

# ON GLOBAL ASYMPTOTIC STABILITY OF FULLY CONNECTED RECURRENT NEURAL NETWORKS

Danilo P. Mandic

Jonathon A. Chambers

Milorad M. Bozic

School of Information Systems,  
University of East Anglia,  
Norwich, UK.  
d.mandic@uea.ac.uk

Dept. of Electrical Engineering,  
Imperial College,  
London, UK.  
j.chambers@ic.ac.uk

Dept. of Electrical Engineering  
University of Banjaluka  
Banjaluka, Bosnia–Herzegovina.  
mbozic@etf-bl.rstel.net

## ABSTRACT

Conditions for Global Asymptotic Stability (GAS) of a non-linear relaxation process realized by a Recurrent Neural Network (RNN) are provided. Existence, convergence, and robustness of such a process are analyzed. This is undertaken based upon the Contraction Mapping Theorem (CMT) and the corresponding Fixed Point Iteration (FPI). Upper bounds for such a process are shown to be the conditions of convergence for a commonly analyzed RNN with a linear state dependence.

## 1. INTRODUCTION

Numerous problems in engineering can be formulated as optimization problems with or without constraints. The real-world problems cast into such a framework include the travelling salesman, transportation, network flow, and associative memory problems [1]. Neural networks have also been successfully applied in: solving systems of linear equations, nonlinear programming, dynamic programming, and network flow computations [2, 3].

In most applications, there is a need for a robust method that works efficiently across a broad spectrum of optimization problems. Global Asymptotic Stability (GAS) and robust GAS theory, which have been widely considered in the context of linear systems, have been shown to provide such robustness. In engineering, formulation of the GAS statement stems from a general Autoregressive Moving Average (ARMA) linear system equation for the case of a zero exogenous input vector.

Within nonlinear systems, such as Recurrent Neural Networks (RNN)s, the corresponding general Nonlinear ARMA, NARMA(p,q) system is given by [4]

$$x(k) = e(k) + h(x(k-1), \dots, x(k-p), e(k-1), \dots, e(k-q)) \quad (1)$$

where  $p$  denotes the order of the Autoregressive (AR) part,  $q$  denotes the order of the Moving Average (MA) part, with

some nonlinear function  $h(\cdot)$ .

Traditionally, the analysis of neural networks for optimization has been based on the Lyapunov stability theory of dynamic systems. Recent developments in neural networks for constrained optimization include Lagrangian relaxation, and a connection with the theory of chaos.

Here, we embark upon the results from [5], and provide the conditions of convergence of a relaxation process when realized by a recurrent neural network. Uniformity, and inherent robustness, of such a GAS process is preserved based upon the contractivity of the logistic activation function of a neuron and the fixed point iteration around a current point in the state space of a recurrent neural network.

## 2. RECURRENT NEURAL NETWORKS

The structure of a fully connected RNN is shown in Figure 1. The  $N$  neurons (nodes) are depicted by circles, and incorporate the operation  $\Phi(\text{sum of inputs})$ . For the  $n$ th neuron, its weights form a  $(p + N + 1) \times 1$  dimensional weight vector  $\mathbf{w}_n^T = [w_{n,1}, \dots, w_{n,p+N+1}]$ , where  $p$  is the number of external inputs  $s$ , and  $N$  is the number of feedback connections, one remaining element of the weight vector  $\mathbf{w}$  being the bias input weight. The weight vectors are encompassed in the weight matrix  $\mathbf{W}$ . Such a network is called a Fully Connected Recurrent Neural Network (FCRNN) [6]. The following equations fully describe the FCRNN

$$\begin{aligned} y_n(k) &= \Phi(v_n(k)), \quad n = 1, 2, \dots, N \\ v_n(k) &= \sum_{l=1}^{p+N+1} w_{n,l}(k)u_l(k) \\ \mathbf{u}_n^T(k) &= [y_1(k-1), \dots, y_N(k-1), 1, s(k-1), \dots, s(k-p)] \end{aligned} \quad (2)$$

where the  $(p + N + 1) \times 1$  dimensional vector  $\mathbf{u}$  comprises both the external and feedback inputs to a neuron, with vector  $\mathbf{u}$  having “unity” for the constant bias input.

