Existence and Learning of Oscillations in Recurrent Neural Networks *

S. Townley[‡¶] A. Ilchmann[‡] M. G. Weiß[§||] W. Mcclements[‡]
A. C. Ruiz[†] D. H. Owens[†] and D. Prätzel-Wolters[§]

† Centre for Systems and Control Engineering, University of Exeter, North Park Road, Exeter EX4 4QF, UK.

‡ School of Mathematical Sciences, University of Exeter, North Park Road, Exeter EX4 4QE, UK.

§ Graduiertenkolleg Technomathematik, Universität Kaiserslautern, Erwin-Schrödinger-Straße, 67663 Kaiserslautern, Germany.

September 1998

Abstract

In this paper we study a particular class of n-node recurrent neural networks (RNNs). In the 3-node case we use monotone dynamical systems theory to show, for a well-defined set of parameters, that, generically, every orbit of the RNN is asymptotic to a periodic orbit. Then, within the usual 'learning' context of Neural Networks, we investigate whether RNNs of this class can adapt their internal parameters so as to 'learn' and then replicate autonomously certain external periodic signals. Our learning algorithm is similar to identification algorithms in adaptive control theory. The main feature of the adaptation algorithm is that global exponential convergence of parameters is guaranteed. We also obtain partial convergence results in the n-node case.

Keywords: Recurrent neural networks, Learning systems, Nonlinear dynamics, Monotone dynamical systems.

1 Introduction

Recently, there has been considerable interest in Recurrent Neural Networks (RNNs) which exhibit periodic or chaotic dynamics. For example, RNNs which generate stable oscillations have been used to model certain biological phenomena. Indeed, Li and Hopfield (1989) have proposed such RNNs as models for the dynamics of the olfactory bulb. See also Atiya and Baldi (1989) for an account of 'oscillatory' RNN models in a biological context. RNNs which generate chaotic dynamics can be used to model oscillations in the cortex and for controlling chaotic dynamical systems, see Babloyantz et

*Supported in part by the UK EPSRC under grant GRJ8095 and an EU Human Capital and Mobility programme, project CHRX-CT-93-0402.

||Supported by Deutsche Forschungsgemeinschaft and Graduiertenkolleg Technomathematik, Universität Kaiserslautern.

¶Corresponding author

al. (1995), Doyon et al. (1993), Sole et al. (1995) and references therein.
In this paper we are mainly interested in determining whether a class of RNNs can maintain a periodic orbit, and, if so, can they be forced to learn such orbits. Such periodic orbits are meant to capture the idea that certain activities or motions are learnt by repetition. In the literature, there are essentially three approaches to this problem:
The first approach considers the behaviour of RNNs from a computational point of view. Pearlmutter (1995) has shown that a fully interconnected 5-dimensional RNN can generate a stable limit cycle. This empirical approach uses a dynamic version of the well-known steepest descent adaptation algorithm to adapt the parameters or weights of the RNN so that, after a training period, the network replicates a predetermined periodic signal. See also Doya and Yoshizawa (1989) for similar results. This approach does not analyse the mechanism by which the periodic signal is generated nor does it make any attempt to characterise the set of parameter values for which the RNN has periodic solutions. Consequently, there is no guarantee that such a set of parameters exists and finding suitable values for the parameters is left to the 'steepest descent' algorithm.
The second approach uses Hopf Bifurcation techniques to prove that certain classes of RNN generate stable limit cycles. Whilst this approach can be used to determine parameter ranges for which such limit cycles exist, by the very nature of the Hopf theorem, these existence results are local, both in parameter and phase space. Hence, these results are not comprehensive enough. However, in view of the highly nonlinear dynamics inherent to RNNs, they do represent a step forward in understanding the dynamic behaviour of RNNs. For results in this direction see Atiya and Baldi (1989), Ruiz et al. (1998), and the references therein.
The third approach considers an RNN as a monotone dynamical system. Whilst the theory of monotone dynamical systems has its motivation in the analysis of partial differential equations arising in mathematical biology (see Matano (1986), Hirsch (1988), and references therein) it has recently been recognised that this theory has potential for analysing systems which arise in Neural Network applications. In fact, many classes of RNNs can be regarded as monotone dynamical systems and we believe that this theory, as developed by Hirsch (1988), Smith (1995) and more recently by Mallet-Paret and Sell (1996), provides a powerful tool for analysing the dynamics of RNNs. To our knowledge, specific developments in this direction have been limited. Smith (1991) has studied a RNN with a cyclic structure and the work of Mallet-Paret and Smith (1990) on general 'cyclic' dynamical systems, can be applied to classes of cyclic RNNs.

In this paper we study a particular class of n-node RNNs. In the 3-node case we use monotone dynamical systems theory to show, for a well-defined set of parameters, that, generically, every orbit of the RNN is asymptotic to a period orbit. Then, within the usual 'learning' context of Neural Networks, we investigate whether RNNs of this class can adapt their internal parameters so as to 'learn' and then replicate autonomously certain external periodic signals. Our learning algorithm is similar to identification algorithms in adaptive control theory. The main feature of the adaptation algorithm is that global exponential convergence of parameters is guaranteed. This is in contrast to 'steepest descent'-based adaptation algorithms which only find local minima of the parameter error cost functionals. We also obtain partial convergence results in the n-node case. Note that whilst we use an identical network structure to that in Ruiz et al. (1998), our results differ on two accounts. First, in Ruiz et al. convergence of a generic orbit of the RNN to a periodic orbit is only proved to $\mathcal{O}(\epsilon)$ on time scales of $\mathcal{O}(1/\epsilon)$. Secondly, in Ruiz et al. the learning algorithm is based on steepest descent techniques. For this type of learning algorithm only local asymptotic convergence can be guaranteed.

The paper is organised as follows. In Section 2 we specify the class of RNNs under consideration and make precise the notions of learning and replication. To do so we introduce the so-called Teaching Network and Learning RNN. The Teaching Network provides the external periodic signal which is to be learnt. In Section 3 we prove that the orbits of the Teaching Network are, generically, asymptotic to a periodic orbit. In Section 4 we develop the parameter adaptation algorithm by which learning is achieved. In the 3-node case our adaptation algorithm guarantees exponential convergence. We also prove partial convergence results for the n-node case. In Section 5 we comment on the difficulties encountered in proving exponential convergence in the n-node case.

2 Structure of the Teaching Network and Learning RNN

We consider the recurrent neural network (RNN) shown in Figure 1. In the figure, $u(t)$ is the scalar

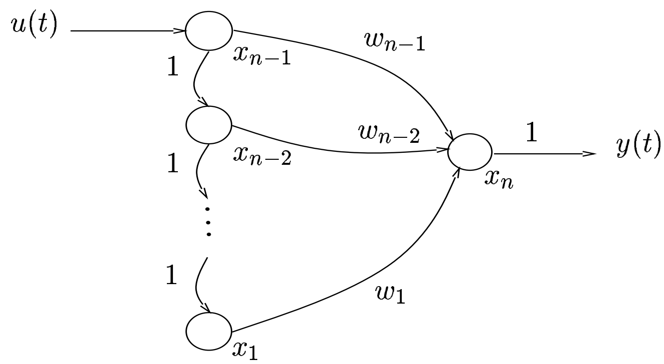

Figure 1: Class of recurrent neural networks.

input and $y(t)$ is the scalar output of the network. The RNN depicted in Figure 1 is described formally by the system of differential equations:

$$
\begin{aligned}
\dot{x}_1(t) &= -x_1(t) && + \tanh x_2(t), \\
&\vdots \\
\dot{x}_{n-2}(t) &= -x_{n-2}(t) && + \tanh x_{n-1}(t), \\
\dot{x}_{n-1}(t) &= -x_{n-1}(t) && + u(t), \\
\dot{x}_n(t) &= -x_n(t) && + w_1(t)\tanh x_1(t) + \cdots + w_{n-1}(t)\tanh x_{n-1}(t), \\
\\
y(t) &= \tanh x_n(t),
\end{aligned}
\tag{1}
$$

where $x(t) := (x_1(t), \ldots, x_n(t))^T \in \mathbb{R}^n$ is the state, $x(0) = x_0 \in \mathbb{R}^n$, $w(t) := (w_1(t), \ldots, w_{n-1}(t))^T \in \mathbb{R}^{n-1}$ is the network parameter or weight vector, $u(\cdot)$ is the input, i.e. teaching signal, and $y(\cdot)$ is the output. In (1) we have taken the nonlinear triggering function of the neurons equal to hyperbolic tangent. However, any triggering function, with similar properties of oddity, boundedness, monotonicity and smoothness, could have been considered. These properties are used in the proof of Lemma 3.1. In Proposition 5.1 we use a triggering function $a \mapsto \tanh \lambda a$.

