96-44

NL_q Theory: Checking and Imposing Stability of Recurrent Neural Networks for Nonlinear Modeling

Johan A. K. Suykens, Joos Vandewalle, *Fellow, IEEE*, and Bart L. R. De Moor

***Abstract*—It is known that many discrete-time recurrent neural networks, such as e.g., neural state space models, multilayer Hopfield networks, and locally recurrent globally feedforward neural networks, can be represented as NL_q systems. Sufficient conditions for global asymptotic stability and input/output stability of NL_q systems are available, including three types of criteria:**

1) **diagonal scaling;**
2) **criteria depending on diagonal dominance;**
3) **condition number factors of certain matrices.**

In this paper, it is discussed how Narendra's dynamic backpropagation procedure, which is used for identifying recurrent neural networks from I/O measurements, can be modified with an NL_q stability constraint in order to ensure globally asymptotically stable identified models. An example illustrates how system identification of an internally stable model corrupted by process noise may lead to unwanted limit cycle behavior and how this problem can be avoided by adding the stability constraint.

***Index Terms*— Dynamic backpropagation, global asymptotic stability, LMI's, multilayer recurrent neural networks, NL_q systems.**

I. Introduction

Recently, NL_q theory has been introduced as a model-based neural control framework with global asymptotic stability criteria [20], [24]. It consists of recurrent neural network models and controllers in state-space form, for which the closed-loop system can be represented in so-called NL_q system form. NL_q systems are discrete-time nonlinear state-space models with q layers of alternating linear and static nonlinear operators that satisfy a sector condition. It has been shown, then, how Narendra's dynamic backpropagation, which is classically used to learn a controller track a set of specific reference inputs, can be modified with NL_q stability constraints. Furthermore, several types of nonlinear behavior, including systems with a unique equilibrium, multiple equilibria, (quasi)-periodic behavior, and chaos have been stabilized and controlled using the stability criteria [20].

In this paper, we focus on nonlinear modeling applications of NL_q theory instead of control applications. For instance, for tracking problems, where Narendra's dynamic backpropagation [9], [10] has been modified with a closed-loop stability constraint [20], we will modify dynamic backpropagation for system identification with a stability constraint in order to obtain identified recurrent neural networks that are globally asymptotically stable. In addition, for linear filters (e.g., IIR filters [16], [17]), this has been an important issue. We will consider the class of discrete-time recurrent neural networks, which is representable as NL_q systems. Examples are, e.g., neural state-space models and locally recurrent globally feedforward neural networks that are models consisting of global and local feedback, respectively.

In order to check stability of identified models, sufficient conditions for global asymptotic stability of NL_qs are applied [20]. A first condition is called diagonal scaling, which is closely related to diagonal scaling criteria in robust control theory [3], [13]. Checking stability can be formulated then as an linear matrix inequality (LMI) problem, which corresponds to solving a convex optimization problem. A second condition is based on diagonal dominance of certain matrices. Certain results of digital filters with overflow characteristic [8] can be considered as a special case for $q = 1$ (one layer NL_q).

Finally, we demonstrate how global asymptotic stability can be imposed on the identified models. This is done by modifying dynamic backpropagation with stability constraints. Besides the diagonal scaling condition, criteria based on condition numbers of certain matrices [20] are proposed for this purpose. In many applications, one has indeed the *a priori* knowledge that the true system is globally asymptotically stable or one is interested in a stable approximator. It is illustrated with an example that process noise can indeed cause identified models that show limit cycle behavior, instead of global asymptotic stability, as for the true system. We show how this problem can be avoided by applying the modified dynamic backpropagation algorithm.

This paper is organized as follows. In Section II, we present two examples of discrete time recurrent neural networks that are representable as NL_q systems: neural state space models and locally recurrent globally feedforward neural networks. In Section III, we review the classical dynamic backpropagation paradigm. In Section IV, we discuss sufficient conditions for

Manuscript received June 5, 1997. This work was carried out at the ESAT Laboratory and the Interdisciplinary Center of Neural Networks ICNN of the Katholieke Universiteit Leuven in the framework of the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology, and Culture (IUAP-17, IUAP-50) and in the framework of a Concerted Action Project Modelbased Information Processing Systems (MIPS) of the Flemish Community, Identification and Control of Complex Systems (ICCoS), Human Capital and Mobility Network (SIMONET), and SCIENCE-ERNSI (SC1-CT92-0779).
The authors are with the Department of Electrical Engineering, Katholieke Universiteit Leuven, ESAT-SISTA, Leuven, Belgium (e-mail: johan.suykens@esat.kuleuven.ac.be; joos.vandewalle@esat.kuleuven.ac.be; bart.demoor@esat.kuleuven.ac.be).
Publisher Item Identifier S 1053-587X(97)08056-2.

global asymptotic stability of identified models. In Section V, Narendra's dynamic backpropagation is modified with NL_q stability constraints. In Section VI, an example is given for a system corrupted by process noise.

II. NL_q Systems

The following discrete time nonlinear state space model is called an NL_q system [20]

$$\begin{cases} p_{k+1} = \Gamma_1\{V_1\,\Gamma_2[V_2 \cdots \Gamma_q(V_q p_k + B_q w_k) \cdots \\ \qquad + B_2 w_k] + B_1 w_k\} \\ e_k = \Lambda_1\{W_1\,\Lambda_2[W_2 \cdots \Lambda_q(W_q p_k + D_q w_k) \cdots \\ \qquad + D_2 w_k] + D_1 w_k\} \end{cases} \tag{1}$$

with state vector $p_k \in \mathbb{R}^{n_p}$, input vector $w_k \in \mathbb{R}^{n_w}$, and output vector $e_k \in \mathbb{R}^{n_e}$. The matrices V_i, B_i, W_i, and D_i ($i = 1, \cdots, q$) are constant with compatible dimensions, and the matrices $\Gamma_i = \operatorname{diag}\{\gamma_i\}$ and $\Lambda_i = \operatorname{diag}\{\lambda_i\}$ ($i = 1, \cdots, q$) are diagonal with diagonal elements $\gamma_i(p_k, w_k)$ and $\lambda_i(p_k, w_k) \in [0, 1]$ for all p_k, w_k. The term "NL_q" stands for the alternating sequence of linear and nonlinear operations in the q-layered system description.

In this section, we first explain the link between NL_qs and multilayer Hopfield networks and then discuss two examples: neural state space models and locally recurrent globally feedforward networks, which are models with global feedback and local feedback, respectively. Other examples of NL_qs are, e.g., (generalized) cellular neural networks, the discrete time Lur'e problem, and linear fractional transformations with real diagonal uncertainty block [20], [23].

A. Multilayer Hopfield Networks

NL_q systems are closely related to multilayer recurrent neural networks of the form

$$\begin{cases} p_{k+1} = \sigma_1\{V_1\,\sigma_2[V_2 \cdots \sigma_q(V_q p_k + B_q w_k) \cdots \\ \qquad + B_2 w_k] + B_1 w_k\} \\ e_k = \eta_1\{W_1\,\eta_2[W_2 \cdots \eta_q(W_q p_k + D_q w_k) \cdots \\ \qquad + D_2 w_k] + D_1 w_k\} \end{cases} \tag{2}$$

with $\sigma_i(.)$, $\eta_i(.)$ vector-valued nonlinear functions that belong to sector [0, 1] [28].