Depending on the particular application, the network has

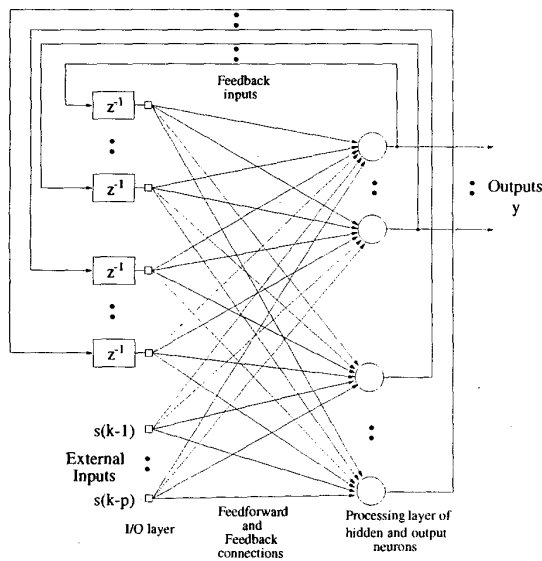


Figure 1: Single recurrent neural network

one or more output neurons. In signal processing, dynamic neural networks are mostly designed to have only a unique equilibrium point, being globally stable, so as to avoid the risk of spurious responses or the problem of local minima. The system given in (2), with only one output, can be written in the state space form as

$$\begin{aligned} \mathbf{Y}(k+1) &= \Phi(\mathbf{Y}(k), \mathbf{W}(k), \mathbf{s}(k)), \\ y_1(k) &= \mathbf{H}\mathbf{Y}(k) \end{aligned} \quad (3)$$

and hence (1) can be approximated by an RNN with  $h$  becoming an activation function of the neuron, which is most commonly the logistic function denoted by

$$\Phi(v) = \frac{1}{1 + e^{-\beta v}} \quad (4)$$

and will be assumed in (2). The Bounded Input Bounded Output (BIBO) stability of (2) for a finite gain  $\beta$  is preserved due to the saturation type nonlinearity in (4).

When applying the RNN given in Figure 1 in the GAS relaxation, the goal is to re-iterate (3) for a fixed external input vector  $\mathbf{s}$  and weight matrix  $\mathbf{W}$ , until reaching the fixed point. In order to provide the conditions of stability and convergence of such a procedure, it is important to introduce the concepts of contraction mapping and fixed point iteration.

### 3. LOGISTIC ACTIVATION FUNCTION AND CONTRACTION MAPPING

The logistic nonlinear activation function of a neuron is shown to provide a **contraction mapping** on  $[a, b] \in \mathbb{R}$  for  $0 <$

$\beta < 4$  [5]. Using the Contraction Mapping Theorem (CMT), an optimization problem  $F(\xi) = 0$  can be iteratively solved as  $\xi(k) = f(\xi(k-1))$ , whose solution is a fixed point  $\xi^* = f(\xi^*)$ . Such a process is called a fixed point iteration (FPI).

**Definition 1**  $\mathbf{x}^* \in \mathbb{R}^d$  is said to be a fixed point of  $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$  if  $F(\mathbf{x}^*) = \mathbf{x}^*$ .

The dynamics of either an associative or hierarchical neural network can boil down to the discovery of a fixed point (a stable equilibrium) or contraction to the already established fixed point in a discrete dynamical system. The domain of attraction of a stable equilibrium point of a dynamical system and the convergence rates are also of interest in control theory, and associative memories. Based upon the domain of attraction, it is possible to analyze various types of stability, such as asymptotic stability and global asymptotic stability.

### 4. CONDITIONS FOR CONVERGENCE OF RNN RELAXATION

GAS relaxation for a recurrent perceptron is provided in [5], where the condition was derived based upon the  $\|\cdot\|_1$  norm of the weight vector and the value of the slope  $\beta$  of the nonlinear activation function  $\Phi$ . An RNN optimization model describes the process by which a system changes its state, e.g.  $\mathbf{Y}(k) \rightarrow \mathbf{Y}(k+1)$ . Let  $\Phi(k, k_0, \mathbf{Y}_0)$  denote the trajectory of a GAS state change for all  $k \geq k_0$ , with  $\Phi(k_0, k_0, \mathbf{Y}_0) = \mathbf{Y}_0$ . If  $\Phi(k, k_0, \mathbf{Y}^*) = \mathbf{Y}^*$  for all  $k \geq 0$ , then  $\mathbf{Y}^*$  is called an *equilibrium point*. The largest set  $\mathcal{D}(\mathbf{Y}^*)$  for which this is true is called the *domain of attraction* of the equilibrium  $\mathbf{Y}^*$ . If  $\mathcal{D}(\mathbf{Y}^*) = \mathbb{R}^n$  and if  $\mathbf{Y}^*$  is asymptotically stable, then  $\mathbf{Y}^*$  is said to be asymptotically stable in large or *globally asymptotically stable*.