We are interested in whether, by adapting the weights, the RNN (1) can learn and then replicate a periodic teaching signal $u(\cdot)$. The use of a periodic teaching signal is motivated by the idea that most learning systems need repetition. To make the problem solvable we restrict the class of signals that are to be learnt. In fact, we assume that the signal $u(\cdot)$ to be learnt is given by

$$u(t) = \tanh z_n(t),$$

where

$$
\begin{array}{rcrcl}
\dot{z}_1(t) & = & -z_1(t) & + & \tanh z_2(t), \\
 & \vdots & & & \\
\dot{z}_{n-2}(t) & = & -z_{n-2}(t) & + & \tanh z_{n-1}(t), \\
\dot{z}_{n-1}(t) & = & -z_{n-1}(t) & + & \tanh z_n(t), \\
\dot{z}_n(t) & = & -z_n(t) & + & w_1^*\tanh z_1(t) + \cdots + w_{n-1}^*\tanh z_{n-1}(t) .
\end{array} \tag{2}
$$

We refer to (2) as the Teaching Network. The Teaching Network, with state $z(t) := (z_1(t), \ldots, z_n(t))^T \in \mathbb{R}^n$ and $z(0) := z_0 \in \mathbb{R}^n$, has a similar structure to (1) but the corresponding weight vector $w^* := (w_1^*, \ldots, w_{n-1}^*)^T$ is fixed and the loop from $y(t)$ to $u(t)$ is closed with unity feedback. We will see in Section 3 that the Teaching Network can have periodic solutions, which we can then use as periodic teaching signals.

The RNN (1) will operate in two modes - as a Learning RNN in the Learning Phase and as a Replicating RNN in the Replicating Phase.

1. As a Learning RNN (1) has time-varying weights and the input $u(t)$ is equal to the output $\tanh z_n(t)$ of the Teaching Network. The time-varying weights of this Learning RNN are adapted so as to enable learning of the periodic teaching signal $\tanh z_n(t)$ and unknown weights of the Teaching Network.

2. As a Replicating RNN (1) has fixed weights and operates in a unity feedback configuration. The output of this Replicating RNN is meant to agree with the output of the Teaching Network.

The overall process of learning/replication is described as follows: The Teaching Network produces at its output an unknown periodic teaching signal. In the Learning Phase this signal is fed as input into the Learning RNN. The weights of the Learning RNN are then adapted. We use a weight adaptation algorithm which is similar to identification algorithms in adaptive control theory. In the case $n = 3$ we can prove that the weights and states of the Learning RNN converge exponentially to those of the Teaching RNN. After some finite time T, assumed long enough for the convergence to be adequate, we switch from the Learning Phase to the Replication Phase so that weight adaptation is terminated and the output of the Teaching Network is removed as input to the Learning RNN to be replaced with its own output. The resulting Replicating RNN, with fixed weight vector $w(T)$, then reproduces (approximately), as its output, the periodic teaching signal. As mentioned in the Introduction, it has been shown experimentally that a class of recurrent networks with configurations similar to the one considered here, are indeed able to learn and replicate certain types of periodic signals, see Doya et al. (1989), Pearlmutter (1995) and Yang et al. (1994). We are interested in proving that such learning and replication has taken place.

In the context of our learning/replication process, there are two crucial aspects. We must prove that the Teaching Network produces periodic signals as its output and we must be able to prove that the adapted weights of the Learning RNN converge to the fixed weights of the Teaching RNN. These issues are dealt with separately in Sections 3 and 4, respectively.

3 Existence of attractive periodic solutions in the Teaching Network

In this section we consider properties of a 3-node version of Teaching Network (2):

$$
\begin{aligned}
\dot{z}_1(t) &= -z_1(t) + \tanh z_2(t), \\
\dot{z}_2(t) &= -z_2(t) + \tanh z_3(t), \\
\dot{z}_3(t) &= -z_3(t) + w_1^* \tanh z_1(t) + w_2^* \tanh z_2(t) .
\end{aligned} \tag{3}
$$

We prove, for a range of weight values, that each trajectory of (3) which does not converge to the equilibrium $z = 0$, converges to a periodic orbit. We do so by regarding the system (3) as a monotone dynamical system and by using techniques from monotone dynamical systems theory.

We begin with a lemma which is proved in Ruiz et al. (1998).

Lemma 3.1 *Consider the system described by (3). If the fixed weights w_1^* and w_2^* satisfy*

$$2w_2^* - w_1^* \geq 8, \qquad 27{w_1^*}^2 - 4{w_2^*}^3 > 0, \tag{4}$$

then

(i) the linearization of (3) about zero has one negative real eigenvalue and a pair of complex conjugate eigenvalues with positive real part,
(ii) the origin is a unique equilibrium of (3).

Now it is easy to see, using the boundedness of $a \mapsto \tanh a$, that all solutions of (3) are bounded. Combining this with the properties (i) and (ii) in Lemma 3.1, it is reasonable to expect that (3) will admit periodic solutions for some, if not all, weight values satisfying (4). This result is proved in the following weak sense in Ruiz et al. (1998): For weight values satisfying (4), the solution $z(\cdot)$ converges to a periodic function to $\mathcal{O}(\epsilon)$ on time scales of $\mathcal{O}(1/\epsilon)$. Our aim is to strengthen this result to asymptotic convergence. First we introduce the concept of a competitive dynamical system.

Definition 3.2 Let $f : V \to \mathbb{R}^n$ be continuously differentiable on some open set $V \subset \mathbb{R}^n$. A system $\dot{x} = f(x)$ is said to be *competitive* on V if, and only if,

$$\frac{\partial f_i}{\partial x_j}(x) \leq 0, \quad \text{for all} \quad i \neq j\,, \quad \text{and for all} \quad x \in V. \tag{5}$$

Competitive systems are special cases of *monotone* systems. For a detailed study of monotone dynamical systems see Smith (1995) and the references therein. Note that we have defined the notion of a competitive system with respect to the positive orthant in $\mathbb{R}^n$ and the usual partial ordering: $x \ll y$ if and only if $x_i \leq y_i$ for all $i \neq j$, and $x \neq y$. The notion can be extended in an obvious way by considering other orthants in $\mathbb{R}^n$.
The main result on competitive systems we need is the following proposition from Smith (1995).

Proposition 3.3 *Let the system*

$$\dot{x} = f(x), \qquad x(0) = x_0 \in \mathbb{R}^3\,,$$

with f continuously differentiable on an open set $V \subset \mathbb{R}^3$, be a competitive system on V. Suppose that V contains a unique equilibrium point x_e which is hyperbolic. Suppose further that $\mathcal{W}^s(x_e)$, the stable manifold at x_e, is one dimensional and tangential at x_e to a non-negative vector v. If $x_0 \notin \mathcal{W}^s(x_e)$ and the positive semi-orbit $\gamma^+(x_0) := \{x(t) : t \geq 0\}$ has compact closure in V, then the ω-limit set, $\omega(x_0)$, of x_0 is a nontrivial periodic orbit.

Note that this result, sometimes referred to as Poincaré-Bendixson theorem for 3-dimensional systems, does not generalize to higher dimensions, except in the special case of cyclic systems, see Mallet-Paret and Smith (1990). Furthermore, monotone systems theory does not provide us with any information concerning the uniqueness and stability of these periodic orbits. Nevertheless, Proposition 3.3 is a useful tool for establishing the existence of attractive periodic orbits in the Teaching Network (3).

