Pergamon

Neural Networks, Vol. 10, No. 4, pp. 615–637, 1997

Printed in Great Britain
0893–6080/97 $17.00+.00

PII: S0893-6080(96)00104-9

*CONTRIBUTED ARTICLE*

# $NL_q$ Theory: A Neural Control Framework with Global Asymptotic Stability Criteria

JOHAN A.K. SUYKENS, BART L.R. DE MOOR AND JOOS VANDEWALLE

Katholieke Universiteit Leuven

*(Received 5 December 1994; accepted 22 August 1996)*

**Abstract**—*In this paper a framework for model-based neural control design is presented, consisting of nonlinear state space models and controllers, parametrized by multilayer feedforward neural networks. The models and closed-loop systems are transformed into so-called $NL_q$ system form. $NL_q$ systems represent a large class of nonlinear dynamical systems consisting of q layers with alternating linear and static nonlinear operators that satisfy a sector condition. For such $NL_q$s sufficient conditions for global asymptotic stability, input/output stability (dissipativity with finite $L_2$-gain) and robust stability and performance are presented. The stability criteria are expressed as linear matrix inequalities. In the analysis problem it is shown how stability of a given controller can be checked. In the synthesis problem two methods for neural control design are discussed. In the first method Narendra's dynamic backpropagation for tracking on a set of specific reference inputs is modified with an $NL_q$ stability constraint in order to ensure, e.g., closed-loop stability. In a second method control design is done without tracking on specific reference inputs, but based on the input/output stability criteria itself, within a standard plant framework as this is done, for example, in $H_\infty$ control theory and $\mu$ theory. © 1997 Elsevier Science Ltd.*

**Keywords**—Multilayer recurrent neutral networks, Neural control, State space models, Lyapunov function, Global asymptotic stability, Input/output stability, Dissipativity, Linear matrix inequalities, Dynamic backpropagation.

## 1. INTRODUCTION

Multilayer neural networks possess a number of interesting properties which motivate their use for control applications, such as the universal approximation ability, the parallel network structure and the possibility for on- and off-line learning (Hunt et al., 1992; Zurada, 1992; Miller et al., 1994). The fact that dynamical models, parametrized by multilayer feedforward neural networks, are nonlinear makes them more powerful, e.g., with respect to linear models because a larger class of phenomena can be modeled, such as limit cycles, systems with many equilibria, saturation and hysteresis effects, chaos etc. Neural networks for control may lead to an improved performance. However, despite the many reported successes, there are still a number of open problems such as the lack of general stability results for model-based neural control systems. In classical control theory internal stability of a closed-loop system is indeed considered to be a first desirable property.

One of the main objectives of this paper is precisely to tackle this problem, by presenting sufficient conditions for global asymptotic stability, input/output stability and robust performance of neural control schemes. The proposed framework in this paper is model-based, like the well-known emulator approach proposed in Nguyen and Widrow (1990). Nonlinear state space models and nonlinear dynamic output feedback controllers are parametrized by multilayer perceptrons (simply called "neural state space models" and "neural state space controllers"). Both process noise and measurement noise can be taken into account and full state information is not assumed for the plant.

All systems treated in the paper will be transformed

Acknowledgement: This research 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) and in the framework of a Concerted Action Project MIPS (Modelbased Information Processing Systems) of the Flemish Community.

Requests for reprints should be sent to Johan A.K. Suykens, Katholieke Universiteit Leuven, Department of Electrical Engineering, ESAT-SISTA, Kardinaal Mercierlaan 94, B-3001 Leuven (Heverlee), Belgium; Tel: 32/16/32 11 11 Fax: 32/16/32 19 70; Email: johan.suykens,bart.demoor,joos.vandewalle@esat.kuleuven.ac.be.

into so-called $NL_q$ system form, for which sufficient stability criteria will be given. It turns out that many problems arising in classical, modern and neural control theory and recurrent neural networks are related to $NL_q$s. Examples include neural state space control problems, the discrete time Lur'e problem, linear fractional transformations with real diagonal $\Delta$ uncertainty block, (multilayer) Hopfield networks, (generalized) cellular neural networks, locally recurrent globally feedforward neural networks, and digital filters with overflow characteristic.

With respect to dissipativity of $NL_q$s it will be shown that $NL_q$ theory is a straightforward extension of linear $H_\infty$ control theory and $\mu$ theory (theories is modern control, see for example Maciejowski, 1989) towards general nonlinear systems. Like for many problems arising in systems and control (see Boyd et al., 1994), linear matrix inequalities (LMIs) play an important role in $NL_q$ theory too. As a consequence, checking stability for a given controller involves solving a convex optimization problem, for which efficient polynomial-time algorithms exist. This leads to the formulation of a modified dynamic backpropagation algorithm, in order to assess global asymptotic stability or dissipativity of the closed loop system.

Hence the scope of this paper is:

- to present a model-based framework for neural control design with conditions for closed loop stability, based upon $NL_q$ theory;
- to explain the links between $NL_q$ theory and modern control theories such as $H_\infty$ control theory and $\mu$ theory;
- to discuss learning algorithms within $NL_q$ theory;
- to show that $NL_q$ theory may serve as a tool for the analysis and synthesis of recurrent neural network architectures.

This paper is organized as follows. In Section 2 the framework for neural control design is proposed with neural state space models and controllers. In Section 3 $NL_q$ systems are introduced and it is explained how neural state space control systems can be transformed into $NL_q$ form. In Section 4 sufficient conditions for global asymptotic stability of $NL_q$s are given. Input/output stability criteria and their relation to global asymptotic stability criteria are discussed in Section 5. In Section 6 it is shown in what sense $NL_q$ theory can be considered as an extension towards $\mu$ robust control theory. Learning and neural control design within $NL_q$ theory is presented in Section 7.

## 2. A FRAMEWORK FOR NEURAL CONTROL DESIGN

The framework for neural control design that will be discussed in this paper is *indirect* in the sense that it is assumed that given measured input/output data on a plant, first nonlinear system identification is done and the control design will be based on a model (emulator) for the plant. Once a controller is designed it will be assumed that the certainty equivalence principle holds which means that the controller can be applied to the plant itself instead of to the model (Fig. 1).

Let us consider plants of the form:

$$\begin{cases} x_{k+1} = f_0(x_k, u_k) + \varphi_k \\ y_k = g_0(x_k, u_k) + \psi_k \end{cases} \quad (1)$$

with $f_0(.)$ and $g_0(.)$ continuous nonlinear mappings, $u_k \in \mathbb{R}^m$ the input vector, $y_k \in \mathbb{R}^l$ the output vector, $x_k \in \mathbb{R}^n$ the state vector and $\varphi_k \in \mathbb{R}^n, \psi_k \in \mathbb{R}^l$ process noise and measurement noise respectively, assumed to be zero mean white Gaussian with covariance matrices

$$E\{\begin{bmatrix} \varphi_k \\ \psi_k \end{bmatrix} [\varphi_s^T \psi_s^T]\} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \delta_{ks}.$$

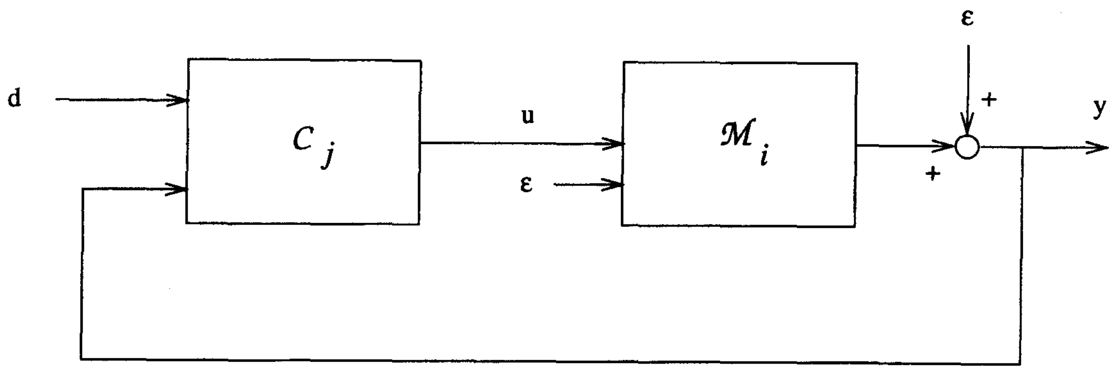


**FIGURE 1. Control scheme with neural state space model (emulator) $\mathcal{M}_i$ and a neural state space controller $C_j$. The model $\mathcal{M}_i$ is described by means of a nonlinear state space model of the form $\hat{x}_{k+1} = f(\hat{x}_k, y_k, \epsilon_k; \theta_m)$, $u_k = g(\hat{x}_k, u_k, \epsilon_k; \theta_m)$ and $C_j$ is a dynamic output feedback controller with nonlinear state space model $z_{k+1} = h(z_k, y_k, d_k; \theta_c)$, $u_k = s(z_k, y_k, d_k; \theta_c)$. The functions $f(.)$, $g(.)$, $h(.)$, $s(.)$ are parametrized by multilayer feedforward neural networks, with interconnection weights $\theta_m$ for $f,g$ and $\theta_c$ for $h,s$ respectively. Several parametrizations are considered, resulting in a certain neural control problem $\mathbb{Z}_j^i$ related to a model $\mathcal{M}_i$ and a controller $C_j$. The signal $y_k$ represents the output of the plant, $d_k$ the reference input, $u_k$ the control signal and $\epsilon_k$ a white noise innovations signal.**

In connection to eqn (1), nonlinear dynamic output feedback controllers of the form

$$\begin{cases} z_{k+1} = h(z_k, y_k, d_k) \\ u_k = s(z_k, y_k, d_k) \end{cases} \quad (2)$$

will be considered with $h(.)$, $s(.)$ continuous nonlinear mappings, $z_k \epsilon \mathbb{R}^{n_z}$ the state of the controller, and $d_k \epsilon \mathbb{R}^l$ the reference input. The plant models and the controllers will be parametrized by multilayer perceptrons.

### 2.1. Neural State Space Models

Nonlinear system identification for systems of the form in eqn (1) is discussed in Suykens et al. (1995a), both for the case of deterministic identification ($\varphi_k = 0, \psi_k = 0$) and for the stochastic case ($\varphi_k \not\equiv 0, \psi_k \not\equiv 0$). The neural network models of Suykens et al. (1995a) are based on the following predictor in innovations form, proposed, e.g., in Goodwin and Sin (1984):

$$\mathcal{M}(\theta): \begin{cases} \hat{x}_{k+1} = \Phi(\hat{x}_k, u_k\, \theta) + K(\theta)\epsilon_k\, \hat{x}_0 = x_0 \text{ given} \\ \hat{y}_k = \Psi(\hat{x}_k, u_k\, \theta). \end{cases} \quad (3)$$

Here $\epsilon_k(\theta) = y_k - \hat{y}_k(\theta)$ is the prediction error, $\mathcal{M}$ denotes the model structure that is parametrized by the parameter vector $\theta$. Given $N$ input/output data $Z^N$, a prediction error algorithm (Ljung, 1987) aims at minimizing the cost function

$$\min_{\theta} V_N(\theta, Z^N) = \frac{1}{N}\sum_{k=1}^{N} l(\epsilon_k(\theta)). \quad (4)$$

A typical choice for $l(\epsilon_k)$ is $\frac{1}{2}\epsilon_k^t\epsilon_k$. The predictor eqn (3) makes use of a direct parametrization of the Kalman gain $K(\theta)$ and does not assume a complicated dependency through a Riccati equation as for extended Kalman filtering. More details on these aspects can be found in Suykens et al. (1995a). For the model structure eqn (3) parametrizations by feedforward neural nets are considered. Such parametrizations make sense because any continuous nonlinear function can be approximated arbitrarily well on a compact interval by a multilayer perceptron with one or more hidden layers (see, for example, Cybenko, 1989; Funahashi, 1989; Hornik et al., 1989; Leshno et al., 1993). The reason for this specific choice of parametrizations is that it will enable us to set up a control theory with stability criteria.

Definition 1 (neural state space models). *The following parametrizations by feedforward neural nets for $\Phi(.)$ and $\Psi(.)$ in the predictor eqn (3) are made and are called $\mathcal{M}_i(i\epsilon\{0,\ldots,3\})$ depending on the specific parametrization*:

$$\mathcal{M}_0(\theta): \begin{cases} \hat{x}_{k+1} = A\hat{x}_k + Bu_k + K\epsilon_k \\ y_k = C\hat{x}_k + Du_k + \epsilon_k \end{cases}$$

$$\mathcal{M}_1(\theta): \begin{cases} \hat{x}_{k+1} = W_{AB}\tanh(V_A\hat{x}_k + V_Bu_k + \beta_{AB}) + K\epsilon_k \\ y_k = C\hat{x}_k + Du_k + \epsilon_k \end{cases}$$

$$\mathcal{M}_2(\theta): \begin{cases} \hat{x}_{k+1} = W_{AB}\tanh(V_A\hat{x}_k + V_Bu_k + \beta_{AB}) + K\epsilon_k \\ y_k = W_{CD}\tanh(V_C\hat{x}_k + V_Du_k + \beta_{CD}) + \epsilon_k \end{cases}$$

$$\mathcal{M}_3(\theta): \begin{cases} \hat{x}_{k+1} = W_{AB}\tanh(V_{AB}\tanh(V_A\hat{x}_k + V_Bu_k + \beta_{AB_v}) \\ \qquad + \beta_{AB_w}) + K\epsilon_k \\ y_k = W_{CD}\tanh(V_{CD}\tanh(V_C\hat{x}_k + V_Du_k + \beta_{CD_v}) \\ \qquad + \beta_{CD_w}) + \epsilon_k. \end{cases} \quad (5)$$

Remark 1. *Here tanh(.) has to be applied elementwise (diagonal nonlinearity). The linear model in innovations form $\mathcal{M}_0$ is the Kalman filter (see, for example, Ljung, 1979, 1987) and can be interpreted as a specific neural state space model having a parametrization without hidden layers and a linear activation function for the output neurons. In eqn (5) the process is linear in $\epsilon_k$. The case of deterministic identification corresponds to a zero Kalman gain K. The reason why the linear model $\mathcal{M}_0$ is included as one of the neural state space models is to make a comparison between* $\mathrm{NL}_q$ *theory and linear control theory possible in the sequel.*

In dynamic backpropagation (Narendra & Parthasarathy, 1990, 1991) the gradient of the cost function eqn (4) is obtained from a sensitivity model, which is in itself a dynamical system. Sensitivity models for neural state space models have been derived in Suykens et al. (1995a). For aspects such as model validation, pruning and regulation see, for example, Billings et al. (1992), Reed (1993) and Sjöberg and Ljung (1992).

### 2.2. Neural State Space Controllers

For eqn (2) we consider the parametrizations:

$$\begin{cases} z_{k+1} = h(z_k, y_k, d_k; \theta_c) \\ u_k = s(z_k, y_k, d_k; \theta_c), \end{cases} \quad (6)$$

parametrized by the controller parameter vector $\theta_c \in \mathbb{R}^{P_c}$.

Definition 2 (neural state space controllers). *The following parametrizations by feedforward neural nets for h(.) and s (.) are made in eqn (2) and are called neural state space controllers $C_i(i\epsilon\{0,\ldots,5\})$ depending on the specific parametrization*:

$$C_0(\theta_c): \begin{cases} z_{k+1} = Ez_k + Fy_k + F_2d_k \\ u_k = Gz_k + Hy_k + H_2d_k \end{cases}$$

$$C_1(\theta_c): \begin{cases} z_{k+1} = Ez_k + Fy_k + F_2d_k \\ u_k = \tanh(Gz_k + Hy_k + H_2d_k) \end{cases}$$

$$C_2(\theta_c):\begin{cases} z_{k+1}=W_{EF}\tanh(V_E z_k+V_F y_k+V_{F_2}d_k+\beta_{EF})\\ u_k=W_{GH}\tanh(V_G z_k+V_H y_k+V_{H_2}d_k+\beta_{GH})\end{cases}$$

$$C_3(\theta_c):\begin{cases} z_{k+1}=W_{EF}\tanh(V_E z_k+V_F y_k+V_{F_2}d_k+\beta_{EF})\\ u_k=\tanh(W_{GH}\tanh(V_G z_k\\ \quad +V_H y_k+V_{H_2}d_k+\beta_{GH}))\end{cases}$$

$$C_4(\theta_c):\begin{cases} z_{k+1}=W_{EF}\tanh((V_{EF}\tanh V_E z_k+V_F y_k\\ \quad +V_{F_2}d_k+\beta_{EF_v})+\beta_{EF_w})\\ u_k=W_{GH}\tanh(V_{GH}\tanh(V_G z_k+V_H y_k\\ \quad +V_{H_2}d_k+\beta_{GH_v})+\beta_{GH_w})\end{cases}$$

$$C_5(\theta_c):\begin{cases} z_{k+1}=W_{EF}\tanh(V_{EF}\tanh(V_E z_k+V_F y_k\\ \quad +V_{F_2}d_k+\beta_{EF_v})+\beta_{EF_w})\\ u_k=\tanh(W_{GH}\tanh(V_{GH}\tanh(V_G z_k+V_H y_k\\ \quad +V_{H_2}d_k+\beta_{GH_v})+\beta_{GH_w})).\end{cases} \tag{7}$$

REMARK 2. *The controllers $C_1$, $C_3$, $C_5$ correspond respectively to the controllers $C_0$, $C_2$, $C_4$ but with saturated output $u_k$. $C_0$ is a classical linear dynamic output feedback controller (see e.g. Boyd and Barratt, 1991; Maciejowski, 1989)* (Fig. 2).