The following theorem gives the necessary condition for the existence of a fixed point.

**Lemma 1 (Brower's Fixed Point Theorem [7])** Let  $H^n = [a, b]^n$  be a closed subset of  $\mathbb{R}^n$  and  $f : H^n \rightarrow H^n$  be a continuous vector-valued function. Then  $f$  has at least one fixed point in  $H^n$ .

For the system under consideration (3), the fixed points of the function  $\mathbf{Y}(k+1) = \Phi(\mathbf{Y}_0(k), \mathbf{W}, \mathbf{s}) = \Phi(\mathbf{Y}(k))$  are the equilibrium points of the system for a given input  $\mathbf{s}$  and the connection weight matrix  $\mathbf{W}$ . Since the logistic function is a continuous, bounded function, the existence of the equilibrium points of the system is then obtained using the above Lemma. In the sequel, we provide the conditions of convergence of the GAS process (3) towards a fixed point.

## 5. GAS FOR FULLY CONNECTED RECURRENT NEURAL NETWORKS

Let  $\mathbf{Y}_i = [y_1^i, \dots, y_N^i]^T$  be a vector comprising the outputs of an RNN at iteration  $i$ . Then, by a CMT in  $\mathbb{R}^N$ , an iterative process applied to the general RNN converges, if there is a hypercube  $\mathbf{M} = [a, b]^N \subset \mathbb{R}^N$  such that

i)  $\Phi : \mathbf{M} \rightarrow \mathbf{M}$

ii) If for some norm  $\|\cdot\|$ ,  $\exists \gamma < 1$ , such that  $\|\Phi(\mathbf{x}) - \Phi(\mathbf{y})\| \leq \gamma \|\mathbf{x} - \mathbf{y}\|$ ,  $\forall \mathbf{x}, \mathbf{y} \in \mathbf{M}$  the equation  $\mathbf{x} = \Phi(\mathbf{x})$  has a unique solution in  $\mathbf{M}$ , denoted by  $\mathbf{x}^*$ , and the iteration  $\mathbf{x}_{i+1} = \Phi(\mathbf{x}_i)$  converges to  $\mathbf{x}^*$  for any starting value  $\mathbf{x}_0 \in \mathbf{M}$ .

Actually, since the function  $\Phi$  in this case is a multivariate function,  $\Phi = \text{diag}(\Phi_1, \dots, \Phi_N)$ , for the  $i$ th iteration of (3), we have a set of mappings

$$\begin{aligned} y_1^i &= \Phi_1(\mathbf{u}_{i-1}^T \mathbf{w}_1) \\ &\dots \dots \\ y_N^i &= \Phi_N(\mathbf{u}_{i-1}^T \mathbf{w}_N) \end{aligned} \quad (5)$$

As the convergence under contraction mapping depends on the choice of norm, to circumvent that, we look for the condition of convergence based upon the features of the nonlinear activation function  $\Phi$ .

Extending the approach from [5], let us denote the Jacobian of  $\Phi$  by  $\mathbf{J}$ . If  $\mathbf{M} \in \mathbb{R}^N$  is a convex set, and  $\Phi$  is continuously differentiable on  $\mathbf{M} = [a, b]^N \in \mathbb{R}^N$ , and satisfies the conditions of the CMT, we have

$$\max_{\mathbf{z} \in \mathbf{M}} \|\mathbf{J}(\mathbf{z})\| \leq \gamma < 1 \quad (6)$$

This stems from the scalar case, where by the intermediate value theorem, for the conditions of the CMT on  $[a, b] \in \mathbb{R}$ , there is a point  $\xi \in \langle a, b \rangle$  such that  $|\Phi(b) - \Phi(a)| \leq \Phi'(\xi)|b - a|$ . As for the logistic function  $\max_{x \in \mathbb{R}} (\Phi'(x)) = \frac{\beta}{4}$ , the conditions for a GAS relaxation for a recurrent perceptron [5] can be extended to the multidimensional case, as follows.