Theorem 3.4 *Consider the Teaching Network (3). Suppose that the weights satisfy (4) and, in addition, $w_2^* > 0$. Then for each $z_0 \notin \mathcal{W}^s(0)$, $\omega(z_0)$ is a non-trivial periodic orbit.*

Proof In order to apply Proposition 3.3 we need some preliminary results:
(i) We show for all $z_0 \in \mathbb{R}^3$, that the positive semi-orbit $\gamma^+(z_0)$ has compact closure. Indeed, since $a \mapsto \tanh a$ is bounded, we can view (3) as an exponentially stable linear system driven with a bounded input. Therefore every positive semi-orbit is bounded and so has compact closure in $\mathbb{R}^3$. In fact the set $\{z \in \mathbb{R}^3 \mid \|z\| \leq 4(1 + \|w_1^*\|) + \|w_2^*\|\}$ attracts all solutions.

(ii) It follows from (4) and Lemma 3.1, parts (i) and (ii), that $z_e = 0$ is a unique hyperbolic equilibrium of (3) with a one-dimensional stable manifold $\mathcal{W}^s(0)$.

(iii) In order to apply Proposition 3.3, we transform (3) into a competitive system on $\mathbb{R}^3$. We do so by using a change of coordinates $\xi_1 = -z_1$, $\xi_2 = z_2$, $\xi_3 = -z_3$ which, when applied to (3), gives

$$\begin{aligned} \dot{\xi}_1(t) &= -\xi_1(t) - \tanh \xi_2(t) \\ \dot{\xi}_2(t) &= -\xi_2(t) - \tanh \xi_3(t), \\ \dot{\xi}_3(t) &= -\xi_3(t) + w_1^* \tanh \xi_1(t) - w_2^* \tanh \xi_2(t)\,. \end{aligned} \tag{6}$$

Since the weights satisfy (4), so that in particular $w_1^* < 0$, and by assumption $w_2^* > 0$, it follows that the right hand side of (6) satisfies (5). Hence the system (6) is competitive.

(iv) Let P be the Jacobian matrix of the right hand side of (6) evaluated at zero. Clearly $-P$ is non-negative and $(-P)^2$ is a positive matrix (i.e $(-P)$ is a primitive matrix). It follows (see, for example Section 8.5 in Horn and Johnson (1985)) that $-P$ has a positive eigenvector corresponding to the (unique by Lemma 3.1) negative real eigenvalue of P. Hence the stable manifold for the zero equilibrium of (6) is tangential at zero to a positive vector.

The proof is now complete since we can use (i) - (iv) to apply Proposition 3.3 to (6). □

Remark 3.5 (i) The conclusions of Theorem 3.4 will hold if the triggering function $a \mapsto \tanh a$ is replaced by any other function σ with similar properties of oddity, boundedness, monotonicity and smoothness provided that the inequalities (4) are scaled appropriately to account for $\sigma'(0) \neq 1$.
(ii) Note that in the case w_1^* and w_2^* satisfying (4) but with $w_2^* < 0$, we have not been able to find a change of coordinates which converts (3) into a competitive system in the sense of Definition 3.2. This has prevented us from extending Theorem 3.4 to the case $w_2^* < 0$. However in Ruiz et al. (1998) we have shown, for such weight parameters, a weaker convergence to a periodic function to $\mathcal{O}(\epsilon)$ on time scales of $\mathcal{O}(1/\epsilon)$.

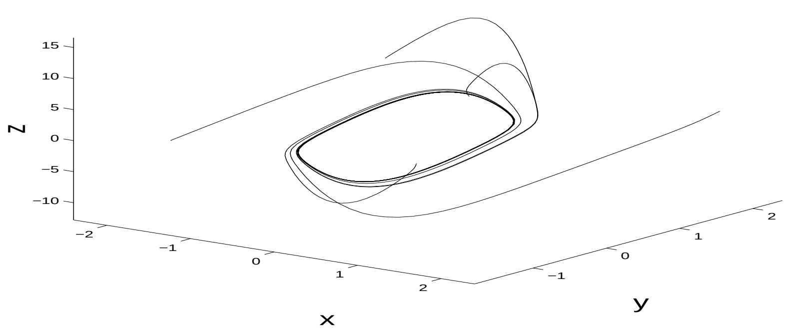

Figure 2: Attractive limit cycle with $w_2^* > 0$

We illustrate the result of Theorem 3.4 by a simulation. We set $w_1^* = -20$ and $w_2^* = 10$. These weights satisfy the conditions required in Theorem 3.4. Figure 2 shows the solution $z(\cdot)$ for a variety of initial conditions for $z(0)$. For comparison, Figure 3 shows simulations in the case $w_1^* = -34$ and $w_2^* = -7$ where the inequalities (4) are satisfied but where $w_2^* < 0$.

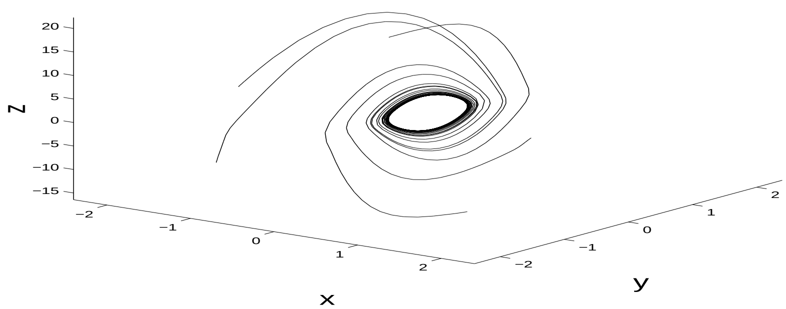

Figure 3: Attractive limit cycle with $w_2^* < 0$

Note that for any weight values satisfying (4), not only does every simulation we have tried produce solutions converging to a periodic function, but that for each pair of weights this periodic function is unique, i.e. simulations suggest that for each pair of weights satisfying (4), (3) has a limit cycle which attracts all solutions except those starting in $\mathcal{W}^s(0)$. This is illustrated by Figures 2 and 3.

It is easy to see that the solution $z(\cdot)$ of the Teaching Network (2) is constant if, and only if, one of its components $z_i(\cdot)$ is. We conclude this section with a lemma concerning the linear independency of the functions $\tanh z_1(\cdot)$ and $\tanh z_2(\cdot)$ in the case of the 3-node teaching network.
Let

$$B(t) := (\tanh z_1(t), \tanh z_2(t))^T . \tag{7}$$

Lemma 3.6 *Assume that $z(\cdot)$ is a non-trivial periodic solution of (3). Then the functions $\tanh z_1(\cdot)$ and $\tanh z_2(\cdot)$ are linearly independent or equivalently*

$$\int_0^\tau B(t)B(t)^T dt > 0\,, \tag{8}$$

where τ is the period of $z(\cdot)$.

Proof Suppose that $\tanh z_1(\cdot)$ and $\tanh z_2(\cdot)$ are dependent. Since neither of $\tanh z_1(\cdot)$ and $\tanh z_2(\cdot)$ are constant, we can find $a \neq 0$ so that

$$\tanh z_2(\cdot) \equiv a \tanh z_1(\cdot)\,.$$

Substituting for z_2 in the first equation in (3) gives

$$\dot{z}_1(t) = -z_1(t) + a\ \tanh z_1(t)\,.$$

It then follows that $z_1(\cdot)$ is constant since the only periodic solution of a first order equation is a constant solution. This implies that $z(\cdot)$ is constant which is a contradiction. Therefore $\tanh z_1(\cdot)$ and $\tanh z_2(\cdot)$ are linearly independent. Now (8) holds if, and only if $\int_0^\tau (v^T B(t))^2 dt > 0$ for all non-zero $v \in \mathbb{R}^2$, i.e.

$$v^T B(\cdot) \not\equiv 0, \quad \text{for all non-zero} \quad v \in \mathbb{R}^2\,,$$

i.e. if, and only if, $\tanh z_1(\cdot)$ and $\tanh z_2(\cdot)$ are linearly independent. □

Remark 3.7 (i) The condition given by (8) states that $B(\cdot)$ is persistently exciting, see Morgan and Narendra (1977).
(ii) We have been unable to obtain necessary and sufficient conditions for persistency of excitation of $(\tanh z_1(\cdot), \ldots, \tanh z_{n-1}(\cdot))^T$, in the case of an n-node Teaching Network. In Proposition 5.1 we show, for a Teaching Network with triggering function $a \mapsto \tanh \lambda a$ in place of $a \mapsto \tanh a$, that persistency of excitation can fail even though the Teaching Network has a limit cycle. However, all our simulations suggest that persistency of excitation holds generically amongst those weight vectors w^* yielding periodic solutions.