Let us illustrate the link between (1) and (2) for the autonomous NL_q system (zero external input)

$$p_{k+1} = \left[\prod_{i=1}^{q} \Gamma_i(p_k) V_i\right] p_k$$

by means of the following autonomous Hopfield network with synchronous updating

$$x_{k+1} = \tanh(W x_k). \tag{3}$$

This can be written as

$$x_{k+1} = \Gamma(x_k) W x_k \tag{4}$$

with $\Gamma = \operatorname{diag}\{\gamma_i\}$ and $\gamma_i = \tanh(w_i^T x_k)/(w_i^T x_k)$, which follows from the element-wise notation

$$\begin{aligned} x^i &:= \tanh\left(\sum_j w_j^i x^j\right) \\ &:= \frac{\tanh\left(\sum_j w_j^i x^j\right)}{\sum_j w_j^i x^j} \cdot \sum_j w_j^i x^j \\ &:= \gamma_i^i \sum_j w_j^i x^j. \end{aligned} \tag{5}$$

The time index is omitted here because of the assignment operator ":=." The notation γ_i^i means that this corresponds to the diagonal matrix $\Gamma(x_k)$. In case $w_i^T x_k = 0$, de l' Hospital's rule can be applied, or a Taylor expansion of $\tanh(.)$ can be taken, leading to $\gamma_i = 1$.

In a similar way the multilayer Hopfield neural network

$$x_{k+1} = \tanh[V \tanh(W x_k)] \tag{6}$$

can be written as

$$x_{k+1} = \Gamma_1(x_k) V \Gamma_2(x_k) W x_k \tag{7}$$

because

$$\begin{aligned} x^i &:= \tanh\left[\sum_j v_j^i \tanh\left(\sum_l w_l^j x^l\right)\right] \\ &:= \tanh\left(\sum_j v_j^i \gamma_{2_j^j} \sum_l w_l^j x^l\right) \\ &:= \gamma_{1_i^i} \sum_j v_j^i \gamma_{2_j^j} \sum_l w_l^j x^l. \end{aligned} \tag{8}$$

B. Neural State Space Models

Neural state space models (Fig. 1) for nonlinear system identification have been introduced in [19] and are of the form

$$\begin{cases} \hat{x}_{k+1} = W_{AB} \tanh(V_A \hat{x}_k + V_B u_k + \beta_{AB}) + K\epsilon_k \\ y_k = W_{CD} \tanh(V_C \hat{x}_k + V_D u_k + \beta_{CD}) + \epsilon_k \end{cases} \tag{9}$$

with estimated state vector $\hat{x}_k \in \mathbb{R}^n$, input vector $u_k \in \mathbb{R}^m$, output vector $y_k \in \mathbb{R}^l$, and zero mean white Gaussian noise input ϵ_k (corresponding to the prediction error $y_k - \hat{y}_k$). W_* and V_* are the interconnection matrices with compatible dimensions, β_* are bias vectors, and K is a steady Kalman gain. If the multilayer perceptrons in the state equation and output equation are replaced by linear mappings, the neural state-space model corresponds to a Kalman filter. Defining $p_k = x_k$, $w_k = [u_k;\ \epsilon_k;\ 1]$, and $e_k = y_k$ in (1), the neural state space model is an NL_2 system with $\Gamma_1 = I$, $V_1 = W_{AB}$, $V_2 = V_A$, $B_2 = [V_B\ 0\ \beta_{AB}]$, $B_1 = [0\ K\ 0]$, $\Lambda_1 = I$, $W_1 = W_{CD}$, $W_2 = V_C$, $D_2 = [V_D\ 0\ \beta_{CD}]$, and$D_1 = [0\ I\ 0]$. For the autonomous case and zero bias terms, the neural state-space model becomes

$$\hat{x}_{k+1} = W_{AB} \tanh(V_A \hat{x}_k). \tag{10}$$

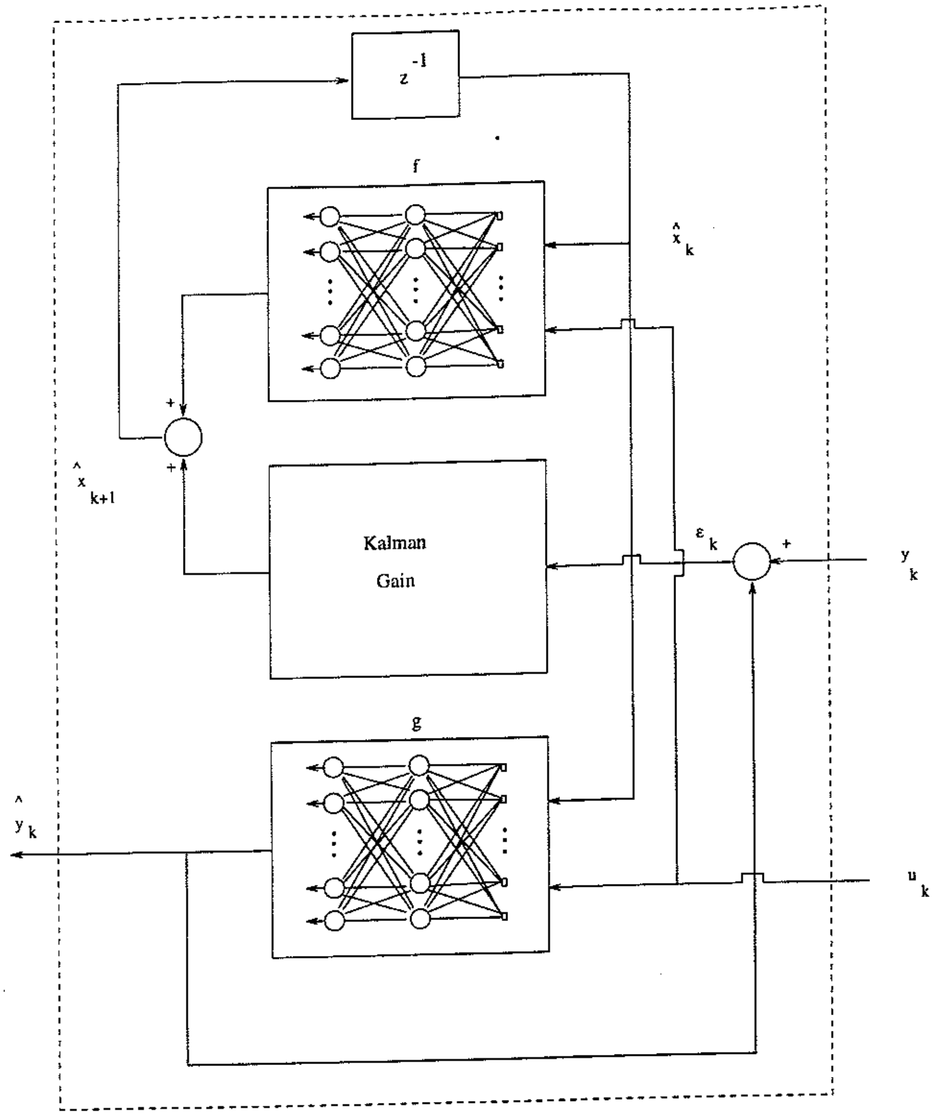


Fig. 1. Neural state space model, which is a discrete time recurrent neural network with multilayer perceptrons for the state and output equation and a Kalman gain for taking into account process noise.

By introducing a new state variable $\xi_k = \tanh(V_A \hat{x}_k)$, this can be written as the NL_1 system

$$\begin{cases} \hat{x}_{k+1} = W_{AB}\, \xi_k \\ \xi_{k+1} = \tanh(V_A W_{AB}\, \xi_k). \end{cases} \tag{11}$$

C. Locally Recurrent Globally Feedforward Networks

In [26], the locally recurrent globally feedforward (LRGF) network has been proposed, which aims at unifying several existing recurrent neural network models. Starting from the McCulloch–Pitts model, many other architectures have been proposed in the past with local synapse feedback, local activation feedback and local output feedback, time-delayed neural networks, etc., e.g., by Frasconi–Gori–Soda, De Vries–Principe, and Poddar–Unnikrishnan. The architecture of Tsoi and Back (Fig. 2) includes most of the latter architectures.

