Bounds of the Induced Norm and Model Reduction Errors for a Class of Nonlinear Systems

Yun-Chung Chu[1] and Keith Glover
Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ, England
ycc@eng.cam.ac.uk kg@eng.cam.ac.uk

Abstract

The class of nonlinear systems described by a discrete-time state equation containing a diagonal nonlinear term as in recurrent neural networks is considered. Sufficient conditions are derived for the stability and induced norm of such systems using positive definite diagonally dominant Lyapunov functions or storage functions, satisfying appropriate Linear Matrix Inequalities. Preliminary results are also presented for model reduction errors for such systems.

Keywords: discrete-time nonlinear systems, Linear Matrix Inequalities, diagonally dominant matrices, recurrent neural networks, state-dependent perturbations, diagonal stability, ℓ_2-ℓ_2 induced gain, model reduction, balanced truncation, Linear Fractional Transformations.

1 Introduction

In this paper we consider the discrete-time nonlinear system described by

$$x^+ = A\underline{\phi}(x) + Bu \qquad (1)$$
$$y = C\underline{\phi}(x) + Du \qquad (2)$$

(superscript + denotes time-shift) where $u \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$, $\underline{\phi}(x) \in \mathbb{R}^n$,

$$\underline{\phi}(x) = \begin{bmatrix} \phi(x_1) \\ \phi(x_2) \\ \vdots \\ \phi(x_n) \end{bmatrix} \qquad (3)$$

and $\phi : \mathbb{R} \to \mathbb{R}$ is nonlinear. Results on four global properties of the system, namely the stability, the ℓ_2-ℓ_2 induced gain, the observability function and the controllability function are presented, and so is a proposition analogous to the balanced model reduction result for linear systems. Note that a discrete-time recurrent artificial neural network is simply a special case of (1,2,3), where ϕ is the activation function of the neuron.

When the function ϕ for each state is not necessarily identical, i.e.

$$\underline{\phi}(x) = \begin{bmatrix} \phi_1(x_1) \\ \phi_2(x_2) \\ \vdots \\ \phi_n(x_n) \end{bmatrix} \qquad (4)$$

and satisfies

$$\forall s \in \mathbb{R}, \quad |\phi_i(s)| \le |s| \qquad (5)$$

the system has been studied by Kaszkurewicz and Bhaya[1]. They used a diagonal Lyapunov function to prove the global stability of the system (1,2,4,5). This is very powerful if ϕ_i are interpreted as state-dependent perturbations. Since then they have also realized the system's close connection to artificial neural networks[2, 3]. However, if the system is interpreted as a neural network, the activation function of each neuron is usually chosen to be the same, and therefore a diagonal Lyapunov function should be too restrictive. In this paper (3) instead of (4) is assumed. In fact, if ϕ is invertible, the system (1,2,3) is (input-output) equivalent to

$$x^+ = \underline{\phi}(Ax + Bu) \qquad (6)$$
$$y = Cx + Du \qquad (7)$$

which is the neural network that has been studied by Albertini and Sontag *et al.*, who have already published some results on the identifiability, the minimality, the observability[4, 5, 6] and the forward accessibility[7] based on different assumptions on ϕ. (In their papers

[1] Y.-C. Chu greatly appreciates the financial support of the Committee of Vice-Chancellors and Principals "Overseas Research Students Awards Scheme" (92–95), the Friends of Cambridge University in Hong Kong and the Cambridge Commonwealth Trust (92–93), the Croucher Foundation (93–95), the Cambridge Philosophical Society (96), and the Department of Engineering at University of Cambridge (96) in his Ph.D. studies during which the work presented in this paper has been done.

the symbol σ has been used instead of ϕ here. However, the authors have also been working on the model reduction of the system in concern and σ will be used to denote the Hankel singular value in future publications.)