Like in modern control theory we will consider the standard plant configuration of Fig. 3 (see Boyd & Barratt, 1991) instead of the classical control scheme. This standard plant is a reorganized scheme with a so-called exogenous input $w_k$ (consisting of the reference input and disturbance signals), regulated output $e_k$ (consisting of the tracking error $d_k-\hat{y}_k$ and other to be regulated variables of interest), sensed output $y_k$ and actuator input $u_k$. An augmented plant $\mathcal{S}_i$ is considered with inputs $w_k$, $u_k$ and output $e_k$, $y_k$.

DEFINITION 3 (neural state space control problem $\Xi_j^i$). *A neural control problem $\Xi_j^i$ is the control problem related to the control scheme of Fig. 1. with a neural state space controller $C_j(j\epsilon\{0,1,\ldots,5\})$ and a neural state space model $\mathcal{M}_i(i\epsilon\{0,1,\ldots,3\})$.*

The family of problems $\Xi_j^i$ given in Table 1 is considered throughout this paper.

## 3. $\mathrm{NL}_q$ SYSTEMS

An essential concept in the further derivation of global asymptotic stability criteria for the family of neural control problems $\Xi_j^i$ is the so-called $\mathrm{NL}_q$ system (Fig. 4).

DEFINITION 4 ($\mathrm{NL}_q$ system). *The following discrete time nonlinear state space model is called an* $\mathrm{NL}_q$ *system*:

$$\begin{cases} p_{k+1}=\Gamma_1(V_1\Gamma_2(V_2\ldots\Gamma_q(V_q p_k+B_q w_k)\ldots\\ \quad +B_2 w_k)+B_1 w_k)\\ e_k=\Lambda_1(W_1\Lambda_2(W_2\ldots\Lambda_q(W_q p_k+D_q w_k)\ldots\\ \quad +D_2 w_k)+D_1 w_k)\end{cases} \tag{8}$$

*where* $\Gamma_i(p_k,w_k),\Lambda_i(p_k,w_k)(i=1,\ldots,q)$ *are diagonal*

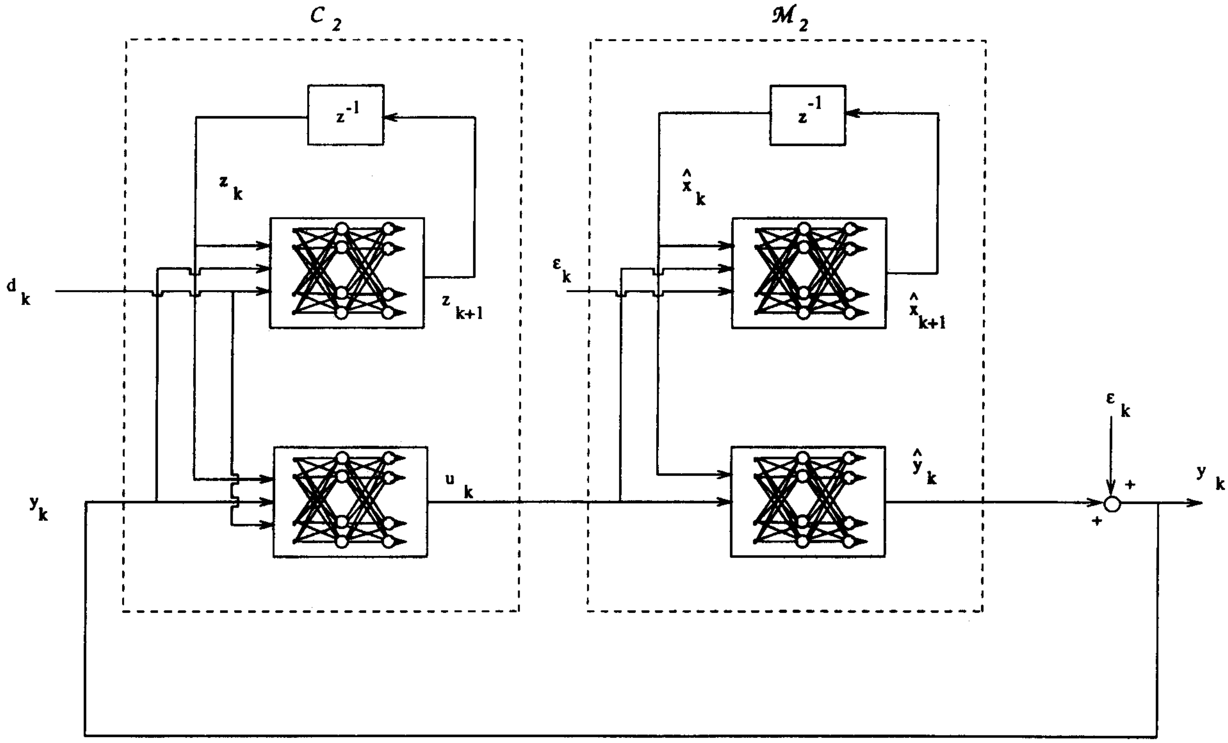


FIGURE 2. Specific case $\Xi_2^2$ of Fig. 1 with neural state space model $\mathcal{M}_2$ and neural state space controller $C_2$.

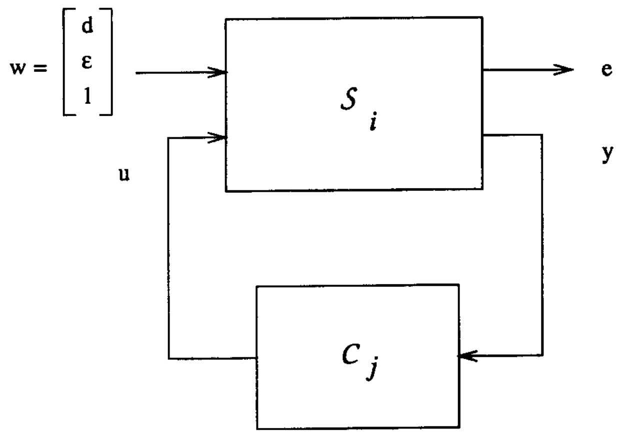


**FIGURE 3. Standard plant representation as it is defined in modern control theory with $e_k$ the regulated outputs, $y_k$ the sensed outputs, $u_k$ the actuator inputs and $w_k = [d_k; \epsilon_k; 1]$ the exogenous input consisting of the reference inputs and disturbances. The constant input 1 is due to the bias in the neural network architectures. $S_i$ is called the augmented plant model.**

*matrices with diagonal elements $\gamma_j(p_k, w_k), \lambda_j(p_k, w_k) \epsilon [0,1]$ for all $p_k$, $w_k$ and depending continuously on the variables $p_k$, $w_k$.*

REMARK 3. *The term "$NL_q$" system refers to the alternating sequence of nonlinear and linear operators acting on the input arguments $p_k$, $w_k$: the diagonal matrices $\Gamma_i, \Lambda_i$ depend on $p_k$, $w_k$ and refer to "N" in "$NL_q$", the constant matrices $V_i$, $W_i$, $B_i$, $D_i$, refer to "L" in "$NL_q$". The index q refers to the number of nonlinear operators in this alternating sequence.*

Such $NL_q$ systems (or shortly $NL_q$s) will then be considered for the standard plant representation of Fig. 3. For $q = 1,2$ one has:

$$q=1: \begin{cases} p_{k+1} = \Gamma_1(V_1 p_k + B_1 w_k) \\ e_k = \Lambda_1(W_1 p_k + D_1 w_k) \end{cases}$$

$$q=2: \begin{cases} p_{k+1} = \Gamma_1(V_1 \Gamma_2(V_2 p_k + B_2 w_k) + B_1 w_k) \\ e_k = \Lambda_1(W_1 \Lambda_2(W_2 p_k + D_2 w_k) + D_1 w_k) \end{cases}$$

with $p_k \epsilon \mathbb{R}^{n_p}, w_k \epsilon \mathbb{R}^{n_w}, e_k \epsilon \mathbb{R}^{n_e}$. The $NL_q$ system eqn (8) is closely related to multilayer recurrent neural networks of the form:

$$\begin{cases} p_{k+1} = \sigma_1(V_1 \sigma_2(V_2 \ldots \sigma_q(V_q p_k + B_q w_k) \ldots \\ \qquad + B_2 w_k) + B_1 w_k) \\ e_k = \eta_1(W_1 \eta_2(W_2 \ldots \eta_q(W_q p_k + D_q w_k) \ldots \\ \qquad + D_2 w_k) + D_1 w_k) \end{cases} \tag{10}$$

with $\sigma_i(.)\eta_i(.)$ vector valued static nonlinearities that belong to sector [0,1] (see, for example, Vidyasagar, 1993). This is illustrated here for the Hopfield network with synchronous updating:

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

In Suykens et al. (1995a) it is shown that the latter can be written as

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

with $\Gamma = diag\{\gamma_i\}$ and $\gamma_i = \tanh(w_i^T x_k)/(w_i^T x_k)$. This is obtained by using an elementwise notation and based on the fact that $\tanh(\cdot)$ is a diagonal nonlinearity. One has

$$\begin{aligned} x^i &:= \tanh(\textstyle\sum_j w_j^i x^j) \\ &:= \frac{\tanh(\sum_j w_j^i x^j)}{\sum_j w_j^i x^j} . \textstyle\sum_j w_j^i x^j \\ &:= \gamma_i^i \textstyle\sum_j w_j^i x^j. \end{aligned}$$

The time index is omitted here because of the assignment operator ": = ". The notation $\gamma_i^i$ 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 $\tanh(\cdot)$ can be taken, leading to $\gamma_i = 1$.

Also if an additional layer is considered

$$x_{k+1} = \tanh(V \tanh(W x_k)),$$

this can be written as

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

where

$$\Gamma_1 = diag\{\gamma_{1_i}\}$$

with

$$\gamma_{1_i} = \tanh(v_i^T \tanh(W x_k))/(v_i^T \tanh(W x_k))$$

**TABLE 1**
**This Table Shows the Several Model-based Neural Control Strategies $\Xi_j^i$. The Rows Represent Possible Models in Increasing Level of Complexity (Hidden Layers). The Columns Represent the Several Possible Neural Controllers in Increasing Level of Complexity.**

| | $C_0$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
|---|---|---|---|---|---|---|
| $M_0$ | $\Xi_0^0$ | $\Xi_1^0$ | - | - | - | - |
| $M_1$ | $\Xi_0^1$ | $\Xi_1^1$ | $\Xi_2^1$ | $\Xi_3^1$ | - | - |
| $M_2$ | $\Xi_0^2$ | $\Xi_1^2$ | $\Xi_2^2$ | $\Xi_3^2$ | - | - |
| $M_3$ | $\Xi_0^3$ | $\Xi_1^3$ | $\Xi_2^3$ | $\Xi_3^3$ | $\Xi_4^3$ | $\Xi_5^3$ |

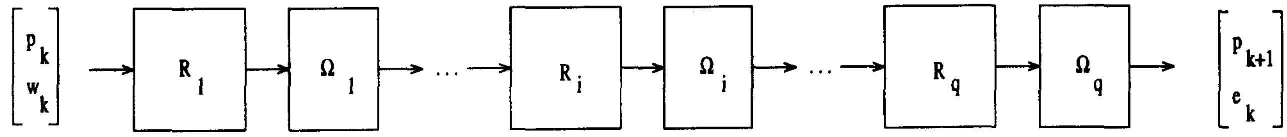


**FIGURE 4. $NL_q$ system with respect to the standard plant representation with $e_k$ the regulated output, $w_k$ the exogenous input and $p_k$ the state of the $NL_q$ system. $NL_q$ stands for an alternating sequence of linear operations $R_i$ and nonlinear operations $\Omega_i$ on the input arguments $[p_k; w_k]$. The matrices $\Omega_i$ are diagonal, the diagonal elements depend on $p_k$, $w_k$ but satisfy the condition $\|\Omega_i\| \leq 1$.. All $\Xi_j^i$ neural control problems are formulated as $NL_q$s.**

and

$$\Gamma_2 = diag\{\gamma_{2_i}\}$$

with

$$\gamma_{2_i} = \tanh(w_i^T x_k)/(w_i^T x_k).$$

This becomes clear by writing:

$$x^i := \tanh(\textstyle\sum_j v_j^i \tanh(\sum_l w_l^j x^l)$$

$$:= \tanh(\textstyle\sum_j v_j^i \gamma_{2_j}^j \sum_l w_l^j x^l)$$

$$:= \gamma_{1_i}^i \textstyle\sum_j v_j^i \gamma_{2_j}^j \sum_l w_l^j x^l.$$

The following Lemma states that the neural state space models $\mathcal{M}_i$ are special cases of $NL_q$ systems.

LEMMA 1 ($\mathcal{M}_i$s are $NL_q$s). *Neural state space models $\mathcal{M}_i (i \in \{0, 1, \ldots, 3\})$ can be written as $NL_q$ systems with $q = 1$ if $i = 0$ and $q = 2$ if $i = 1, 2$ and finally $q = 3$ if $i = 3$. The input vector of the $NL_q$ is $w_k = [u_k; \epsilon_k; 1]$.*

*Proof.* See Appendix A. By taking an elementwise notation, multiplying and dividing by the arguments of the activation functions, the alternating sequence of nonlinear and linear operators in eqn (8) is obtained. The elements $\gamma_i, \lambda_i$ belong to [0,1] because the activation function tanh is a nonlinearity belonging to sector [0,1]. □

REMARK 4. *The constant input 1 in $w_k$ is due to the bias vector of the neural network. It is well known that the bias vector can be formally treated as part of the interconnection matrix by introducing an additional constant input to the network.*

LEMMA 2 ($\Xi_j^i$s are $NL_q$s). *All members of the family of neural control problems $\Xi_j^i$ can be written as $NL_q$ systems.* □

In Table 2 the $q$ values for the several neural space control problems are shown. Two examples, $\Xi_1^0$ and $\Xi_2^2$ are given here.

EXAMPLE 1. *Given the Kalman filter ($\mathcal{M}_0$) and a linear dynamic output feedback controller with saturation ($C_1$)*

$$M_0 : \begin{cases} \hat{x}_{k+1} = A\hat{x}_k + Bu_k + K\epsilon_k \\ y_k = C\hat{x}_k + Du_k + \epsilon_k \end{cases}$$

$$C_1 : \begin{cases} z_{K+1} = Ez_k + Fy_k + F_2 d_k \\ u_k = \tanh(Gz_k) \end{cases}$$

*the state equation for the closed loop system is*

$$\begin{cases} \hat{x}_{k+1} = A\hat{x}_k + B\tanh(Gz_k) + K\epsilon_k \\ z_{k+1} = FC\hat{x}_k + Ez_k + FD\tanh(Gz_k) + F\epsilon_k + F_2 d_k. \end{cases}$$

*A new state variable is introduced:*

$$\xi = \tanh(Gz_k).$$

*This state augmentation leads to the following representation*

$$\begin{cases} \hat{x}_{k+1} = A\hat{x}_k + B\xi_k + K\epsilon_k \\ z_{k+1} = Ez_k + FC\hat{x}_k + FD\xi_k + F\epsilon_k + F_2 d_k \\ \xi_{k+1} = \tanh(GEz_k + GFC\hat{x}_k + GFD\xi_k \\ \qquad + GF\epsilon_k + GF_2 d_k) \end{cases}$$

*and can be written as an $NL_1$ system*

$$p_{k+1} = \Gamma_1(V_1 p_k + B_1 w_k),$$

*by taking*

$$p_k = [\hat{x}_k \; z_k \; \xi_k], w_k = [d_k \; \epsilon_k \; 1]$$

*and matrices*

$$V_1 = \begin{bmatrix} A & 0 & B \\ FC & E & FD \\ GFC & GE & GFD \end{bmatrix}, B_1 = \begin{bmatrix} 0 & K & 0 \\ F_2 & F & 0 \\ GF_2 & GF & 0 \end{bmatrix}$$

*and $\Gamma_1 = diag\{I, I, \Gamma_G(\xi_k, z_k, \hat{x}_k, d_k, \epsilon_k)\}$, where $\Gamma_G$ is a diagonal matrix with elements $\gamma_{G_i} \in [0,1]$ which follows from Lemma 1.*

EXAMPLE 2. *Given the model $\mathcal{M}_2$ and the controller $C_2$* (Fig. 2)

$$M_2 : \begin{cases} \hat{x}_{k+1} = W_{AB}\tanh(V_A\hat{x}_k + V_B u_k + \beta_{AB}) + K\epsilon_k \\ y_k = W_C\tanh(V_C\hat{x}_k) + \epsilon_k \end{cases}$$

$$C_2 : \begin{cases} z_{k+1} = W_{EF}\tanh(V_E z_k + V_F y_k + V_{F_2} d_k + \beta_{EF}) \\ u_k = W_G \tanh(V_G z_k) \end{cases}$$

*the state equation for the closed loop system is*

$$\begin{cases} \hat{x}_{k+1} = W_{AB}\tanh(V_a \hat{x}_k + V_B W_G \tanh(V_G z_k) \\ \qquad + \beta_{AB}) + K\epsilon_k \\ z_{k+1} = W_{EF}\tanh(V_E z_k + V_F W_C \tanh(V_C \hat{x}_k) \\ \qquad + V_F \epsilon_k + V_{F_2} d_k + \beta_{EF}). \end{cases}$$

*State augmentation is done by defining*

$$\xi_k = \tanh(V_C \hat{x}_k)$$

$$\eta_k = \tanh(V_G z_k),$$

*leading to the representation*

$$\begin{cases} \hat{x}_{k+1} = W_{AB}\tanh(V_A \hat{x}_k + V_B W_G \eta_k + \beta_{AB}) + K\epsilon_k \\ z_{k+1} = W_{EF}\tanh(V_E z_k + V_F W_C \xi_k \\ \quad + V_F \epsilon_k + V_{F_2} d_k + \beta_{EF}) \\ \xi_{k+1} = \tanh(V_C W_{AB}\tanh(V_A \hat{x}_k + V_B W_G \eta_k + \beta_{AB}) \\ \quad + V_C K \epsilon_k) \\ \eta_{k+1} = \tanh(V_G W_{EF}\tanh(V_E z_k + V_F W_{CD} \xi_k \\ \quad + V_F \epsilon_k + V_{F_2} d_k + \beta_{EF})) \end{cases}$$

*which can be written as an* $\mathrm{NL}_2$ *system*

$$p_{k+1} = \Gamma_1(V_1 \Gamma_2(V_2 p_k + B_2 w_k) + B_1 w_k)$$

*by taking* $p_k = [\hat{x}_k\ z_k\ \xi_k\ \eta_k]$, $w_k = [d_k\ \epsilon_k\ 1]$ *and*

$$V_1 = \begin{bmatrix} W_{AB} & & & \\ & W_{EF} & & \\ & & V_C W_{AB} & \\ & & & V_G W_{EF} \end{bmatrix},$$

$$V_2 = \begin{bmatrix} V_A & 0 & 0 & V_B W_G \\ 0 & V_E & V_F W_C & 0 \\ V_A & 0 & 0 & V_B W_G \\ 0 & V_E & V_F W_C & 0 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0 & 0 & \beta_{AB} \\ V_{F_2} & V_F & \beta_{EF} \\ 0 & 0 & \beta_{AB} \\ V_{F_2} & V_F & \beta_{EF} \end{bmatrix}, B_1 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & V_C k & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

## 4. GLOBAL ASYMPTOTIC STABILITY CRITERIA FOR $\mathrm{NL}_q$s

In this Section autonomous $\mathrm{NL}_q$ systems are considered for eqn (8) with $w_k = 0$. Sufficient conditions for global asymptotic stability are presented. Three types of criteria are given: diagonal scaling and criteria based on diagonal dominance or condition numbers of certain matrices.