4 Learning the output from the Teaching Network

In this section we construct weight adaptation algorithms which enable the Learning RNN to learn the output of the Teaching Network. More accurately, in the case $n = 3$ the learning algorithms guarantee that the state $x(t)$ and weight vector $w(t)$ of the Learning RNN converge exponentially to the state $z(t)$ and fixed weight vector w^* of the Teaching Network. The weight adaptation algorithms are similar to identification algorithms in adaptive control. The convergence proofs use the persistency of excitation property (8). We also obtain partial convergence results for the n-node case.

Theorem 4.1 *Consider the 3–node Teaching Network (3) and the corresponding 3–node Learning RNN*

$$\begin{aligned} \dot{x}_1(t) &= -x_1(t) + \mathit{tanh}\, x_2(t), \\ \dot{x}_2(t) &= -x_2(t) + u(t), \\ \dot{x}_3(t) &= -x_3(t) + w_1(t)\, \mathit{tanh}\, x_1(t) + w_2(t)\, \mathit{tanh}\, x_2(t), \end{aligned} \tag{9}$$

where

$$u(t) = \mathit{tanh}\, z_3(t)\,. \tag{10}$$

Define the weight adaptation algorithm by

$$\begin{aligned} \dot{w}_1(t) &= -(x_3(t) - z_3(t))\mathit{tanh}\, x_1(t), \\ \dot{w}_2(t) &= -(x_3(t) - z_3(t))\mathit{tanh}\, x_2(t)\,. \end{aligned} \tag{11}$$

Then for arbitrary initial conditions $z(0), x(0) \in \mathbb{R}^3$ and $w(0) \in \mathbb{R}^2$, the closed-loop system (3), (9), (10) and (11) has a unique solution defined on $[0, \infty)$. Furthermore, if $z(\cdot)$ is a nontrivial periodic solution of (3), then there exist $M, \lambda > 0$ independent of $x(0) \in \mathbb{R}^3$ and $w(0) \in \mathbb{R}^2$, so that

$$\|x(t) - z(t)\| \le Me^{-\lambda t} \qquad \text{and} \qquad \|w(t) - w^*\| \le Me^{-\lambda t}\,, \quad \text{for all} \quad t \ge 0. \tag{12}$$

Proof Existence and uniqueness of solutions on $[0, \infty)$ is guaranteed because the right hand side of the closed-loop system (3), (9), (10) and (11) is continuous and affine linearly bounded.
Let $e(t) = x(t) - z(t)$. Then

$$\begin{aligned} \dot{e}_1(t) &= -e_1(t) + \beta_2(t), \\ \dot{e}_2(t) &= -e_2(t), \\ \dot{e}_3(t) &= -e_3(t) + \tilde{w}_1(t)\tanh x_1(t) + \tilde{w}_2(t)\tanh x_2(t) + w_1^*\beta_1(t) + w_2^*\beta_2(t) \end{aligned} \tag{13}$$

where, for $i = 1$ or 2,

$$\beta_i(t) := \tanh x_i(t) - \tanh z_i(t) \qquad \text{and} \qquad \tilde{w}_i(t) := w_i(t) - w_i^*\,.$$

Clearly, from (13),

$$e_2(t) = e^{-t}e_2(0), \qquad \text{for all } t \ge 0. \tag{14}$$

Since $a \mapsto \tanh a$ has a global Lipschitz constant equal to 1 we have that

$$|\beta_i(t)| \le |e_i(t)|, \quad \text{for all} \quad t \ge 0 \quad \text{and} \quad i = 1,2\,. \tag{15}$$

It follows, using Variation of Contants in the first equation in (13), and then taking estimates, that

$$\begin{aligned} |e_1(t)| &\le e^{-t}|e_1(0)| + \textstyle\int_0^t e^{-(t-s)}|\beta_2(s)|ds \\ &\le e^{-t}|e_1(0)| + \textstyle\int_0^t e^{-(t-s)}|e_2(s)|ds \\ &\le e^{-t}|e_1(0)| + e^{-t}t|e_2(0)| \\ &\le e^{-t}|e_1(0)| + 2e^{-t/2}e^{-1}|e_2(0)|\,. \end{aligned}$$

Hence there exists $M_1, \lambda_1 > 0$ such that

$$|e_i(t)| \le M_1 e^{-\lambda_1 t} \quad \text{for all} \quad t \ge 0 \quad \text{and} \quad i = 1,2\,. \tag{16}$$

It remains to show that

$$\eta(t) := (e_3(t), \tilde{w}_1(t), \tilde{w}_2(t))^T$$

decays to zero exponentially. First we show that the weights are bounded. Indeed, differentiating

$$V(t) = \frac{1}{2}\|\eta(t)\|^2$$

along solutions and using (15) and (16), we obtain, for all $t \ge 0$, that

$$\dot{V}(t) = -e_3^2(t) + e_3(t)\,(w_1^*\beta_1(t) + w_2^*\beta_2(t)) \le -\frac{1}{2}e_3^2(t) + \frac{1}{2}(|w_1^*| + |w_2^*|)^2\, M_1^2 e^{-2\lambda_1 t}\,. \tag{17}$$

Integrating (17) from 0 to t gives

$$V(t) = \frac{1}{2}\|\eta(t)\|^2 \le V(0) - \frac{1}{2}\int_0^t e_3^2(s)ds + \frac{1}{2}\int_0^\infty (|w_1^*| + |w_2^*|)^2\, M_1^2 e^{-2\lambda_1 t}dt < \infty\,. \tag{18}$$

It follows from (18) that $\eta(\cdot)$, and in particular $w(\cdot)$, is bounded. Next we look at the differential equation which describes the evolution of $\eta(\cdot)$. This can be written in the form

$$\dot{\eta}(t) = A(t)\eta(t) + P(t)\eta(t) + D(t) \tag{19}$$

where

$$A(t) := \begin{pmatrix} -1 & B(t)^T \\ -B(t) & 0 \end{pmatrix}, \quad P(t) := \begin{pmatrix} 0 & 0 & 0 \\ -\beta_1(t) & 0 & 0 \\ -\beta_2(t) & 0 & 0 \end{pmatrix},$$

$D(t) := ([w_1(t)\beta_1(t) + w_2(t)\beta_2(t)], 0, 0)^T$ and $B(\cdot)$ is given by (7). Notice that $P(t)$ and $D(t)$, which we consider as perturbation terms in (19) decay to zero exponentially. In analysing (19) we first consider the 'unperturbed', homogeneous system

$$\dot{\xi}(t) = A(t)\xi(t), \qquad \xi(0) := \xi_0, \tag{20}$$

with $\xi(t) \in \mathbb{R}^3$. Since $B(\cdot)$ satisfies the persistency of excitation condition (8) it follows from Corollary 2.3 in Narendra and Annaswamy (1989), that the system given by (20) is uniformly exponentially stable. Now, using the fact that $\lim_{t\to\infty}\|P(t)\| = 0$, it follows from standard perturbation results, (see for example Rugh (1996), p. 134), that for the transition matrix $\Psi(\cdot,\cdot)$ of

$$\dot{\nu}(t) = (A(t) + P(t))\nu(t)\,,$$

there exists $M_2, \lambda_2 > 0$ so that

$$\|\Psi(t,s)\| \leq M_2 e^{-\lambda_2(t-s)} \tag{21}$$

for all $t \geq s$, and $s \geq 0$. Using Variation of Constants in (19) and estimating, using (21), gives, for all $t \geq 0$,

$$\|\eta(t)\| \leq M_2 e^{-\lambda_2 t}\eta(0) + \int_0^t M_2 e^{-\lambda_2(t-s)}\|D(s)\|ds\,. \tag{22}$$

Now using the boundedness of $w(\cdot)$ and the exponential decay to zero of $\beta_1(\cdot)$ and $\beta_2(\cdot)$ given by (15) and (16) we have that $\|D(t)\|$ converges to zero exponentially as t tends to ∞. It then follows that there exists $M, \lambda > 0$ so that (12) holds. $\square$

Remark 4.2 Note that the learning algorithm (11) is realizable since $z_3(\cdot)$ can be obtained from the teaching signal $\tanh z_3(\cdot)$ using the invertibility of $a \mapsto \tanh a$.