2 ℓ_2-ℓ_2 Induced Gain

Notation 1 We say that a matrix P is pdd if $P > 0$ (positive definite, i.e. P is symmetric such that $x^*Px > 0$ for all $x \neq 0$, where superscript $*$ denotes transpose) and (row) diagonally dominant:

$$\forall i,\ \ |p_{ii}| \geq \sum_{j \neq i} |p_{ij}|.$$

$\triangleleft$

Lemma 1 Assume that ϕ satisfies

$$\forall s, t \in \mathbb{R},\ \ |\phi(s) + \phi(t)| \leq |s + t|. \tag{8}$$

Then if P is pdd, P will satisfy

$$\forall x,\ \ \underline{\phi}(x)^* P \underline{\phi}(x) - x^* P x \leq 0. \tag{9}$$

$\triangleleft$

Remark 1 (8) is equivalent to the condition that ϕ is odd (by putting $t = -s$) and 1-Lipschitz:

$$\forall s, t \in \mathbb{R},\ \ |\phi(s) - \phi(t)| \leq |s - t| \tag{10}$$

(by putting $t = -t$). So for example, the semilinear, the hyperbolic tangent (the most popular activation function used in neural networks) and the sine function all fall in the class. Certainly (8) implies (5). $\otimes$

Proof of Lemma 1: (8,10) implies that $\forall s, t \in \mathbb{R}$,

$$(\phi(s) \pm \phi(t))^2 \leq (s \pm t)^2$$

and hence

$$2|\phi(s)\phi(t) - st| \leq -\left((\phi(s)^2 - s^2) + (\phi(t)^2 - t^2)\right). \tag{11}$$

Note that

$$\begin{aligned} & \underline{\phi}(x)^* P \underline{\phi}(x) - x^* P x \\ = & \sum_{ij} \phi(x_i) p_{ij} \phi(x_j) - \sum_{ij} x_i p_{ij} x_j \\ = & \sum_{ij} p_{ij} (\phi(x_i)\phi(x_j) - x_i x_j). \end{aligned}$$

Then the remainder of the proof is standard: Write $\phi(x_i)\phi(x_j) - x_i x_j$ as α_{ij}. Then we have $\alpha_{ii} \leq 0$ from (5) and

$$2|\alpha_{ij}| \leq -(\alpha_{ii} + \alpha_{jj})$$

from (11). So

$$\begin{aligned} 2\sum_{i \neq j} p_{ij}\alpha_{ij} & \leq 2\sum_{i \neq j} |p_{ij}||\alpha_{ij}| \\ & \leq -\sum_{i \neq j} |p_{ij}|(\alpha_{ii} + \alpha_{jj}) \\ & = -2\sum_{i \neq j} |p_{ij}|\alpha_{ii} \\ & \leq -2\sum_{i} p_{ii}\alpha_{ii}. \end{aligned}$$

$\blacksquare$

Proposition 2 If there exists a pdd P such that

$$A^* P A - P < 0 \tag{12}$$

then (the origin of) the system (1,2,3,8) is globally asymptotically stable. $\triangleleft$

The proof is omitted since we shall outline the proof of Proposition 3.

From now on we assume that the system (1,2,3,8) is globally asymptotically stable.

Proposition 3 If there exists a pdd P such that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^* \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} - \begin{bmatrix} P & 0 \\ 0 & \gamma^2 I \end{bmatrix} \leq 0 \tag{13}$$

then the ℓ_2-ℓ_2 induced gain of the system (1,2,3,8) $\leq \gamma$. More precisely, if $x(0) = 0$, then

$$\sum_{k=0}^{\infty} y(k)^* y(k) \leq \gamma^2 \sum_{k=0}^{\infty} u(k)^* u(k).$$

$\triangleleft$

Proof: Let $V(x) = x^* P x$. Pre- and post-multiply (13) by $\begin{bmatrix} \underline{\phi}(x) \\ u \end{bmatrix}^*$ and respectively $\begin{bmatrix} \underline{\phi}(x) \\ u \end{bmatrix}$ to get

$$\begin{aligned} & (A\underline{\phi}(x) + Bu)^* P (A\underline{\phi}(x) + Bu) - \underline{\phi}(x)^* P \underline{\phi}(x) \\ \leq\ & \gamma^2 u^* u - (C\underline{\phi}(x) + Du)^* (C\underline{\phi}(x) + Du) \end{aligned}$$

i.e.

$$V(x^+) - \underline{\phi}(x)^* P \underline{\phi}(x) \leq \gamma^2 u^* u - y^* y.$$

Then apply Lemma 1 to give

$$y^* y \leq x^* P x - V(x^+) + \gamma^2 u^* u$$

and adding over k gives the result. $\blacksquare$

Notation 2 We write γ_l, γ_{dd} and γ_d to denote the smallest γ such that (13) is satisfied by a positive definite P, a pdd P and a positive diagonal P respectively. Then γ_l is actually the ℓ_2-ℓ_2 induced gain of the corresponding linear system, i.e. when $\phi =$ the identity, and γ_d is actually extending Kaszkurewicz and Bhaya's idea of a diagonal Lyapunov function to deducing an upper bound of the ℓ_2-ℓ_2 induced gain. $\triangleleft$

