2004 International Symposium on Nonlinear Theory and its Applications (NOLTA2004)
Fukuoka, Japan, Nov. 29 - Dec. 3, 2004

Blinking connections enhance locally connected networks

Martin Hasler and Igor Belykh

Laboratory of Nonlinear Systems
School of Computer and Communication Sciences
Swiss Federal Institute of Technology Lausanne (EPFL)
Station 14, CH-1015 Lausanne, Switzerland
Email: martin.hasler@epfl.ch, igor.belykh@epfl.ch

Abstract– Locally connected networks such as cellular neural networks are advantageous for integrated circuit implementations. Certain information processing tasks, however, suffer from the restriction to local connections. This restriction can be circumvented by introducing connections that are switched on and off in a random fashion (blinking connections). Such connections do not have to be wired but can be realized by sending packets on a communication network that is needed in any case for bringing information in and out of a locally connected network. We show under what conditions the network with blinking connections has the same asymptotic behavior as the averaged network, where the blinking connections are replaced by fixed connections.

1. Introduction

Analog circuits can speed up computations considerably. They are used e.g. for massively parallel processing of visual information. Often they consist of 1-d or 2-d arrays of simple circuit modules (cells) that are interconnected by wires. Examples are cellular neural networks where usually first order dynamical systems are placed in a regular 2-d array and next nearest neighbors are connected by wires [1].

Many functions can be computed by arrays of locally connected dynamical systems, but others require global interactions of the elementary modules for efficient computation. Hard-wired all-to-all connections of n x n modules would require n^4 wires which is not realistic in most cases. In this paper we show that connections can actually be switched on and off without changing the computational function the network performs, if the (temporal) mean strength of each interaction between two cells is the same as in the non-switched network and if the switching is rapid enough. We call such connections *blinking* connections. The blinking interaction can be realized by information packages routed on a communication network that is needed anyway to transport information to and from the cells.

The switching is performed randomly at times 0, τ, 2τ, … Each blinking connection is turned on and off randomly, independently in each time interval and independently of other connections. The fact that the rapidly switched system has the same behavior as the averaged system is intuitively clear, but its proof needs a few technical assumptions. While averaging is a classical technique in the study of nonlinear oscillators, averaging for blinking system needs some special mathematical techniques for obtaining rigorous convergence proofs. Such techniques have been used in [3] for synchronization of blinking networks of chaotic dynamical systems. The details are different in this paper, but the basic idea is the same.

2. Example.

In order to illustrate the topic, we use the example of a Winner-take-all (WTA) CNN (cellular neural network). We do not claim that this is a particularly good example from an application point of view, but it illustrates well the subject of the paper. A WTA network is designed to find, using the network dynamics, the largest among a set of numbers. In the realization we refer to, each number is associated with a cell of the network. More precisely, the initial condition of each cell of the CNN is set to the value of the corresponding number. The time-evolution of the winner is such that the output of the cell corresponding to the largest number converges to 1 and all other outputs converge to -1, in normalized variables.

It is not difficult to see that in an only locally connected CNN it is not possible to realize the WTA function, at least not in this way. In [4], the design equations for a globally connected WTA CNN are given. Actually, the point is made in [4] that one does not need all-to-all connections, but only an additional sum module which has inputs and outputs that are connected to all cells. This reduces the number of wired connections from n^4 to $2n^2$, which is even less than the $5n^2$ connections in a nearest neighbor connected CNN. However most of these $2n^2$ connections go across a large part of the circuit and thus still pose problems of realization. In our blinking WTA CNN the n^4-$5n^2$ switched connections can just as well be reduced, but for the simplicity and generality of exposition, we do not want to take advantage of the peculiarities of the example.

$$\frac{dx_i}{dt} = -x_i + (1+\delta)y_i - \alpha\sum_{j=1}^{n} y_j + \kappa$$
$$y_i = f(x_i) = \begin{cases} 1 & for \;\; x_i > 1 \\ x_i & for \;\; |x_i| \le 1 \\ -1 & for \;\; x_i < -1 \end{cases} \qquad (1)$$