**Claim 1** *The GAS relaxation based upon a RNN realization converges, if for the  $k$ th neuron ( $k = 1, \dots, N$ ), and  $i$ th iteration, and  $w_{k,j} \in \mathbf{W}$ ,  $k, j = 1, \dots, N$*

$$\max_{k=1, \dots, N} \Phi'(w_{k,k} y_k^i) < 1 \Leftrightarrow w_{k,k} < \frac{4}{\beta} \quad (7)$$

The proof follows naturally from the above discussion, and the fact that the condition of convergence for the  $\Phi'_i$   $i = 1, \dots, N$  is  $\Phi' \leq \gamma < 1$ .

However, although the previous Claim gives the conditions of convergence, it does not establish the conditions for uniform convergence. Notice that it is the uniform convergence that preserves robustness of an iterative approach.

**Theorem 1** *For the FPI process (3) to converge uniformly, it is necessary that*

$$\max_{i=1, \dots, N} \Phi'_i \sum_{k=1}^N |w_{k,k}| < 1 \quad (8)$$

which for the logistic activation function corresponds to

$$\sum_{k=1}^N |w_{k,k}| < \frac{4}{\beta}, \quad \beta = \max_{i=1, \dots, N} \beta_i \quad (9)$$

and

$$\max_{1 \leq k \leq N} |w_{k,k}| < \frac{4}{N \max_{1 \leq i \leq N} \beta_i} \quad (10)$$

**Proof:**

In order to prove Theorem 1, recall that for the case of a single recurrent perceptron, the condition for a uniform GAS was  $\sum_{j=1}^N |w_j| < \frac{4}{\beta} \Leftrightarrow \|\mathbf{w}\|_1 < \frac{4}{\beta} = \frac{1}{\Phi'_{\max}}$ . However, for a network of  $N$  neurons, it is possible to have a convergent FPI, even if some of the neurons violate the previous conditions. When it comes to the uniform convergence, it is important that the processes at every neuron converge uniformly. This means that  $\sum_{j=1}^N |w_{i,j}| < \frac{1}{\Phi'_{\max}}$  for each neuron  $i = 1, 2, \dots, N$ . As the FPI process (3) is auto-regressive, for the uniform convergence at every particular neuron, it is only the diagonal weights which are related to the slope  $\beta_i$  in the particular nonlinear activation function  $\Phi_i$   $i = 1, \dots, N$  and have an influence on the convergence in the FPI sense.  $\square$

### 5.1. Influence of the Number of External Input Signals

It is important to see whether the number of external input signals has an influence on the GAS.

**Claim 1** *The rate of convergence of relaxation in RNNs does not depend on the number of external input signals.*

**Proof:**

Notice that for every neuron, during the FPI (3), for the  $(k+1)$ th iteration, we have

$$y_i(k+1) = \Phi_i(y_i(k) + \zeta), \quad i = 1, 2, \dots, N \quad (11)$$

where  $\zeta$  embodies all the terms  $\{w_{i,j} * s_j\}$  and  $\{w_{i,j} y_j\}$ ,  $j \neq i$ . Since the external inputs  $\mathbf{s}$  and weights  $\mathbf{W}$  are kept constant, the number of the external inputs does not have an influence on the convergence process of the GAS relaxation. It only determines the value of the fixed point, but not the rate of convergence. Therefore, the convergence properties of the iteration

$$\mathbf{Y}_{i+1} = \Phi(\mathbf{Y}_i, \mathbf{s}, \mathbf{W}) \quad (12)$$

do not depend on the number of external input samples.  $\square$  The rate of convergence is defined by the ratio  $\frac{\|\mathbf{Y}(k) - \mathbf{Y}^*\|}{\|\mathbf{Y}(k-1) - \mathbf{Y}^*\|}$ . From the CMT, the rate of convergence of uniform GAS is linear with the convergence rate defined by the norm of the Jacobian of  $\Phi$ . This means that, say for neuron  $j$ , its convergence rate is defined by  $\Phi'(y_j^*)$ .

