Loughborough
University

This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

GLOBAL ASYMPTOTIC CONVERGENCE OF NONLINEAR RELAXATION EQUATIONS REALISED THROUGH A RECURRENT PERCEPTRON

Danilo P. Mandic and Jonathon A. Chambers

Signal Processing Section, Department of Electrical and Electronic Engineering
Imperial College, London, UK
d.mandic@ic.ac.uk, j.chambers@ic.ac.uk

ABSTRACT

Conditions for Global Asymptotic Stability (GAS) of a nonlinear relaxation equation realised by a Nonlinear Autoregressive Moving Average (NARMA) recurrent perceptron are provided. Convergence is derived through Fixed Point Iteration (FPI) techniques, based upon a contraction mapping feature of a nonlinear activation function of a neuron. Furthermore, nesting is shown to be a spatial interpretation of an FPI, which underpins a recently proposed Pipelined Recurrent Neural Network (PRNN) for nonlinear signal processing.

1. INTRODUCTION

Global Asymptotic Stability (GAS) and robust GAS have been widely considered in the context of linear systems. Recently, the problem of GAS and robust GAS of a time–variant m–th order linear difference equation

$$
\begin{aligned}
y(k) &= \mathbf{a}^T(k)\mathbf{y}(k-1) \\
&= a_1(k)y(k-1) + \cdots + a_m(k)y(k-m) \quad (1)
\end{aligned}
$$

has been extensively investigated [1, 2]. Equation (1) is, in fact, a relaxation equation, which represents an autonomous system, which under certain conditions converges [1]. Equation (1) stems from a general Autoregressive Moving Average (ARMA) linear system equation

$$
\mathbf{Y}(k+1) = \mathbf{A}(k)\mathbf{Y}(k) + \mathbf{B}(k)\mathbf{u}(k) \quad (2)
$$

for the case of a zero exogenous input vector $\mathbf{u}(k) = \mathbf{0}$, $\forall k$ [3]. It is, in fact, an iterative processing technique, which runs on fixed (zero) external data applied to the system. The GAS analysis for the linear system (1) is naturally based entirely upon the values of the parameter vector $\mathbf{a}(k)$, and is shown to converge for $\| \mathbf{a}(n) \|_1 < 1$ [1].

When it comes to the corresponding general Nonlinear Autoregressive Moving Average (NARMA) equation,

the NARMA(p,q) system is given by [4]

$$
\begin{aligned}
x(k) &= e(k) + h\left(x(k-1), \cdots, x(k-p),\right. \\
&\qquad \left. e(k-1), \cdots, e(k-q)\right) \quad (3)
\end{aligned}
$$

where p denotes the order of the Autoregressive (AR) part, q denotes the order of the Moving Average (MA) part, with some nonlinear function $h(\cdot)$; the GAS analysis of the relaxation of (3) is still in its infancy. Here, we provide the conditions of convergence of (3) when realised by a NARMA recurrent perceptron, and show that based upon nesting, a Pipelined Recurrent Neural Network (PRNN) represents a spatial realisation of an iterative, relaxive procedure based on (3).

2. NARMA RECURRENT NEURAL NETWORKS

A number of stochastic signal models have been developed so as the estimate $\hat{x}(k) \approx E\left(x(k)\right)$ in (3) exhibits certain behaviour. Since the innovation process $\{e(k)\}$ is not observable, the residual $\hat{e}(k) = x(k) - \hat{x}(k)$, is an approximation which can be used instead of $e(k)$ in (3). The NARMA scheme from (3), can be further approximated as [5]

$$
\begin{aligned}
y(k) &= \hat{x}(k) = h(x(k-1), \cdots, x(k-p), \\
&\qquad \hat{e}(k-1), \cdots, \hat{e}(k-q)) \\
&= H(x(k-1), \cdots, x(k-p), \\
&\qquad y(k-1), \cdots, y(k-q)) \quad (4)
\end{aligned}
$$

where H is some new, nonlinear smooth function. The last equation in (4) is now suitable for the RNN implementation, with H becoming an activation function of the neuron, which is typically the logistic function denoted by

$$
\Phi(v) = \frac{1}{1 + e^{-\beta v}} \quad (5)
$$

and will be assumed in (4). The structure to realise (3) by a recurrent perceptron is shown in Figure 1. The
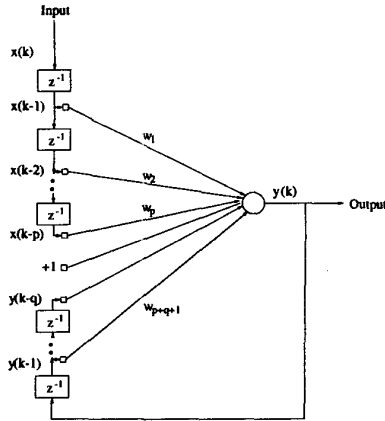
Figure 1: NARMA(p,q) recurrent perceptron