Following [4], we consider the piecewise linear dynamical system, whose state equations are given in (1), for i = 1, … , N=n[2]. If the parameters α,δ and κ are chosen suitably, there are exactly N asymptotically stable equilibrium points, one in each linear region where one of the output signals y_i is 1 and the other N-1 are -1. All other equilibrium points are unstable. The basin of attraction of the asymptotically stable equilibrium point with outputs

$$y_i = 1, \quad y_j = -1, \; j \neq i \tag{2}$$

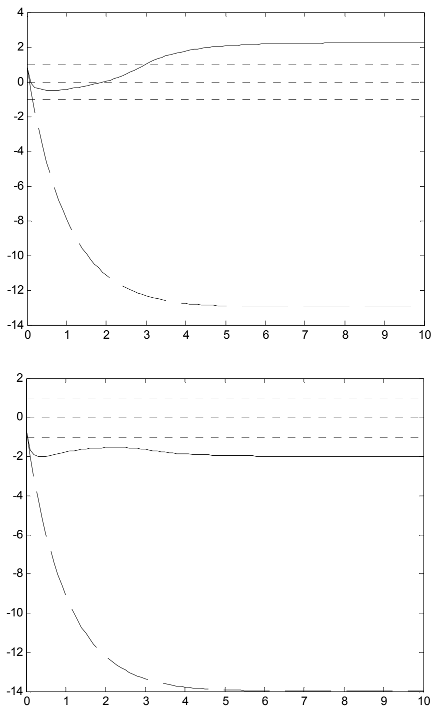
is composed of the state vectors $\boldsymbol{x}$ with

$$x_i \; > \; x_j \qquad for\ all\ \ j \neq i \tag{3}$$

Hence, indeed the CNN realizes the WTA function by its internal dynamics. In Fig.1, two components of the state trajectory $\mathbf{x}(t)$ of a 4 x 4 WTA CNN are shown, for the initial conditions

$$\begin{array}{cccc} 0.3644 & 0.3958 & 0.1871 & 0.2898 \\ -0.3945 & -0.2433 & -0.0069 & 0.6359 \\ 0.0833 & 0.7200 & 0.7995 & 0.3205 \\ -0.6983 & 0.7073 & 0.6433 & -0.3161 \end{array} \tag{4}$$

For the fully connected CNN (solid line), one can see that the state of the cell with largest initial condition (Fig.1a) converges to a value higher than 1 (output 1), whereas the state of any other cell (Fig.1b) converges to a value lower than -1 (output -1). When only nearest neighbor connections are used (dashed line), the network fails to detect the largest initial value.



a)

b)

Fig.1. Trajectory component $x_i(t)$ for a 4x4 WTA CNN, with a = , d = , k = (solid line). Fig.1a): component of the cell (3,3) whose initial condition has the maximal value. Fig.1b): Component of the cell (4,1). Dashed line: trajectory components of the CNN with the same parameters, but only nearest neighbor connections.

3. Convergence to an asymptotically stable equilibrium

Consider a dynamical system of the more general form

$$\begin{aligned} \frac{dx_i}{dt} &= -x_i + \sum_{i=1}^{N} A_{ij} y_j + I_i y \\ y_i &= f(x_i) \end{aligned} \tag{5}$$

where f is continuously differentiable, and strictly increasing from -1 at -∞ to +1 at +∞ (“sigmoid function”).

It is well-known that for a symmetric coupling matrix A almost all trajectories converge to an asymptotically stable equilibrium point. This fact is established using the Lyapunov function

$$W(\mathbf{y}) = \sum_{i=1}^{N} f^{-1}(y_i) - \frac{1}{2} \sum_{i,j=1}^{N} A_{ij} y_i y_j - \sum_{i=1}^{N} I_i y_i \tag{6}$$

It is not difficult to see that along any solution of (5)

$$\frac{d}{dt} W(\mathbf{y}(t)) = -\sum_{i=1}^{N} \frac{df}{dx}\left(f^{-1}(y_i(t))\right) \left[\frac{\partial W}{\partial y_i}(\mathbf{y}(t))\right]^2 \leq 0 \tag{7}$$

which implies the convergence property.

Actually, the CNN with the piecewise-linear nonlinearity (1) does not fit exactly into this framework, because f is not invertible. The reasoning would have to be adapted to this case, but for the sake of simplicity, we stick to the case with sigmoid f.

4. Blinking connections

We divide the time-axis into intervals of length τ and number them consecutively. Thus, interval k is defined by $(k-1)\tau \le t < k\tau$. During each time interval, the circuit connections remain constant, but for different time intervals, in general the circuit connections are different. In the spirit of CNN’s, however, nearest neighbor connections remain constant all the time. For each non-nearest-neighbor pair of cells i,j we introduce a discrete time binary signal s_{ij}^k whose value is 1 if the connection is switched on during the k-th time interval, and 0 otherwise. We call these signals *switching sequences*. We extend the switching sequences to continuous time switching signals by setting $s_{ij}(t) = s_{ij}^k$ in the k-th time interval.

