$, such that for any solution $u(t)$ of the differential equation 2.1, there holds
\[
\lim_{t \to \infty} |u(t) - u^*| = O(e^{-\alpha t}).
\]  

In fact, under either one of previous three conditions, the matrix $(D - \alpha I_n) - MG$ is an M-matrix (see Berman & Plemmons, 1979), where $M = (m_{ij})_{i,j=1}^{N}$ with entries
\[
\begin{cases}
m_{ii} = A_{ii} \\
m_{ij} = |A_{ij}| & i \neq j.
\end{cases}
\]  
By a property of M-matrix, $(D - \alpha I_n)G^{-1} - M$ is also an M-matrix. Therefore, there exists $P = \text{diag}[p_1, \ldots, p_n]$, with $p_i > 0$, $i = 1, \ldots, n$ such that
\[
p_i[d_i - \alpha]G_i^{-1} - \frac{1}{2} \sum_{j=1, j \neq i}^{N} \{ p_i |A_{ij}| + p_j |A_{ji}| \} > 0 \quad i = 1, \ldots, N.
\]  
By Gershgorin’s theorem (see Golub and Van Loan, 1990), all the eigenvalues of the matrix $P[(D - \alpha I_n)G^{-1} - M]$ are positive. Therefore, $P[(D - \alpha I_n)G^{-1} - A]$ is positive definite. This corollary is a direct consequence of theorem 1.
Remark 4. The results in corollary 1 are generalizations of those given in Chen and Amari (2001a). Moreover, it gives a sharper convergence rate.

If the activation functions are sigmoidal functions $g_i(x) = \tanh(G_i x)$, we have the following stronger results, and the proof is much simpler.

Theorem 2. Let $A = (a_{ij})_{i,j=1}^n$, $P = \text{diag}[p_1, \ldots, p_n]$, where $p_i > 0$, $i = 1, \ldots, n$, $g_i(x) = \tanh(G_i x)$. If there is a positive constant $\alpha > 0$ such that $\alpha < d_i$ for $i = 1, \ldots, N$ and the least eigenvalue $\lambda_{\text{min}} = \lambda_{\text{min}}[P[(D - \alpha I_n)G^{-1} - A]]^\ast > 0$. Then the dynamical system 2.1 has a unique equilibrium point $u^\ast$, such that

$$|u_i(t) - u_i^\ast| = O(e^{-(\alpha + \delta - \epsilon)t}) \quad i = 1, \ldots, n$$

(2.31)

where

$$\delta = \lambda_{\text{min}} \min_{i=1,\cdots,n} \left\{ \frac{g_i'(u_i^\ast)}{P_i} \right\},$$

(2.32)

and $\epsilon > 0$ is any fixed small number.

Proof. Using the same notations as in the proof of the theorem 1, define

$$L_1(t) = \sum_{i=1}^n P_i \int_0^{v_i(t)} g_i^\ast(\xi) d\xi$$

(2.33)

and

$$L_2(t) = \frac{1}{2} \sum_{i=1}^n |v_i(t)|^2.$$  

(2.34)

Under the assumptions, $v_i(x)$ $i = 1, \ldots, n$ are bounded. Therefore, there is a constant $C$ such that $G_i \geq g_i^\ast(u_i(t)) > C$ for all $t > 0$ and $i = 1, \ldots, n$. Thus,

$$\frac{1}{2} C |v_i(t)|^2 \leq \int_0^{v_i(t)} g_i^\ast(\xi) d\xi \leq \frac{1}{2} G_i |v_i(t)|^2.$$  

(2.35)

Then we can find two positive constants $C_1$ and $C_2$ such that

$$C_1 L_1(t) \leq L_2(t) \leq C_2 L_1(t).$$  

(2.36)