equations which describe the recurrent perceptron are

$$
\begin{aligned}
y(k) &= \Phi\left(v(k)\right) \\
v(k) &= \mathbf{W}(k)^T \cdot \mathbf{X}(k),
\end{aligned} \tag{6}
$$

where $\mathbf{X}(k) = [y(k-1), \ldots, y(k-q), 1, x(k-1), \ldots, x(k-p)]^T$, and $\mathbf{W}(k) = [w_1(k), \ldots, w_{p+q+1}(k)]^T$, where $(\cdot)^T$ denotes the transpose operator.

The Bounded Input Bounded Output (BIBO) stability of (6) is preserved due to the saturation type nonlinearity in (5).

3. LOGISTIC ACTIVATION FUNCTION AND CONTRACTION MAPPING

It has been shown that the logistic nonlinear activation function of a neuron provides a **contraction mapping** on $[a, b] \in \mathcal{R}$ for $0 < \beta < 4$ [6]. By the Contraction Mapping Theorem (CMT), function K is a contraction on $[a, b] \in \mathcal{R}$ if:

i) $x \in [a, b] \Rightarrow K(x) \in [a, b]$

ii) $\exists \gamma < 1 \in \mathcal{R}^+$ s.t. $|K(x) - K(y)| \leq \gamma|x - y|, \quad \forall x, y \in [a, b]$

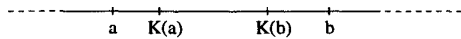as illustrated in Figure 2. Using the Mean Value The-

Figure 2: The contraction mapping

orem (MVT), due to differentiability of (5), for $\forall x, y \in$

$[a, b]$, $\exists \xi \in (a, b)$ such that

$$
|\Phi(x) - \Phi(y)| = |\Phi'(\xi)(x - y)| = |\Phi'(\xi)||x - y| \tag{7}
$$

Now, the clause $\gamma < 1$ in $ii)$ becomes $\gamma \geq |\Phi'(\xi)|$, $\xi \in (a, b)$. The first derivative of the logistic function (5) is

$$
\Phi'(x) = \frac{\beta e^{-\beta x}}{\left(1 + e^{-\beta x}\right)^2} \tag{8}
$$

which is strictly positive, and whose maximum value is $\beta/4$ for $x = 0$. Hence, for $\beta \leq 4$, the first derivative $\Phi' \leq 1$. Finally, for $\gamma < 1 \Leftrightarrow \beta < 4$, function Φ given in (5) is a contraction on $\forall[a, b] \in \mathcal{R}$.

4. CONDITIONS FOR CONVERGENCE OF RNN RELAXATION

In order to derive the conditions of GAS convergence for (3), we apply a contraction mapping and corresponding Fixed Point Iteration (FPI) to a recurrent perceptron (Figure 1). Recall that if function f is contractive, then an equation $F(\xi) = 0$ can be iteratively solved as $\xi(k) = f(\xi(k-1))$, which converges towards a fixed point $\xi^* = f(\xi^*)$. Such a process is called an FPI. Unlike in the linear case (1), the external input data to (3) do not need to be a zero-vector, but simply kept constant. Let us therefore denote $\{x(k-1), \ldots, x(k-p)\}$ by $\{x_1, \ldots, x_p\}$.

Theorem 1 *GAS relaxation for a recurrent perceptron given by*

$$
y(k+1) = \Phi\left(\mathbf{X}(k)^T \mathbf{W}(k)\right) \tag{9}
$$

where $\mathbf{X}_k^T = [y(k), \ldots, y(k-q+1), 1, x_1, \cdots, x_p]$, *is a contraction mapping and converges to some value* $y^* \in (0, 1)$ *for* $\beta \cdot \sum_{j=1}^q |w_j(k)| < 4$.