Assuming in Fig. 2 that the transfer functions $G_i(z)$ $(i = 1, \cdots, n)$ (which may have both poles and zeros) have a state-space realization (A_i, B_i, C_i), an LRGF network can be described in state-space form as

$$\begin{cases} \xi_{k+1}^{(i)} = A^{(i)} \xi_k^{(i)} + B^{(i)} u_k^{(i)}, \qquad i = 1, \cdots, n-1 \\ z_k^{(i)} = C^{(i)} \xi_k^{(i)} \\ \xi_{k+1}^{(n)} = A^{(n)} \xi_k^{(n)} + B^{(n)} f\left[\sum_{j=1}^{n} z_k^{(j)}\right] \\ z_k^{(n)} = C^{(n)} \xi_k^{(n)} \\ y_k = f\left[\sum_{j=1}^{n} z_k^{(j)}\right]. \end{cases} \tag{12}$$

Here, $u_k^{(i)}, z_k^{(i)} \in \mathbb{R}$ $(i = 1, \cdots, n-1)$ are the inputs and local outputs of the network, $y_k \in \mathbb{R}$ is the output of the network, and $z_k^{(n)} \in \mathbb{R}$ the filtered output of the network. $f(.)$ is a static nonlinearity belonging to sector [0, 1]. Applying the state augmentation $\eta_k = f[\sum_{j=1}^{n} z_k^{(j)}]$, an NL_1 system is obtained with state vector $p_k = [\xi_k^{(1)}; \xi_k^{(2)}; \cdots; \xi_k^{(n-1)}; \xi_k^{(n)}; \eta_k]$, input

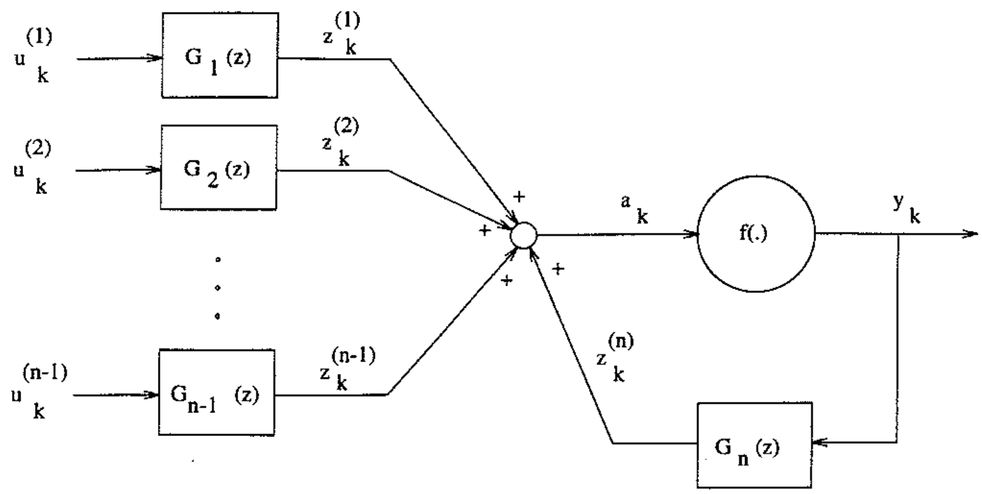


Fig. 2. Locally recurrent globally feedforward network of Tsoi and Back, consisting of linear filters $G_i(z)(i = 1, \cdots, n-1)$ for local synapse feedback and $G_n(z)$ for local output feedback.

vector $w_k = [u_k^{(1)}; \cdots; u_k^{(n-1)}]$, and matrices in (12a), shown at the bottom of the page.

III. Classical Dynamic Backpropagation

Dynamic backpropagation, according to Narendra and Parthasarathy [9], [10], is a well-known method for training recurrent neural networks. In this section, we shortly review this method for models in state-space form because this fits into the framework of NL_q representations.

Let us consider discrete-time recurrent neural networks that can be written in the form

$$\begin{cases} \hat{x}_{k+1} = \Phi(\hat{x}_k, u_k, \epsilon_k; \alpha); \; \hat{x}_0 = x_0 \text{ given} \\ \hat{y}_k = \Psi(\hat{x}_k, u_k; \beta) \end{cases} \tag{13}$$

where $\Phi(.)$, $\Psi(.)$ are twice continuously differentiable nonlinear mappings, and the weights α, β are elements of the parameter vector θ to be identified from a number of N input/output data $Z^N = \{u_k, y_k\}_{k=1}^{k=N}$

$$\min_{\theta} J_N(\theta, Z^N) = \frac{1}{N} \sum_{k=1}^{N} l[\epsilon_k(\theta)]. \tag{14}$$

A typical choice for $l(\epsilon_k)$ in this prediction error algorithm is $\frac{1}{2} \epsilon_k^T \epsilon_k$ with prediction error $\epsilon_k = y_k - \hat{y}_k$. For gradient-based optimization algorithms, one computes the gradient

$$\frac{\partial J_N}{\partial \theta} = \frac{1}{N} \sum_{k=1}^{N} \epsilon_k^T \left(-\frac{\partial \hat{y}_k}{\partial \theta} \right). \tag{15}$$

Dynamic backpropagation [9], [10] then makes use of a sensitivity model

$$\begin{cases} \dfrac{\partial \hat{x}_{k+1}}{\partial \alpha} = \dfrac{\partial \Phi}{\partial \hat{x}_k} \cdot \dfrac{\partial \hat{x}_k}{\partial \alpha} + \dfrac{\partial \Phi}{\partial \alpha} \\ \dfrac{\partial \hat{y}_k}{\partial \alpha} = \dfrac{\partial \Psi}{\partial \hat{x}_k} \cdot \dfrac{\partial \hat{x}_k}{\partial \alpha} \\ \dfrac{\partial \hat{y}_k}{\partial \beta} = \dfrac{\partial \Psi}{\partial \beta} \end{cases} \tag{16}$$

in order to generate the gradient of the cost function. The sensitivity model is a dynamical system with state vector $\partial \hat{x}_k / \partial \alpha \in \mathbb{R}^n$ driven by the input vector consisting of $\partial \Phi / \partial \alpha \in \mathbb{R}^n$, $\partial \Psi / \partial \beta \in \mathbb{R}^l$, and at the output, $\partial \hat{y}_k / \partial \alpha \in \mathbb{R}^l$, $\partial \hat{y}_k / \partial \beta \in \mathbb{R}^l$ are generated. The Jacobians $\partial \Phi / \partial \hat{x}_k \in \mathbb{R}^{n \times n}$ and $\partial \Psi / \partial \hat{x}_k \in \mathbb{R}^{l \times n}$ are evaluated around the nominal trajectory.

$$V_1 = \begin{bmatrix} A^{(1)} & & & & 0 & 0 \\ & A^{(2)} & & & & \\ & & \ddots & & & \vdots \\ & & & A^{(n-1)} & 0 & 0 \\ 0 & 0 & \cdots & 0 & A^{(n)} & B^{(n)} \\ C^{(1)}A^{(1)} & C^{(2)}A^{(2)} & \cdots & C^{(n-1)}A^{(n-1)} & C^{(n)}A^{(n)} & C^{(n)}B^{(n)} \end{bmatrix}$$

$$B = \begin{bmatrix} B^{(1)} & & & \\ & B^{(2)} & & \\ & & \ddots & \\ & & & B^{(n-1)} \\ 0 & 0 & \cdots & 0 \\ C^{(1)}B^{(1)} & C^{(2)}B^{(2)} & \cdots & C^{(n-1)}B^{(n-1)} \end{bmatrix}. \tag{12a}$$

Examples and applications of dynamic backpropagation applied to neural state-space models, are discussed in [19]–[21]. For aspects of model validation, regularization, and pruning of neural network models, see, e.g., [1], [15], and [18].

