ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 369-374

# Maximisation of stability ranges for recurrent neural networks subject to on-line adaptation

Jochen J. Steil and Helge Ritter

University of Bielefeld, Department of Computer Science,
Neuroinformatics Group,
P.O.-Box 10 01 31, D-33501 Bielefeld, Germany,
{jsteil,helge}@techfak.uni-bielefeld.de

**Abstract.** We present conditions for absolute stability of recurrent neural networks with time-varying weights based on the Popov theorem from non-linear feedback system theory. We show how to maximise the stability bounds by deriving a convex optimisation problem subject to linear matrix inequality constraints, which can efficiently be solved by interior point methods with standard software.

## 1 Introduction

One of the most exciting properties of recurrent neural networks (RNN) is their ability to model the time-behaviour of arbitrary dynamical systems [6]. With a number of schemes available which incrementally adapt a network using time-dependent error signals [13] recurrent networks can solve identification and adaptive control tasks in larger systems [8,14]. In such applications the proper functioning of the control system then crucially depends on the the dynamical behaviour of the network.

Thus one of the most investigated issues in RNN theory is stability, especially the existence and uniqueness of a global asymptotically stable (GAS) equilibrium (for instance [12,11,7] and the references therein). If it exists we obtain a unique mapping from constant inputs to the equilibrium which is used in optimisation applications to avoid spurious responses and local minima [5]. Further in control applications with time-varying input a GAS equilibrium of the unforced dynamics ensures the existence of a unique mapping from the input to the output function space. As already pointed out in [7] a GAS equilibrium also avoides the need to reset activations when changing inputs because the location of the equilibrium does not depend on the initial conditions.

Most of the stability results derived so far do not account for time-variation in the weights, which can be caused as well by adaptation as by delays or perturbations in the connections eg. noise in hardware implementations. The introduction of such uncertainty in the parameters causes the dynamical equations describing the corresponding recurrent network to become non-autonomous and then naturally they are much harder to analyse. The corresponding stability results must become absolute stability results: they must hold for all possible variations in some specified range. In Section 2 we discuss the common approach to deal with this complexity and show that it leads to computationally intractable conditions. In Section 3 we present a different scheme to reduce this complexity with the help of the Popov stability criterion. In Section 4 we show how to formulate some resulting optimisation problems and finally add some discussion.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 369-374

## 2 Time-varying weights in recurrent networks

A recurrent neural network subject to adaptation can generally be described as system

$$\dot{\mathbf{x}}(t) = \Leftrightarrow \mathsf{A}\mathbf{x}(t) + (\mathsf{W} + \varDelta\mathsf{W})\varPhi(\mathbf{x}(t)), \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\mathsf{A} = \mathrm{diag}\{a_i\} > 0$ a positive diagonal matrix, the time-stationary weight matrix is denoted by $\mathsf{W} \in \mathbb{R}^{n\times n}$, and $\varPhi(\mathbf{x})$ is a non-linear vector function, which is *sector bounded*, i.e. can be rewritten as $(\varPhi(\mathbf{x}))_i = \varphi_i(x_i) = k_i(x_i)x_i$ with $k_i(x_i) \in [0, 1]$ (e.g. $\varPhi(\mathbf{x}) = \tanh(\mathbf{x})$). Note that the $k_i$ are autonomous parameters as they depend on time only through the state $x_i$. To account for the time changes of the weights we introduce in (1) the additional interval matrix $\varDelta\mathsf{W}$ with unknown, independently varying parameters $\varDelta\omega_{ij}(t) \in [\underline{\varDelta}_{ij}, \overline{\varDelta}_{ij}]$. As opposed to the autonomous $k_i$ these parameters are non-autonomous as they depend on time in some unknown way. Below we will use the Popov theorem to draw on this difference and to address the following absolute stability problem: *Find conditions on the weight matrix* $\mathsf{W}$*, the non-linear functions* $\varphi_i$*, and the parameters* $\underline{\varDelta}_{ij}, \overline{\varDelta}_{ij}$ *such that for all variations of* $\varDelta w_{ij}(t)$ *within its given range the network (1) is GAS.*