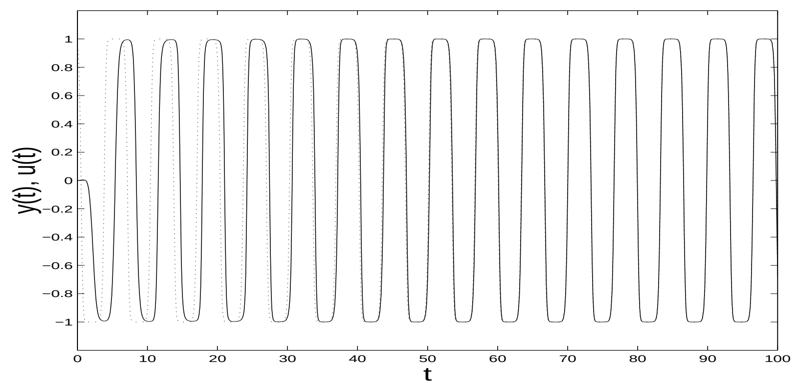

Figure 4: Reference signal $u(t)$ (dotted), learning signal $y(t)$ (continuous)

Figure 4 shows the results from a simulation of the learning algorithm where the weights of the teaching signal are $w^* = (-20, 10)$, $z(0) = (0.4736, 0.8745, 1.8497)$, $x(0) = (0,0,0)$ and $w(0) = (0,0)$. Figure 5 demonstrates the exponential convergence of the output error as proved in Theorem 4.1.

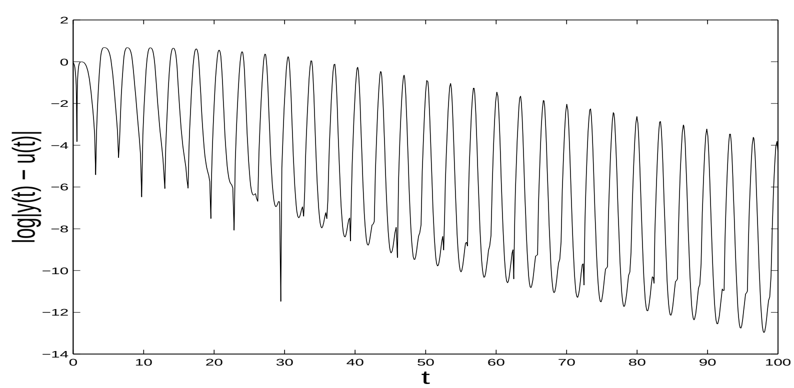

Figure 5: Output error on a logarithmic scale: $\log|y(t) - u(t)|$

The convergence of the weight parameters is shown in Figure 6.

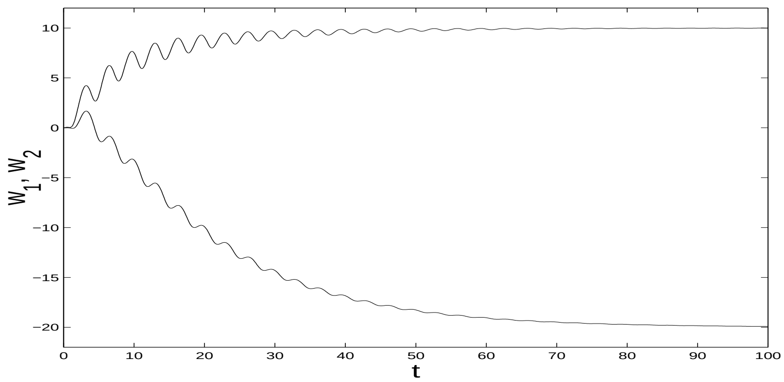

Figure 6: Weight dynamics w_1, w_2

The weight adaptation (11) is chosen to make the right hand side of (17) semi-negative definite, except for a term which decays to zero exponentially. Note that the algorithm (11) guarantees not only local, but global convergence. A similar construction can be used in the n-node case to obtain the following partial extension of Theorem 4.1.

Theorem 4.3 *Let $x(t)$ and $z(t)$ be given by (1) and (2) respectively. Define the weight adaptation algorithm by*

$$\dot{w}_i(t) = -[x_n(t) - z_n(t)]\,tanh\, x_i(t), \qquad \text{for} \quad i = 1, \dots, n-1\,. \tag{23}$$

Then

(i) there exist $M, \lambda > 0$ so that

$$|x_i(t) - z_i(t)| \leq M e^{-\lambda t}, \quad \text{for all} \quad t \geq 0 \quad \text{and} \quad i = 1, \dots, n-1\,,$$

(ii) the weights $w_i(\cdot)$ are bounded,

(iii) $\lim_{t\to\infty} |x_n(t) - z_n(t)| = 0$,

(iv) for any $t_1 > 0$ we have

$$\lim_{t\to\infty} \tilde{w}(t)^T \left(\int_t^{t+t_1} B(s)B(s)^T ds \right) \tilde{w}(t) = 0\,, \tag{24}$$

where $B(t) := (tanh\, z_1(t), \dots, tanh\, z_{n-1}(t))^T$.

Proof As in the proof of Theorem 4.1, introduce

$$e(t) := x(t) - z(t), \qquad \beta_i(t) := \tanh x_i(t) - \tanh z_i(t), \qquad \tilde{w}(t) := w(t) - w^*\,.$$

Then

$$\dot{e}_n(t) = -e_n(t) + \sum_{i=1}^{n-1} (\tilde{w}_i(t) \tanh x_i(t) + w_i^* \beta_i(t))\,. \tag{25}$$

Part (i) follows analogously to the corresponding result in Theorem 4.1. To prove parts (ii) and (iii) let

$$V(t) = \frac{1}{2}\|\eta(t)\|^2\,,$$

where $\eta(t) := (e_n(t), \tilde{w}_1(t), \dots, \tilde{w}_{n-1}(t))^T$. Then, as in the 3-node case,

$$\dot{V}(t) = -e_n(t)^2 + e_n(t)\left[w_1^*\beta_1(t) + \dots + w_{n-1}^*\beta_{n-1}(t)\right] \leq -\frac{1}{2}e_n(t)^2 + \frac{1}{2}(|w_1^*| + \dots + |w_{n-1}^*|)^2 M^2 e^{-2\lambda t}$$

and

$$V(t) \leq V(0) - \frac{1}{2}\int_0^t e_n(s)^2 ds + \frac{1}{2}\int_0^\infty M^2(|w_1^*| + \ldots + |w_{n-1}^*|)^2 e^{-2\lambda t} dt < \infty .$$

It follows that $e_n(\cdot)$ and $\tilde{w}(\cdot)$ are bounded, and $e_n(\cdot) \in L^2(0,\infty)$. Then, from (25), it follows that $\dot{e}_n(\cdot)$ is bounded. Hence, using Barbălat's Lemma (see Corollary 2.9 in Narendra and Annaswamy (1987)), $\lim_{t\to\infty} e_n(t) = 0$. This proves (ii) and (iii). All that remains is to prove the partial convergence (iv). To do this we borrow techniques from partial convergence proofs in adaptive control (see Theorem 2.7.4 in Sastry and Bodson 1979). Now for each $i = 1, \ldots, n-1$,

$$\lim_{t\to\infty} \dot{\tilde{w}}_i(t) = -\lim_{t\to\infty} (e_n(t) \tanh x_i(t)) = 0 . \tag{26}$$