Computing γ_l or γ_d is obviously a minimization of a linear objective subject to a Linear Matrix Inequality (LMI) constraint. (For a general reference of LMI, see [8].) The important fact is that so is computing γ_{dd}, since we can specify a matrix to be pdd by a set of LMI whose "size" only grows reasonably with n. (See Lemma 4 below for the precise statement.) Hence Lemma 1 and all the propositions are guaranteed to be useful both theoretically and practically.

Lemma 4 P is pdd if and only if $P > 0$ and there exists a symmetric R such that

$$\begin{aligned} \forall i \neq j, \qquad & r_{ij} \geq 0, \;\; p_{ij} + r_{ij} \geq 0 \\ \forall i, \qquad & p_{ii} \geq \sum_{j \neq i} (p_{ij} + 2r_{ij}). \end{aligned}$$

Hence there would be only $n(n-1)/2$ additional variables and n^2 additional inequalities. $\triangleleft$

Proof: Sufficiency is obvious since

$$\begin{aligned} p_{ii} \;\; & \geq \;\; \sum_{j \neq i} (p_{ij} + 2r_{ij}) \\ & = \;\; \sum_{j \neq i} (|p_{ij} + r_{ij}| + |-r_{ij}|) \\ & \geq \;\; \sum_{j \neq i} |p_{ij}| \end{aligned}$$

Necessity follows from choosing

$$r_{ij} = \begin{cases} -p_{ij} & \text{if} \;\; p_{ij} < 0 \\ 0 & \text{if} \;\; p_{ij} \geq 0 \end{cases}$$

■

As an example to see the improvement of using a pdd P instead of a positive diagonal P when all ϕ_i are identical, we show how much γ_{dd} is tighter than γ_d. We use the `drmodel` function in the Control toolbox of MATLAB to generate 22 discrete-time random systems, with $m = 1$, $n = 4$, $p = 2$, and then use the LMI Control toolbox[9] of MATLAB to compute all γ_d, γ_{dd} and γ_l. The results are summarized in Table 1. Certainly

$$\gamma_d \geq \gamma_{dd} \geq \gamma_l.$$

Note that if ϕ is locally the identity at the origin ($\phi'(0) = 1$), then the actual ℓ_2-ℓ_2 induced gain of the system (1,2,3,8) is lower bounded by γ_l and upper bounded by γ_{dd}, and we see that in many cases in Table 1, both bounds coincide, i.e. the ℓ_2-ℓ_2 induced gain of the system (1,2,3,8) is certain and the same for all ϕ satisfying (8), even though we do not actually know what ϕ is. Clearly this will occur when the set of solutions to the LMI for γ_l contains a solution that is pdd.

Table 1: Comparison between γ_d and γ_{dd}

	γ_d	γ_{dd}	γ_l
1	4.2612	2.9819	2.9819
2	2.6979	2.2819	2.2819
3	1.2141	1.2141	1.2141
4	2.4801	2.4801	2.4801
5	1.7715	1.7516	1.7516
6	1.7257	1.7211	1.7211
7	4.7369	4.7361	4.7361
8	1.3429	0.8037	0.7127
9	2.8785	2.8209	2.8201
10	2.0328	1.9478	1.9478
11	1.4438	1.3225	1.3202
12	1.8311	1.8290	1.8290
13	1.0145	1.0145	1.0145
14	1.8812	0.8222	0.8161
15	2.6264	2.4879	2.4735
16	4.7092	3.3271	3.3271
17	0.7911	0.7888	0.7888
18	1.7156	1.1278	1.1278
19	4.6011	4.5416	4.5408
20	1.8608	1.8608	1.8608
21	2.6772	2.3915	2.3915
22	9.1276	6.2394	6.2393

3 Model Reduction

The use of a pdd P instead of a positive diagonal P does not merely give stronger results. The need is actually more basic. Suppose we consider two systems G_1, G_2 of the form (1,2,4,8) with the same dimension (m, n, p) and nonlinearity $\underline{\phi}$, with ϕ_i all different for different i. Sometimes we would like to know how "close" they are, by examining the ℓ_2-ℓ_2 induced gain of their difference $G_1 - G_2$. Note that $G_1 - G_2$ is of the form (1,2,4,8) with dimension $(m, 2n, p)$ but in addition ϕ_i pairwise identical. Using a diagonal P does not take the latter into account and therefore does not provide useful conclusions. (The requirement that two systems be necessarily close even if the two sets of ϕ_i are not necessarily identical is too strong to be useful practically.)