THEOREM 1 (diagonal scaling). *Consider the autonomous* $\mathrm{NL}_q$ *system*

$$p_{k+1} = \left(\prod_{i=1}^{q} \Gamma_i(p_k) V_i\right) p_k \tag{11}$$

*and let*

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

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

**TABLE 2**
**Neural State Space Control Problems $\Xi^i$**

| | $C_0$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
|---|---|---|---|---|---|---|
| Corresponding $q$ value for the $\mathrm{NL}_q$ related to the problem $\Xi_j^i$ | | | | | | |
| $M_0$ | 1 | 1 | – | – | – | – |
| $M_1$ | 2 | 2 | 2 | 3 | – | – |
| $M_2$ | 2 | 2 | 2 | 3 | – | – |
| $M_3$ | 4 | 4 | 4 | 4 | 4 | 5 |
| The number of new state variables to be introduced in order to obtain the $\mathrm{NL}_q$ form for a given $\Xi_j^i$ | | | | | | |
| $M_0$ | 0 | 1 | – | – | – | – |
| $M_1$ | 0 | 1 | 1 | 1 | – | – |
| $M_2$ | 1 | 2 | 2 | 2 | – | – |
| $M_3$ | 1 | 2 | 2 | 2 | 2 | 2 |

*A sufficient condition for global asymptotic stability of eqn (11) is to find a diagonal matrix $D_{tot}$ such that*

$$\|D_{tot}V_{tot}D_{tot}^{-1}\|_2^q = \beta_D < 1 \qquad (12)$$

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

*Proof.* See Appendix A. The Theorem is based on the Lyapunov function $V = \|D_1 p\|_2$ which is radially unbounded. The proof makes use of the properties of induced norms. □

REMARK 5.

- *The equilibrium point $p = 0$ in eqn (11) is unique because the Lyapunov function is radially unbounded.*
- *The minimum over $D_{tot}$ yields a maximal contraction rate of the flow for the* $\mathrm{NL}_q$ *system and shows how such a feasible point $D_{tot}$ is found in practice*:

$$\min_{D_{tot}} \|D_{tot}V_{tot}D_{tot}^{-1}\|_2. \qquad (13)$$

- *Given a constant matrix $V_{tot}$ the criterion eqn (12) in the unknown diagonal matrix log ($D_{tot}$) is a convex optimization problem. This implies that the problem has only one minimum, which is the global one. Moreover this minimum can be found in polynomial time. One also meets criteria of the form eqn (12) in the field of modern robust control theory (e.g. Boyd et al., 1994; Packard & Doyle, 1993; Kaszkurewicz & Bhaya, 1993). This already suggests there are links between that field and the neural state space model framework that is outlined in this paper. The precise links will be pointed out in Section 6.*

In order to obtain conditions that are possibly "sharper" (less conservative) a Lyapunov function $V(p) = \|P_1 p\|_2$ will be proposed where $P_1$ is a matrix in which nonzero off-diagonal elements are allowed. Diagonality of $P_1$ is then relaxed to diagonal dominance of $P_1^T P_1$. In order to prove the next theorem, the following definition is introduced first.

DEFINITION 5 (level of diagonal dominance). *A matrix $Q \in \mathbb{R}^{n \times n}$ is called diagonal dominant of level $\delta_Q \geq 1$ if the following property holds*:

$$q_{ii} > \delta_Q \cdot \sum_{j=1 (j \neq i)}^{n} |q_{ij}|, \forall i = 1, \ldots, n. \qquad (14)$$

REMARK 6. *This definition is consistent with the original definition of diagonal dominance (see, for example, Liu & Michel, 1992), which corresponds to the special case $\delta_Q = 1$.*

THEOREM 2 (diagonal dominance). *A sufficient condition for global asymptotic stability of the autonomous* $\mathrm{NL}_q$ *system eqn (11) is to find matrices $P_i$ $N_i$ such that*

$$c_\alpha \beta_P < 1 \qquad (15)$$

*where* $c_\alpha = \prod_{i=1}^{q} (i + \alpha_i)^{1/2}$, $\beta p = \|P_{tot}V_{tot}P_{tot}^{-1}\|_2^q$ *with* $P_{tot} = \mathrm{blockdiag}\{P_2, P_3, \ldots, P_q, P_1\}$, *and $P_i \in \mathbb{R}^{n_{h_i} \times n_{h_i}}$ of full rank. The matrices $Q_i = P_i^T P_i N_i$ are diagonal dominant with $\delta_{Q_i} = (i + \alpha_i)/\alpha_i \geq 1$ ($\alpha_i \geq 0$) and $N_i$, are diagonal matrices with positive diagonal elements.*

*Proof.* See Appendix A. The proof is based on the Lyapunov function $V(p) = \|P_1 p\|_2$ which is radially unbounded and makes use of the properties of induced norms. The levels of diagonal dominance $\delta_{Qi}$, are derived from Gerschgorin's Theorem (see Wilkinson, 1965). □

The following Lemma gives an equivalent expression for diagonal dominance of a matrix in terms of the diagonal and the off-diagonal part of the matrix:

LEMMA 3. *Given a matrix $Q \in \mathbb{R}^{n \times n}$, the condition of diagonal dominance of level $\delta_Q$* eqn (14)

$$q_{ii} > \delta_Q \cdot \sum_{j=1 (j \neq i)}^{n} |q_{ij}|, \forall i = 1, \ldots, n$$

*is equivalent to the condition*

$$\|X_Q\|_\infty < 1/\delta_Q \qquad (16)$$

*with $X_Q = D_Q^{-1} H_Q$, where $Q = D_Q + H_Q$ and $D_Q$ is the diagonal and $H_Q$* the off-diagonal part of $Q$.

*Proof.* See Appendix A. It follows immediately from the definition of $\|X_Q\|_\infty$. □

REMARK 7.

- *Again the equilibrium point $p = 0$ is unique because the Lyapunov function is radially unbounded.*
- *The solution to the optimization problem*

$$\min_{P_i, N_i} c_\alpha \beta_P \qquad (17)$$

  *yields a maximal contraction rate of the flow of the autonomous* $\mathrm{NL}_q$ *system with respect to the condition eqn (15).*
- *Theorem 1 is a special case of Theorem 2 corresponding to $\alpha_i = 0$ (or $\delta_{Q_i} \to \infty$), making the matrices $P_i$ diagonal.*
- *The problem eqn (15) can be interpreted as follows: find matrices $P_i$, $N_i$ such that*

$$\|P_{tot}V_{tot}P_{tot}^{-1}\|_2^q < 1/c_\alpha$$

  *and*

$$\|X_{Q_i}\|_\infty < 1\delta_{Q_i}.$$

  *The upper bounds are plotted with respect to $\alpha_i$ for $q = 1$ in Fig. 5. There exists a trade-off between $\delta_Q$ and $c_\alpha$: the lower the level of diagonal dominance on the*

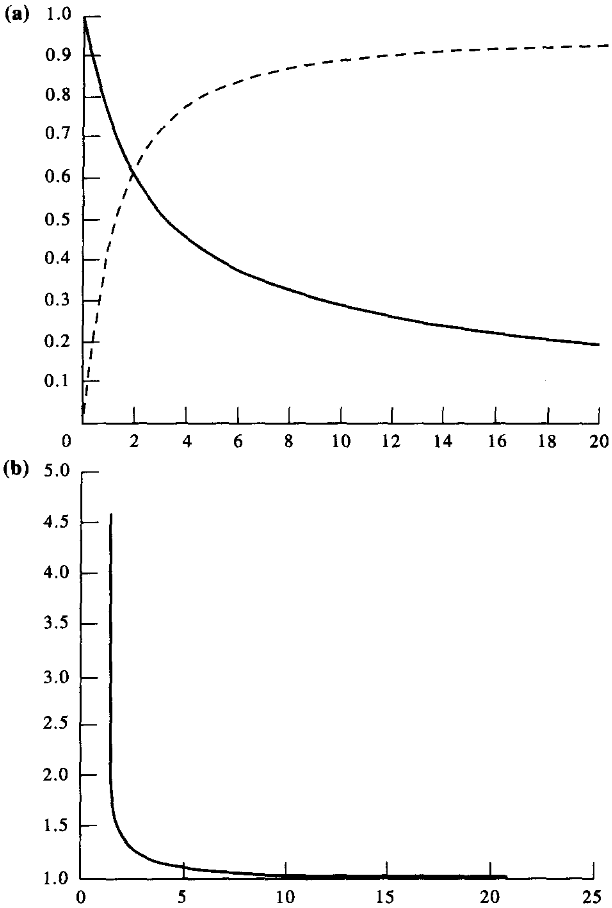


**FIGURE 5. Interpretation of Theorem 2 for $q = 1$: (a) upper bounds $1/c_\alpha$ (full line) and $1/\delta_Q$ (dashed line) as a function of $\alpha$. (b) trade-off curve between $c_\alpha$ and $\delta_Q$. The curve $c_\alpha \delta_Q$ shows that a large correction factor $c\alpha$ is needed if the level of diagonal dominance of $Q = P^T P$ becomes low. Large diagonal dominance of $Q$ implies a small correction factor $c_\alpha$. The limiting case $\delta_Q \rightarrow \infty$ corresponds to diagonal scaling.**

*$Q_i$s is, the stronger the condition on $\|P_{tot} V_{tot} P_{tot}^{-1}\|_2$ becomes. Low diagonal dominance together without a large correction factor $c_\alpha$ is typically unfeasible.*

- *Elimination of $\alpha_i$ in eqn (15), yields the criterion*

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

- *For all values $l \in \{1, 2, \ldots, q\}$ for which $\Gamma_l = I$ there is no condition of diagonal dominance on $Q_l$, but only on the remaining $Q_i$s ($i \neq \ell$). This follows immediately from the proof of the Theorem. Examples of this are the neural state space control problems $\Xi_0^1$ and $\Xi_1^1$ for which $\Gamma_1 = I$. Another example is the linear system*

$$p_{k+1} = A p_k,$$

*for which a sufficient condition for global asymptotic stability becomes*

$$\min_{P \in \mathbb{R}^{n \times n}} \|PAP^{-1}\|_2 < 1$$

*because $\Gamma_1 = I$ and by taking the Lyapunov function $V(p) = \|P_p\|_2$. It is also well known that this corresponds to*

$$\rho(A) < 1$$

*where $\rho(A)$ denotes the spectral radius of $A$ (see Boyd et al., 1994). In this case the condition (27) is also necessary.*

- *In the case of a* sat(.) *activation function (defined as $sat(x) = x$ if $|x| \leq 1$ and $sat(x) = sign(x)$ if $|x| \geq 1$) it is proven in Liu and Michel (1992) for $q = 1$ that a necessary and sufficient condition for*

$$\mathrm{sat}(p)^T Q\, \mathrm{sat}(p) < p^T Q_p$$

*to hold $\forall_p$ is $\delta_Q = 1$, taking $N_1 = I$. Hence for that specific activation function eqn (15) becomes for $q = 1$*

$$\|PVP^{-1}\|_2 < 1 \tag{19}$$

*subject to*

$$\|X_Q\|_\infty \leq 1.$$

*Hence $c_\alpha = 1$ for this particular activation function and the criterion is sharper. This problem was investigated in the context of the stability of digital filters with overflow characteristic in Liu and Michel (1992).*

The following Theorem also allows full instead of diagonal matrices $P_i$, but a correction factor is expressed in terms of the condition numbers of the matrices $P_i$.

THEOREM 3 (condition number factor). *A sufficient condition for global asymptotic stability of the autonomous* $\mathrm{NL}_q$ *is to find matrices $P_i$ such that*

$$\prod_{i=1}^{q} \kappa(P_i)\, \|P_{tot} V_{tot} P_{tot}^{-1}\|_2^q < 1 \tag{20}$$

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

*Proof.* See Appendix A.

## 5. INPUT/OUTPUT STABILITY CRITERIA FOR $\mathrm{NL}_q$S

In this section input/output properties of the non-autonomous $\mathrm{NL}_q$ system are studied. First some equivalent input/output representations for $\mathrm{NL}_q$ are given. Then input/output stability criteria are presented with respect to these representations. Again three types of criteria are derived: diagonal scaling and criteria based on diagonal dominance or condition number factors of certain matrices. The link between input/output stability and internal stability is clarified through the concept of dissipativity.

### 5.1. Equivalent I/O Representations for NL$_q$s

LEMMA 4. *Equivalent input/output representations for the NL$_q$ system eqn (8) are*

$$\begin{cases} p_{k+1} &= \prod_{i=1}^{q} \Gamma_{i,e}(p_k, w_k) M_i \begin{bmatrix} p_k \\ w_k \end{bmatrix} \\ e_k &= \prod_{i=1}^{q} \Lambda_{i,e}(p_k, w_k) N_i \begin{bmatrix} p_k \\ w_k \end{bmatrix} \end{cases} \tag{21}$$

*with*

$$\Gamma_{1,e} = \Gamma_1, \Gamma_{i,e} = \begin{bmatrix} \Gamma_i & 0 \\ 0 & I \end{bmatrix}, M_1 = [V_1\ B_1],$$

$$M_i = \begin{bmatrix} V_i & B_i \\ 0 & I \end{bmatrix} (i = 2, \ldots, q)$$

$$\Lambda_{1,e} = \Lambda_1, \Lambda_{i,e} = \begin{bmatrix} \Lambda_i & 0 \\ 0 & I \end{bmatrix}, N_1 = [W_1\ D_1],$$

$$N_i = \begin{bmatrix} W_i & D_i \\ 0 & I \end{bmatrix} (i = 2, \ldots, q)$$

*and*

$$\begin{bmatrix} p_{k+1} \\ e_k^{ext} \end{bmatrix} = \prod_{i=1}^{q} \Omega_i(p_k, w_k) R_i \begin{bmatrix} p_k \\ w_k \end{bmatrix} \tag{22}$$

*with*

$$\Omega_i = \begin{bmatrix} \Gamma_{i,e} & 0 & 0 \\ 0 & \Lambda_{i,e} & 0 \\ 0 & 0 & 0 \end{bmatrix}, R_i = \begin{bmatrix} M_i & 0 & 0 \\ 0 & N_i & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$R_q = \begin{bmatrix} M_q \\ N_q \\ 0 \end{bmatrix} (i = 1, \ldots, q-1).$$

$e_k^{ext}$ is defined such that its dimension is equal to the dimension of $w_k$($e_k^{ext}$ corresponds to $e_k$, augmented with zero elements), in order to make the matrix $\Pi_{i=1}^{q} R_i$ square (Fig. 4).