Proof:
Equation (9) can be written as

$$
y(k+1) = \Phi\left(\sum_{j=1}^{p+q+1} w_j(k) z_j\right) \tag{10}
$$

where z_j is the jth element of $\mathbf{X}(k)$. In other words, the iteration (9) is biased, and can be expressed as

$$
y(k+1) = \Phi\left(y(k), \ldots, y(k-q+1)\right) \tag{11}
$$

The Existence, Uniqueness, and Convergence features of mapping (9), follow from contractiveness of the logistic function. Iteration (9) converges to a fixed point $y^* = \Phi(y^* + constant)$, where the *constant* is given by $constant = \left(\sum_{j=p+1}^{p+q+1} w_j(k) z_j\right)$. It is assumed that

1038

the weights which correspond to the external input signals are not time–variant. Since the condition for convergence of the logistic function to a fixed point is $0 < \beta < 4$, it follows that the slope in the logistic function β and the weights $w_1(k), \ldots, w_q(k)$ in the weight vector $\mathbf{W}(k)$ are not independent and that the effective slope in the logistic function now becomes the product $\beta \cdot \sum_{j=1}^{p} w_j(k)$. Therefore

$$|\beta \cdot \sum_{j=1}^{q} w_j(k)| \le \beta \sum_{j=1}^{q} |w_j(k)| < 4 \qquad (12)$$

is the condition of GAS convergence of (3) realised through a recurrent NARMA perceptron. **QED**
A comparison of the nonlinear GAS result (12) with its linear counterpart shows that they are both based upon the $\| \cdot \|_1$ norm of the appropriate coefficient vector. In the nonlinear case, however, the measure of nonlinearity is also included.

Corollary 1 *In the case of the realisation of (3) by a NARMA recurrent perceptron, convergence towards a point in the FPI sense* **does not** *depend on the number of external input signals (i.e. p–the AR part of the NARMA(p,q) process (4)), nor on their values, as long as they are finite.*

Corollary 2 *The FPI convergence for a NARMA(p,q) recurrent perceptron of order q lasts a least q steps longer than for a recurrent perceptron with $q = 1$.*

4.1. Convergence Rate of the FPI

Convergence rate is the ratio between the distances between the current and previous iterate of an FPI and a fixed point y^*, i.e. as $\frac{y(k)-y^*}{y(k-1)-y^*}$. This reveals how quickly an FPI process converges towards a point.

Corollary 3 *A realisation of an iterative process (9) by a recurrent perceptron converges towards a fixed point y^* exhibiting linear convergence with convergence rate $\Phi'(y^*)$.*

5. NESTING IN THE RNN FRAMEWORK

Nesting corresponds to the procedure of reducing the interval size in set theory. However, in signal processing, nesting is essentially a nonlinear spatial structure which corresponds to the cascaded structure in linear signal processing [7, 8]. The RNN based nested sigmoid scheme can be written as [9]

$$F(W, X) = \Phi \left(\sum_n w_n \Phi(\sum_i v_i \Phi(\cdots \Phi(\sum_j u_j X_j) \cdots)) \right) \qquad (13)$$

where Φ is a sigmoidal function. This corresponds to a multilayer network of units that sum their inputs with "weights" $W = \{w_n, v_i, \cdots, u_j, \cdots\}$ and then perform a sigmoidal transformation of this sum. This scheme is uncommon in classical theory of approximation of continous functions. Let us now, for the sake of clarity, consider a recurrent perceptron with only one feedback output signal $(q = 1)$.

Theorem 2 *The nested function*

$$\begin{aligned} y(k+1) &= \Phi(y(k)) = \\ &= \Phi(\Phi(y(k-1))) = \\ &= \Phi(\Phi(\Phi(\cdots(\Phi(y(1)))\cdots) \qquad (14) \end{aligned}$$

provides a contraction mapping which converges towards a point y^* in the FPI sense.

Proof:
Notice that the nesting process (14) represents an implicitly written fixed point iteration process

$$\begin{aligned} y(k+1) &= \Phi(y(k)) \Leftrightarrow y(k+1) = \Phi(\Phi(y(k-1))) = \\ &= \Phi(\Phi(\Phi(\cdots(\Phi(y(1)))\cdots) \qquad (15) \end{aligned}$$

Hence, nesting provides a finite–length fixed point iteration. Hence, it is expected that a nested structure with m stages (14), converges towards a point $y^* \in [|\Phi'(y^*)|^m a, |\Phi'(y^*)|^m b]$, for the initial values in the interval $[a, b] \in \mathcal{R}$. For m small, the fixed point iteration achieved through a nesting process (14) may not reach its fixed point. **QED**
The same analogy between the spatial nesting and temporal FPI holds for a general NARMA(p,q) network. The FPI convergence towards a point is shown in Example 1 in Section 7.