Therefore a method has to be developed to handle the situation when certain ϕ_i are known to be identical, and our solution is to introduce a (block-diagonal) pdd P. (See Remark 5 below.)

The need to compare how close two systems are arose when the authors started to consider the model reduction of the system (1,2,4,8) or (1,2,3,8). In this section we give some results in this direction. For clarity the results are stated for the system (1,2,3,8) only, and still assuming that it is globally asymptotically stable.

Proposition 5 If there exists a pdd Q such that

$$A^*QA - Q + C^*C \leq 0 \tag{14}$$

then the output energy generated by the state $x(0)$ is upper-bounded by $x(0)^*Qx(0)$. More precisely, if $u(k) = 0$ for $k \geq 0$, then

$$\sum_{k=0}^{\infty} y(k)^*y(k) \leq x(0)^*Qx(0).$$

$\triangleleft$

Proof: Pre- and post-multiply (14) by $\underline{\phi}(x)^*$ and respectively $\underline{\phi}(x)$ and then apply Lemma 1. $\blacksquare$

Proposition 6 If there exists P such that P^{-1} is pdd and

$$APA^* - P + BB^* \leq 0 \tag{15}$$

then the input energy needed to reach the state $x(0)$ from 0 is lower-bounded by $x(0)^*P^{-1}x(0)$. More precisely, if $x(-\infty) = 0$, then

$$\sum_{k=-\infty}^{-1} u(k)^*u(k) \geq x(0)^*P^{-1}x(0).$$

$\triangleleft$

Proof: Pre- and post-multiply (15) by P^{-1} and use Schur complements to get

$$\begin{bmatrix} A & B \end{bmatrix}^* P^{-1} \begin{bmatrix} A & B \end{bmatrix} - \begin{bmatrix} P^{-1} & 0 \\ 0 & I \end{bmatrix} \leq 0. \tag{16}$$

Then pre- and post-multiply (16) by $\begin{bmatrix} \underline{\phi}(x) \\ u \end{bmatrix}^*$ and respectively $\begin{bmatrix} \underline{\phi}(x) \\ u \end{bmatrix}$ and apply Lemma 1. $\blacksquare$

The above two propositions are related to the observability function and the controllability function defined in [10], which generalize the observability gramian and the controllability gramian for linear systems to nonlinear systems.

Proposition 7 Let G be a system of the form (1,2,3,8) with n states, and Q, P^{-1} satisfy (14), (16) respectively such that

$$\begin{bmatrix} P^{-1} + Q & P^{-1} - Q \\ P^{-1} - Q & P^{-1} + Q \end{bmatrix}$$

is pdd. Suppose that

$$\begin{aligned} Q &= T^*\Sigma T \\ P^{-1} &= T^*\Sigma^{-1}T \end{aligned}$$

and T, Σ can be partitioned as

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & I \end{bmatrix}$$

respectively, where the dimension of the 11-block is $r \times r$, T_{12} or $T_{21} = 0$ (i.e. T is block-triangular), and Σ_{11} is positive diagonal. Let G_r be a reduced order model, of the same form (1,2,3,8) and with the same nonlinearity ϕ, but have only r states and be composed of the following data

$$\begin{aligned} A_r &= (T_{11})^{-1}[TAT^{-1}]_{11}T_{11} \\ B_r &= (T_{11})^{-1}[TB]_1 \\ C_r &= [CT^{-1}]_1T_{11} \\ D_r &= D. \end{aligned}$$

Then γ_{dd} and hence the ℓ_2-ℓ_2 induced gain of the error system $G - G_r \leq 2$. $\triangleleft$

Remark 2 The condition that

$$\begin{bmatrix} P^{-1} + Q & P^{-1} - Q \\ P^{-1} - Q & P^{-1} + Q \end{bmatrix}$$

is pdd is stronger than the condition that both Q and P^{-1} are pdd as required in Proposition 5 and 6. In fact, it is equivalent to the condition that the matrix being composed of the smaller diagonal entries and the larger off-diagonal entries of Q and P^{-1} is still pdd. $\otimes$