Using equation 2.19 and direct differentiating leads to

$$\dot{L}_1(t) = -\{g^\ast(t)\}^T P D v(t) + \{g^\ast(t)\}^T P A \{g^\ast(t)\}$$

$$- \alpha \{g^\ast(t)\}^T P v(t) - \{g^\ast(t)\}^T P (D - \alpha I_n) v(t)$$

$$+ \{g^\ast(t)\}^T PA \{g^\ast(t)\}$$

(2.37)
\begin{align*}
&\leq \alpha (g^*(t))^T P v(t) - \{g^*(t)\}^T \{P[D - \alpha I_n]G^{-1} - A]\} s^* g(t) \quad (2.38) \\
&\leq -\alpha (g^*(t))^T P v(t) - \lambda_{\min}(g^*(t))^T g^*(t) \quad (2.39) \\
&\leq -2\alpha \sum_{i=1}^{n} P_i \int_{0}^{v_i(t)} g_i^*(\xi) d\xi \\
&\quad - \lambda_{\min} \sum_{i=1}^{n} (g_i'(u_i^*))^2 |v_i(t)|^2 + o \left( \sum_{i=1}^{n} |v_i(t)|^2 \right) \quad (2.40) \\
&\leq -2\alpha \sum_{i=1}^{n} P_i \int_{0}^{v_i(t)} g_i^*(\xi) d\xi \\
&\quad - 2\lambda_{\min} \sum_{i=1}^{n} g_i'(u_i^*) \int_{0}^{v_i(t)} g_i^*(\xi) d\xi + o \left( \sum_{i=1}^{n} |v_i(t)|^2 \right) \quad (2.41) \\
&\leq -2 \left( \alpha + \lambda_{\min} \min_{i=1,\ldots,n} \left\{ \frac{g_i'(u_i^*)}{P_i} \right\} + o(1) \right) L_1(t) \quad (2.42) \\
&\leq -2(\alpha + \delta - \varepsilon)L_1(t) \quad (2.43)
\end{align*}

holds for sufficiently large $t$. Therefore,

\begin{equation}
\frac{1}{2} \sum_{i=1}^{n} |v_i(t)|^2 = L_2(t) \leq C_2 L_1(t) \leq C_2 L_1(0)e^{-2(\alpha + \delta - \varepsilon)t} \quad (2.44)
\end{equation}

and

\begin{equation}
v_i(t) = O(e^{-(\alpha + \delta - \varepsilon)t}) \quad i = 1, \ldots, n, \quad (2.45)
\end{equation}

which proves theorem 2.

**Remark 5.** Under the assumptions that the activation functions $g_i$ satisfy $g \in \tilde{G}$ and $g_i'(x) > C$, $i = 1, \ldots, n$ in any finite interval for some constant $C$. The conclusions in theorem 2 are valid too.

**Remark 6.** Liang and Si (2001) claim that the lower bound $\min_i d_i$ is attainable for a special degenerate system (linear). Theorem 2 shows that for a wide class of activation functions, the convergence rate of the nonlinear system, equation 2.1, can be even better than the lower bound $\min_i d_i$ given in Liang and Si (2001).

### 3 Comparisons

In this section, we compare our results with other work that most closely relates to ours. In this other work, the activation functions are assumed to satisfy $0 \leq \frac{g_i'(x+h) - g_i'(x)}{h} \leq G_i$. 
Zhang et al. (1999) estimated convergence rate. The main result is as follows: If there exists $0 < \alpha < \min_{1 \leq i \leq n} d_i$ such that the matrix

$$\lambda_{\text{max}}[-(D - \alpha I_n)G^{-1} - A]^s \leq 0,$$

(3.1)

then dynamical system 2.1 has a unique equilibrium point $u^*$, such that

$$|u_i(t) - u_i^*| = O(e^{-\frac{\alpha}{2}t}) \quad i = 1, \ldots, n. \quad (3.2)$$

Instead, in theorem 2, condition $0 < \alpha < \min_{1 \leq i \leq n} d_i$ can be relaxed to $\alpha \leq d_i$, and the convergence rate is

$$|u_i(t) - u_i^*| = O(e^{-\alpha \delta - \epsilon}t) \quad i = 1, \ldots, n. \quad (3.3)$$

Liang and Wu (1999) and Liang and Wang (2000) addressed absolute exponential stability under various assumptions. In most of these articles, the lower bound of the convergence rate is $O(e^{-\alpha t/2})$. Recently Liang and Si (2001) discussed the lower bound of the convergence rate and showed that in some case (when all $g_j(x) = 0$), the lower bound $\alpha = \min_i d_i$ is attainable. However, in this case, the system is degenerate and has no great interest. Theorem 2 shows that if the activation functions are sigmoidal functions, then under the condition

$$\lambda_{\text{min}}[-(D - \min_i d_i I_n)G^{-1} - A]^s > 0,$$

(3.4)

the convergence rate is

$$|u_i(t) - u_i^*| = O(e^{-(\min_i d_i + \delta - \epsilon)t}) \quad i = 1, \ldots, n, \quad (3.5)$$

which is even better than $O(e^{-(\min_i d_i)t})$.

4 Conclusion

We investigated the absolute exponential stability of a class of recurrent connected neural networks. The convergence rate given here is much sharper than that in the existing literature.

Acknowledgments

This project is supported by the National Science Foundation of China 69982003 and 60074005. This work was completed when T. Chen visited the RIKEN Brain Science Institute in Japan.
References


Received September 27, 2001; accepted May 24, 2002.