Let $t_1 > 0$ be arbitrary. We claim that

$$\lim_{t\to\infty} \int_t^{t+t_1} |\tilde{w}(s)^T B(s)|^2 ds = 0. \tag{27}$$

Indeed, using

$$\dot{e}_n(t) = -e_n(t) + \tilde{w}(t)^T B(t) + \sum_{i=1}^{n-1} w_i(t)\beta_i(t)$$

and the facts that $e_n(\cdot) \in L^2(0,\infty)$, $w(\cdot) \in L^\infty(0,\infty)$, and from (i) and the global Lipschitz continuity of $a \mapsto \tanh a$, that $|\beta_i(t)| \leq Me^{-\lambda t}$ for each $i = 1, \ldots, n-1$, (27) holds if

$$\lim_{t\to\infty} \int_t^{t+t_1} \dot{e}_n(s)^2 ds = 0.$$

Now, from (25), $\dot{e}_n(\cdot) \in L^\infty(0,\infty)$ and

$$\ddot{e}_n(t) = -\dot{e}_n(t) + \sum_{i=1}^{n-1} \left(\dot{\tilde{w}}_i(t) \tanh x_i(t) + w_i(t)(1 - (\tanh x_i(t))^2)\dot{x}_i(t) + w_i^*(1 - (\tanh z_i(t)^2))\dot{z}_i(t) \right) ,$$

yields $\ddot{e}_n(\cdot) \in L^\infty(0,\infty)$. Then, using part (iii), we have

$$\lim_{t\to\infty} \int_t^{t+t_1} \dot{e}_n(s)^2 ds = \lim_{t\to\infty} \left(e_n(t+t_1)\dot{e}_n(t+t_1) - e_n(t)\dot{e}_n(t) - \int_t^{t+t_1} e_n(s)\ddot{e}_n(s) ds \right) = 0 ,$$

and therefore (27) holds as claimed. Let $M_1 \geq \max(\|B(\cdot)\|_\infty, \|\tilde{w}(\cdot)\|_\infty)$. Then, following the techniques in the proof of Theorem 2.7.4, in Sastry and Bodson (1979), we have

$$\begin{aligned}
&\tilde{w}(t)^T \left(\int_t^{t+t_1} B(s)B(s)^T ds \right) \tilde{w}(t) \\
&= \int_t^{t+t_1} \tilde{w}(s)^T B(s)B(s)^T \tilde{w}(s) ds + \int_t^{t+t_1} (\tilde{w}(t) - \tilde{w}(s))^T B(s)B(s)^T (\tilde{w}(t) + \tilde{w}(s)) ds \\
&\leq \int_t^{t+t_1} \tilde{w}(s)^T B(s)B(s)^T \tilde{w}(s) ds + 2M_1^3 \sup_{s\in[t,t+t_1]} \|\dot{\tilde{w}}(s)\| \int_t^{t+t_1} |t-s|\, ds \\
&= \int_t^{t+t_1} |\tilde{w}(s)^T B(s)|^2 ds + M_1^3 t_1^2 \sup_{s\in[t,t+t_1]} \|\dot{\tilde{w}}(s)\| .
\end{aligned} \tag{28}$$

Taking limits in (28) as t tends to ∞, using (26) and (27), gives (iv). □

Corollary 4.4 *Under the assumptions of Theorem 4.3:*

(i) If $z(\cdot)$ is periodic and $B(\cdot)$ is persistently exciting, then $e(\cdot)$ and $\tilde{w}(\cdot)$ converge to zero exponentially;

(ii) *If $z(\cdot)$ is periodic with period τ, then*

$$\lim_{t\to\infty} \tilde{w}(t)^T \left(\int_0^\tau B(s)B(s)^T ds\right) \tilde{w}(t) = 0\,, \tag{29}$$

i.e. the weight vector converges to Ker $\left(\int_0^\tau B(s)B(s)^T ds\right)$.

Proof

(i) This follows using the same techniques as in the proof of Theorem 4.1.

(ii) This follows by taking $t_1 = \tau$ in Theorem 4.3, part (iv). □

Whilst the conclusions of Corollary 4.4 give us exact generalisations of the results we obtained in the 3-node case, Corollary 4.4 is unsatisfactory because the additonal conditions of periodicity and persistency of excitation are, in general, uncheckable, except by simulations. Note that Weiß (1997) has obtained similar results to Corollary 4.4 for a slightly more general RNN structure including self-connections of the neurons. More recently, one of the authors, see Weiß (1998), has shown that if $z(\cdot)$ is periodic, then convergence of the weight error given in (24), and of $e_n(t)$ to zero, is exponential.

To illustrate the algorithm in the n-node case we consider an example with $n = 5$ and

$$w^* = (-1.0012, -1.0609, 0.9491, 1.4653).$$

In the simulations we choose $z(0) = (-0.4170, -0.4739, -0.5049, -0.4906, -0.4056)$ for the Teaching Network, and $x(0) = (0,0,0,0,0)$ and $w(0) = (0,0,0,0)$ in the Learning RNN and weight adaptation algorithm. The weights of the Teaching Network were chosen so that the linearization of (2) about $z = 0$ has one pair of unstable complex conjugate eigenvalues and three exponentially stable eigenvalues. Whilst we have been unable to prove that this type of eigenvalue configuration produces oscillatory behaviour, our simulations suggest that this is the case. Figures 7, 8 and 9 show the output signals, the output error convergence on a logarithmic scale and the weight convergence. It is clear, in this simulation, that both the state error $x - z$ and the weight error $\tilde{w}$ converge to zero exponentially. This claim is supported by the fact that the matrix $\left(\int_0^{100} B(s)B(s)^T ds\right)$ is positive definite which, combined with the apparent periodicity of $z(\cdot)$, would give us the required persistency of excitation condition needed to apply Corollary 4.4.

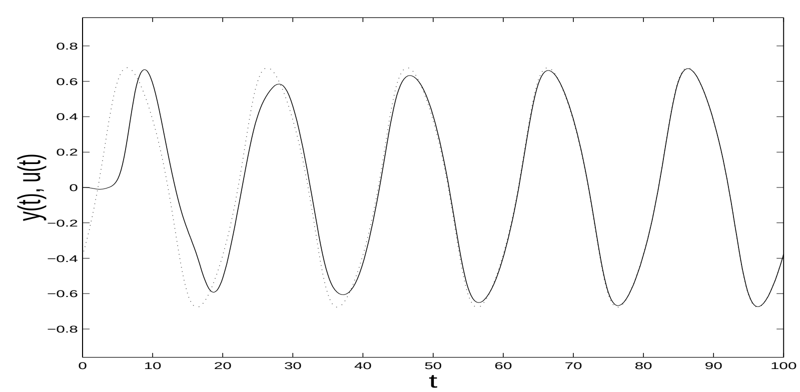

Figure 7: Reference signal $u(t)$ (dotted), learning signal $y(t)$ (continuous)

5 Further comments on weight convergence

In Section 3 we proved that the 3-node Teaching Network has periodic solutions. In Section 4 we proved that the Learning RNN can learn these periodic outputs in the sense that the output

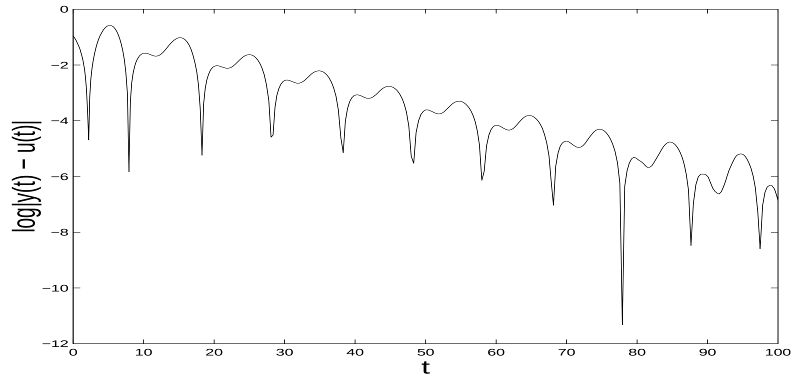

Figure 8: Output error on logarithmic scale: $\log|y(t) - u(t)|$

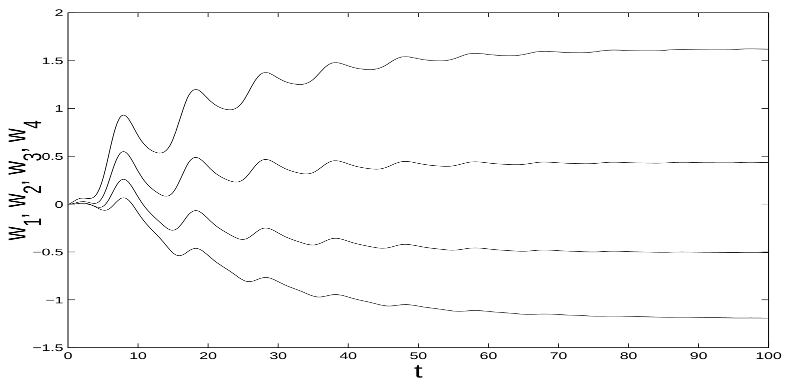

Figure 9: Weight dynamics $w_1, \dots, w_4$

error and weight error converge to zero exponentially. The proof of this exponential convergence used persistency of excitation conditions. For n-node RNNs, we could only obtain partial convergence results. There are two major obstacles to obtaining a general theory in the n-node case. First, we have not been able to determine conditions on the weights of the n-node Teaching Network (2) which guarantee the existence of periodic solutions. The existence of periodic solutions is fundamental to our idea of learning by repetition. Secondly, we have not been able to prove that $(\tanh z_1(\cdot), \dots, \tanh z_{n-1}(\cdot))^T$ is persistently exciting, or, equivalently in the case when $z(\cdot)$ is periodic, that the functions $\tanh z_1(\cdot), \dots, \tanh z_{n-1}(\cdot)$ are linearly independent.