*Proof.* Straightforward calculation. Remark that $\|\Omega_i\| \leq 1$ because $\|\Gamma_i\| \leq 1$ and $\|\Lambda_i\| \leq 1$. □

In order to fix the ideas, Lemma 4 is illustrated for $q = 2$, which states that the following representations are equivalent:

$$\begin{cases} p_{k+1} &= \Gamma_1(V_1 \Gamma_2(V_2 p_k + B_2 w_k) + B_1 w_k) \\ e_k &= \Lambda_1(W_1 \Lambda_2(W_2 p_k + D_2 w_k) + D_1 w_k) \end{cases}$$

with

$$\begin{cases} p_{k+1} &= \Gamma_1 [V_1\ B_1] \begin{bmatrix} \Gamma_2 & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} V_2 & B_2 \\ 0 & I \end{bmatrix} \begin{bmatrix} p_k \\ w_k \end{bmatrix} \\ e_k &= \Lambda_1 [W_1\ D_1] \begin{bmatrix} \Lambda_2 & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} W_2 & D_2 \\ 0 & I \end{bmatrix} \begin{bmatrix} p_k \\ w_k \end{bmatrix} \end{cases}$$

and

$$\begin{bmatrix} p_{k+1} \\ e_k^{ext} \end{bmatrix} = \begin{bmatrix} \Gamma_1 & & \\ & \Lambda_1 & \\ & & 0 \end{bmatrix} \begin{bmatrix} V_1 & B_1 & & & \\ & & W_1 & D_1 & \\ & & & & 0 \end{bmatrix} \begin{bmatrix} \Gamma_2 & & & & \\ & I & & & \\ & & \Lambda_2 & & \\ & & & I & \\ & & & & 0 \end{bmatrix} \begin{bmatrix} V_2 & B_2 \\ & I \\ W_2 & D_2 \\ & I \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p_k \\ w_k \end{bmatrix}.$$

### 5.2. Main Theorems

The following Theorems will give conditions for input/output stability of NL$_q$ systems.

THEOREM 4 (I/O stability-diagonal scaling). *Given the* NL$_q$ *representation eqn (22), let*

$$R_{tot} \begin{bmatrix} 0 & R_2 & & & 0 \\ & 0 & R_3 & & \\ & & \ddots & & \\ & & & 0 & R_q \\ R_1 & & & & 0 \end{bmatrix},$$

$$R_i \in \mathbb{R}^{n_{r_i} \times n_{r_{i+1}}}, n_{r_1} = n_{r_{q+1}} = n_p + n_w.$$

*If there exists a matrix $D_{tot}$ such that*

$$\|D_{tot} R_{tot} D_{tot}^{-1}\|_2^q = \beta_D < 1, \tag{23}$$

*then the* NL$_q$ *system is input/output stable in the sense that there exist positive constants $c_1$, $c_2$ such that*

$$c_2(1 - \beta_D^2)\|p\|_2^2 + \|e\|_2^2 \leq \beta_D^2 \|w\|_2^2 + c_1 \|p_0\|_2^2 \tag{24}$$

*provided that $\{w_k\}_{k=0}^{\infty} \in l_2$*[1]*. The matrix $D_{tot}$ corresponds to $D_{tot} = \mathrm{diag}\{D_2, D_3, \ldots, D_q, D_{s_1}\}$ with $D_{s_1} = \mathrm{diag}\{D_1, I_{n_w}\}$ and $D_1 \in \mathbb{R}^{np \times np}, D_i \in \mathbb{R}^{n_{r_i} \times n_{r_i}}$ diagonal matrices with nonzero diagonal elements.*

[1] $l_2^n$ denotes the set of square summable sequences in $\mathbb{C}^n$. The $l_2$ norm of a sequence $e \in l_2^n$, denoted as $\|e\|_2$, is defined as $\|e\|_2^2 = \sum_{k=1}^{\infty} \|e_k\|_2^2$

*Proof.* See Appendix A. The proof is similar to the proof given in Packard and Doyle (1993) related to the state space upper bound test for the robust performance problem in $\mu$ theory. □

THEOREM 5 (I/O stability-diagonal dominance). *Given the representation eqn (22), if there exist matrices $P_i$, $N_i$ such that*

$$c_\alpha \beta_P < 1 \tag{25}$$

*with $c_\alpha = \prod_{i=1}^{q}(1+\alpha_i)^{1/2}$ and $\beta_P = \|P_{tot} R_{tot} P_{tot}^{-1}\|_2^q$, then there exist positive constants $c_1$, $c_2$ such that*

$$c_2(1 - c_\alpha^2 \beta_P^2)\|p\|_2^2 + \|e\|_2^2 \le c_\alpha^2 \beta_P^2 \|w\|_2^2 + c_1 \|p_0\|_2^2 \tag{26}$$

*provided that $\{w_k\}_{k=0}^{\infty} \in l_2$. One has $P_{tot} = \text{blockdiag}\{P_2, P_3, \ldots, P_q, P_{S_1}\}$ with $P_{S_1} = \text{blockdiag}\{P_1, I_{n_w}\}$ and $Q_i = P_i^T P_i N_i$ diagonal dominant matrices with $\delta_{Q_i} = (1+\alpha_i)/\alpha_i \ge 1, (\alpha_i \ge 0)$. $N_i$ are diagonal matrices with positive diagonal elements and $P_i$ are of full rank.*

*Proof.* See Appendix A. □

The following theorem gives a condition for I/O stability in terms of condition numbers of certain matrices.

THEOREM 6 (I/O stability-condition number factor). *Given the representation eqn (22), if there exist matrices $P_i$ such that*

$$c_P \beta_P < 1 \tag{27}$$

*with $c_p = \kappa(P_{S_1})\Pi_{i=2}^{q}\kappa(P_i)$ and $\beta_p = \|P_{tot} R_{tot} P_{tot}^{-1}\|_2^q$ then there exist positive constants $c_1$, $c_2$ such that*

$$c_2(1 - c_P^2 \beta_P^2)\|p\|_2^2 + \|e\|_2^2 \le c_p^2 \beta_P^2 \|w\|_2^2 + c_1 \|p_0\|_2^2 \tag{28}$$

*provided that $\{w_k\}_{k=0}^{\infty} \in l_2$. Here $P_{tot} = \text{blockdiag}\{P_2, P_3, \ldots, P_q, P_{S_1}\}$ with $P_{S_1} = \text{blockdiag}\{P_1, I_{n_w}\}$ and $P_i$ are full rank matrices.*

*Proof.* See Appendix A.

REMARK 8. *For a linear control system $\Xi_0^0$ Theorems 5 and 6 are related to the $H_\infty$ control problem with dynamic output feedback (see, for example, Stoorvogel, 1992). In that case $\Omega_1 = I$ and then it follows immediately from the proof of the Theorems that conditions eqns (25) and (27) become*

$$\left\| \begin{bmatrix} P_1 & 0 \\ 0 & I \end{bmatrix} R_1 \begin{bmatrix} P_1 & 0 \\ 0 & I \end{bmatrix}^{-1} \right\|_2 < 1. \tag{29}$$

*This condition corresponds to the $H_\infty$ norm of a linear dynamical system (see, e.g., Packard & Doyle, 1993).*

*A comparison between Theorems 1 & 4, 2 & 5 and 3 & 6 shows that there is a close connection between internal stability (autonomous case) and the property of finite $L_2$-gain. The latter becomes clear through the concept of dissipativity. This was already stated, e.g., in Hill and Moylan (1980) and Willems (1972).*

DEFINITION 6 (Dissipativity). *The dynamic system eqn (8) with input $w_k$ and output $e_k$ and state vector $p_k$ is called dissipative if there exists a nonnegative function $V(p) : \mathbb{R}^{n_p} \to \mathbb{R}$ with $V(0) = 0$, called the storage function, such that $\forall w \in \mathbb{R}^{n_w}$ and $\forall k \ge 0$:*

$$V(p_{k+1}) - V(p_k) \le W(e_k, w_k) \tag{30}$$

*where $W(e_k, w_k)$ is called the supply rate.*

LEMMA 5. *The $NL_q$ system eqn (8) is dissipative under the condition of Theorem 4, with storage function $V(p) = \|D_1 p\|_2^2$, supply rate $W(e_k, w_k) = \beta_D^2 \|w_k\|_2^2 - \|e_k\|_2^2$ with finite $L_2$-gain $\beta_D < 1$ or under the condition of Theorem 5 with storage function $V(p) = \|P_1 p\|_2^2$, supply rate $W(e_k, w_k) = c_\alpha^2 \beta_P^2 \|w_k\|_2^2 - \|e_k\|_2^2$ with finite $L_2$-gain $c_\alpha \beta_P < 1$ or finally under the condition of Theorem 6 with storage function $V(p) = \|P_1 p\|_2^2$, supply rate $W(e_k, w_k) = c_P^2 \beta_P^2 \|w_k\|_2^2 - \|e_k\|_2^2$ with finite $L_2$-gain $c_P \beta_P < 1$.*

*Proof.* See Appendix A. □

REMARK 9. *Several kinds of dissipativity exist in general, depending on the type of supply rate. A system with supply rate $W = e_k^T w_k$ is called passive, while a supply rate $W = \gamma^2 \|w_k\|_2^2 - \|e_k\|_2^2$ stands for finite $L_2 - \text{gain}\ \gamma$. (see Hill & Moylan, 1976, 1977, 1980).*

## 6. ROBUST PERFORMANCE PROBLEM

In this section we consider $NL_q$ systems with parametric uncertainties upon the system matrices, e.g., due to parametric uncertainties on the neural network model or controller. for such perturbed $NL_q$s, conditions for robust stability and robust performance are derived. The main goal of this Section is however to show the precise links between $NL_q$ theory and $\mu$ robust control theory (see, e.g., Maciejowski, 1989 and Packard & Doyle, 1993).

### 6.1. Perturbed $NL_q$s

The sufficient conditions for I/O stability with finite $L_2$-gain of $NL_q$ systems will be extended here with respect to parametric uncertainties on the system matrices (Fig. 6). The motivation for assuming such uncertainties is to study the influence of changes around the nominal values of the weights on the stability and the $L_2$-gain of the system.

Assume a perturbation of the form $R_l = R_l^{nom} + \Delta R_l$ on the $NL_q$ eqn (22)

$$\begin{bmatrix} p_{k+1} \\ e_k^{ext} \end{bmatrix} = \left(\prod_{i=1}^{l-1} \Omega_i R_i\right) \Omega_l (R_l^{nom} + \Delta R_l) \left(\prod_{j=l+1}^{q} \Omega_j R_j\right) \begin{bmatrix} p_k \\ w_k \end{bmatrix}. \tag{31}$$

Suppose that this perturbation on $R_l$ is through one of the

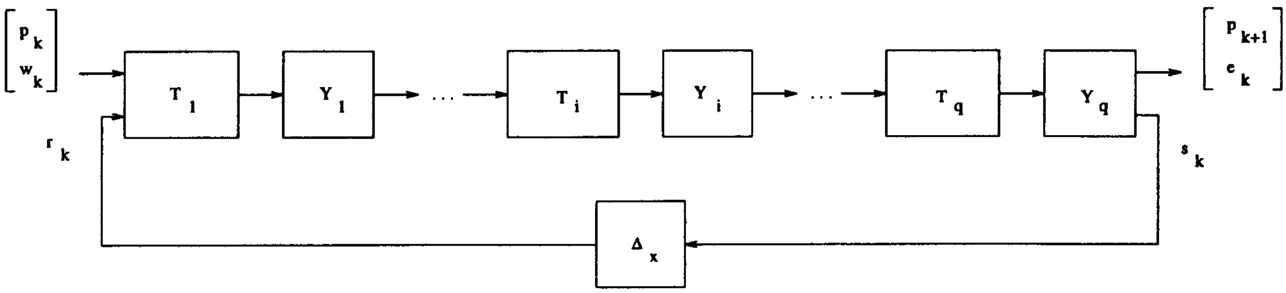


**FIGURE 6. Perturbed $NL_q$s: this representation takes into account parametric uncertainties on a certain matrix $R_l(l \in \{1,..,q\})$ through a matrix $X$ on which $R_l$ depends.**

interconnection matrices $X \in \mathbb{R}^{n \times m}$ and that the elements $x_{ij}$ belong to bounded intervals: $x_{ij} \in [\underline{x}_{ij}\bar{x}_{ij}]$. According to Steinbuch et al. (1992), by defining

$$x_{ij}^{nom} = (\bar{x}_{ij} + \underline{x}_{ij})/2$$

$$r_{ij} = (\bar{x}_{ij} - \underline{x}_{ij})/2$$

one obtains

$$x_{ij} = x_{ij}^{nom} + \delta_{ij} r_{ij}, \delta_{ij} \in [-1, 1].$$

For $X = X^{nom} + \Delta X$ one has

$$\Delta X = S_x \Delta_x R_x$$

where $\Delta_x \in \mathbb{R}^{nm \times nm}$ is a real diagonal matrix satisfying the property $\|\Delta_x\| \leq 1$ and

$$\Delta_x = \mathrm{diag}\{\delta_{11}, \ldots, \delta_{1m}, \delta_{21}, \ldots, \delta_{2m}, \delta_{n1}, \ldots, \delta_{nm}\}$$

$$S_x = \mathrm{blockdiag}\{1_{1\times m}, \ldots, 1_{1\times m}\}$$

$$R_x = [\mathrm{diag}\{r_{1i}\}; \mathrm{diag}\{r_{2i}\}; \ldots; \mathrm{diag}\{r_{ni}\}], (i = 1, \ldots, m).$$

It is supposed then that there exist matrices $U_{R_l}, V_{R_l}$, such that $R_l(X)$ with a perturbed $X$ can be written as

$$R_l(X) = R_l^{nom} + U_{R_l} \Delta_x V_{R_l}.$$

The perturbed system eqn (31) can then be written as (Fig. 6)

$$\left\{ \begin{array}{lcl} \begin{bmatrix} p_{k+1} \\ e_k^{ext} \\ s_k \end{bmatrix} & = & (\prod_{i=1}^{q} \Upsilon_i(T_i)) \begin{bmatrix} p_k \\ w_k \\ r_k \end{bmatrix} \\ r_k & = & \Delta_x s_k, \quad \|\Delta_x\| \leq 1 \end{array} \right. \tag{32}$$

with

$$\Upsilon_i = \begin{bmatrix} \Omega_i & \\ & I \end{bmatrix}, T_i = \begin{bmatrix} R_i & \\ & I \end{bmatrix} (i \neq l), T_l = \begin{bmatrix} R_l^{nom} & U_{R_l} \\ V_{R_l} & 0 \end{bmatrix}$$

where $\|\Upsilon_i\| \leq 1 (i = 1, \ldots, q)$. As stated in the following Lemma, such a perturbed $NL_q$ is nothing else but an $NL_{q+1}$.

LEMMA 6 (perturbed $NL_q$s as $NL_{q+1}$s). *The perturbed $NL_q$ eqn (32) can be written as an $NL_{q+1}$ system with $T_{q+1} = I$ and $\Upsilon_{q+1} = \mathrm{diag}\{I, I, \Delta_x\}$ as*

$$\begin{bmatrix} p_{k+1} \\ e_k^{ext} \\ s_k \end{bmatrix} = (\prod_{i=1}^{q} \Upsilon_i T_i \begin{bmatrix} I & & \\ & I & \\ & & \Delta_x \end{bmatrix} I) \begin{bmatrix} p_k \\ w_k \\ s_k \end{bmatrix} \tag{33}$$

*with $\Delta_x$ diagonal and $\|\Delta_x\|_2 \leq 1$.*

*Proof.* Eliminate $r_k$ in eqn (32). □

The following example illustrates the perturbed $NL_q$ eqn (32).

EXAMPLE 3. *The $\Xi_1^0$ system (autonomous case) with uncertainties on the elements of $C (C = C^{nom} + \Delta C$ with $\Delta C = S_c \Delta_c R_c)$ is considered:*

$$\begin{bmatrix} x_{k+1} \\ z_{k+1} \\ \xi_{k+1} \end{bmatrix} = \Gamma_1 \begin{bmatrix} A & 0 & B \\ F(C+\Delta C) & E & FD \\ GF(C+\Delta C) & GE & GFD \end{bmatrix} \begin{bmatrix} x_k \\ z_k \\ \xi_k \end{bmatrix}$$

*or in the form eqn (32)*

$$\left\{ \begin{array}{lcl} \begin{bmatrix} x_{k+1} \\ z_{k+1} \\ \xi_{k+1} \\ s_k \end{bmatrix} & = & \begin{bmatrix} \Gamma_1 & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 & B & 0 \\ FC & E & FD & FS_c \\ GFC & GE & GFD & GFS_c \\ R_c & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ z_k \\ \xi_k \\ r_k \end{bmatrix} \\ r_k & = & \Delta_c s_k, \quad \|\Delta_c\| \leq 1 \end{array} \right.$$

*with*

$$U_{R_l} = \begin{bmatrix} 0 \\ FS_c \\ GFS_c \end{bmatrix}, V_{R_l} = [R_c \; 0 \; 0] \; (l = 1).$$

*This is easily verified by straightforward calculation.*

The following Theorem holds then for input/output stability of the perturbed $NL_q$.