IV. Checking Global Asymptotic Stability of Identified Models

In many applications, recurrent neural networks have been used in order to model systems with a unique or multiple equilibria, (quasi)-periodic behavior or chaos. On the other hand, one is often interested in obtaining models that are globally asymptotically stable, e.g., in case one has this *a priori* knowledge about the true system or one is interested in such an approximator. In this section, we present two criteria that are sufficient in order to check global asymptotic stability of the identified model, which is represented as NL_q system.

Theorem 1—Diagonal Scaling [20]: Consider the autonomous NL_q system

$$p_{k+1} = \left[\prod_{i=1}^{q} \Gamma_i(p_k) V_i\right] p_k \tag{17}$$

and let

$$V_{\text{tot}} = \begin{bmatrix} 0 & V_2 & & & 0 \\ & 0 & V_3 & & \\ & & \ddots & & \\ & & & 0 & V_q \\ V_1 & & & & 0 \end{bmatrix}$$

$$V_i \in \mathbb{R}^{n_{h_i} \times n_{h_{i+1}}},\ n_{h_1} = n_{h_{q+1}} = n_p.$$

A sufficient condition for global asymptotic stability of (17) is to find a diagonal matrix D_{tot} such that

$$\|D_{\text{tot}} V_{\text{tot}} D_{\text{tot}}^{-1}\|_2^q = \beta_D < 1 \tag{18}$$

where $D_{\text{tot}} = \text{diag}\{D_2, D_3, \cdots, D_q, D_1\}$, and $D_i \in \mathbb{R}^{n_{h_i} \times n_{h_i}}$ are diagonal matrices with nonzero diagonal elements. □

Proof: See [20].

The condition is based on the Lyapunov function $V(p) = \|D_1 p\|_2$, which is radially unbounded in order to prove that the origin is a unique equilibrium point. Finding such a diagonal matrix D_{tot} for a given matrix V_{tot} can be formulated as the linear matrix inequality (LMI) in D_{tot}^2

$$V_{\text{tot}}^T D_{\text{tot}}^2 V_{\text{tot}} < D_{\text{tot}}^2. \tag{19}$$

It is well known that this corresponds to solving a convex optimization problem [3], [12], [27]. Similar criteria are known in the field of control theory as "diagonal scaling" [3], [7], [13]. However, it is also known that such criteria can be conservative.

Sharper criteria are obtained by considering a Lyapunov function $V(p) = \|P_1 p\|_2$ with a nondiagonal matrix P_1 instead of D_1. The next theorem is expressed then in terms of diagonal dominance of certain matrices. Let us recall that a matrix $Q \in \mathbb{R}^{n \times n}$ is called diagonally dominant of level $\delta_Q \geq 1$ if

$$q_{ii} > \delta_Q \sum_{j=1 (j \neq i)}^{n} |q_{ij}|, \qquad \forall i = 1, \cdots, n. \tag{20}$$

holds [20]. The following Theorem then holds.

Theorem 2—Diagonal Dominance [20]: A sufficient condition for global asymptotic stability of the autonomous NL_q system (17) is to find matrices P_i, N_i such that

$$\prod_{i=1}^{q} \left(\frac{\delta_{Q_i}}{\delta_{Q_i} - 1}\right)^{1/2} \|P_{\text{tot}} V_{\text{tot}} P_{\text{tot}}^{-1}\|_2^q < 1 \tag{21}$$

with $P_{\text{tot}} = \text{blockdiag}\{P_2, P_3, \cdots, P_q, P_1\}$, and $P_i \in \mathbb{R}^{n_{h_i} \times n_{h_i}}$ full rank matrices. The matrices $Q_i = P_i^T P_i N_i$ are diagonally dominant with $\delta_{Q_i} > 1$, and N_i are diagonal matrices with positive diagonal elements. □

Proof: See [20].

In order to check stability of an identified model, i.e., for a given matrix V_{tot}, one might formulate an optimization problem in P_i and N_i such that (21) is satisfied. LMI conditions that correspond to (21) are derived in [20].

For the case of neural networks with a sat(.) activation function (linear characteristic with saturation), Theorem 2 can be formulated in a sharper way [20]. In that case, it is sufficient to find matrices P_i, N_i with

$$\|P_{\text{tot}} V_{\text{tot}} P_{\text{tot}}^{-1}\|_2 < 1 \qquad \text{such that } \delta_{Q_i} = 1 \quad (i = 1, \cdots, q). \tag{22}$$

The latter follows from a result on stability of digital filters with overflow characteristic by Liu and Michel [8], which corresponds to the NL_1 system

$$x_{k+1} = \text{sat}(V x_k). \tag{23}$$

A sufficient condition for global asymptotic stability is then to find a matrix $Q = P^T P$ with

$$\|P V P^{-1}\|_2 < 1 \quad \text{such that} \quad \delta_Q = 1. \tag{24}$$

We also remark that for a linear system

$$x_{k+1} = A x_k \tag{25}$$

one obtains the condition

$$\|P A P^{-1}\|_2 < 1 \tag{26}$$

from the Lyapunov function $V(x) = \|Px\|_2$, where P is a full-rank matrix. The spectral radius of A corresponds to

$$\rho(A) = \min_{P \in \mathbb{R}^{n \times n}} \|P A P^{-1}\|_2. \tag{27}$$

V. Modified Dynamic Backpropagation: Imposing Stability

In this section, we discuss modified dynamic backpropagation algorithms by adding an NL_q stability constraint to the cost function (14). In this way, it will be possible to identify recurrent neural network models that are guaranteed to be globally asymptotically stable.

Based on the condition of Theorem 1, one may solve the constrained nonlinear optimization problem

$$\min_{\theta, D_{\text{tot}}} J_N(\theta, Z^N) = \frac{1}{N} \sum_{k=1}^{N} l[\epsilon_k(\theta)] \qquad \text{such that} \quad V_{\text{tot}}(\theta)^T D_{\text{tot}}^2 V_{\text{tot}}(\theta) < D_{\text{tot}}^2. \tag{28}$$