Our simulations suggest that linear independency, and hence the persistency of excitation condition, holds generically amongst those weight parameters for which the Teaching Network (2) has periodic solutions. However, there do exist Teaching Networks which have periodic solutions but which violate the linear independency condition. To construct such an example we need to modify (2) slightly by replacing the triggering function $a \mapsto \tanh a$ with $a \mapsto \tanh \lambda a$ for some $\lambda > 0$.

Proposition 5.1 *Let $z(t)$ be given by (2) but with triggering function $a \mapsto \tanh \lambda a$ for some $\lambda > 0$. Assume $w_k^* = -1$ for some $k \in \{1, \dots, n-1\}$ and $w_j^* = 0$ otherwise.*

(i) If

$$\lambda > 1\,, \tag{30}$$

then the modified Teaching Network possesses a limit cycle.

(ii) If $z(\cdot)$ is a periodic solution of the modified Teaching Network, then $z_{n-j} \equiv -z_{k-1-j}$ for $j = 0, \dots, k-2$. In particular, for $n \geq 4$, if $k \geq 3$, then the functions $\sigma(z_1(\cdot)), \dots, \sigma(z_{n-1}(\cdot))$ are linearly dependent.

Proof

(i) Let $F(z) = (f_1(z), \ldots, f_n(z))^T$ denote the right hand side of the modified Teaching Network. Consider first the case $k = 1$. Then the Teaching Network forms a cyclic system (where components are indexed modulo n) with

$$\prod_{j=1}^{n} \left|\frac{\partial f_j}{\partial z_{j+1}}(0)\right| = \lambda^n > 1 .$$

The existence of a limit cycle follows from Atiya and Baldi (1989) using techniques from Hastings et. al. (1977).

Let $k > 1$. Then the dynamics for components $z_k, \ldots, z_n$ are given by a cyclic $n - k + 1$-dimensional sub-system which, as in the case $k = 1$, has a limit cycle. For the other $k - 1$ components let

$$X = \{f : \mathbb{R} \to \mathbb{R}, f \text{ continuous and periodic}\}$$

and define a nonlinear operator $P : X \to X$ by

$$(Pf)(t) = \int_{-\infty}^{t} \exp(-(t - \tau)\sigma(f(\tau))d\tau ,$$

with $f \mapsto P^j(f)$ denoting P composed with itself j times. If $z_k, \ldots, z_n$ are the component functions of the limit cycle for the $n - k + 1$-dimensional sub-system extended to $\mathbb{R}$ by periodicity, then the periodic functions $z_{k-1} = P(z_k), \ldots, z_1 = P^{k-1}(z_k)$, restricted to $[0, \infty)$, determine the remaining $k-1$ components of the solution and hence the required limit cycle for the modified Teaching Network.

(ii) By the structure of the modified Teaching Network

$$z_{j-1}(\cdot) = (Pz_j)(\cdot) \text{ for } j = 2, \ldots, n , \tag{31}$$

and

$$z_j(\cdot) = (P^{n-j} z_n)(\cdot) \text{ for } j = 1, \ldots, n . \tag{32}$$

On the other hand using $P(-f) = -Pf$ in the nth equation of the modified Teaching Network gives,

$$z_n(\cdot) = (P(-z_k))(\cdot) = -(Pz_k)(\cdot) = -(P^{n-k+1} z_n)(\cdot) = -z_{k-1}(\cdot) . \tag{33}$$

Applying P^j to (33), and using (31), (32) and (33), we obtain

$$z_{n-j}(\cdot) = (P^j z_n)(\cdot) = (P^j(-z_{k-1}))(\cdot) = -z_{(k-1)-j}(\cdot) \text{ for } j = 0, \ldots, k - 2 .$$

This gives a very simple linear dependence of the z_j. Oddity of tanh $(\cdot)$ yields

$$\sigma(z_{n-j}(\cdot)) = -\sigma(z_{k-1-j)}(\cdot)) \text{ for } j = 0, \ldots, k - 2 .$$

□

For a modified Teaching Network of dimension $n \geq 4$ with $k \geq 3$ the resulting dependency of the functions $\sigma(z_1(\cdot)), \ldots, \sigma(z_{n-1}(\cdot))$ means that the corresponding persistency of excitation condition fails. This in turn means that exponential convergence of the weights cannot be guaranteed. To actually find suitable parameters by which failure of exponential convergence of weights is observed can be quite delicate. The failure of exponential convergence does occur in the case $n = 5$ and $k = 3$ with $\lambda = 3$, for which, by Proposition 5.1, $z_2(\cdot) = -z_5(\cdot)$ and $z_1(\cdot) = -z_4(\cdot)$. For the simulation we choose $z(0) = (0.3975, 0.4463, -0.0064, -0.3975, -0.4463)^T$, $x(0) = (0.3, 0.1, 0.2, -0.1, 0.3)^T$ and $w(0) = (-0.05, 0.1, -0.3, 0.1)^T$. In this simulation the components $\tilde{w}_2$ and $\tilde{w}_3$ converge to zero exponentially and, on the time scale of the simulation, are almost indistinguishable. The components $\tilde{w}_1$ and $\tilde{w}_4$ do not converge to zero. Notice the strange behaviour of $|\tilde{w}_4(t)|$ which appears to converge, to zero but then, after $t = 500$, rises to a non zero value.

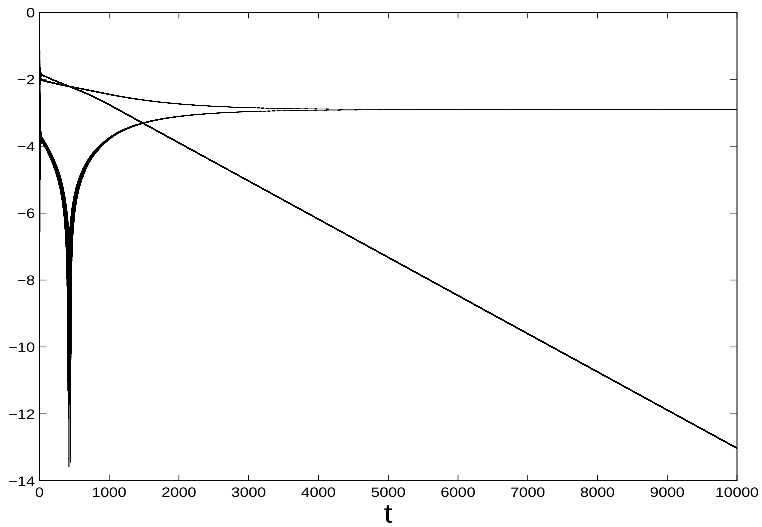

Figure 10: Logarithm of weight error for a 5-dimensional cyclic system

Remark 5.2 (i) For $w_k^* = +1$ in Proposition 5.1, the modified Teaching Network is a *cooperative* system and therefore it cannot have a limit cycle. See Hirsch (1988).

(ii) Whilst the modified cyclic Teaching Network is useful in illustrating that failure of the persistency of excitation can lead to non-convergence of the weights, from the point of view learning its significance is limited. This is due to the fact that the components $z_1(\cdot), \ldots, z_{k-1}(\cdot)$ do not contribute to the dynamics of the output neuron z_n in the Teaching Network. Hence the same output can be generated by a cyclic Teaching Network of dimension $n-k+1$ which, in all our simulations, yields a linearly independent set of functions $\{\tanh z_1(\cdot), \ldots, \tanh z_{n-1}(\cdot)\}$.