So far this problem has mainly been studied as a special case of a time-varying linear system $\dot{\mathbf{x}} = \mathsf{B}(t)\mathbf{x}$, where $\mathsf{B}(t)$ is an element of the convex polytope $\mathbb{B}$ of $n \times n$ matrices obeying the $n^2$ interval restrictions on their elements $b_{ij}(t)$ that result from the constraints on the $k_i$ and $\varDelta w_{ij}$. In [16] only the parametrisation of the $n$ non-linear functions $\varphi_i(x_i) = k_i(x_i)x_i$ is studied and criteria relying on the parallel solution of $2^n$ Lyapunov equations for the $2^n$ vertices of the corresponding smaller matrix polytope are given together with a very conservative sufficient condition $||\mathsf{A}^{-1}\mathsf{W}|| < 1$. Inclusion of $N$ time-varying weights in this approach then requires to solve simultaneously the corresponding $2^N$ Lyapunov equations.

The same approach of parametrising only the non-linearity is taken also in [3], where conditions based on matrix measures are given and lead to simple conditions if also the slopes $\dot{\varphi}_i(x_i)$ are bounded. As shown in [4] inclusion of $N$ time-varying weights in this approach requires to prove $[\mathsf{B}_i]^S = \mathsf{B}_i + \mathsf{B}_i^T < 0$ for all $2^N$ symmetric parts $[\mathsf{B}_i]^S$, where $\mathsf{B}_i$ is a vertex of the polytope $\mathbb{B}$. Though this can in principle be done by interior point methods [1] (as also the solution of Lyapunov equations in [16]) the large number of vertex matrices makes this intractable in practice and in both approaches it is not distinguished between intervals resulting from time-varying (non-autonomous) $\varDelta w_{ij}$ and time-invariant (autonomous) $k_i$.

The only conditions which use this additional information are derived in [5], where the matrix $\varDelta\mathsf{W}$ is regarded as disturbance matrix added to the time-invariant system (1) with $\varDelta\mathsf{W} = 0$, which is shown to be GAS if there exists a positive diagonal matrix $\mathsf{D}$ such that

$$[(\mathsf{D}(\Leftrightarrow\mathsf{W} + \mathsf{A}))]^S > 0. \tag{2}$$

Then there can either be used a conservative bound on the maximal allowed total deviation $\sum_{ij} |\varDelta w_{ij}|$, which depends on the matrix $\mathsf{D}$ and needs the computation of a factorisation of $\mathsf{D}(\Leftrightarrow\mathsf{W} + \mathsf{A}) = \mathsf{L}^T\mathsf{L}$, or we have to compute once more eigenvalues for the $2^{n^2}$ vertex matrices derived from $\varDelta\mathsf{W}$. Also in this approach it is intractable to evaluate or maximise the stability bounds on $\varDelta\mathsf{W}$ in an computationally efficient way.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 369-374

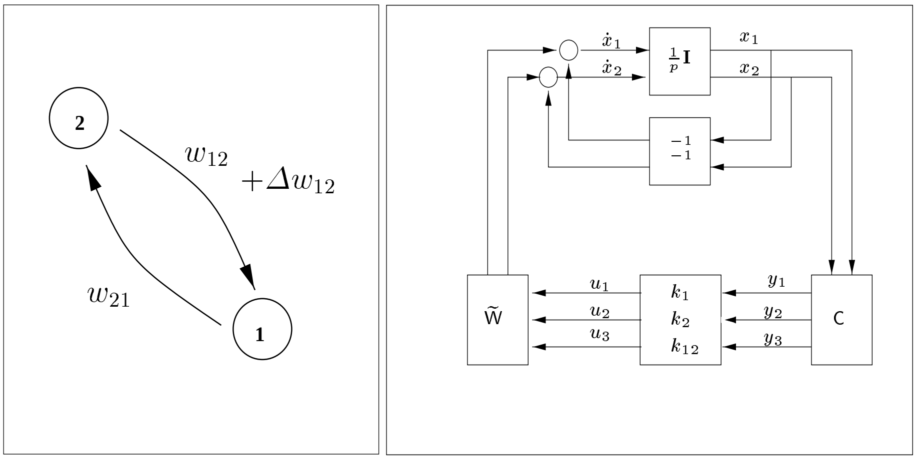


**Fig. 1.** (a) The minimal recurrent network with time-varying weight $w_{12} + \Delta\omega_{12}$. (b) The corresponding feedback circuit according to (3) with modified weight matrix $\widetilde{\mathsf{W}}$ and additional feedback $k_{12}$ to account for the time-varying $\Delta w_{12}$.