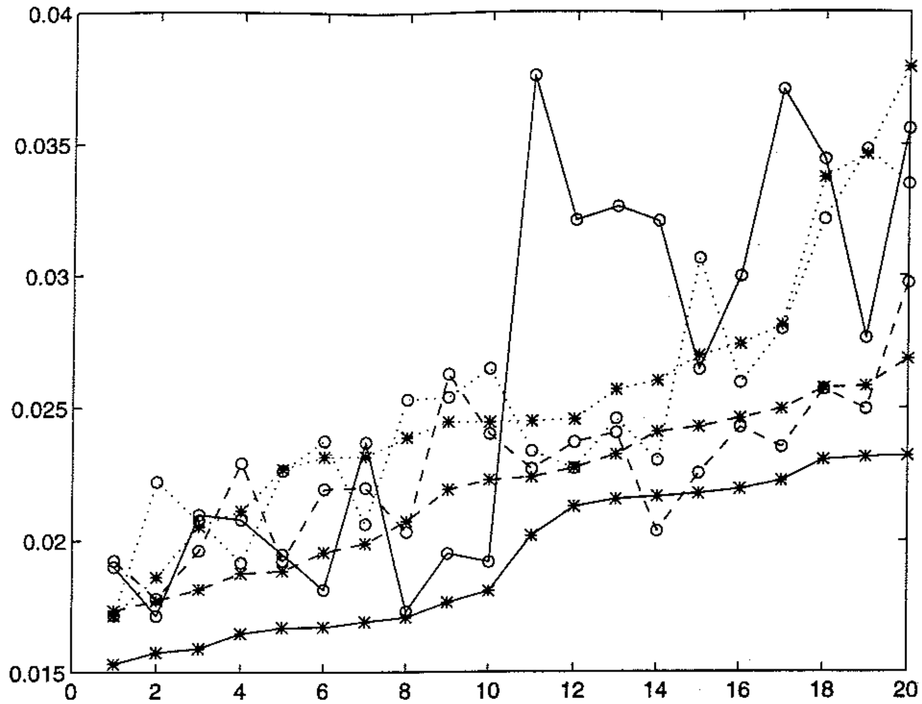


Fig. 3. Comparison between dynamic backpropagation (–), modified dynamic backpropagation with diagonal scaling (– –) and condition number constraint (···) for the error on the training data (*) and test data (o). The errors are plotted with respect to the experiment number. Twenty random initial parameter vectors were chosen. The experiments were sorted with respect to the error on the training set.

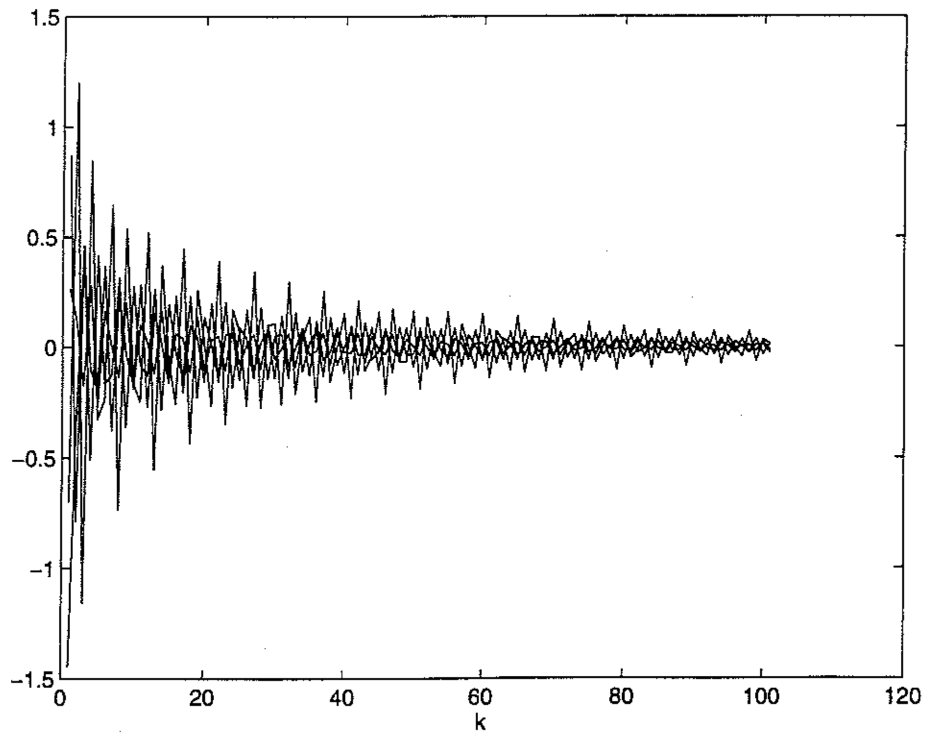


Fig. 4. Behavior of the true system to be identified. State variables of the autonomous system for some randomly chosen initial state and zero noise are shown.

The cost function is differentiable, but the constraint becomes nondifferentiable when the two largest eigenvalues of the matrix $V_{\text{tot}}(\theta)^T D_{\text{tot}}^2 V_{\text{tot}}(\theta) - D_{\text{tot}}^2$ coincide [14]. Convergent algorithms for such nonconvex nondifferentiable optimization problems have been described, e.g., by Polak and Wardi [14]. The gradient-based optimization method makes use of the concept of a generalized gradient [2] for the constraint. An alternative formulation of the problem is

$$\min_{\theta} J_N(\theta, Z^N) = \frac{1}{N} \sum_{k=1}^{N} l[\epsilon_k(\theta)] \quad \text{such that} \quad \min_{D_{\text{tot}}} \|D_{\text{tot}} V_{\text{tot}}(\theta) D_{\text{tot}}^{-1}\|_2 < 1. \tag{29}$$

The evaluation of the constraint corresponds then to solving a convex optimization problem.

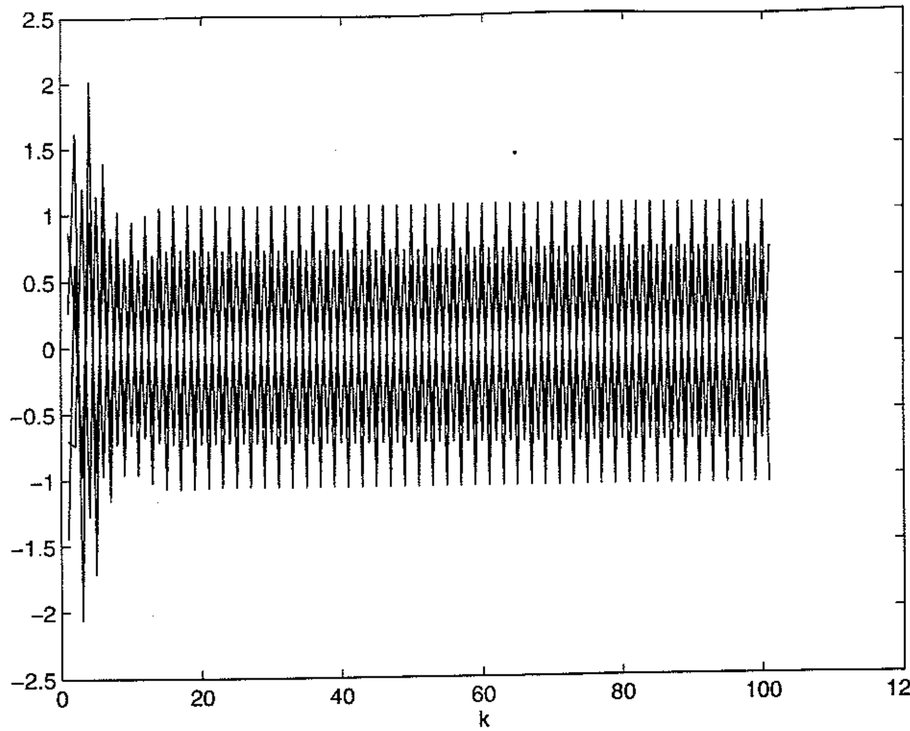


Fig. 5. Unwanted limit cycle behavior of the identified model after applying classical dynamic backpropagation. The I/O data were generated by a globally asymptotically stable system corrupted by process noise. State variables of the identified model for the autonomous case are shown.