6. PRNN AS A REALISATION OF NESTING

The realisation of process (14) is the so–called Pipelined Recurrent Neural Network (PRNN) [7], given in Figure 3, which provides a spatial form of the iteration (9). The PRNN consists of a number of small–scale RNNs, which are nested, and share the same weight matrix, and a number of external input signals [7]. Therefore, instead of having a temporal FPI on a recurrent perceptron (Figure 1), it suffices, for a finite–length FPI, to consider a PRNN spatial structure. As the rate of GAS convergence for a recurrent perceptron does not depend on the number of external input signals (Corollary 1), and the convergence of the NARMA(p,q) network lasts at least q steps longer than for a NARMA(p,1) network (Corollary 2), it is desirable to choose a recurrent perceptron with $q = 1$ as a module in the PRNN.
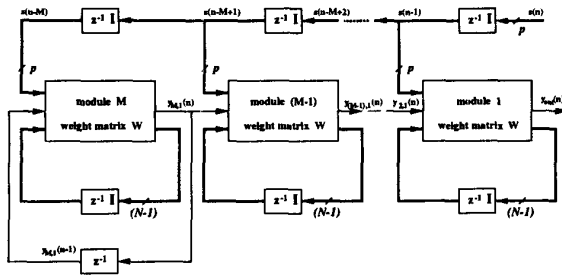
1039

Figure 3: Pipelined Recurrent Neural Network

7. EXAMPLE

We support our analysis by a simple example with a NARMA(p,1) recurrent perceptron. For the sake of clarity, all the constant parameters are embodied in the unity–valued constant.

Example 1 *Show that the iteration*

$$y(k) = \Phi(y(k-1)) = \frac{1}{1 + e^{-y(k-1)+1}} \qquad (16)$$

with initial values $y_0 = 10$ and $y_0 = 10$ converges towards a point $y^* \in [-10, 10]$.

The numerical values for iteration (16) are given in Table 1. Indeed, the iterates from either starting point

Table 1: Fixed point iterates for the NARMA perceptron

Starting value y_0	-10	10
First iterate	0.0000	0.9999
Second iterate	0.2689	0.5000
Third iterate	0.3250	0.3775
Fourth iterate	0.3374	0.3492
Fifth iterate	0.3401	0.3428
Sixth iterate	0.3408	0.3414
Seventh iterate	0.3409	0.3410

converge to a value $y^* \in [0.3409, 0.3410] \in [-10, 10]$. It can be shown that after 22 iterations for $y_0 = -10$ and 24 iterations for $y_0 = 10$, the fixed point to which the FPI (16) converges is $y^* = 0.34095393159261$.

8. SUMMARY

Global Asymptotic Stability (GAS) for a class of non-linear relaxive systems realised by a Nonlinear Autoregressive Moving Average (NARMA) recurrent perceptron has been studied. Based upon the Fixed Point Iteration (FPI) technique, it has been shown that these

conditions rest entirely upon the slope of the activation function β, and a measure of the $\| \cdot \|_1$ norm of the weight vector of a recurrent perceptron. In that case, the GAS iteration converges linearly towards an equilibrium point. Moreover, convergence does not depend on the number of external input signals to a recurrent perceptron (the NAR part of NARMA(p,q) network). Connection between nesting and FPI, which is the basis of the GAS convergence, has been established, and a Pipelined Recurrent Neural Network (PRNN) has been shown to be a spatial realisation of the FPI process, for a finite–length FPI. The example provided supports the analysis.

9. REFERENCES

[1] P. Bauer, M. Mansour, and J. Duran, "Stability of polynomials with time–varying coefficients," *IEEE Transactions on Circuits and Systems–I: Fundamental Theory and Applications*, vol. 40, no. 6, pp. 423–426, 1993.

[2] K. Premaratne and M. Mansour, "Robust stability of time–variant difference equations with restricted parameter perturbations: Regions in coefficient–space," in *ICASSP–97*, vol. 3, pp. 2341–2344, 1997.

[3] J. P. LaSalle, *The Stability and Control of Discrete Processes*. Springer–Verlag, 1986.

[4] M. Priestley, *Non–linear and Non–stationary Time Series Analysis*. Academic Press, London, 1991.

[5] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 240–254, 1994.

[6] D. P. Mandic and J. A. Chambers, "A fixed point iteration interpretation of the global asymptotic convergence of recurrent neural networks." Submitted to the IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics, 1998.

[7] S. Haykin and L. Li, "Nonlinear adaptive prediction of nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 526–535, 1995.

[8] J. Baltersee and J. A. Chambers, "Non–linear adaptive prediction of speech signals using a pipelined recurrent neural network." to appear in IEEE Transactions on Signal Processing, August 1998.

[9] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1481–1497, 1990.