Proof of Proposition 7: The proof is a little bit long and therefore only an outline is given. Since T is block-triangular, the r-state system G_r is (input-output) equivalent to the n-state system $\hat{G}_r$ being composed of

$$\begin{aligned} \hat{A}_r &= T^{-1} \begin{bmatrix} [TAT^{-1}]_{11} & 0 \\ 0 & 0 \end{bmatrix} T \\ \hat{B}_r &= T^{-1} \begin{bmatrix} [TB]_1 \\ 0 \end{bmatrix} \\ \hat{C}_r &= \begin{bmatrix} [CT^{-1}]_1 & 0 \end{bmatrix} T \\ \hat{D}_r &= D. \end{aligned}$$

Note that $(G - \hat{G}_r)/2$ can be realized by

$$\begin{aligned} A_e &= \begin{bmatrix} \hat{A}_r & 0 \\ 0 & A \end{bmatrix} \\ B_e &= \begin{bmatrix} \hat{B}_r/2 \\ B/2 \end{bmatrix} \\ C_e &= \begin{bmatrix} -\hat{C}_r & C \end{bmatrix} \\ D_e &= -\hat{D}_r/2 + D/2 = 0. \end{aligned}$$

Let

$$T_{be} = \begin{bmatrix} -\Sigma^{\frac{1}{2}} & \Sigma^{\frac{1}{2}} \\ \Sigma^{-\frac{1}{2}} & \Sigma^{-\frac{1}{2}} \end{bmatrix}.$$

Then it can be shown (as in the linear case) that

$$\begin{bmatrix} T_{be} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} T_b A_e T_b^{-1} & T_b B_e \\ C_e T_b^{-1} & D_e \end{bmatrix} \begin{bmatrix} T_{be} & 0 \\ 0 & I \end{bmatrix}^{-1}$$

is contractive, where

$$T_b = \begin{bmatrix} T & 0 \\ 0 & T \end{bmatrix}.$$

Hence

$$\begin{bmatrix} T_{be}T_b & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A_e & B_e \\ C_e & D_e \end{bmatrix} \begin{bmatrix} T_{be}T_b & 0 \\ 0 & I \end{bmatrix}^{-1}$$

is contractive. Now $(T_{be}T_b)^*(T_{be}T_b)$ being pdd guarantees that Proposition 3 (with $\gamma^2 = 1$) is applicable. ∎

Clearly Proposition 7 is analogous to the balanced model reduction result for linear systems. In particular, if (A, B, C, D) is originally balanced in the linear sense, Proposition 7 just gives the same error bound for the state-truncation of our system in concern. However, if (A, B, C, D) is not originally balanced, the requirement that T should be block-triangular makes the model reduction more difficult. This requirement may be converted into a more tractable condition on $Q - P^{-1}$ by the following Lemma.

Lemma 8 If

$$Q - P^{-1} = \begin{bmatrix} Q_{11} - [P^{-1}]_{11} & 0 \\ 0 & 0 \end{bmatrix} \tag{17}$$

where Q_{11}, $[P^{-1}]_{11}$ are the top-left $r \times r$ submatrices of Q and P^{-1} respectively, or more generally there exists a diagonal Λ with rank$(\Lambda) \leq r$ such that

$$\Lambda \geq Q - P^{-1} \geq -\Lambda \tag{18}$$

then T, Σ may be partitioned as in Proposition 7 with $T_{12} = 0$. ◁

The authors have developed a model reduction algorithm based on Proposition 7 and Lemma 8 and the study of its performance can be found in [11]. In brief, not for all 22 random systems in Section 2 a satisfactory reduced-order model can be found. Only for 7 systems a reduced-order model with a relative error less than 10 percent can be obtained (6 with $r = 3$ and 1 with $r = 2$). It is possibly because the system (1,2,3,8) is nonlinear after all and the chance of doing model reduction for it is far smaller than a linear system in essence.

It should be realized that, for nonlinear systems of the form (1,2,3,8), a system being composed of (A, B, C, D, ϕ) is by no means input-output equivalent to the system being composed of $(TAT^{-1}, TB, CT^{-1}, D, \phi)$ in general. As pointed out by Albertini and Sontag, the "realizations" of such systems are in fact very limited. For linear systems we do model reduction via an input-output equivalent balanced realization. Proposition 7 points out that, although we do not have such an "intermediate" equivalence for our system in concern, we can still arrive at a reduced order model in a similar fashion.