6 Conclusions

We have shown, using a result from monotone dynamical systems theory, that a certain 3-node RNN with fixed weights, the so-called Teaching Network, has periodic solutions. The motivation behind the need for the Teaching Network to have periodic solutions arises from the observation that learning usually requires repetition. We then used the periodic output of the Teaching Network as a teaching signal to be learnt by a 3-node Learning RNN. The Learning RNN has a similar structure to the Teaching Network but with time-varying weights. The algorithm by which the weights are adapted is similar to parameter identification algorithms in adaptive control. We were able to prove global exponential convergence of the state and weights of the Learning RNN to the fixed weights and periodic solution of the Teaching Network. This global and exponential convergence is much sharper than the local and asymptotic convergence which is usually associated with gradient descent adaptation. Note also that the inherently nonlinear nature and the resulting limit cycle-like structure of the periodic solutions of the Teaching Network and Learning/Replicating RNN provides robustness of the learnt signal against external disturbances. This contrasts with the case of linear RNNs as developed by Blach and Owens (1992) and Reinke (1994) which are sensitive to such disturbances. We also obtained partial convergence results in the n-node case by using techniques from adaptive control. Under appropriate persistency of excitation type conditions we obtain global exponential convergence as in the 3-node case.

Techniques for speeding up the exponential convergence of the weights in the case when the persistency of excitation condition is satisfied have been developed in Weiß (1997) for a similar RNN structure.

These techniques could also be applied to our class of RNN.
Applications of our results to the control of a robot arm have been developed in Ruiz et al. (1998), in the case of a gradient descent weight adaptation algorithm, and in Waschler (1997), with our weight adaptation algorithm.
An area of research which requires further work is to make use of monotone dynamical systems theory in studying more general RNN structures. So far our results are restricted, in the main, to a special class 3-node RNNs. Another issue, which we did not address here, is to understand the detailed structure of the class of periodic signals which can be generated by the Teaching Network. Our simulations suggest that the periodic signals are very nearly sinusoidal. This issue would be important if many Learning RNNs were combined in parallel so as to facilitate learning of more complicated signals. See Ruiz et al. (1997) for preliminary simulation-based studies of RNNs comprised of several 3-node networks in parallel.
Finally, we have restricted attention to the problem of learning and then replicating a teaching signal. Another issue of interest is to consider the recall capabilities of RNNs. More precisely, how can we build into the Learning RNN, mechanisms for recognising a previously learnt signal so as to then speed-up, or even by-pass, re-learning. Preliminary results in this direction are reported in Waschler et al. (1998).

References

Asada H., J.J.E. Slotine, (1986), *Robot Analysis and Control*, John Wiley & Sons, New York.

Atiya A., P. Baldi, (1989), 'Oscillations and synchronizations in neural networks: an exploration of the labeling hypothesis', *Int. Journal of Neural Systems*, Vol. 1, pp. 103-124.

Babloyantz A., C. Lourenco, A.J. Sepulchre, (1995), 'Control of chaos in delay differential equations in a network of oscillators and in model cortex', *Physica D*, Vol. 86, pp. 274-283.

Blach R., Owens, D. H., (1992), Replication of periodic signals using a neural-type structured network, Internal Report No. 92/03, Centre for Systems and Control Engineering, University of Exeter.

Boyd S., S. S. Sastry, (1986) 'Necessary and sufficient conditions for parameter convergence in adaptive control', *Automatica*, Vol. 22, pp. 629-639.

Doya K., S. Yoshizawa, (1989), 'Adaptive neural oscillator using continuous time backpropagation learning', *Neural Networks*, Vol. 2, pp. 375-385.

Doyon B., M. Cessac, M. Quoy, M. Samuelides, (1993), 'Control of the transition to chaos in neural networks with random connectivity', *Int. J. of Bifurcation and Chaos*, Vol. 3, pp. 279-291.

Hastings S., J. J. Tyson, D. Webster, (1977), 'Existence of periodic solutions for negative feedback equations', *J. Differential Equations*, Vol. 25, pp. 39-64

Hirsch M., (1988), 'Systems of differential equations which are competitive or cooperative III', *Nonlinearity*, Vol.1, pp. 51-71.

Horn R. A., C. R. Johnson, (1985), *Matrix Analysis*, Cambridge University Press, Cambridge.

Li Z., J.J. Hopfield, (1989), 'Modeling the olfactory bulb and its neural oscillatory processings', *Biol. Cybern.*, Vol. 61, pp. 379-392.

Mallet-Paret J., H. Smith, (1990), 'The Poincaré-Bendixson theorem for monotone cyclic feedback systems', *J. Differential Equations*, Vol. 2, pp. 376-421.

Mallet-Paret J., G.R. Sell, (1996), 'The Poincaré-Bendixson theorem for monotone cyclic feedback systems with delay', *J. Differential Equations*, Vol. 125, pp. 441-489.

Matano H., (1986), *Strongly order-preserving local semi-dynamical systems-Theory and Applications*, in: Res. Notes in Math. Vol. 141, Semigroups, Theory and Applications (H. Brezis, M.G. Crandall and F. Kappel, eds.), Longman Scientific and Technical, London, pp. 178-185.

Morgan A. P., K.S. Narendra, (1977), 'On the stability of nonautonomous differential equations $\dot{x} = [A + B(t)]x$, with skew symmetric matrix $B(t)$', *SIAM J. Control and Optimization*, Vol. 15, pp. 163-176.

Narendra K.S., A. Annaswamy, (1987), *Stable Adaptive Systems*, Prentice Hall, Englewood Cliffs, New York.

Narendra K.S., K. Parthasarathy, (1990), 'Identification and control of dynamical systems using neural networks', *IEEE Transactions on Neural Networks*, Vol.1, pp. 4-26.

Pearlmutter B.A., (1995), 'Gradient calculations for dynamic recurrent neural networks: a survey', *IEEE Transactions on Neural Networks*, Vol. 6, pp. 1212-1228.

Reinke R., (1994), 'Adaptive Regeln zum Lernen und Reproduzieren periodischer Signale mit dynamischen Netzwerken', PhD thesis, Universität Kaiserslautern.

Rugh W. J., (1996), *Linear Systems Theory*, Prentice Hall, Upper Saddle River, New Jersey.

Ruiz A., D.H. Owens, S. Townley, (1997), 'L_2 approximation properties of recurrent neural networks', *Proc. 4th European Control Conference*, Brussels, July 1997, electronic publication WE-E G1.

Ruiz A., D.H. Owens, S. Townley, (1998), 'Existence, learning and replication of periodic motions in recurrent neural networks', *IEEE Transactions on Neural Networks*, Vol. 9, pp. 651-661.

Sastry S., M. Bodson, (1979), *Adaptive Control: Stability, Convergence and Robustness*, Prentice Hall, Englewood Cliffs, New York.

Smith H., (1991), 'Convergent and oscillatory activation dynamics for cascades of neural nets with nearest neighbor competitive or cooperative interactions', *Neural Networks*, Vol. 4, pp. 41-46.

Smith H., (1995), *Monotone Dynamical Systems*, American Mathematical Society, Providence, Rhode Island.

Sole R.V., L.M. de la Prida, (1995), 'Controlling chaos in discrete neural networks', *Phys. Lett. A* Vol. 199, pp. 65-69.

Waschler R., (1997), A learning controller for linearizable nonlinear systems, M. Engg. (Eur) Thesis, University of Exeter.

Waschler R., D. H. Owens, S. Townley, (1998), Learning and recall in a class of recurrent neural networks, Internal Report No. 98 01, School of Engineering, University of Exeter.

Weiß M.G., (1997) 'Learning Oscillations using Adaptive Control', in: *Proc. ICANN 97* (Eds. W. Gerstner et al.), LNCS 1327, Springer-Verlag, pp. 331-336.

Weiß M.G., (1998) 'Adaptive Learning of Oscillations in Continuous and Discrete Time Neural Networks', Berichte der Arbeitsgruppe Technomathematik, Report 197, Universität Kaiserslautern.

Yang H., T.S. Dillon (1994), 'Exponential stability and oscillation of Hopfield graded response neural network', *IEEE Transactions Neural Networks*, Vol. 5, pp. 719-729.