Given a switching signal, we can write the equations of the blinking CNN as

$$\frac{dx_i}{dt} = -x_i(t) + \sum_{j\ nn\ of\ i} A_{ij} y_i(t) + \\ + \sum_{j\ not\ nn\ of\ i} B_{ij} s_{ij}(t) y_j(t) + I_i \qquad (8)$$
$$y_i = f(x_i)$$

The choice of the switching sequences is performed at random. More precisely, we consider the set of identically distributed independent random variables

$$S_{ij}^k, \quad i \text{ and } j \text{ not nearest neighbors}, \quad k = 1, 2, \cdots \qquad (9)$$

which take the value 1 with probability *p* and the value 0 with probability 1-*p*. The switching sequences are instances of this stochastic process.

Due to the random choice of switching, the trajectories of (8) become also random processes. Intuitively, if the switching is much faster than the intrinsic time constants of the trajectories of a corresponding non-switched system, we expect the trajectories of the blinking system to follow closely the trajectories of the non-switched system whose connections have the average value with respect to (8), the *averaged system*:

$$\frac{d\xi_i}{dt} = -\xi_i(t) + \sum_{j\ nn\ of\ i} A_{ij} \eta_i(t) + \\ + \sum_{j\ not\ nn\ of\ i} B_{ij} p \eta_j(t) + I_i \qquad (10)$$
$$\eta_i = f(\xi_i)$$

Since we would like the blinking system to have the same behavior as the fully connected system (5), we have to set

$$B_{ij} = A_{ij} / p \qquad (11)$$

In particular, we expect that the trajectories of the blinking system with (11) get close to the same asymptotically stable equilibria of the averaged system, as $t \to +\infty$, when both start from the same initial condition. We shall sketch the proof that by choosing the switching time τ sufficiently small, we can make the probability that this is not the case arbitrarily small.

5. Asymptotic behavior of the connections of the blinking CNN.

In order to limit the exposition to the essentials, we consider a more general blinking equation:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}(t), \mathbf{s}(t)) \qquad (12)$$

where $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{F} : \mathbb{R}^{N+M} \to \mathbb{R}^N$, $\mathbf{s} \in \mathbb{R}^M$ and **F** is continuously differentiable. Consider the averaged equation

$$\frac{d\xi}{dt} = \mathbf{\Phi}(\xi(t)) \qquad (13)$$

where $\xi \in \mathbb{R}^N$, $\mathbf{\Phi} : \mathbb{R}^N \to \mathbb{R}^N$ and

$$\mathbf{\Phi}(\mathbf{x}) = \sum_{\mathbf{s} \in \{0,1\}^M} \mathbf{F}(\mathbf{x}, \mathbf{s}) P(\mathbf{s}) = \\ = \sum_{\mathbf{s} \in \{0,1\}^M} \mathbf{F}(\mathbf{x}, \mathbf{s}) \prod_{i=1}^{M} \left(p s_i + (1-p)(1-s_i) \right) \qquad (14)$$

Now, suppose $\bar{\xi}$ is an asymptotically stable equilibrium point of (13). Furthermore, we suppose there is a Lyapunov function $W : \mathbb{R}^N \to \mathbb{R}$ with the following properties:

a) W is continuously differentiable
b) W has a local minimum in $\bar{\xi}$, for convenience $W(\bar{\xi}) = 0$
c) If ξ(t) is a solution of (13) then $\frac{d}{dt} W(\xi(t)) \le 0$
d) The level sets $\{\xi \mid W(\xi) \le V\}$ are compact

In addition, suppose there are two constants $0 < V_0 < V$ and $\gamma > 0$ such that

e) The connected component U of $\{\xi \mid W(\xi) < V\}$ that contains $\bar{\xi}$ is contained in the basin of attraction of $\bar{\xi}$.
f) Within U, but outside of $U_0 = \{\xi \mid \xi \in U, W(\xi) < V_0\}$, we have along any solution ξ(t) of (13) the inequality

$$\frac{d}{dt} W(\xi(t)) \le -\gamma \qquad (15)$$

The meaning of the last two conditions is to consider a level set Uof W that has some (small) distance from other local minima of W such that W decreases at a minimal speed except close to $\bar{\xi}$.

Now consider a solution **x**(t) of the blinking system (12) starting from the initial state $\mathbf{x}(0) \in U$ and a solution of the averaged system (13) starting from the same initial condition ξ(0) = **x**(0). Then, by (15),

$$W(\xi(t)) \le W(\xi(0)) - \gamma t \qquad (16)$$

We want to show, that with high probability, W also decreases along the solution **x**(t) of the blinking model.