THEOREM 7 (Robust performance). *Given the perturbed $NL_q$ system eqn (32), let*

$$T_{tot} = \begin{bmatrix} 0 & T_2 & & & 0 \\ & 0 & T_3 & & \\ & & \ddots & & \\ & & & 0 & T_q \\ T_1 & & & & 0 \end{bmatrix}, T_i \in \mathbb{R}^{n_{t_i} \times n_{t_{i+1}}}, n_{t_1} = n_{t_{q+1}}.$$

*If there exist matrices $P_i$, $N_i$ such that*

$$c_\alpha \beta_P < 1 \quad (34)$$

*with $c_\alpha = \Pi_{i=1}^q (1+\alpha_i)^{1/2}$ and $\beta_P = \|P_{tot} T_{tot} P_{tot}^{-1}\|_2^q$ then the perturbed $NL_q$ system is I/O stable for all real diagonal matrices $\Delta_x$ that satisfy the condition $\|\Delta_x\|_2 \leq 1$:*

$$\|e\|_2^2 \leq c_\alpha^2 \beta_P^2 \|w\|_2^2 + c_1 \|p_0\|_2^2 \quad (35)$$

*provided that $\{w_k\}_{k=0}^\infty \in l_2$. Here $P_{tot} = \text{blockdiag}\{P_2, P_3 \ldots, P_q, P_{S1}\}$ with $P_{S1} = \text{blockdiag}\{P_1, I, D_1\}$. $Q_i = P_i^T P_i N_i$ are diagonally dominant matrices with $\delta_{Qi} = (1+\alpha_i)/\alpha_i \geq 1 (\alpha_i \geq 0)$. $N_i$ are diagonal matrices with positive diagonal elements and $P_i$ of full rank. $D_1$ is a diagonal matrix with nonzero diagonal elements.*

*Also, if there exist a matrix $P_{S1}$ and matrices $P_i$ ($i$ = 2,..., $q$)* such that

$$c_P \beta_P < 1 \quad (36)$$

with $c_P = \kappa(P_{S_1}) \prod_{i=2}^q \kappa(P_i)$, then

$$\|e\|_2^2 \leq c_P^2 \beta_P^2 \|w\|_2^2 + c_1 \|p_0\|_2^2 \quad (37)$$

holds for all real diagonal matrices $\Delta_x$ that satisfy $\|\Delta_x\|_2 \leq 1$.

Proof: see Appendix A. The outline of the proof is similar to the proof of Theorems 5 and 6. □

The problem of the influence of a combination of uncertainties on several matrices $X_i$ (instead of one single $X$) is not investigated in this paper.

## 6.2. Connections ith $\mu$ Theory

In $\mu$ theory (see, e.g., Packard & Doyle, 1993; Maciejowski, 1989) one analyses the influence of parametric uncertainties dynamics etc. related to a nominal linear model on the system's performance or one takes into account these uncertainties in the controller synthesis problem. Uncertainty is formulated here by means of linear fractional transformations (LFTs) (Fig. 7). Related

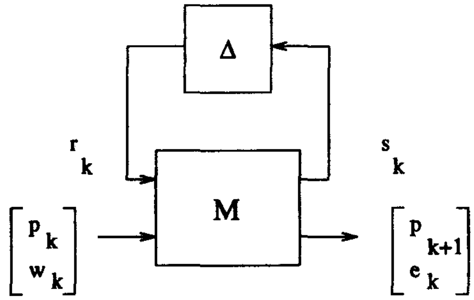


**FIGURE 7. LFT representation as it occurs in robust control theory ($\mu$ theory). $M$ is the augmented plant (nominal system) and the uncertainty on $M$ is represented through the feedback perturbation $\Delta$.**

to the standard plant, a state space formulation for an LFT looks like (Fig. 7)

$$\begin{cases} \begin{bmatrix} x_{k+1} \\ e_k^{ext} \\ s_k \end{bmatrix} = M \begin{bmatrix} x_k \\ w_k \\ r_k \end{bmatrix} \\ r_k = \Delta \, s_k \end{cases} \quad (38)$$

with $M = [M_{11} M_{12} M_{13}; M_{21} M_{22} M_{23}; M_{31} M_{32} M_{33}]$. In general a block uncertainty structure is considered for $\Delta \in \boldsymbol{\Delta} \subset \mathbb{C}^{n_{tot} \times n_{tot}}$ (see Packard & Doyle, 1993)

$$\boldsymbol{\Delta} = \{\text{diag}[\delta_1 I_{r_1}, \ldots, \delta_s I_{r_s}, \Delta_{s+1}, \ldots, \Delta_{s+f}] : \delta_i \in \mathbb{C}, \Delta_{s+j} \in \mathbb{C}^{m_j \times m_j}; i = 1, \ldots, s; j = 1, \ldots f\} \quad (39)$$

where $\Sigma_{i=1}^s r_i + \Sigma_{j=1}^f m_j = n_{tot}$ for consistency. Note that these uncertainty blocks are not necessarily real, but can have complex values. The fact that these are real or complex depends on the specific problem that one studies: parametric uncertainties are real in nature, while unmodeled dynamics may lead to complex values. The uncertainties can be of time-invariant, time-variant or nonlinear nature. For linear time-invariant uncertainties $\Delta$ a necessary and sufficient condition for robust performance (finite $L_2$-gain smaller than one, regardless of any $\Delta$ belonging to $\boldsymbol{B}_\Delta = \{\Delta \in \boldsymbol{\Delta} : \|\Delta\|_2 \leq 1\}$):

$$\|e\|_2 \leq \beta \, \|w\|_2 \quad \beta \in [0, 1) \quad (40)$$

for zero-initial-state-response is

$$\mu_{\Delta_S}(M) < 1 \quad (41)$$

where $\Delta_S$ is an augmented block structure for $\Delta$ and $\mu_{\Delta_S}(M)$ denotes the *structured* singular value of the matrix $M$ related to the uncertainty block structure $\Delta_S$. Some properties of $\mu$ are

$$\rho(M) \leq \max_{Q \in \mathcal{Q}} \rho(QM) \leq \max_{\Delta \in \boldsymbol{B}_\Delta} \rho(\Delta M) = \mu_\Delta(M) \leq \min_{D \in \mathcal{D}} \|DMD^{-1}\|_2 \leq \|M\|_2 \quad (42)$$

where $\mathcal{Q}=\{Q\in\Delta : Q*Q=I_n\}$, $\mathcal{D}=\{\mathrm{diag}[D_1,\ldots,D_s, d_{s+1}I_{m1},\ldots,d_{s+f}I_{m_f}] : D_i\in\mathbb{C}^{r_i\times r_i}, \quad D_i=D_i^*>0, d_{s+j}\in\mathbb{R},\ d_{s+j}>0\}$.

The upper bound of $\mu$ corresponds to diagonal scaling of $M$, which is similar to the diagonal scaling in Theorem 1 or Theorem 4. Working with this upper bound instead of $\mu$, also guarantees robust performance with respect to uncertainties of nonlinear and time-varying nature (sufficient but not necessary condition in that case but might be too conservative). The following Theorem is due to Packard and Doyle.

THEOREM 8 (Packard & Doyle). *Given the LFT eqn (38), if*

$$\min_{D_S}\|D_S M D_S^{-1}\|_2=\beta<1 \tag{43}$$

*with $D_S=\mathrm{diag}\{D_1,d_2I,D\}$ and $D_1, D$ diagonal matrices with positive diagonal elements and $d_2$ a positive constant, then the following holds for zero initial state*

$$\|e\|_2^2\le\beta^2\|w\|_2^2 \tag{44}$$

*if $\{w_k\}_{k=0}^{\infty}\in l_2$.*

*Proof.* see Packard and Doyle (1993). □

The following theorem states that for a real diagonal $\Delta$ uncertainty block the Theorem of Packard and Doyle is a special case of Theorem 7.

THEOREM 9 (connection between theorems 7 and 8). *The LFT eqn (38) is a special case of the perturbed* NL$_q$ *system eqn (32) under the following conditions:*

1. $q=1$ *and* $\Upsilon_1=I$;
2. $\delta_{Q_i}\to\infty$ *(Diagonal scaling case)*;
3. $M_{33}=0$;
4. *uncertainty block* $\Delta$ *is real*;
5. *diagonal* $\Delta$ *block*: $r_i=0(\forall i), m_j=1(\forall j), \quad s=0, f=n_{tot}$ *in eqn (39).*

*Proof.* Readily checked by comparing eqn (33) with eqns (38), (39) and (43). □

## 7. NEURAL CONTROL DESIGN

In the previous sections sufficient conditions for global asymptotic stability and finite $L_2$-gain of $NL_q s$ were derived. The final goal however is to design a neural controller, given an identified neural state space model. In modern *linear* control theory, controller design involves conceptually two major aspects: a first desirable property of a control system is (robust) internal stability, a second one optimal (robust) performance (see, e.g., Boyd & Barratt, 1991; Maciejowski, 1989). For linear models the class of all stabilizing controllers is characterized by the Youla parametrization, which enables searching within this class for a controller with optimal performance. Many problems in linear control theory also lead to *convex* optimization problems, that have a unique minimum that can be found in polynomial time (see Boyd et al., 1994). Hence the assumption on linearity of the plant model has the considerable advantage of keeping mathematics more tractable and to develop efficient numerical algorithms.

In *neural* control theory (indirect approach) on the other hand one is more ambitious with respect to the modeling part: a more accurate nonlinear model (emulator) is obtained that may also lead to improved tracking performance when a neural controller is trained based on this emulator. In many cases the neural controller is trained then with respect to a specific reference input. The disadvantage of such a method is that many open problems exist about stability of the closed loop system or how the controller would generalize to other reference inputs.

The present framework of NL$_q$ theory intends to be a compromise between those two extremes: the neural controller can be trained with respect to a given reference input like in classical neural control using, e.g., Narendra's sensitivity model approach (dynamic backpropagation), but in addition it is also possible to check for global asymptotic stability of this learned controller or to see how the neural control system reacts upon any reference input belonging to the class $l_2$ (as is done in $H_\infty$ control theory where one considers an input signal belonging to a certain class of signals $l_2$). This will be called here the analysis problem. For the synthesis problem a modified dynamic backpropagation algorithm will be proposed where the neural controller learns to track a given reference input, under the constraint that the closed loop system has to be globally asymptotically stable or is dissipative with finite $L_2$-gain smaller than 1. Another method is also proposed where the controller is trained based on the stability criteria only, without tracking on specific reference inputs.

### 7.1. Tracking of a Specific Reference Input

As for the learning algorithm related to the neural state space models $\mathcal{M}_i$ in Section Section 2, Narendra's sensitivity model approach in dynamic backpropagation can also be applied in order to generate the gradient of the cost function in the following tracking problem

$$\min_{\theta_c} J(\theta_c)=\frac{1}{2N}\sum_{k=1}^{N}\{[d_k-\hat{y}_k(\theta_c)]^T[d_k-\hat{y}_k(\theta_c)] + \lambda.u_k(\theta_c)^T u_k(\theta_c)\} \tag{45}$$

where $d_k$ is the specific reference input and $\hat{y}_k$, $u_k$ are the output of the neural state space model and the controller respectively. The parameter vector $\theta_c\in\mathbb{R}^{pc}$ contains the interconnection weights of the neural controller. $\lambda$ is a given positive constant and $N$ is the time horizon. Suppose the closed loop system for a given $\Xi_j^i$ problem of Fig. 1 is expressed as

$$\begin{cases} p_{k+1} &= \zeta(p_k, d_k; \alpha) \\ \hat{y}_k &= \chi(p_k, d_k; \beta) \\ u_k &= \vartheta(p_k, d_k; \gamma) \end{cases} \tag{46}$$

instead of in the standard plant form of Fig. 3. Here $\alpha$, $\beta$, $\gamma$ are elements of $\theta_c$. It is also possible to let $d_k$ be the output of a reference model. In that case the state vector $p_k$ of the closed loop model eqn (46) is augmented with the state of the reference model. A gradient based optimization scheme such as a steepest descent, conjugate gradient or quasi-Newton method then makes use of the gradient

$$\frac{\partial J}{\partial \theta_c} = \frac{1}{N}\sum_{k=1}^{N} \{[d_k - \hat{y}_k(\theta_c)]^T(-\frac{\partial \hat{y}_k}{\partial \theta_c}) + \lambda.u_k(\theta_c)^T\frac{\partial u_k}{\partial \theta_c}\} \tag{47}$$

where the gradients $\partial \hat{y}_k/\partial \theta_c$ and $\partial u_k/\partial \theta_c$ are the output of the sensitivity model

$$\begin{cases} \frac{\partial p_{k+1}}{\partial \alpha} &= \frac{\partial \zeta}{\partial p_k}.\frac{\partial p_k}{\partial \alpha} + \frac{\partial \zeta}{\partial \alpha} \\ \frac{\partial \hat{y}_k}{\partial \alpha} &= \frac{\partial \chi}{\partial p_k}.\frac{\partial p_k}{\partial \alpha} \\ \frac{\partial \hat{y}_k}{\partial \beta} &= \frac{\partial \chi}{\partial \beta} \\ \frac{\partial u_k}{\partial \gamma} &= \frac{\partial \vartheta}{\partial \gamma}, \end{cases} \tag{48}$$

which is a dynamical system with state vector $\partial p_k/\partial \alpha$ driven by $\partial \zeta/\partial \alpha$, $\partial \chi/\partial \beta$ and $\partial \vartheta/\partial \gamma$ (see Narendra & Parthasarathy, 1991).

### 7.2. Analysis Problem

7.2.1. *Problem Statement.* Suppose a neural controller in $\Xi_j^i$ has been trained to track a specific reference input according to Section 7.1, its properties can then be analysed according to the previous theorems. The following conditions are sufficient for global asymptotic stability, finite $L_2$-gain smaller than 1 and robust performance:

- *Diagonal scaling*: find diagonal matrices $D_i$ such that

$$\|D_{tot}Z_{tot}D_{tot}^{-1}\|_2^q < 1. \tag{49}$$

- *Diagonal dominance*: find matrices $P_i$ and diagonal matrices $N_i$ such that

$$c_\alpha \|P_{tot}Z_{tot}P_{tot}^{-1}\|_2^q < 1 \tag{50}$$

and

$$c_\alpha = \prod_{i=1}^{q}(1+\alpha_i)^{1/2}, \alpha_i = \frac{1}{\delta_{Q_i}-1}, Q_i = P_i^T P_i N_i.$$

- *Condition number factor*: find matrices $P_i$ such that

$$c_P \|P_{tot}Z_{tot}P_{tot}^{-1}\|_2^q < 1 \tag{51}$$

with

$$c_P = \kappa(P_{S_1})\prod_{i=2}^{q}\kappa(P_i).$$

- *Choice of $Z_{tot}, D_{tot}, P_{tot}$:*

  $Z_{tot} = V_{tot}$: global asymptotic stability
  $= R_{tot}$: finite $L_2$-gain
  $= T_{tot}$: robust performance

  and $D_{tot} = \mathrm{diag}\{D_2, D_3, \ldots, D_q, D_{S1}\}$ with

  $D_{S_1} = D_1$: global asymptotic stability
  $=$ diag $\{D_1, I\}$: finite $L_2$-gain
  $=$ diag $\{D_{1_a}, I, D_{1_b}\}$: robust performance

  and $P_{tot} = \mathrm{blockdiag}\{P_2, P_3, \ldots, P_q, P_{S_1}\}$ with

  $P_{S_1} = P_1$: global asymptotic stability
  $=$ blockdiag $\{P_1, I\}$: finite $L_2$-gain
  $=$ blockdiag $\{P_1, I, D_1\}$: robust performance

7.2.2. *Formulation as LMI Problem.* The feasibility problems eqns (49)–(51) correspond to non-differentiable optimization problems that involve matrix inequalities of the form $M(x) < 0$ with $M = M^T$ with unknown $x \in \mathbb{R}^n$ because an expression like $\|A\|_2 < \gamma$ is equivalent to $A^TA - \gamma I < 0$[2]. For the special case of linear matrix inequalities (LMIs) this matrix $M$ depends affinely on $x: M(x) = M_0 + \Sigma_{i=1}^m x_i M_i$ with $M_0 = M_0^T, M_i = M_i^T$. A set of LMIs $M_l(x) < 0 (l = 1, \ldots, L)$ is equivalent to a single LMI by considering $M = \mathrm{blockdiag}\{M_1, \ldots, M_L\}$ (see Boyd et al., 1994). Finding a feasible $x$ for such an LMI is a convex problem, which has a unique minimum. LMIs occur frequently in systems and control problems (Boyd et al., 1994). From a computational point of view LMIs are attractive because efficient algorithms exist for solving it. A general theory of interior-point polynomial-time methods for convex programming is presented in Nesterov and Nemirovskii (1994). Other related work is, e.g., by Vandenberghe and Boyd (1995), and Overton (1988). Software for solving LMIs is available in Matlab's LMI lab (see Gahinet & Nemirovskii, 1993). The problems eqns (49)–(51) will now be discussed in more detail:

1. Diagonal Scaling: The LMI corresponding to eqn (49) is

$$Z_{tot}^T D_{tot}^2 Z_{tot} < D_{tot}^2 \tag{52}$$

[2] The notation $A < 0$ means $A$ negative definte and $A < B$ means $A - B$ negative definite, for real symmetric matrices $A$,$B$.

and can be handled in LMI lab. The function *psv* of Matlab's Robust Control Toolbox (The MathWorks Inc., 1994) computes $D_{tot}\|D_{tot}Z_{tot}D_{tot}^{-1}\|_2$ for the case $D_{S_1}=D_1$ (global asymptotic stability).