## 3 Stability with the Popov theorem

In the following we give an alternative approach to avoid the computational explosion using the vertices of the 'uncertainty polytope' $\mathbb{B}$. We achieve this in two steps, (i) we rewrite the system (1) to express the time-dependence of each $\Delta w_{ij}(t)$ by an additional parameter $k_{ij}(t)$, which is also *sector bounded*, and (ii) we regard the resulting system as linear system with non-linear partially time-invariant $k_i(x_i)x_i$ and partially time-varying $k_{ij}(t)x_j$ feedback, which allows to apply the well known Popov theorem from non-linear feedback system theory [2].

We illustrate the main idea with the help of the 'minimal' network shown in Fig. 1(a), where $\mathsf{A} = \mathbf{I}$, $\mathsf{W} = \left(\begin{smallmatrix} 0 & w_{12} \\ w_{21} & 0 \end{smallmatrix}\right)$ and $\Delta\mathsf{W} = \left(\begin{smallmatrix} 0 & \Delta w_{12} \\ 0 & 0 \end{smallmatrix}\right)$, which we rewrite as

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} \Leftrightarrow x_1 \\ \Leftrightarrow x_2 \end{pmatrix} + \begin{pmatrix} 0 & (w_{12} \Leftrightarrow \underline{\Delta}_{12}) & 1 \\ w_{21} & 0 & 0 \end{pmatrix} \begin{pmatrix} k_1 y_1 \\ k_2 y_2 \\ k_{12} y_3 \end{pmatrix}, \quad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{3}$$

where $k_1(x_1), k_2(x_2) \in [0, 1]$ (as in the general case all $k_i(x_i)$), and $k_{12} \in [0, \overline{\Delta}_{12} \Leftrightarrow \underline{\Delta}_{12}]$ characterises the uncertainty in the weight $w_{12}$. The systems in the form (1) and (3) are equivalent because $(w_{12} + \Delta\omega_{12}(t)) = (\omega_{12} \Leftrightarrow \underline{\Delta}_{12} + k_{12}(t))$ for some $k_{12}(t)$ and we will prove stability for all possible $k_{ij}(t)$ in its given range. The resulting feedback loop is shown in Fig. 1(b).

In the general case let $N$ be the number of connections for which $\Delta w_{ij} \neq 0$ , then we have

$$\dot{\mathbf{x}} = \Leftrightarrow \mathsf{A}\mathbf{x} + \widetilde{\mathsf{W}}\mathbf{u}, \quad \mathbf{y} = \mathsf{C}\mathbf{x}, \quad \mathbf{u} = \mathcal{K}(t)\mathbf{y}, \tag{4}$$

where the matrix $\widetilde{\mathsf{W}} \in \mathbb{R}^{n \times (n+N)}$ includes first the $n \times n$ matrix $\mathsf{W} + \underline{\Delta}\mathsf{W}$ and then one column with a single entry 1 for each $\Delta w_{ij}$ which is then multiplied by $k_{ij}(t) \in$

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 369-374

$[0, \bar{k_{ij}}]$ with $\bar{k}_{ij} = \overline{\Delta}_{ij} \Leftrightarrow \underline{\Delta}_{ij}$. We get further $\mathcal{K}(\mathbf{x}, t) = \mathrm{diag}\{k_i(x_i), k_{ij}(t)\} \in \mathbb{R}^{(n+N)^2}$ and $\mathsf{C} \in \mathbb{R}^{(n+N)\times n}$ is easily constructed as appropriate. The systems (1) and (4) are equivalent but (4) is in the standard form of a linear control system with input $\mathbf{u}$ and input matrix $\widetilde{\mathsf{W}}$, output $\mathbf{y}$ and output matrix $\mathsf{C}$, and non-linear sector bounded feedback $\mathbf{u} = \mathcal{K}(\mathbf{x}, t)\mathbf{y}$. Therefore we can apply the multivariable Popov theorem as for example in [2],[10].