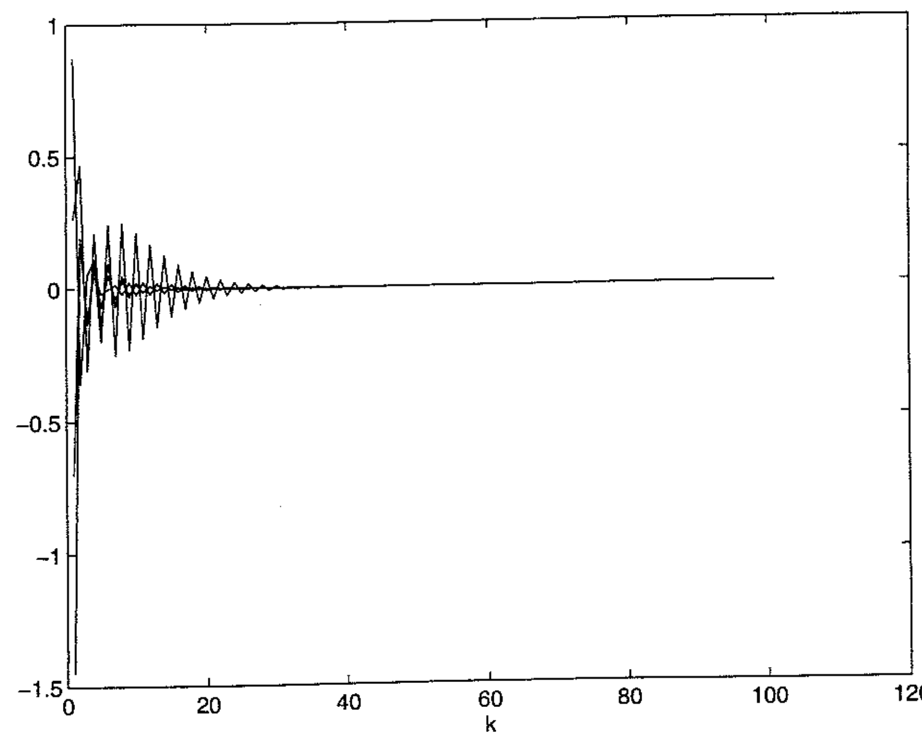


Fig. 6. Model identified by applying modified dynamic backpropagation with diagonal scaling constraint. The model is guaranteed to be globally asymptotically stable as is illustrated for some randomly chosen initial state on this figure. State variables of the identified model for the autonomous case are shown.

Though LMI conditions have been formulated for the condition of diagonal dominance of Theorem 2, the use of these LMI conditions is rather impractical for a modified dynamic backpropagation algorithm. Therefore, we will make use of another theorem in order to impose stability on the identified models.

Theorem 3—Condition Number Factor [20]: A sufficient condition for global asymptotic stability of the autonomous NL_q (17) is to find matrices P_i such that

$$\prod_{i=1}^{q} \kappa(P_i)\, \|P_{\mathrm{tot}} V_{\mathrm{tot}} P_{\mathrm{tot}}^{-1}\|_2^q < 1 \tag{30}$$

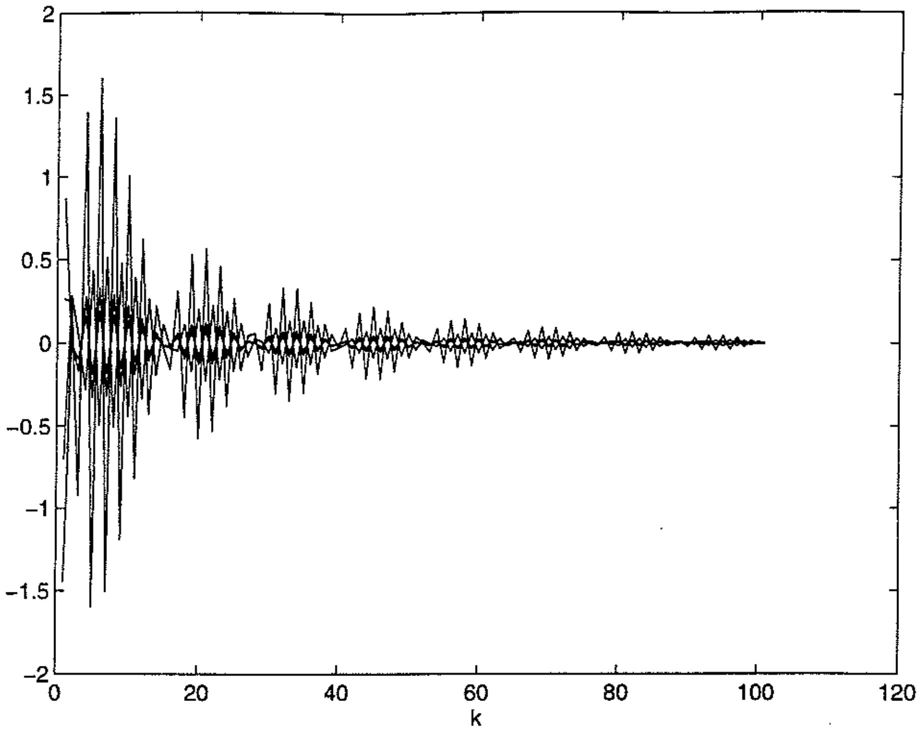


Fig. 7. Model identified by applying modified dynamic backpropagation with condition number constraint. Local stability at the origin is imposed. Its region of attraction is determined by the condition number. State variables of the identified model for the autonomous case are shown.

where $P_{\text{tot}} = \text{blockdiag}\{P_2, P_3, \cdots, P_q, P_1\}$, and $P_i \in \mathbb{R}^{n_{h_i} \times n_{h_i}}$ are full rank matrices. The condition numbers $\kappa(P_i)$ are, by definition, equal to $\|P_i\|_2 \|P_i^{-1}\|_2$.

Proof: See [20]. □

In practice, this theorem is used as

$$\min_{P_{\text{tot}}} \max_i \{\kappa(P_i)\} \qquad \text{such that } \|P_{\text{tot}} V_{\text{tot}} P_{\text{tot}}^{-1}\|_2 < 1. \tag{31}$$

The constraint in (31) imposes local stability at the origin. The basin of attraction of the origin is enlarged by minimizing the condition numbers. Even when condition (30) is not satisfied, the basin of attraction can often be made very large (or probably infinitely large as simulation results suggest). This principle has been demonstrated on stabilizing and controlling systems with one or multiple equilibria, periodic and quasiperiodic behavior, and chaos [20].

Based on (31), dynamic backpropagation is modified then as

$$\min_{\theta, Q_{\text{tot}}} J_N(\theta, Z^N) = \frac{1}{N} \sum_{k=1}^{N} l[\epsilon_k(\theta)] \qquad \text{such that}$$
$$\begin{cases} V_{\text{tot}}(\theta)^T Q_{\text{tot}} V_{\text{tot}}(\theta) < Q_{\text{tot}} \\ I < Q_i < \alpha_i^2 I. \end{cases} \tag{32}$$

The latter LMI corresponds to $\kappa(P_i) < \alpha_i$ and $Q_{\text{tot}} = P_{\text{tot}}^T P_{\text{tot}}$, $Q_i = P_i^T P_i$. An alternative formulation to (32) is

$$\min_{\theta} J_N(\theta, Z^N) = \frac{1}{N} \sum_{k=1}^{N} l[\epsilon_k(\theta)] \qquad \text{such that}$$
$$\min_{P_{\text{tot}}} \|P_{\text{tot}} V_{\text{tot}}(\theta) P_{\text{tot}}^{-1}\|_2 < 1 \text{ with } \max\{\kappa(P_i)\} < c \tag{33}$$

with c a user-defined upper bound on the condition numbers. For (29)–(33), the difference in computational complexity between classical and modified dynamic backpropagation is in the LMI constraint that has to be solved at each iteration step and is $\mathcal{O}(m^4 L^{1.5})$ for L inequalities of size m [27]. One can avoid solving LMI's at each iteration step at the expense of introducing additional unknown parameters to the optimization problem as Q_{tot} in (32). Finally, for the case of NL_1 systems, the problem can be formulated by solving a Lyapunov equation

$$\min_{\theta, 0<\gamma<1} J_N(\theta, Z^N) = \frac{1}{N} \sum_{k=1}^{N} l[\epsilon_k(\theta)] \qquad \text{such that}$$
$$\begin{cases} V_{\text{tot}}(\theta)^T Q V_{\text{tot}}(\theta) + I = \gamma^2 Q \\ \min_{\delta \geq 0} \delta \text{ s.t. } R = Q + \delta I > 0 \\ \kappa(P) < c, \, R = P^T P. \end{cases} \tag{34}$$

VI. Example: A System Corrupted by Process Noise

A. Problem Statement

In this example, we consider system identification of the following neural state space model as true system.

$$\begin{cases} x_{k+1} = W_{AB} \tanh(V_A x_k + V_B u_k) + \varphi_k \\ y_k = W_{CD} \tanh(V_C x_k + V_D u_k) \end{cases}. \tag{35}$$

It is corrupted by zero mean white Gaussian process noise φ_k.