2. Diagonal dominance: A set of LMIs will be associated here to the diagonal dominance case eqn (50). In the sequel we have to assume $N_i=I$ for $i=1,\dots,q$. The condition is formulated then as a convex feasibility problem in the unknown matrices $Q_i, Q_{S_1}$
$$Z_{tot}^T Q_{tot} Z_{tot} < \gamma_1 Q_{tot} \tag{53}$$
with $Q_{tot}=P_{tot}^T P_{tot}=\text{blockdiag}\{Q_2,\dots,Q_q,Q_{S_1}\}$, $Q_i=P_i^T P_i, Q_{S_1}=P_{S_1}^T P_{S_1}$, such that
$$\delta_{Q_i} > \gamma_{2_i}(\forall i|\Gamma_i \neq I),$$
where $\gamma_1 \le 1, \gamma_{2_i} \ge 1$ are user defined constants. In order to obtain a set of LMIs the constraint for diagonal dominance is replaced by
$$\begin{cases} Q_i-\gamma_{3_i}I & > & 0\ (\forall i|\Gamma_i \neq I) \\ \|Q_i\|_2 & < & \gamma_{4_i}(\forall i|\Gamma_i \neq I) \\ Q_i & > & 0, (\forall i) \end{cases} \tag{54}$$
and the Schur complement form is used for $\|Q_i\|_2 < \gamma_{4_i}$ which is
$$\begin{bmatrix} \gamma_{4_i} & Q_i \\ Q_i^T & \gamma_{4_i} I \end{bmatrix} > 0.$$
A qualitative explanation why the set of LMIs eqns (53) and (54) enforces the matrices $Q_i$ to be diagonally dominant can be given based on Gerschgorin's Theorem and Lemma 3. Let us fix the ideas for $q=1$. Gerschgorin's Theorem applied to the matrix $Q-\gamma_3 I$ states that the discs, containing the eigenvalues of the matrix, are centered at $q_{ii}-\gamma_3$ with radii $\Sigma_{j(j\neq i)}|q_{ij}|$. Now it is clear that for $\gamma_3>0$ and if $Q-\gamma_3 I>0$ holds an increasing value $\gamma_3$ enforces $Q$ to be more diagonally dominant. The introduction of an upper bound on $\|Q\|_2$ follows from Lemma 3. Indeed because and $Q=D_Q(I+X_Q)$ and $\|X_Q\|_2 \le \sqrt{n}\,\|X_Q\|_\infty$ (assuming $Q\in\mathbb{R}^{n\times n}$) and $\|X_Q\|_\infty < 1/\delta_Q$) one obtains
$$\|Q\|_2 < \|D_Q\|_2(1+\sqrt{n}/\delta_Q)$$
which means that some upper bound $\gamma_4$ on $\|Q\|_2$ is needed. The choice of $\gamma_3$ influences $\|D_Q\|_2$. One may use the following rule of thumb: choose first $\gamma_3$ and secondly $\gamma_4$ proportionally to $\gamma_3$ (e.g. $\gamma_4=10\gamma_3$). This was affirmed by computer simulations using LMI lab, by trying several combinations of $\gamma_1,\gamma_3,\gamma_4$ on random matrices $Z_{tot}$.

3. Condition number factor: In this case it is easier to formulate a set of LMIs than In the diagonal dominance case. Indeed according to Boyd et al. (1994) an upper bound on the condition number of a matrix $P:\kappa(P)<\alpha$ can be formulated as the LMI
$$I<P^TP<\alpha^2 I.$$
Hence the problem eqn (51) can be interpreted as a feasibility problem in the matrices $Q_i, Q_{S_1}$
$$\begin{cases} Z_{tot}^T Q_{tot} Z_{tot} & < & \gamma^2 Q_{tot} \\ I<Q_{S_1} & < & \beta^2 I \\ I<Q_i & < & \alpha_i^2 I,\ i=2.\dots,q \end{cases} \tag{55}$$
where $Q_{tot}=P_{tot}^TP_{tot}=\text{blockdiag}\{Q_2,\dots,Q_q,Q_{S_1}\}$ and $Q_i=P_i^TP_i, Q_{S1}=P_{S1}^TP_{S1}$. In order to find a feasible point to the set of LMLs, one chooses $\gamma$ and tries to make $\beta$ and $\alpha_i$ as small as possible.

### 7.3. Synthesis Problem

7.3.1. *Problem Statement.* More difficult than the analysis problem is the synthesis problem where the controller parameter vector $\theta_c$ belongs to the unknowns of the optimization problem. In general this leads to non-convex nondifferentiable optimization problems. Two types of synthesis problems are considered.

*7.3.1.1. Type 1.*

- Diagonal scaling: feasibility problem in $\theta_c, D_i$ such that
$$Z_{tot}(\theta_c)^T D_{tot}^2 Z_{tot}(\theta_c) < D_{tot}^2. \tag{56}$$
- Diagonal dominance: feasibility problem in $\theta_c, Q_i=Q_i^T$ such that
$$\begin{cases} Z_{tot}(\theta_c)^T Q_{tot} Z_{tot}(\theta_c) < \gamma_1 Q_{tot} \\ Q_i-\gamma_{3_i} I>0 \\ Q_i>0 \\ \begin{bmatrix} \gamma_{4_i} I & Q_i \\ Q_i^T & \gamma_{4_i} I \end{bmatrix} > 0 \end{cases} \tag{57}$$
where $\gamma_1,\gamma_{3_i},\gamma_{4_i}$, are user defined constants and $Q_i>0$ must hold $\forall i$ and the other conditions $\forall i$ for which $\Gamma_i \neq I$.
- Condition number factor: feasibility problem in $\theta_c$, $Q_i=Q_i^T$ such that
$$\begin{cases} Z_{tot}(\theta_c)^T Q_{tot} Z_{tot}(\theta_c) & < & \gamma^2 Q_{tot} \\ I<Q_{S_1} & < & \beta^2 I \\ I<Q_i & < & \alpha_i^2 I, i=2,\dots,q \end{cases} \tag{58}$$
with $\alpha_i$, $\beta$, $\gamma$ user defined constants.

*7.3.1.2. Type 2.* Feasible controller $\theta_c$ with optimal tracking of a given reference input
$$\min_{\theta_c, Q_i \text{ or } D_i} J(\theta_c) \tag{59}$$
such that the type 1 condition eqn (56), eqn (57) or eqn (58) holds. The cost function $J$ corresponds to eqn (45).

7.3.2. *Non-convex Nondifferentiable Optimization.* A framework for solving non-convex nondifferentiable optimization problems with singular value inequality constraints is proposed, e.g., in Polak and Wardi (1982) in the context of semi-infinite optimization and applies to the two types of problems of Section 7.3.1. Also ellipsoid algorithms have been applied to non-convex problems, although they are normally intended for convex problems (see, e.g., Boyd & Barratt, 1991). A disadvantage of these algorithms is that they are quite slow. In general more work remains to be done on quadratically convergent algorithms for non-convex nondifferentiable optimization.

An essential aspect in non-convex problems is to replace the gradient by a *generalized gradient*. In Polak and Wardi (1982) the generalized gradient is discussed related to the problem

$$\min_{x \in \mathbb{R}^n} \lambda_{\max}[M(x)], M = M^T > 0 \tag{60}$$

where $\lambda_{\max}$ denotes the maximal eigenvalue of $M$. $M$ denotes the matrix inequality related to eqns (56)–(58). Some properties related to $\lambda_{\max}(x)$ are that this function is Lipschitz continuous and differentiable whenever $\lambda_1(x) \neq \lambda_2(x)$. In that case the gradient is equal to

$$[\nabla \lambda_{\max}[M(x)]]_i = u_1(x)^T \frac{\partial M(x)}{\partial x_i} u_1(x) \tag{61}$$

where $u_1(x)$ is the unit eigenvector corresponding to the largest eigenvalue $\lambda_1(x)$ of M(x). In the case of *linear* matrix inequalities the derivative $\frac{\partial M(x)}{\partial x_i}$ corresponds to a constant matrix. At the points where $\lambda_{\max}(x)$ is not differentiable a generalized gradient must be considered

$$\partial \lambda_{\max}[M(x)] = co\{v \in \mathbb{C}^n | v_i = (Uz)^* \frac{\partial M(x)}{\partial x_i}(Uz), \quad z \in \mathbb{C}^{k(x)}, \|z\| = 1\} \tag{62}$$

where $co$ denotes the convex hull of the set in $\{.\}$ and $k(x)$ is such that $\lambda_1(x) = \lambda_2(x) = \ldots = \lambda_{k(x)}(x) > \lambda_{k(x)+1}(x)$. One should also note that the use of a generalized gradient does not necessarily lead to a descent direction like in differentiable problems. The algorithms described in Polak and Wardi (1982) make use of the (generalized) gradients of the objective function and of the constraints. Such algorithms have been applied to the design of linear multivariable feedback systems, e.g., in Polak and Salcudean (1989).

7.3.3. *A Modified Dynamic Backpropagation Algorithm.* Instead of applying methods of non-convex nondifferentiable optimization to the general problem in $\theta_c$ and $Q_i$ or $D_i$, it is possible to exploit the fact that the Type 1 or Type 2 problem of Section 7.3.1 contains a *convex subproblem*. Without loss of generality let us consider the Type 2 problem with diagonal scaling:

$$\min_{\theta_c, D_i} J(\theta_c) \text{ such that } M(\theta_c, D_i) < 0. \tag{63}$$

It can be interpreted as consisting of two nested optimization problems: a non-convex and a convex one. For a given value of $\theta_c$, called $\theta_c^0$, the "inner" feasibility problem

$$M(\theta_c^0, D_i) < 0 \tag{64}$$

is convex. The "outer" optimization problem is then

$$\min_{\theta_c} J(\theta_c) \text{ such that } \lambda_{\max}[M(\theta_c)] < 0. \tag{65}$$

The gradient of $J(\Theta_c)$ is given by eqn (47) and the gradient for the constraint at a given point $\theta_c^0$ is given by

$$[\nabla \lambda_{\max}[M(\theta_c^0)]]_i = u_1(\theta_c^0)^T \frac{\partial M(\theta_c^0, D_i^{(*)})}{\partial \theta_{c_i}^0} u_1(\theta_c^0) \tag{66}$$

with $u_1(\theta_c^0)$ the unit eigenvector corresponding to the largest eigenvalue $\lambda_1(\theta_c^0)$ of $M(\theta_c^0, D_i^{(*)})$ and $D_i^{(*)}$ is the solution to the convex feasibility problem eqn (64).

REMARK 10.

- *It may happen that it is impossible to find a feasible point to the matrix inequality, especially in the case of diagonal scaling and the criteria with condition numbers, because the conditions might be conservative. One may not conclude then that the system is not globally asymptotically stable because the condition is only sufficient. In that case one has either to rely on the classical dynamic backpropagation algorithm or one may impose local stability of the* $\mathrm{NL}_q$ *at* $p = 0$*, which leads again to a matrix inequality constraint. A similar constraint of local stability at a target point was also successfully imposed in order to solve the swinging up problem for an inverted pendulum system (see Suykens et al., 1994). One may solve then one of the following optimization problems:*

$$\min_{\theta_c, P_{tot}} c_\alpha \text{ such that } \|P_{tot} Z_{tot} P_{tot}^{-1}\|_2 < 1 \tag{67}$$

*or*

$$\min_{\theta_c, P_{tot}} c_P \text{ such that } \|P_{tot} Z_{tot} P_{tot}^{-1}\|_2 < 1. \tag{68}$$

*For the case of internal stability* ($Z_{tot} = V_{tot}$) *this means that local stability of the origin is imposed and the basin of attraction of this equilibrium point is maximized.*

- *It is also possible to consider a modified dynamic backpropagation algorithm for the system identification problem (Section Section 2), as for the tracking problem. In that case the objective eqn (4) of the prediction error algorithm is modified with an LMI constraint that expresses global asymptotic stability of the model*

$$\min_{\theta_m, D_i} V_N(\theta_m, Z^N) \text{ such that } M(\theta_m, D_i) < 0. \tag{69}$$

### 7.4. Example 4

A simple example is given here in order to illustrate the control design. We consider the following nonlinear plant $\mathcal{P}$, neural state space model $\mathcal{M}_2$, linear dynamic controller $C_0$ a reference model $\mathcal{L}$:

$$\mathcal{P}: y_{k+1} = 0.3(1+y_k^2)y_k + 0.5u_k$$

$$\mathcal{M}_2: \begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_B u_k + \beta_{AB}) \\ y_k &= W_{CD}\tanh(V_C\hat{x}_k) + \epsilon_k \end{cases}$$

$$C_0: \begin{cases} z_{k+1} &= Ez_k + Fy_k + F_2 d_k \\ u_k &= Gz_k \end{cases}$$

$$\mathcal{L}: d_{k+1} = 0.6d_k + r_k.$$

First nonlinear system identification was done according to Section 2.1. Input/output data were generated by taking white noise for $u_k$ (uniformly distributed in the interval $[-0.5, 0.5]$). The data set consists of 2000 data points (the first 1000 data are the training set and the following 1000 data are the test set). The structure of the neural network model is $n = 1$ and seven hidden neurons for the state equation and for the output equation and $x_0 = y_0 = 0$. A quasi-Newton method with BFGS updating of the Hessian and mixed quadratic and cubic line search was applied (Fletcher, 1987; Gill et al., 1981) to eqn (4) (function *fminu* of Matlab's optimization toolbox) (The MathWorks Inc., 1994). The model and its associated sensitivity model were simulated in C code using Matlab's *cmex* facility. 50 different starting points (randomly chosen according to a normal distribution with zero mean and variance 0.1) for the parameter vector were generated. The selected solution had a fitting error of $V_{fit} = 8.3731e-05$ and a generalization error of $V_{gen} = 7.9801e-05$ and was acceptable according to correlation tests on the training data and the test data.

Then a tracking problem was specified with $r_k = 0.2(\sin 2\pi k/25) + \sin(2\pi k/10))$ for $k = 1, .., 100$. The closed loop system is an $NL_2$ system with state vector $p_k = [\hat{x}_k\, z_k\, \xi_k\, d_k]$, $\xi_k = \tanh(V_C\hat{x}_k)$, input $w_k = [r_k; \epsilon_k; 1]$

$$\begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_B G z_k + \beta_{AB}) \\ z_{k+1} &= Ez_k + FW_{CD}\xi_k + F\epsilon_k + F_2 d_k \\ \xi_{k+1} &= \tanh(V_C W_{AB}\tanh(V_A\hat{x}_k + V_B G z_k + \beta_{AB})) \\ d_{k+1} &= 0.6d_k + r_k \\ e_k &= d_k - \hat{y}_k \end{cases}$$

and the matrices $V_1$, $V_2$ of the $NL_q$ are equal to

$$V_1 = \begin{bmatrix} W_{AB} & & & \\ & I & & \\ & & V_C W_{AB} & \\ & & & 1 \end{bmatrix},$$

$$V_2 = \begin{bmatrix} V_A & V_B G & 0 & 0 \\ 0 & E & FW_{CD} & F_2 \\ V_A & V_B G & 0 & 0 \\ 0 & 0 & 0 & 0.6 \end{bmatrix}.$$

The original dynamic backpropagation algorithm eqn (45) as well as the modified dynamic backpropagation algorithm eqn (63) with global asymptotic stability of the closed loop system (diagonal scaling) were applied. A quasi-Newton method with BFGS updating of the Hessian was used for dynamic backpropagation (function *fminu* of Matlab's optimization toolbox). For the modified dynamic backpropagation algorithm the inner convex optimization problem was solved using the function *psv* of Matlab's robust control toolbox and the outer constrained optimization problem by a sequential quadratic programming (SQP) method, using the function *constr* of Matlab's optimization toolbox with numerical calculation of the gradients and simulation of the closed loop system in C code. This SQP method, which is intended for differentiable optimization problems, is meaningful here as long as the largest eigenvalue related to the LMI does not coincide with any other eigenvalue. In the experiments the order of the controller was chosen equal to 3 in all cases. The results are shown on Fig. 8. A solution to the unconstrained problem is plotted in Fig. 8. In Fig. 8b and c, a constraint of $\|D_{tot}V_{tot}D_{tot}^{-1}\|_2 < 0.99$ and $< 0.85$ was introduced respectively. Starting from the same random initial controller in the three cases the following (feasible) minima were obtained in eqns (45) and (59): $J = 0.0066$ (Fig. 8a), $J = 0.0073$ (Fig. 8b), $J = 0.0124$ (Fig. 8c). Hence the tracking performance for the specific reference input degrades by ensuring a larger contraction rate of the flow for the autonomous control system. The solution related to Fig. 8a did not result in a feasible point for eqn (52).

The stability criteria on diagonal scaling, illustrated in this example, are however mainly useful in order to control systems of which the neural state space model is globally asymptotically stable, according to the diagonal scaling criteria. The use of the other stability criteria for controlling systems with a unique equilibrium, multiple equilibria, periodic, quasi-periodic or chaos has been demonstrated in Suykens et al. (1995b) and Suykens and Vandewalle (1996a), by applying eqns (67) and (68).