**Theorem 1.** *The system (1) is absolute stable for all* $\Delta w_{ij}(t) \in [\underline{\Delta}_{ij}, \overline{\Delta}_{ij}]$ *and all non-linear time-stationary feedback functions* $\varphi_i(x_i) = k_i(x_i)x_i, k_i(x_i) \in [0, 1]$, *if there exist diagonal matrices* $\mathsf{P} = \mathrm{diag}\{p_i, p_{ij}\} > 0$, $\mathsf{Q} = \mathrm{diag}\{q_i, \mathbf{0}\} > 0$ *such that*

$$\mathrm{Re}\,[\mathsf{M}(\imath\omega)] = \mathrm{Re}\left[\mathsf{P}((\mathbf{I} + \imath\omega\mathsf{Q})\mathsf{G}(\imath\omega) \Leftrightarrow \bar{\mathcal{K}}^{-1})\right] < 0, \text{ for all } \omega \in [0, \infty], \qquad (5)$$

*where* $\mathrm{Re}\,[\mathsf{M}] = \frac{1}{2}(\mathsf{M}^* + \mathsf{M})$ *denotes the hermitian part of the complex matrix* $\mathsf{M}$ *and* $\mathsf{G}(\imath\omega) = \mathsf{C}(\mathsf{A} + \imath\omega\mathbf{I})^{-1}\widetilde{\mathsf{W}} \in \mathbb{R}^{(n+N)^2}$ *is the frequency matrix of the system (4) and* $\bar{\mathcal{K}}^{-1} = \mathrm{diag}\{\mathbf{I}, \bar{k}_{ij}^{-1}\}$.

Note that the first $n$ rows in $\mathsf{G}(\imath\omega)$ correspond to the time-invariant $k_i$, which allows to include additional parameters $q_i > 0, i = 1..n$ in this range, whereas the following $N$ rows correspond to the time-varying $k_{ij}(t)$ where we have $q_{ij} = 0$.

**Remarks:**

**1** If $\mathsf{W}$ is normal ($\mathsf{W}^*\mathsf{W} = \mathsf{W}\mathsf{W}^*$) and $\mathsf{A} = \mathbf{I}$ we showed previously that it is possible to derive efficient graphical tests involving the eigenvalues of $\mathsf{W}$ only [15].

**2** The condition (2) is a special case of (5) when $\Delta\mathsf{W} = 0$. Then we get $\mathsf{C} = \mathbf{I}$ and can choose $\mathsf{P} = \mathsf{DA}, \mathsf{Q} = \mathsf{A}^{-1}$ for any positive diagonal $\mathsf{D}$, which removes the frequency dependence in (5) and yields

$$(5) = [\mathsf{D}(\mathsf{W} \Leftrightarrow \mathsf{A})]^S < 0 \Leftrightarrow [\mathsf{D}(\Leftrightarrow\mathsf{W} + \mathsf{A})]^S > 0 = (2).$$

**3** Also the norm bound derived in [16] can be obtained from (5) if $\Delta\mathsf{W} = 0$ and $\mathsf{C} = \mathbf{I}$. Then choose $\mathsf{P} = \mathbf{I}, \mathsf{Q} = 0$ and use that

$$\|\mathsf{A}^{-1}\mathsf{W}\| < 1 \Rightarrow \mathrm{Re}\left[(\mathbf{I} + \mathsf{A}^{-1}\imath\omega)^{-1}\mathsf{A}^{-1}\mathsf{W}\right] < 1 \Rightarrow (5).$$

## 4 Maximisation of the stability range

If maximal bounds on $\Delta w_{ij}$ are given, e.g. we deal with noise or some normalisation in the adaptation scheme, then we can directly use the condition (5) in the frequency domain in two steps. First we prove it for $\omega = 0$ and then evaluate the determinant of $\det \mathrm{Re}\,[\mathsf{M}(\imath\omega)]$ as a function of $\omega$. If it has no real roots, then no eigenvalue changes sign and we can conclude from the case $\omega = 0$ for stability for all $\omega$ ( for an efficient algorithm see [10]).