I/O data were generated by taking as

$$W_{AB} = \begin{bmatrix} 0.4157 & -0.2006 & 0.1260 & -0.0237 \\ 1.1271 & -0.0401 & -0.6084 & 0.4073 \\ -0.2141 & 0.4840 & -0.2966 & -0.0027 \\ -0.1986 & -0.6325 & 0.4208 & -1.0233 \end{bmatrix}$$

$$V_A = \begin{bmatrix} -0.3152 & -0.8392 & -0.2323 & -0.5119 \\ -0.2872 & -0.7385 & -0.4354 & 0.4126 \\ -0.1009 & -0.6593 & -0.5717 & -0.8109 \\ -0.8550 & -0.7005 & -0.0671 & 0.0592 \end{bmatrix}$$

$$V_B = \begin{bmatrix} 0.1256 \\ 1.2334 \\ 1.0599 \\ -1.7554 \end{bmatrix}$$

$$W_{CD} = [-0.5546 \quad -0.2603 \quad 1.3030 \quad -1.3587]$$

$$V_C = \begin{bmatrix} 0.3600 & 0.5972 & 1.7870 & -1.4743 \\ 1.3145 & -0.2945 & 0.0347 & 0.2681 \\ -1.2125 & -0.9360 & 0.3255 & 1.7658 \\ 0.5938 & -0.2655 & 0.5651 & -1.7682 \end{bmatrix}$$

$$V_D = \begin{bmatrix} 1.6275 \\ -2.3663 \\ -0.8700 \\ -0.7000 \end{bmatrix}$$

with zero initial state. The input u_k is zero mean white Gaussian noise with standard deviation 5. The standard deviation of the noise φ_k was chosen equal to 0.1. A set of 1000 data points was generated in this way, with the first 500 data for the training set and the next 500 data for the test set. Some properties of the autonomous system are $\rho(W_{AB}V_A) = 0.98 < 1$ and $\min_{D_{\text{tot}}} \|D_{\text{tot}} V_{\text{tot}} D_{\text{tot}}^{-1}\|_2 = 1.66 > 1$, which means that the origin is locally stable, but global asymptotic stability is not proven by the diagonal scaling condition. However, simulation of the autonomous system for several initial conditions suggests that the system is globally asymptotically stable (Fig. 4).

Because the state space representation of neural state space representation is only unique up to a similarity transformation and sign reversal of the hidden nodes [20], we are interested in identifying the true system, not in the sense of finding back the original matrices, but in order to obtain the same qualitative behavior for the autonomous system of the identified model as for the true system.

B. Application of Classical versus Modified Dynamic Backpropagation

System identification by using classical dynamic backpropagation and starting from randomly chosen parameter vectors θ [deterministic neural state space model with the same number of hidden neurons as (35)] yields identified models with limit cycle behavior for the autonomous system (Fig. 5). This effect is due to the process noise φ_k. No problems of this kind were met for the system (35) in the purely deterministic case or in the case of observation noise. In order to impose global asymptotic stability on the identified models, a modified dynamic backpropagation procedure was applied with diagonal scaling (28) and with condition number constraint (34). The autonomous system of (35) can be represented as the NL_1 system (11). Fig. 3 shows a comparison between classical and modified dynamic backpropagation for the error on the training set and test set, starting from 20 random parameter vectors θ for the three methods. Besides the fact that the identified models appear to be globally asymptotically stable for modified dynamic backpropagation (Figs. 6 and 7), the performance on the test data is often better than with respect to classical dynamic backpropagation. One might argue that by taking a Kalman gain into the predictor, the limit cycle problem might be avoided as well. However, the deterministic part is often identified first, and second, the Kalman gain is identified while keeping the other parameters constant. Moreover, the stability constraint can also be taken into account for stochastic models.

C. Software

In the experiments, a quasi-Newton optimization method with BFGS updating of the Hessian [4], [6] was used for classical dynamic backpropagation (function *fminu* in Matlab [25]). For modified dynamic backpropagation, sequential quadratic programming [4], [6] was used (function *constr* in Matlab). Numerical calculation of the gradients was done. In the experiments, 70 iteration steps were taken for the optimization (Fig. 3). For the case of diagonal scaling the Matlab function *psv* was used in order to calculate (28). Other software for solving LMI problems is, e.g., LMI lab [5]. For the case of condition numbers, $c = 100$ was chosen as upper bound in (34).

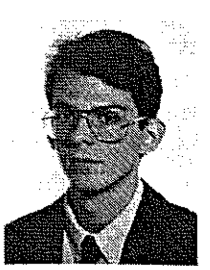
VII. Conclusion

In this paper, we discussed checking and imposing stability of discrete time recurrent neural network for nonlinear modeling applications. The class of recurrent neural networks that is representable as NL_q systems has been considered. Sufficient conditions for global asymptotic stability on diagonal scaling and diagonal dominance have been proposed for checking stability. Dynamic backpropagation has been modified with NL_q stability constraints in order to obtain models that are globally asymptotically stable. Therefore, criteria of diagonal scaling and condition number factors have been used. It has been illustrated on an example of how a globally asymptotically stable system, corrupted by process noise, may lead to unwanted limit cycle behavior for the autonomous behavior of the identified model if one applies classical dynamic backpropagation. The modified dynamic backpropagation algorithm overcomes this problem.

References

[1] S. Billings, H. Jamaluddin, and S. Chen, "Properties of neural networks with applications to modeling nonlinear dynamical systems," *Int. J. Contr.*, vol. 55, no. 1, pp. 193–224, 1992.
[2] S. Boyd and C. Barratt, *Linear Controller Design, Limits of Performance.* Englewood Cliffs, NJ: Prentice-Hall, 1991.
[3] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, "Linear matrix inequalities in system and control theory," *SIAM*, vol. 15, 1994.
[4] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York: Wiley, 1987.
[5] P. Gahinet and A. Nemirovskii, "General-purpose LMI solvers with benchmarks," in *Proc. Conf. Decision Contr.*, 1993, pp. 3162–3165.