## 8. CONCLUSIONS

In this paper we have proposed a modelbased neural control framework, consisting of neural state space models and controllers, with sufficient conditions for global asymptotic stability, I/O stability and robust performance. Three types of criteria were presented: diagonal scaling and criteria in terms of diagonal dominance and condition numbers of certain matrices. The criteria were formulated for $NL_q$ systems, to which closed-loop

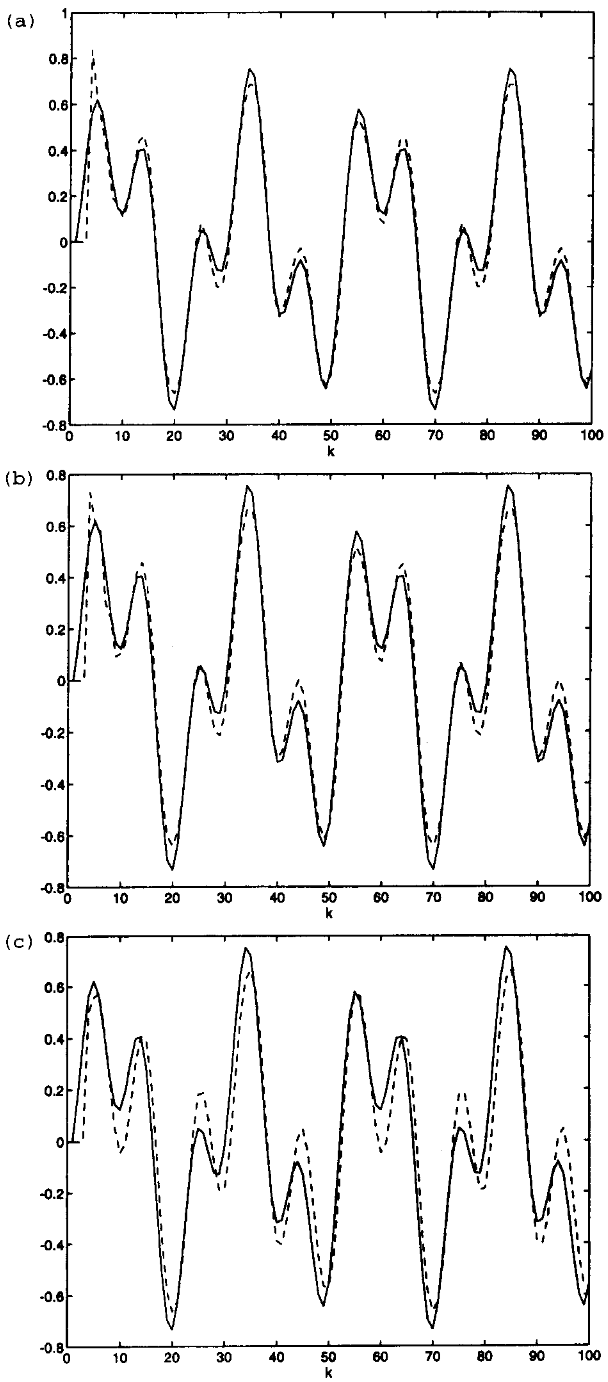


**FIGURE 8. Example of the modified dynamic backpropagation algorithm with respect to classical dynamic backpropagation. (a) Classical dynamic backpropagation: best tracking performance on a specific reference input. (b) and (c) Modified dynamic backpropagation with guaranteed global asymptotic stability of the closed loop system according to the diagonal scaling case. The tracking performance for the specific reference input degrades for decreasing upper bound on $\|D_{tot}V_{tot}D_{tot}^{-1}\|_2$: upper bounds are 0.99 and 0.85 for (b) and (c) respectively. (Full line: output of reference model, dashed line: output of controlled plant.)**

system forms have been transformed. All systems that have been considered are in state space form: in this way a given system was represented in $NL_q$ form by means of state augmentation. Vector norms as Lyapunov functions and properties of induced norms served as tools to derive the main Theorems. Furthermore, the neural controller analysis and synthesis problem were discussed, i.e. checking stability of a trained controller and designing the neural controller on the basis of the stability criteria or by means of a modified dynamic backpropagation algorithm for ensuring closed-loop stability. The emphasis in this paper was on control applications of $NL_q$ systems. Other recurrent neural network architectures such as LRGF networks (Tsoi & Back, 1994) and generalized CNNs (Guzelis & Chua, 1993) can also be represented as $NL_q$ systems (see Suykens et al., 1995b; Suykens & Vandewalle, 1996b). More examples and case studies on controlling nonlinear systems can be found in Suykens et al. (1995b).

However, open problems exist, such as, e.g., necessary and sufficient conditions for global asymptotic stability (the conditions in this paper are only sufficient) and conditions for asymptotic stability (which allows the existence of more than one equilibrium point, relevant, e.g., for associative memory applications). The present framework of $NL_q$ theory is in discrete time. A continuous time version has been studied in Suykens and Vandewalle (1996c), corresponding to a Lur'e problem with multilayer perceptron nonlinearity. Another open problem is, e.g., whether it is possible to find a parametrization of all stabilizing neural controllers for a specific $\Xi_j^i$ problem. Such a parametrization could simplify the synthesis problem. With respect to control design many criteria can be expressed as matrix inequalities, containing convex subproblems. The overall controller design problem is however non-convex and possibly non-differentiable. More progress has to be made in general in this area in order to obtain superlinear or quadratically convergent local optimization algorithms.

Hence $NL_q$ theory may contribute as a tool for the analysis and design of recurrent neural networks in modelling and control applications.

## REFERENCES

Billings, S., Jamaluddin, H., & Chen, S. (1992). Properties of neural networks with applications to modelling non-linear dynamical systems. *International Journal of Control*, **55 (1)**, 193–224.

Boyd, S., & Barratt C. (1991). *Linear controller design, limits of performance*. Englewood Cliffs, New Jersey: Prentice-Hall.

Boyd, S., El Ghaoui, L., Feron, E., & Balakrishnan, V. (1994). *Linear matrix inequalities in system and control theory*, SIAM (Studies in Applied Mathematics), Vol. 15. Philadelphia: SIAM.

Cybenko, G. (1989). Approximations by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, **2**, 183–192.

Fletcher, R. (1987). *Practical methods of optimization*, second edition, Chichester and New York: John Wiley and Sons.

Funahashi, K.-I. (1989). On the Approximate Realization of continuous Mappings by Neural Networks. *Neural Networks*, **2**, 183–192.

Gahinet, P., & Nemirovskii, A. (1993). General-Purpose LMI Solvers with Benchmarks. *Proceedings IEEE Conference on Decision and Control* (pp. 3162–3165).

Gill, P. E., Murray, W., & Wright, M. H. (1981) *Practical optimization*. London: Academic Press.

Goodwin, G., & Sin, K. (1984). *Adaptive filtering, prediction and control*. Englewood Cliffs, New Jersey: Prentice-Hall.

Guzelis, C., & Chua, L.O. (1993). Stability analysis of generalized cellular neural networks. *International Journal of Circuit Theory and Applications*, **21**, 1–33.

Hill, D.J., & Moylan, P.J. (1976). The stability of nonlinear dissipative systems. *IEEE Transactions on Automatic Control*, **21 (AC)**, 708–711.

Hill, D.J., & Moylan, P.J. (1977). Stability results for nonlinear feedback, systems. *Automatica*, **13**, 377–382.

Hill, D.J., & Moylan, P.J. (1980). Connections between finite-gain and asymptotic stability. *IEEE Transactions on Automatic Control, AC*, **25 (5)**, 931–936.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**, 359–366.

Hunt, K.J., Sbarbaro, D., Zbikowski, R., & Gawthrop, P.J. (1992). Neural networks for control systems-a survey. *Automatica*, **28 (6)**, 1083–1112.

Kaszkurewicz, E., & Bhaya, A. (1993). Robust stability and diagonal Liapunov functions. *SIAM Journal on Matrix Analysis and Applications*, **14 (2)**, 508–520.

Leshno, M., Lin, V.Y., Pinkus, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, **6**, 861–867.

Liu, D., & Michel, A.N. (1992). Asymptotic stability of discrete time systems with saturation nonlinearities with applications to digital filters. *IEEE Transactions on Circuits and Systems I*, **39 (10)**, 798–807.

Ljung, L. (1979). Asymptotic behavior of the extended Kalman flter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, **24 (AC)**, 36–50.

Ljung, L. (1987). *System Identification: Theory for the User*. Englewood Cliffs, New Jersey: Prentice-Hall.

Maciejowski, J. M. (1989). *Multivariable feedback design*. New York: Addison-Wesley.

Miller, W. T., Sutton, R. S., & Werbos, P. J., (1990). *Neural networks for control*. Cambridge, MA: MIT Press.

Narendra, K.S., & Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, **1 (1)**, 4–27.

Narendra, K.S., & Parthasarathy, K. (1991). Gradient methods for the optimization of dynamical systems containing neural networks. *IEEE Transactions on Neural Networks*, **2 (2)**, 252–262.

Nesterov, Y., & Nemirovskii, A. (1994). *Interior point polynomial algorithms in convex programming*. SIAM (Studies in Applied Mathematics), Vol. 13. Philadelphia: SIAM.

Nguyen, D., & Widrow, B. (1990). Neural networks for self-learning control systems. *IEEE Control Systems Magazine*, **10 (3)**, 18–23.

Overton, M.L. (1988). On minimizing the maximum eigenvalue of a symmetric matrix. *SIAM Journal on Matrix Analysis and Applications*, **9 (2)**, 256–268.

Packard, A., & Doyle, J. (1993). The complex structured singular value. *Automatica*, **29 (1)**, 71–109.

Polak, E., & Salcudean, S. (1989). On the design of linear multivariable feedback systems via constrained nondifferentiable optimization in H∞ spaces. *IEEE Transactions on Automatic Control*, **34 (3)**, 268–276.

Polak, E., & Wardi, Y. (1982). Nondifferentiable optimization algorithm for designing control systems having singular value inequalities. *Automatica*, **18 (3)**, 267–283.

Reed, R. (1993). Pruning algorithms – a survey. *IEEE Transactions on Neural Networks*, **4 (5)**, 740–747.

Sjöberg, J., & Ljung, L. (1992). Overtraining, regularization and searching for minimum in neural networks. *4th IFAC International Symposium on adaptive systems in control and signal processing*, ACASP 92, 669–674, Grenoble, France.

Steinbuch, M., Terlouw, J., Bosgra, O., & Smit, S. (1992). Uncertainty modelling and structured singular-value computation applied to an electromechanical system. *IEE Proceedings – D*, **139 (3)**, 301–307.

Stoorvogel, A.A. (1992). The discrete time $H_\infty$ control problem with measurement feedback. *SIAM Journal on Control and Optimization*, **30 (1)**, 182–202.

Suykens, J.A.K., De Moor, B., & Vandewalle, J. (1994). Static and dynamic stabilizing neural controllers, applicable to transition between equilibrium points. *Neural Networks*, **7 (5)**, 819–831.

Suykens, J.A.K., De Moor, B., & Vandewalle, J. (1995). Nonlinear system identification using neural state space models, applicable to robust control design. *International Journal of Control*, **62 (1)**, 129–152.

Suykens, J. A. K., Vandewalle, J. P. L., & De Moor, B. L. R. (1995b). *Artificial neural networks for modelling and control of non-linear systems*. Boston: Kluwer Academic Publishers.

Suykens, J.A.K., & Vandewalle, J. (1996). Control of a recurrent neural network emulator for the double scroll. *IEEE Transactions on Circuits and Systems – I*, **43 (6)**, 511–514.

Suykens, J.A.K., & Vandewalle, J. (1996). Discrete time Interconnected Cellular Neural Networks within $NL_q$ theory. *International Journal of Circuit Theory and Applications (Special Issue on Cellular Neural Networks)*, **24**, 25–36.

Suykens, J.A.K., & Vandewalle, J. (1996c). *Absolute stability criterion for a Lur'e problem with mutilayer perceptron nonlinearity*. NDES'96 International Workshop on Nonlinear dynamics of electronic systems, Sevilla Spain, 117–122.

The MathWorks Inc. Matlab, (1994) *Optimization Toolbox* (Version 4.1), User's guide, Nov. 1992 & *Robust Control Toolbox* (Version 4.1), User's guide, Aug. 1992.

Tsoi, A.C., & Back, A.D. (1994). Locally recurrent globally feedforward networks: a critical review of architectures. *IEEE Transactions on Neural Networks*, **5 (2)**, 229–239.

Vandenberghe, L., & Boyd, S. (1995). A primal–dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming*, **69**, 205–236.

Vidyasagar, M. (1993). *Nonlinear systems analysis*. Englewood Cliffs, New Jersey: Prentice-Hall. (1993).

Wilkinson, J. H. (1965). *The algebraic eigenvalue problem*. Berlin: Oxford University Press.

Willems, J.C. (1972). Dissipative dynamical systems I: General theory. II: Linear systems with quadratic supply rates. *Archive for Rational Mechanics and Analysis*, **45**, 321–343.

Zurada, J. M., (1992). *Introduction to Artificial Neural Systems*. St Paul: West Publishing Company.

# APPENDIX

## Proof of Lemma 1

This is shown for each model $\mathcal{M}_i$

1. Model $\mathcal{M}_0 : q=1$ by taking $\Gamma_1=I,\ V_1=A,\ B_1=[B\,K\,0],\ \Lambda_1=I,\ W_1=C,\ D_1=[D\,I\,0], p_k=x_k, v_k=[u_k;\epsilon_k;1]$.
2. Model $\mathcal{M}_1 : q=2$ by taking $\Gamma_1=I,\ \Gamma_2=\Gamma_{AB},\ V_1=W_{AB},\ V_2=V_A,\ B_2=[V_B\,0\,\beta_{AB}],\ B_1=[0\,K\,0],\ \Lambda_1=I,\ \Gamma_2=I,\ W_1=I,\ W_2=C,\ D_2=[D\,I\,0], D_1=I$. The meaning of $\Gamma_{AB}$ follows from the proof for model $\mathcal{M}_1$.
3. Model $\mathcal{M}_2 : q=2$ by taking $\Gamma_1=I,\ \Gamma_2=\Gamma_{AB},\ V_1=W_{AB},\ V_2=V_A,\ B_2=[V_B\,0\,\beta_{AB}],\ B_1=[0\,K\,0],\ \Lambda_1=I,\ \Lambda_2=\Gamma_{CD},\ W_1=W_{CD},\ W_2=V_C,\ D_2=[V_D\,0\,\beta_{CD}],\ D_1=[0\,I\,0]$.
4. Model $\mathcal{M}_3 : q=3$ by taking $\Gamma_1=I,\ \Gamma_2=\Gamma_{AB_w},\ \Gamma_3=\Gamma_{AB_v},\ V_1=W_{AB},\ V_2=V_{AB},\ V_3=V_A,\ B_3=[V_B\,0\,\beta_{AB_v}],\ B_2=[0\,0\,\beta_{AB_w}],\ B_1=[0\,K\,0],\ \Lambda_1=I,\ \Lambda_2=\Gamma_{CD_w},\ \Lambda_3=\Gamma_{CD_v},\ W_1=W_{CD},\ W_2=V_{CD},\ W_3=V_C,\ D_3=[V_D\,0\,\beta_{CD_v}],\ D_2=[0\,0\,\beta_{CD_w}],\ D_1=[0\,I\,0]$

## Proof of Theorem 1

Propose a Lyapunov function $V(p) = \|D_1 p\|_2$ (positive, radially unbounded and V(0) = 0) with $D_1 \in \mathbb{R}^{n_p} \times {}^{n_p}$ a diagonal matrix with nonzero diagonal elements. Hence $V_k = \|D_1 p_k\|_2$ and $V_{k+1} = \|D_1 \Gamma_1 V_1 \Gamma_2 V_2 \ldots \Gamma_q V_q p_k\|_2$. The following procedure is followed then: (1) insert $D_i^{-1} D_i$ after $\Gamma_i$ where $D_i \in \mathbb{R}^{n_{h_i}} \times {}^{n_{h_i}} (i = 1, \ldots q-1)$ and insert $D_1^{-1} D_1$ after $V_q$; (2) the diagonal matrices $\Gamma_i$ and $D_i^{-1}$ commute; (3) use the properties of induced norms and the fact that $\|\Gamma_i\|_2 \leq 1$. Hence

$$V_{k+1} = \|D_1 \Gamma_1 D_1^{-1} D_1 V_1 \ldots \Gamma_q D_q^{-1} D_q V_q D_1^{-1} D_1 p_k\|_2$$

$$\leq \|\Gamma_1\|_2 \|D_1 V_1 D_2^{-1}\|_2 \|\Gamma_2\|_2 \|D_2 V_2 D_3^{-1}\|_2 \ldots \|\Gamma_q\|_2 \|D_q V_q D_1^{-1}\|_2 V_k \quad (70)$$

$$\leq \|D_1 V_1 D_2^{-1}\|_2 \|D_2 V_2 D_3^{-1}\|_2 \ldots \|D_q V_q D_1^{-1}\|_2 V_k$$

Defining $\beta_D = \Pi_{i=1}^q (mod q) \|D_i V_i D_{i+1}^{-1}\|_2$, a sufficient condition for global asymptotic stability of the $\mathrm{NL}_q$ system is $\beta_D$, because then $\Delta V_k = V_{k+1} - V_k < (\beta_D - 1) V_k < 0$.

This condition $\beta_D < 1>$ is satisfied if

$$\max_i \{\|D_i V_i D_{i+1}^{-1}\|_2 : i = 1, \ldots, q(mod q)\} \leq \beta_D^{1/q} < 1 \quad (71)$$

or if

$$\bar{\sigma}(\begin{bmatrix} D_1 V_1 D_2^{-1} & & & \\ & D_2 V_2 D_3^{-1} & & \\ & & \ddots & \\ & & & D_q V_q D_1^{-1} \end{bmatrix} 0 \leq \beta_D^{1/q} < 1 \quad (72)$$

or if

$$\bar{\sigma}(\begin{bmatrix} & I_{n_{h_2}} & & \\ & & \ddots & \\ & & & I_{n_{h_q}} \\ I_{n_{h_1}} & & & \end{bmatrix} \cdot \begin{bmatrix} D_1 V_1 D_2^{-1} & & & \\ & D_2 V_2 D_3^{-1} & & \\ & & \ddots & \\ & & & D_q V_q D_1^{-1} \end{bmatrix}) \leq \beta_D^{1/q} < 1 \quad (73)$$

The condition on this permuted block diagonal matrix is then satisfied if

$$\|D_{tot} V_{tot} D_{tot}^{-1}\|_2^q \leq \beta_D < 1 \quad (74)$$

where $D_{tot} = diag\{D_2, D_3, \ldots, D_q, D_1\}$.