The first step can be put into a form suited for application of interior point polynomial methods, which efficiently solve convex optimisation problems subject to linear

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 369-374

matrix inequalities (LMI) constraints [1] (for links to standard software see [9]). Before setting $\omega$ to zero we choose $\mathsf{Q}$ to remove the frequency dependence in the first $n$ coordinates, i.e. $\mathsf{Q} = \mathsf{A}^{-1}$, and $\mathsf{P} = \mathrm{diag}\{p_i a_i, p_{ij}\}$, which results in the condition (2) if only the first $n \times n$ block of $\mathsf{G}(\imath\omega)$ is considered (see Remark [2], Sec. 3). We obtain the following LMI feasibility problem

$$\text{find } \mathsf{P} \text{ subject to} \quad \mathrm{Re}\left[\mathsf{P}\,\mathrm{diag}\{\mathsf{A},\mathbf{I},\mathbf{I}\}(\mathsf{CA}^{-1}\mathsf{W} \Leftrightarrow \bar{\mathcal{K}}^{-1})\right] < 0, \mathrm{diag}\{\mathsf{P}\} > 0.$$

In the general case we want also to maximise the sector bounds $\bar{k}_{ij}$. In this case we can use the famous Kalman-Yakubovich-Lemma to show that (5) is equivalent to the problem to find matrices $\mathsf{H} = \mathsf{H}^T > 0 \in \mathbb{R}^{n \times n}$, $\mathsf{P}$, and $\mathsf{Q}$ such that

$$\begin{pmatrix} \mathsf{AH} + \mathsf{H}^T\mathsf{A} & \mathsf{H}\widetilde{\mathsf{W}} + \mathsf{AC}^T\mathsf{Q} + \mathsf{C}^T\mathsf{P} \\ \widetilde{\mathsf{W}}^T\mathsf{H} + \mathsf{QCA} + \mathsf{PC} & \mathsf{QC}\widetilde{\mathsf{W}} + \mathsf{W}^T\mathsf{C}^T\mathsf{Q} \Leftrightarrow 2\mathsf{P}\bar{\mathcal{K}}^{-1} \end{pmatrix} < 0. \tag{6}$$

In case of fixed $\bar{\mathcal{K}}^{-1}$ we can now use (6) as LMI constraint on the variables $\mathsf{H}, \mathsf{P}, \mathsf{Q}$ and solve the corresponding feasibility problem, but regarding also $\bar{k}_{ij}$ as a variable for optimisation leads in the constraint matrix (6) to the term $\mathsf{P}\bar{\mathcal{K}}^{-1}$ where both $\mathsf{P}$ and $\bar{\mathcal{K}}^{-1}$ are optimisation variables. To avoid such a quadratic term which destroys the linear dependence of the matrix on the variables we define new variables $r_i = p_i, r_{ij} = p_{ij}/\bar{k}_{ij}$, $\mathsf{R} = \mathrm{diag}\{r_i, r_{ij}\}$ and substitute in (6) $\mathsf{R}$ instead of $\mathsf{P}\bar{\mathcal{K}}^{-1}$. Then we can solve the optimisation problem

$$\text{maximise} \quad \sum_{ij} p_{ij} \Leftrightarrow r_{ij} \quad \text{subject to}$$
$$(6), \mathsf{H} = \mathsf{H}^T > 0, \mathsf{P} > 0, \mathsf{Q} = \mathrm{diag}\{q_i, \mathbf{0}, \mathbf{0}\} > 0, \mathsf{R} > 0,$$

which leads to large values of the $\bar{k}_{ij}$ because if $p_{ij}$ is large and $r_{ij} = p_{ij}/\bar{k}_{ij}$ is small then the corresponding $\bar{k}_{ij}$, which yields the upper stability bound, must be large as well. This formulation also allows to include easily constraints like removing or freezing a connection $w_{ij}$, which corresponds to set $r_{ij}$ to the corresponding value and exclude it from the optimisation. The optimal solution of the previous problem then is a good starting point for the new one and in general yields fast convergence.