Since W is continuously differentiable, it has a global Lipschitz L_W constant on U and thus

$$\left|W\left(\mathbf{x}(t)\right)-W\left(\xi(t)\right)\right| \;\le\; L_W\left\|\mathbf{x}(t)-\xi(t)\right\| \qquad (17)$$

On the other hand

$$\begin{aligned}
\left\|\mathbf{x}(t)-\xi(t)\right\| \;&\le\; \left\|\int_0^t \mathbf{F}\left(\mathbf{x}(\rho),\mathbf{s}(\rho)\right)d\rho \;-\; \int_0^t \mathbf{\Phi}\left(\xi(\rho)\right)d\rho\right\| \\
&\le\; \left\|\int_0^t \left[\mathbf{F}\left(\mathbf{x}(\rho),\mathbf{s}(\rho)\right)-\mathbf{F}\left(\mathbf{x}(0),\mathbf{s}(\rho)\right)\right]d\rho\right\| \;+ \\
&\quad+\; \left\|\int_0^t \mathbf{F}\left(\mathbf{x}(0),\mathbf{s}(\rho)\right)d\rho \;-\; \mathbf{\Phi}\left(\mathbf{x}(0)\right)\right\| \;+ \\
&\quad+\; \left\|\int_0^t \mathbf{\Phi}\left(\xi(\rho)\right)d\rho \;-\; \mathbf{\Phi}\left(\xi(0)\right)\right\|
\end{aligned} \qquad (18)$$

One can show with similar methods as in [3] that the probability P_τ that the second term in (18) is larger than λt is converging to 0 as the switching time $\tau \to 0$, for any $\lambda,t>0$. Using Lipschitz constants L_F and L_Φ and bounds C_F and C_Φ for the functions **F** and **Φ** on U, we obtain with probability $1 - P_\tau$ that

$$\begin{aligned}
W\left(\mathbf{x}(t)\right) \;&\le\; W\left(\xi(t)\right) \;+\; \left|W\left(\mathbf{x}(t)\right) \;-W\left(\xi(t)\right)\right| \\
&\le\; W\left(\mathbf{x}(0)\right) \;-\; \gamma t \;+\; L_W\left(\frac{L_F C_F + L_\Phi C_\Phi}{2}t^2 + \lambda t\right)
\end{aligned} \qquad (19)$$

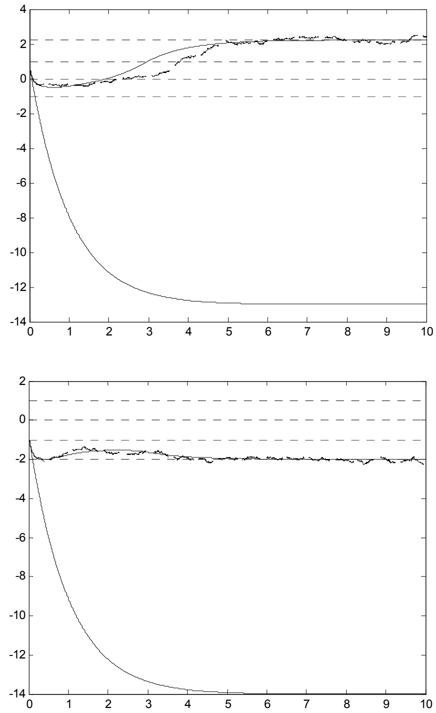
Choosing t and λ sufficiently small we obtain that with probability P_τ we have $W(\mathbf{x}(t)) < W(\mathbf{x}(0)) - \gamma t/2$. With some additional reasoning similar to [3] this proves

Theorem:

With the above definitions and conditions, any solution of the blinking model starting in U converges to the neighborhood U_0 of the equilibrium point of the averaged equation in U, with a probability that converges to 1 as the switching period converges to zero.

6. Back to the example

The various conditions can be verified for our example system, with a sigmoid nonlinearity. In Fig.2 we show the trajectory for the blinking system (with piecewise nonlinearity), starting from the same initial conditions as in (4). Indeed, the blinking network performs the WTA function correctly for this switching sequence. According to the theorem, only a very small portion of the switching sequences could lead to a misidentification of the largest initial value.



a)

b)

Fig.2. Same Figure as Fig.1, but in addition trajectory of the blinking network with a switching time τ = 0.0005 and p = 0.1.

Acknowledgments

This work has been financially supported by the grant nr. 2100-065268 of the Swiss National Science Foundation.

References

[1] L.O.Chua and T. Roska, *Cellular Neural Networks and Visual Computing,* Cambridge University Press, Cambridge, UK, 2002.

[2] N.M.Bogoliubov and Y.A.Mitropolsky, *Asymptotic Methods in the Theory of Non-Linear Oscillators*, Hindustan Publishing Co., 1961.

[3] I. Belykh, V. Belykh, M. Hasler,"Blinking model and synchronization in small-world networks with a time-varying coupling", *Physica D*, Vol. 195, Nr 1-2, pp 188-206, 2004.

[4] G.Seiler, J.A.Nossek, "Winner-take-all cellular neural networks", *IEEE Transactions on Circuits and Systems II*, pp. 184-190, 1993.