4 Discussion

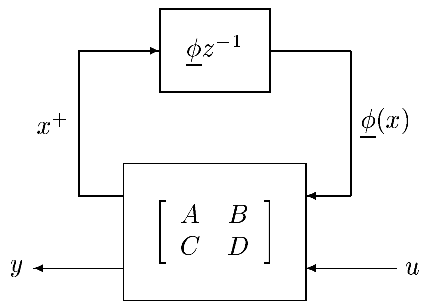
In this section we add a few remarks.

Remark 3 Our results do not depend on the exact form of ϕ but only require that it satisfies (8). Therefore it may be a little bit conservative when ϕ is exactly known but robust when ϕ is uncertain (for example, interpreted as an uncertain activation function of the neuron or state-dependent perturbations of a linear system). In fact we can allow ϕ to be time-varying (or even dynamic; see [11] for details). ⊗

Remark 4 Although γ_{dd} is typically tight for the random systems in Section 2 it may not be tight in all cases. If ϕ is exactly known one may wish to consider (9) directly instead of the sufficient condition that P is pdd. The authors have developed an ellipsoid algorithm-like algorithm to handle (9) directly but it is computationally very expensive and much less reliable. (See the appendix in [11].) ⊗

Remark 5 Our approach is not restricted to the case that all ϕ_i must be identical. When certain but not all ϕ_i are identical (for example the situation mentioned at the beginning of Section 3), we may just use a block-diagonal P, with each block pdd corresponding to each set of identical ϕ_i, to deduce similar results. If certain ϕ_i are the identity, the corresponding block can be made simply positive definite instead of pdd. ⊗

Remark 6 $\underline{\phi}$ may be regarded as a diagonal nonlinear operator on the vector x, and the system (1,2,3,8) or (1,2,4,8) may be represented by an upper Linear Fractional Transformation (LFT) with respect to a diagonal odd 1-Lipschitz nonlinear operator:



Indeed this paper has presented an approach that can expand the LFT framework for uncertain linear systems[12] into the nonlinear case. For details and more results see [11]. $\otimes$

5 Conclusions

It has been shown that solving the standard induced-norm LMIs with pdd matrices gives an effective method to determine a bound on the gain of this class of nonlinear systems. The situation in model reduction is less straight-forward but an algorithm has been developed along the lines of the "balance and truncate" result of Section 3.

References

[1] E. Kaszkurewicz and A. Bhaya. Robust stability and diagonal Liapunov functions. *SIAM Journal Matrix Anal. Appl.*, 14(2):508–520, April 1993.

[2] E. Kaszkurewicz and A. Bhaya. On a class of globally stable neural circuits. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, 41(2):171–174, February 1994.

[3] E. Kaszkurewicz and A. Bhaya. Comments on "Necessary and sufficient condition for absolute stability of neural networks". *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, 42(8):497–499, August 1995.

[4] F. Albertini and E. D. Sontag. For neural networks, function determines form. *Neural Networks*, 6:975–990, 1993.

[5] F. Albertini and E. D. Sontag. Uniqueness of weights for recurrent nets. In *Proceedings of the International Symposium Math. Theory of Networks Syst.*, volume II, pages 599–602, Regensburg, Germany, August 1993.

[6] F. Albertini and E. D. Sontag. State observability in recurrent neural networks. *Systems & Control Letters*, 22(4):235–244, 1994.

[7] F. Albertini and P. Dai Pra. Forward accessibility for recurrent neural networks. *IEEE Transactions on Automatic Control*, 40(11):1962–1968, November 1995.

[8] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, 1994.

[9] P. Gahinet, A. Nemirovskii, A. J. Laub, and M. Chilali. *The LMI Control Toolbox.* The MathWorks Inc., November 1994. Beta-Release.

[10] J. M. A. Scherpen. Balancing for nonlinear systems. *Systems & Control Letters*, 21(2):143–153, August 1993.

[11] Y.-C. Chu. *Control of Systems with Repeated Scalar Nonlinearities.* PhD thesis, submitted to University of Cambridge, 1996.

[12] C. Beck, R. D'Andrea, F. Paganini, W.-M. Lu, and J. Doyle. A state-space theory of uncertain systems. In *Proceedings of the 13th Triennial IFAC World Congress*, pages 291–296, San Francisco, California, July 1996.