## 5 Discussion

We give a new approach to the computation of stability ranges for time-varying weights which leads to convex optimisation problems tractable with interior point methods. Though we discuss optimisation of the upper bound of the stability range only, we could minimise the lower bound as well, and also, adding additional $N$ parameters, upper and lower bounds simultaneously. In the contrary to the theoretical derivation, which uses well known but advanced concepts from non-linear system theory, the resulting conditions are 'ready to apply' with standard software.

It is interesting that the to our knowledge best known condition for time-invariant networks (2) derived in [5] appears to be a special case of the frequency approach given

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 369-374

here, though the Lyapunov function used in [5] is not a Lur'e-Postnikov function which can be constructed from the the LMI (6). It seems that the condition (2) may be the best possible if the non-linearities are sector bounded in [0,1] and time-invariant.

However, usually the squashing functions used in neural networks have additional properties, for instance they are bounded which in turn bounds $\|\mathbf{x}(t)\|$ and leads to smaller sectors $[\underline{k_i}, 1]$ for the parameters $k_i(x_i)$. Further they are slope restricted and even the sign of the second derivative may be known. The flexible and quite universal frequency conditions can also incorporate such information to improve the stability bounds, which will be subject to future work.

**Acknowledgements:** J.J. Steil is supported from the DFG (German Research Foundation) under grant RI 621/2-1. We thank Prof. Dr. I.B. Junger for fruitful discussions.

## References

1. S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *SIAM Studies in Applied Mathematics*. SIAM, Philadelphia, 1994.
2. C. Desoer and M. Vidyasagar. *Feedback Systems: input-output properties*. Academic Press, New York, 1975.
3. Y. Fang and T. G. Kincaid. Stability analysis of dynamical neural networks. *IEEE Tansactions on Neural Networks*, 7(4):996–1005, 1996.
4. Y. Fang, K. A. Lopardo, and X. Feng. A sufficient condition for staiblity of a polytope of matrices. *Systems & Control Letters*, 23:237–245, 1994.
5. M. Forti and A. Tesi. New conditions for global stability of neural networks with application to linear and quadratic programming problems. *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, 42(7):354–366, 1995.
6. K. Funahashi and Y. Nakamura. Approximation of dynamical systems by continous time recurrent neural networks. *Neural Networks*, 6:801–806, 1993.
7. W. Hirsch. Convergent activation dynamics in continuous time networks. *Neural Networks*, 1:331–349, 1989.
8. K. J. Hunt, D. Sbarbaro, R. Zbikowski, and P. J. Garthrop. Neural networks for control systems – a survey. *Automatica*, 28(6):1083–1111, 1992.
9. Interior point methods online. http://www.mcs.anl.gov/otc/InteriorPoint/.
10. I. B. Junger and O. I. Gerasimov. Algebraic analysis of absolute stability for uncertain dynamical systems with nonlinear time-varying properties. *Automatica*, 29(3):763–766, 1993.
11. X. Liang and L. Wu. Global exponential stability of hopfield-type neural network and its applications. *Science in China (Series A)*, 38(6):757–768, 1995.
12. K. Matsuoka. Stability conditions for nonlinear continuous neural networks with asymmetric connection weights. *Neural Networks*, 5:495–500, 1992.
13. B. A. Pearlmutter. Gradient calculations for dynamic recurrent neural networks: A survey. *IEEE Tansactions on Neural Networks*, 6(5):1212–1228, 1995.
14. J.-J. C. Slotine and R. M. Sanner. Stable adaptive control of robot manipulators using neural networks. *Neural Computation*, 7(4):753–790, 1995.
15. J. J. Steil and H. Ritter. Input-output stability of recurrent neural networks with delays using circle criteria. In *Proc. Int. ICSC/IFAC Symp. NEURAL COMPUTATION, NC'98*, pages 519–525. ICSC Academic Press, 1998.
16. K. Wang and A. M. Michel. Stability analysis of differential inclusions in banach space with applications to nonlinear systems with time delays. *IEEE Transactions on Circuits and Systems-I:Fundamental Theory and Applications*, 43(8):617–626, 1996.