**Claim 2** *The convergence of a uniform GAS, when performed by a FCRNN, is linear, with the convergence rate defined by the norm of the Jacobian of the nonlinear activation function  $\Phi$ .*

## 6. UPPER BOUNDS FOR GAS RELAXATION WITHIN AN FCRNN

Systems of the form

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\sigma[\mathbf{W}\mathbf{x}(k) + \mathbf{s}] = f(\mathbf{x}(k)) \quad (13)$$

have been widely considered [8]. Here,  $\mathbf{x}$  is the state vector of the network, and  $\sigma(\cdot)$  is a vector of nonlinear activation functions. The weight matrix  $\mathbf{W}$  of (3) can be split up into the feedback part (index  $a$ ) and the feedforward part (index  $b$ ), which gives  $\mathbf{Y}(k+1) = \Phi(\mathbf{W}_a \mathbf{Y}(k) + \mathbf{W}_b \mathbf{s}(k))$ . Obviously, this can degenerate into the form (13). Namely, for a contractive activation function  $\Phi$ , the relation

$$\Phi(a+b) < \Phi(a) + \Phi(b) < a + \Phi(b) \quad (14)$$

holds, which is also used in the analysis of *a posteriori* algorithms [9]. Therefore, widely available results for system (13) provide the upper bound for the system described in this article, providing also the measure of robustness of the GAS relaxation.

**Example 1** *Show that the GAS relaxation for a fully connected RNN with three neurons and six external input signals converges to a fixed point in the FPI sense.*

If the initial values  $\mathbf{s}_0$  and  $\mathbf{W}$  are chosen at random, the iteration process converges towards a unique fixed point  $\mathbf{Y} = [0.473205, 0.501625, 0.553336]$ , as shown in Table 1. The

Table 1: Fixed point iterates for a fully connected RNN

States	$y_1$	$y_2$	$y_3$
$\mathbf{Y}_0$	0.211518	0.999116	0.153604
$\mathbf{Y}_1$	0.478783	0.506919	0.564501
$\mathbf{Y}_2$	0.476030	0.503071	0.553442
$\mathbf{Y}_3$	0.474018	0.501753	0.553435
$\mathbf{Y}_4$	0.473603	0.501644	0.553351
$\mathbf{Y}_5$	0.473534	0.501627	0.553338
$\mathbf{Y}_6$	0.473522	0.501625	0.553336
$\mathbf{Y}_7$	0.473521	0.501625	0.553336

iteration process for each of the three neurons in the RNN is shown in an appropriate column in Table 1. For every neuron, the iteration process converges, and the output neuron reaches its fixed point last.

## 7. CONCLUSIONS

Understanding the fixed point potential of recurrent neural networks (the number, stability, and bifurcations of fixed points) brings some light into practical realizations of optimization problems. We have provided the relationships between the number of neurons in the RNN, the slope in the activation function  $\beta$ , and some features of  $\mathbf{W}$ , which guarantee convergence of a relaxation process realized by fully connected recurrent neural networks. The results obtained can be applied when recurrent neural networks are used as computational models, in particular, as optimization models. The results can also be used as stability analysis tools for some classes of nonlinear control systems.

## 8. REFERENCES

- [1] Y. Takahashi, "A neural network theory for constrained optimization," *Neurocomputing*, vol. 24, pp. 117–161, 1999.
- [2] A. Cichocki and R. Unbehauen, *Neural networks for optimization and signal processing*. Wiley, 1993.
- [3] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-dynamic programming*. Belmont, MA: Athena Scientific, 1996.
- [4] M. B. Priestley, *Non-linear and Non-stationary Time Series Analysis*. Academic Press, London, 1991.
- [5] D. P. Mandic and J. A. Chambers, "Global asymptotic stability of nonlinear relaxation equations realised through a recurrent perceptron," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP-99)*, vol. 2, pp. 1037–1040, 1999.
- [6] R. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Computation*, vol. 1, pp. 270–280, 1989.
- [7] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*. Addison-Wesley, 1989.
- [8] L. Jin, P. N. Nikiforuk, and M. M. Gupta, "Absolute stability conditions for discrete-time recurrent neural networks," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 954–964, 1994.
- [9] D. P. Mandic and J. A. Chambers, "Relations between the a priori and a posteriori errors in nonlinear adaptive neural filters." Accepted for *Neural Computation*, 1999.