[6] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization.* London, U.K.: Academic, 1981.
[7] E. Kaszkurewicz and A. Bhaya, "Robust stability and diagonal Liapunov functions," *SIAM J. Matrix Anal. Applicat.*, vol. 14, no. 2, pp. 508–520, 1993.
[8] D. Liu and A. N. Michel, "Asymptotic stability of discrete time systems with saturation nonlinearities with applications to digital filters," *IEEE Trans. Circuits Syst.*, vol. 39, pp. 798–807, Oct. 1992.
[9] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Networks,* vol. 1, pp. 4–27, Jan. 1990.
[10] ——, "Gradient methods for the optimization of dynamical systems containing neural networks," *IEEE Trans. Neural Networks,* vol. 2, pp. 252–262, Mar. 1991.
[11] Y. Nesterov and A. Nemirovskii, "Interior point polynomial algorithms in convex programming," *SIAM*, vol. 13, 1994.
[12] M. L. Overton, "On minimizing the maximum eigenvalue of a symmetric matrix," *SIAM J. Matrix Anal. Applicat.*, vol. 9, no. 2, pp. 256–268, 1988.
[13] A. Packard and J. Doyle, "The complex structured singular value," *Automatica,* vol. 29, no. 1, pp. 71–109, 1993.
[14] E. Polak and Y. Wardi, "Nondifferentiable optimization algorithm for designing control systems having singular value inequalities," *Automatica,* vol. 18, no. 3, pp. 267–283, 1982.
[15] R. Reed, "Pruning algorithms—A survey," *IEEE Trans. Neural Networks,* vol. 4, pp. 740–747, Sept. 1993.
[16] P. Regalia, *Adaptive IIR Filtering in Signal Processing and Control.* New York: Marcel Dekker, 1995.
[17] J. Shynk, "Adaptive IIR filtering," *IEEE Acoust., Speech, Signal Processing Mag.*, pp. 4–21, 1989.
[18] J. Sjöberg *et al.*, "Nonlinear black-box modeling in system identification: A unified overview," *Automatica,* vol. 31, no. 12, pp. 1691–1724, 1995.
[19] J. A. K. Suykens, B. De Moor, and J. Vandewalle, "Nonlinear system identification using neural state space models, applicable to robust control design," *Int. J. Contr.*, vol. 62, no. 1, pp. 129–152, 1995.
[20] J. A. K. Suykens, J. P. L. Vandewalle, and B. L. R. De Moor, *Artificial Neural Networks for Modeling and Control of Nonlinear Systems.* Boston: Kluwer, 1995.
[21] J. A. K. Suykens and J. Vandewalle, "Learning a simple recurrent neural state space model to behave like Chua's double scroll," *IEEE Trans. Circuits Syst.*, vol. 42, pp. 499–502, Aug. 1995.
[22] ——, "Control of a recurrent neural network emulator for the double scroll," *IEEE Trans. Circuits Syst.*, vol. 43, pp. 511–514, June 1996.
[23] ——, "Discrete time interconnected cellular neural networks within NL_q theory," *Int. J. Circuit Theory Applicat. (Special Issue on Cellular Neural Networks)*, vol. 24, pp. 25–36, 1996.
[24] J. A. K. Suykens, B. L. R. De Moor, and J. Vandewalle, "NL_q theory: A neural control framework with global asymptotic stability criteria," *Neural Networks*, vol. 10, no. 4, pp. 615–637, 1997.
[25] The MathWorks Inc., "Optimization toolbox (version 4.1), user's guide;" also "Robust control toolbox (version 4.1), user's guide," *Matlab*, 1996.
[26] A. C. Tsoi and A. D. Back, "Locally recurrent globally feedforward networks: A critical review of architectures," *IEEE Trans. Neural Networks*, vol. 5, pp. 229–239, Apr. 1994.
[27] L. Vandenberghe and S. Boyd, "A primal-dual potential reduction method for problems involving matrix inequalities," *Math. Programming*, vol. 69, pp. 205–236, 1995.
[28] M. Vidyasagar, *Nonlinear Systems Analysis.* Englewood Cliffs, NJ: Prentice-Hall, 1993.



Johan A. K. Suykens was born in Willebroek, Belgium, on May 18, 1966. He received the B.S. degree in electro-mechanical engineering in 1989 and the doctoral degree in applied sciences in 1995, both from the Katholieke Universiteit Leuven (KU Leuven), Belgium.

He is currently a Postdoctoral Researcher at the ESAT-SISTA group of the KU Leuven. His research interests are mainly in nonlinear systems and neural networks, both theory and applications. He is author of the book *Artificial Neural Networks for Modeling and Control of Non-Linear Systems* (Boston, MA: Kluwer). He was a Visiting Postdoctoral Researcher at the University of California, Berkeley, in 1996.

Dr. Suykens is currently serving as an associate editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, PART I.

Joos Vandewalle (F'92) was born in Kortrijk, Belgium, in August 1948. He received the electrical engineering degree and a doctorate in applied sciences, both from the Katholieke Universiteit Leuven (KU Leuven), Belgium, in 1971 and 1976, respectively.

From 1976 to 1978, he was a Research Associate and from July 1978 to July 1979, he was Visiting Assistant Professor at the University of California, Berkeley. Since July 1979, he has benn with the ESAT Laboratory of the KU Leuven, where he has been a Full Professor since 1986. He has been an Academic Consultant since 1984 at the VSDM Group of Interuniversity Microelectronics Center (IMEC), Leuven. Since August 1996, he has been Chairman of the Department of Electrical Engineering at the KU Leuven. He teaches courses in linear algebra, linear and nonlinear system and circuit theory, signal processing, and neural networks. His research interests are mainly in mathematical system theory and its applications in circuit theory, control, signal processing, cryptography, and neural networks. In 1991 to 1992, he held the Francqui chair on Artificial Neural Networks at the University of Liége, Belgium. He is one of the three coordinators of the Interdisciplinary Center for Neural Networks (ICNN) that was set up in 1993 in order to stimulate the interaction and cooperation among about 50 researchers on neural networks at the KU Leuven. He has authored or co-authored more than 200 papers in these areas. He is the co-author, with S. Van Huffel, of the book *The Total Least Squares Problem* (Philadelphia, PA: SIAM, 1991) and the co-editor, with T. Roska, of the book *Cellular Neural Networks* (New York: Wiley, 1993).

Dr. Vandewalle is a member of the editorial board of *Journal A*, which is a quarterly journal of automatic control, of the *International Journal of Circuit Theory and its Applications*, and of the *Journal of Circuits Systems and Computers*. From 1989 to 1991, he was Associate Editor at the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS in the area of nonlinear and neural networks.

Bart L. R. De Moor was born in Halle, Belgium, on July 12, 1960. He received the Doctoral degree in applied sciences in 1988 from the Katholieke Universiteit Leuven (KU Leuven), Belgium. From 1988 to 1989, he was a Visiting Research Associate with the Departments of Computer Science and Electrical Engineering of Stanford University, Stanford, CA. Since October 1989, he has been a Senior Research Associate of the Fund for Scientific Research Flanders (FWO-Vlaanderen) and Associate Professor at the Department of Electrical Engineering of KU Leuven.

His research interests include numerical linear algebra (generalized svd, complementarity problems, structured total least squares, tensor algebra), system identification (subspace methods, structured total least squares), control theory (robust control, neural networks), and signal processing. He has more than 200 papers in international journals and conference proceedings. From 1991 to 1992, he was the Chief of Staff of the Belgian Federal Minister of Science, and later, he was Chief of Staff of the Belgian Prime Minister. Since 1994, he has been the main advisor on science and technology policy of the Flemish Minister-President

Dr. De Moor received the Leybold–Heraeus Prize (1986), the Leslie Fox Prize (1989), the Guillemin–Cauer Best Paper Award (1990) of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, the biannual Siemens Prize (1994), the triannual best paper award of *Automatica*, which is a journal of the International Federation of Automatic Control (at the IFAC World Congress, San Francisco, CA, 1996), and became a Laureate of the Belgian Royal Academy of Sciences (1992). He is a member of several boards of administrators of (inter)national scientific, cultural, and commercial organizations, including the Flemish Interuniversity Institute for Biotechnology, the Flemish Center for Agricultural Conservation Techniques and Intelligent Systems, Modeling and Control (ISMC NV), which is a spin-off company specializing in modeling and control of multivariable industrial processes and mechatronics analysis and design. He is also a founding member of the philosophical think tank "Worldviews."