## Proof of Theorem 2

Propose a Lyapunov function $V(p) = \|P_1 p\|_2$ (positive, radially unbounded and V(0) = 0) with $P_1 \in \mathbb{R}^{n_p} \times {}^{n_p}$. Hence $V_k = \|P_1 p_k\|_2$ and $V_{k+1} = \|P_1 \Gamma_1 V_1 \Gamma_2 V_2 \ldots \Gamma_q V_q p_k\|_2$. Defining $q_k = V_1 \Gamma_2 V_2 \ldots \Gamma_q V_q p_k$, we will study first under what conditions there exist a matrix $P_1$ and a positive scalar $\alpha_1$ such that the following is satisfied $\forall q$ and $\forall \Gamma_1 (\gamma_{1_i} \in [0, 1])$:

$$\|P_1 \Gamma_1 q_k\|_2^2 \leq (1 + \alpha_1) \|P_1 q_k\|_2^2 \quad (75)$$

or $\lambda_{\min}[(1+\alpha)U - \Gamma U \Gamma] > 0$, where $S = P_1^T P_1$, $\Gamma = \Gamma_1$ $U = N^{-1} S N$ (index 1 is omitted for notational reasons). Defining $R = (1+\alpha)U - \Gamma U \Gamma$ with elements $r_{ij} = (1 - \gamma_i \gamma_j + \alpha) n_i^{-1} n_j s_{ij}$, Gerschgorin's Theorem applied to the matrix $R$ states that all eigenvalues $\lambda_l$ of $R$ are lying in at least one of the circular disks with centers $r_{ii}$ and radii $\Sigma_{j(j \neq i)} |r_{ij}|$ :

$$|\lambda_l - r_{ii}| \leq \sum_{j(j \neq i)} |r_{ij}|, \forall l, i \quad (76)$$

or $|\lambda_l - (1 - \gamma_i^2 + \alpha) s_{ii}| \leq n_i^{-1} \Sigma_{j(j \neq i)} |(1 - \gamma_i \gamma_j + \alpha) n_j s_{ij}$. Because $\gamma_i \in [0, 1]$, if

$$\alpha q_{ii} > (1 + \alpha) \sum_{j(j \neq i)} |q_{ij}|, \forall i \quad (77)$$

then $R$ has only positive eigenvalues. Hence under that condition

$$V_k = \|P_1 \Gamma_1 q_k\|_2 \quad (78)$$

$$\leq (1 + \alpha_1)^{1/2} \|P_1 q_k\|_2$$

$$\leq (1 + \alpha_1)^{1/2} \|P_1 V_1 P_2^{-1}\|_2 \|P_2 \Gamma_2 r_k\|_2$$

where $r_k = \|P_2 \Gamma_2 V_2 \ldots \Gamma_q V_q p_k\|_2$. The same procedure can now be repeated for $\|P_2 \Gamma_2 r_k\|_2$ as was done for $\|P_1 \Gamma_1 q_k\|_2$ etc. Hence finally

$$V_{k+1} \leq \prod_{i=1(mod q)} (1 + \alpha_i)^{1/2} \|P_i V_i P_{i+1}^{-1}\|_2 V_k \quad (79)$$

with $P_{1+1} = P_1$ and diagonal dominant matrices $Q_i = P_i^T P_i$ with level of diagonal dominance equal to $\delta_{Q_i} = (1 + \alpha_i)/\alpha_i$.

Defining $\beta_P = \Pi_{i=1(mod q)}^q \|P_i V_i P_{i+1}^{-1}\|_2$ and $c_\alpha = \Pi_{i=1}^q (1 + \alpha_i)^{1/2}$, global asymptotic stability of the $\mathrm{NL}_q$ system is obtained for

$$\beta_P < 1/c_\alpha \quad (80)$$

or if

$$\max_i \{\|P_i V_i P_{i+1}^{-1}\|_2 : i = 1, \ldots, q(mod q)\} \leq \beta_P^{1/q} < 1/c_\alpha \quad (81)$$

or if

$$\|P_{tot} V_{tot} P_{tot}^{-1}\|_2^q \leq 1/c_\alpha \quad (82)$$

where $P_{tot} = blockdiag\{P_2, P_3, \ldots, P_q, P_1\}$.

## Proof of Lemma 3

By definition $\|X_Q\|_\infty = \max_i \Sigma_j |x_{Q_{ij}}| = \max_i \Sigma_j d_{Qi}^{-1} |h_{Q_{ij}}|$. Assuming $d_{Q_i} > 0$ the condition $\|X_Q\|_\infty < 1/\delta_Q$ corresponds to $\max_i \Sigma_j d_{Qi}^{-1} |h_{Qij}| < 1/\delta_Q$ or $d_{Qi} > \delta_Q \Sigma_{j(j \neq i)} |h_{Q_{ij}}| (\forall i)$ because $h_{Q_{ii}} = 0$ by definition. Hence the latter means $q_{ii} > \delta_Q \Sigma_{j(j \neq i)} |q_{ii}| (\forall i)$.

## Proof of Theorem 3

Propose a Lyapunov function $V(p) = \|P_1 p\|_2$ (positive, radially unbounded and $V(0) = 0$) with $P_1 \in \mathbb{R}^{n_p \times n_p}$. Following a similar procedure as in the proof of Theorem 2 one obtains

$$V_{k+1} = \|P_1 \Gamma_1 P_1^{-1} P_1 V_1 \ldots \Gamma_q P_q^{-1} P_q V_q P_1^{-1} P_1 p_k\|_2$$

$$\leq \|P_1 \Gamma_1 P_1^{-1}\|_2 \|P_1 V_1 P_2^{-1}\|_2 \|P_2 \Gamma_2 P_2^{-1}\|_2 \|P_2 V_2 P_3^{-1}\|_2 \ldots$$

$$\|P_q \Gamma_q P_q^{-1}\|_2 \|P_q V_q P_1^{-1}\|_2 V_k$$

$$\leq \prod_{i=1}^q \kappa(P_i) \|P_1 V_1 P_2^{-1}\|_2 \|P_2 V_2 P_3^{-1}\|_2 \ldots \|P_q V_q P_1^{-1}\|_2 V_k.$$

because $\|\Gamma_i\|_2 \leq 1$.

## Proof of Theorem 4

Defining $D_{S_1} = diag\{D_1, I_{n_w}\}$ it can be shown that

$$\|D_{S_1} \begin{bmatrix} p_{k+1} \\ e_k^{ext} \end{bmatrix} \|_2^2 = \|D_1 p_{k+1}\|_2^2 + \|e_k\|_2^2 \quad (83)$$

$$\leq \|D_{S_1} R_1 D_2^{-1}\|_2^2 \|D_2 R_2 D_3^{-1}\|_2^2 \ldots \|D_q R_q D_{S_1}^{-1}\|_2^2 (\|D_1 p_k\|_2^2 + \|w_k\|_2^2)$$

according to the proof of Theorem 1, with $D_i$ ($i$ = 2,...,q) diagonal with nonzero diagonal elements.

Defining $\beta_D = \|D_{S_1} R_1 D_2^{-1}\|_2 \Pi_{i=2}^{q-1} \|D_i R_i D_{i+1}^{-1}\|_2 \|D_q R_q D_{S_1}^{-1}\|_2$ it follows from the proof in Theorem 1 that $\beta_D < 1$ is satisfied if $\|D_{tot} R_{tot} D_{tot}^{-1}\|_2^q < 1$ where $D_{tot}$ = diag$\{D_2, D_3, \ldots, D_q, D_{S1}\}$.

Defining $r_k = \|D_1 p_k\|_2$, the following holds, provided that $\beta_D < 1$;

$$r_{k+1}^2 + \|e_k\|_2^2 \le \beta_D^2 r_k^2 + \beta_D^2 \|w_k\|_2^2 \tag{84}$$

$$\Rightarrow \sum_{k=0}^{N-1} r_{k+1}^2 + \sum_{k=0}^{N-1} \|e_k\|_2^2 \le \beta_D^2 \sum_{k=0}^{N-1} r_k^2 + \beta_D^2 \sum_{k=0}^{N-1} \|w_k\|_2^2$$

$$\Rightarrow r_N^2 + (1-\beta_D^2) \sum_{k=0}^{N-1} r_{k+1}^2 + \sum_{k=0}^{N-1} \|e_k\|_2^2 \le r_0^2 + \beta_D^2 \sum_{k=0}^{N-1} \|w_k\|_2^2$$

$$\Rightarrow \sum_{k=0}^{N-1} \|e_k\|_2^2 \le r_0^2 + \beta_D^2 \sum_{k=0}^{N-1} \|w_k\|_2^2$$

Furthermore for $N \rightarrow \infty$ : $(1-\beta_D^2)\|r\|_2^2 + \|e\|_2^2 \le r_0^2 + \beta_D^2 \|w\|_2^2$. Defining constants $c_1 = \bar{\sigma}_{D_1}^2$, $c_2 = \underline{\sigma}_{D_1}^2$ one obtains the result.

## Proof of Theorem 5

Defining $P_{S1}$ = blockdiag $\{P_1, I_{nw}\}$ and according to the proofs of Theorems 2 and 4

$$\|P_{S_1} \begin{bmatrix} p_{k+1} \\ e_k^{ext} \end{bmatrix} \|_2^2 = \|P_1 p_{k+1}\|_2^2 + \|e_k\|_2^2 \tag{85}$$

$$\le \prod_{i=1}^q (1+\alpha_i) \|P_{S_1} R_1 P_2^{-1}\|_2^2 \ldots \|P_i R_i P_{i+1}^{-1}\|_2^2 \ldots \|P_q R_q P_{S_1}^{-1}\|_2^2$$

$$(\|P_1 p_k\|_2^2 + \|w_k\|_2^2)$$

with diagonal dominant matrices $Q_i = P_i^T P_i N$: $\delta_{Qi} = (1+\alpha_i)/\alpha_i$. Defining

$$c_\alpha = \prod_{i=1}^q (1+\alpha_i)^{1/2},$$

$$\beta_p = \|P_{S_1} R_1 P_2^{-1}\|_2^2 \prod_{i=2}^{q-1} \|P_i R_i P_{i+1}^{-1}\|_2^2 P_q R_q P_{S_1^{-1}} \|_2^2$$

and according to Theorem 2, $c_\alpha \beta_P < 1$ holds if $\|P_{tot} R_{tot} P_{tot}^{-1}\|_2^q < 1/c_\alpha$ where $P_{tot}$ = blockdiag$\{P_2, P_3, \ldots, P_q, P_{S1}\}$. Putting $r_k = \|P_1 p_k\|_2$ then

$$r_{k+1}^2 + \|e_k\|_2^2 \le c_\alpha^2 \beta_P^2 r_k^2 + \beta_P^2 \|w_k\|_2^2 \tag{86}$$

$$\Rightarrow r_N^2 + (1 - c_\alpha^2 \beta_P^2) \sum_{k=0}^{N-1} r_k^2 + \sum_{k=0}^{N-1} \|e_k\|_2^2 \le r_0^2 + c_\alpha^2 \beta_P^2 \sum_{k=0}^{N-1} \|w_k\|_2^2$$

$$\Rightarrow \sum_{k=0}^{N-1} \|e_k\|_2^2 \le r_0^2 + c_\alpha^2 \beta_P^2 \sum_{k=0}^{N-1} \|w_k\|_2^2$$

Defining constants $c_1 = \bar{\sigma}_{P_1^2}$, $c_2 = \underline{\sigma}_{P_1^2}$ and letting $N \rightarrow \infty$ one obtains the result.

## Proof of Theorem 6

Defining again $P_{S_1} = blockdiag\{P_1, I_{n_w}\}$ one obtains

$$\|P_{S_1} \begin{bmatrix} p_{k+1} \\ e_k^{ext} \end{bmatrix} \|_2^2 = \|P_1 p_{k+1}\|_2^2 + \|e_k\|_2^2$$

$$\le \kappa(P_{S_1}) \prod_{i=1}^q \kappa(P_i) \|P_{S_1} R_1 P_2^{-1}\|_2^2 \ldots \|P_i R_i P_{i+1}^{-1}\|_2^2 \ldots$$

$$\|P_q R_q P_{S_1}^{-1}\|_2^2 (\|P_1 p_k\|_2^2 + \|w_k\|_2^2)$$

## Proof of Lemma 5

This follows immediately from the proofs of Theorems 4, 5 and 6, because, e.g., take the diagonal dominance case: if $c_\alpha \beta_P < 1$ then $r_{k+1}^2 + \|e_k\|_2^2 \le c_\alpha^2 \beta_P^2 r_k^2 + c_\alpha^2 \beta_P^2 \|w_k\|_2^2 \le r_k^2 + c_\alpha^2 \beta_P^2 \|w_k\|_2^2$ or $r_{k+1}^2 - r_k^2 \le c_\alpha^2 \beta_P^2 \|w_k\|_2^2 - \|e_k\|_2^2$ which proves the Lemma.

## Proof of Theorem 7

Consider the perturbed NL$_q$ written as an NL$_{q+1}$ system. Defining $P_{S_1} = blockdiag\{P_1 I, D_1\}$ and $\Upsilon_{q+1}$ = diag $\{I, I, \Delta_x\}$ one obtains, as in the proofs of Theorems 5 and 6:

$$\|P_{S_1} \begin{bmatrix} p_{k+1} \\ e_k^{ext} \\ s_k \end{bmatrix} \|_2^2 = \|P_1 p_{k+1}\|_2^2 + \|e_k\|_2^2 + \|D_1 s_k\|_2^2$$

$$\le \prod_{i=1}^q (1+\alpha_i) \|P_{S_1} T_1 P_2^{-1}\|_2^2 \ldots \|P_i T_i P_{i+1}^{-1}\|_2^2$$

$$\ldots \|P_q T_q P_{S_1}^{-1} P_{S_1} \Upsilon_{q+1} \begin{bmatrix} p_k \\ w_k \\ s_k \end{bmatrix} \|_2^2$$

$$\le \prod_{i=1}^q (1+\alpha_i) \|P_{S_1} T_1 P_2^{-1}\|_2^2 \ldots \|P_i T_i P_{i+1}^{-1}\|_2^2 \ldots \|P_q T_q P_{S_1}^{-1}\|_2^2$$

$$(\|P_1 p_k\|_2^2 + \|w_k\|_2^2 + \|D_1 s_k\|_2^2)$$

because $\Upsilon_{q+1}$ commutes with $P_{S_1}$ and $T_{q+1} = I$. The matrices $Q_i = P_i^T P_i N_i$ are again diagonal dominant. Defining

$$c_\alpha = \prod_{i=1}^q (1+\alpha_i)^{1/2},$$

$$\beta_P = \|P_{S_1} T_1 P_2^{-1}\|_2 \prod_{i=2}^{q-1} \|P_i T_i P_{i+1}^{-1}\|_2 \|P_q T_q P_{S_1^{-1}}\|_2$$

and according to Theorem 5, $c_\alpha \beta_P < 1$ holds if $\|P_{tot} T_{tot} P_{tot}^{-1}\|_2^q < 1/c_\alpha$ where $P_{tot} = blockdiag\{P_2, P_3, \ldots, P_q, P_{S_1}\}$. Putting $r_k = \|P_{1pk}\|_2$ then

$$r_{k+1}^2 + \|e_k\|_2^2 + \|D_1 s_k\|_2^2 \le c_\alpha^2 \beta_P^2 (r_k^2 + \|w_k\|_2^2 + \|D_1 s_k\|_2^2)$$

$$\le c_\alpha^2 \beta_P^2 r_k^2 + c_\alpha^2 \beta_P^2 \|w_k\|_2^2 + \|D_1 s_k\|_2^2$$

$$\Rightarrow r_{k+1}^2 + \|e_k\|_2^2 \le r_k^2 + \|w_k\|_2^2$$

because $c_\alpha \beta_P$.

The second case follows immediately from

$$\|P_{S_1} \begin{bmatrix} p_{k+1} \\ e_k^{ext} \\ s_k \end{bmatrix} \|_2^2 \le \kappa(P_{S_1}) \prod_{i=2}^q \kappa(P_i) \|P_{S_1} T_1 P_2^{-1}\|_2^2 \ldots \|P_i T_i P_{i+1}^{-1}\|_2^2$$

$$\ldots \|P_q T_q P_{S_1}^{-1}\|_2^2 \|P_{S_1} \begin{bmatrix} p_k \\ w_k \\ s_k \end{bmatrix} \|_2^2.$$

# NOMENCLATURE

## Systems

$\mathcal{M}_i$ Neural state space model ($i \in \{0,1,\ldots,3\}$)

$\mathcal{C}_j$ Neural state space controller ($j \in \{0,1,\ldots,5\}$)

| | |
|---|---|
| $\Xi_j^i$ | Member of the family of neural control problems $\{\Xi_j^i \mid i \in \{0,1,\ldots,3\}, j \in \{0,1,\ldots,5\}\}$, model $\mathcal{M}_i$ and controller $C_j$ |
| $\mathcal{S}_i$ | Standard plant model related to model $\mathcal{M}_i$ |
| $NL_q$ | Nonlinear state space model consisting of a $q$ times alternating sequence of linear and nonlinear operators that satisfy a sector condition [0,1] |
| LFT | Linear fractional transformation |

**Signals**

| | |
|---|---|
| $k$ | discrete time index, $k \in \mathbb{N}$ |
| $w_k$ | exogenous input |
| $d_k$ | reference input |
| $\epsilon_k$ | white noise input |
| $e_k$ | regulated output |
| $u_k$ | control signal |
| $y_k$ | sensed output |
| $\hat{x}_k$ | state of model $\mathcal{M}_i$ |
| $z_k$ | state of controller $C_j$ |
| $p_k$ | state of $NL_q$ system for standard plant configuration |