Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 1994-58

Neural state space model framework for identification and control with global asymptotic stability criteria[1]

Johan A.K. Suykens, Bart L.R. De Moor and Joos Vandewalle[2]

Sept 1994

[1]This report is available by anonymous ftp from *gate.esat.kuleuven.ac.be* in the directory *pub/SISTA/suykens/reports*

[2]ESAT - Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, 3001 Leuven (Heverlee), Belgium, tel 32/16/22 09 31, fax 32/16/22 18 55, email: *johan.suykens@esat.kuleuven.ac.be*. 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 a Concerted Action Project of the Flemish Community, entitled *Applicable Neural Networks*, and within 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 IUAP-50). The scientific responsibility rests with its authors.

Neural state space model framework for identification and control with global asymptotic stability criteria*

Johan A.K. Suykens
Bart L.R. De Moor†
Joos Vandewalle

Katholieke Universiteit Leuven
Department of Electrical Engineering, ESAT-SISTA
Kardinaal Mercierlaan 94
B-3001 Leuven (Heverlee), Belgium
tel: 32/16/22 09 31 fax: 32/16/22 18 55
e-mail: Johan.Suykens@esat.kuleuven.ac.be

July 19, 1994

*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 a Concerted Action Project of the Flemish Community, entitled *Applicable Neural Networks* and within 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. The scientific responsibility rests with its authors.

†Research Associate of the Belgian National Fund for Scientific Research

Abstract

A framework for system identification and model based control is proposed with models and controllers described in state space form and both parametrized by multilayer feedforward neural networks. Prediction error learning algorithms for system identification are discussed with respect to Narendra's sensitivity model approach. Main results are sufficient conditions for global asymptotic stability of the models and closed loop systems, together with conditions for dissipativity with finite l_2-gain smaller than one (input/output stability) and conditions for robust performance taking into account parametric uncertainties on the models. For control design the tracking of a given reference input is discussed using Narendra's sensitivity model approach in dynamic backpropagation. In the analysis problem it is shown how one can check global asymptotic stability, dissipativity or robust performance of the learned controller by solving linear matrix inequalities, for which efficient polynomial-time algorithms exist. In the synthesis problem the dynamic backpropagation paradigm is modified by adding constraints to the original problem in order to ensure global asymptotic stability or dissipativity of the closed loop system. An essential and new concept in the present neural state space control theory is the so called NL_q model. Based on NL_qs connections are revealed with problems in classical and modern control theory (H_∞ control theory, μ-theory and the Lur'e problem) and architectures in neural network theory (Hopfield network, locally recurrent globally feedforward neural networks, multilayer cellular neural networks).

Keywords. state space models, NL_qs, sensitivity models, nonlinear programming, Lyapunov function, global asymptotic stability, standard plant, dissipativity, l_2-gain, diagonal scaling, diagonal dominance, linear matrix inequalities, nondifferentiable optimization.

1 Preliminaries and notation

Throughout this paper we use the following notations, definitions and basic results and we refer here to the following books: Golub & Van Loan (1989), Wilkinson (1965), Boyd *et al.* (1994), Luenberger (1979). $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}^n$ is the set of real column vectors with n components, $\mathbb{R}^{n\times m}(\mathbb{C}^{n\times m})$ is the set of real (complex) $n \times m$ matrices. The ij-th entry of a matrix A is denoted by a_j^i or a_{ij}. A^T is the transpose of A, A^* is the complex conjugate transpose of A. $|.|$ denotes the absolute value of a scalar. For vector norms on $x \in \mathbb{R}^n$ one has $\|x\|_2 = (\sum_{i=1}^n |x_i|^2)^{1/2}$, $\|x\|_\infty = \max_i |x_i|$, $\|x\|_1 = \sum_{i=1}^n |x_i|$ and in general $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p} (1 \le p < \infty)$. For induced matrix norms on $A \in \mathbb{R}^{n\times m}$ one has $\|A\|_2 = \overline{\sigma}(A)$, $\|A\|_\infty = \max_i \sum_{j=1}^m |a_{ij}|$, $\|A\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$. $\overline{\sigma}(A) = (\lambda_{max}(A^*A))^{1/2}$ is the maximal singular value of A, $\underline{\sigma}(A)$ is the minimal singular value of A. $\sigma_i(A)$ is the i-th singular value of A, $\lambda_i(A)$ is the i-th eigenvalue of A. $\rho(A) = \max_i |\lambda_i(A)|$ is the spectral radius of A. The following holds $\|A+B\| \le \|A\| + \|B\|$, $\|AB\| \le \|A\|\|B\|$, $\|Ax\| \le \|A\|\|x\|$. Furthermore $\|A\|_2 \le \sqrt{\|A\|_1\|A\|_\infty}$, $\frac{1}{\sqrt{n}}\|A\|_\infty \le \|A\|_2 \le \sqrt{m}\|A\|_\infty$, $\frac{1}{\sqrt{m}}\|A\|_1 \le \|A\|_2 \le \sqrt{n}\|A\|_1$. We will also make use of Gerschgorin's Theorem that states that every eigenvalue of a matrix A lies in at least one of the circular discs with centers a_{ii} and radii $\sum_{j(j\neq i)} |a_{ij}|$ or $|\lambda - a_{ii}| \le \sum_{j(j\neq i)} |a_{ij}|$.

The l_2 norm of a sequence $e \in l_2^n$, where l_2^n denotes the set of square summable sequences in $\mathbb{C}^n$ is by definition $\|e\|_2^2 = \sum_{k=1}^\infty \|e_k\|_2^2$. A column vector x which is the concatenation of other vectors x_1 and x_2 is denoted as $x = [x_1; x_2]$. A matrix A consisting of the block rows $[A_{11}\ A_{12}]$ and $[A_{21}\ A_{22}]$ is denoted as $[A_{11}\ A_{12}; A_{21}\ A_{22}]$ (Matlab notation). $A > 0$ means a A positive definite. $A > B$ means $A - B$ positive definite. A Matrix Inequality is $M(x) < 0$ with $M = M^T$ and $x \in \mathbb{R}^n$. In a Linear Matrix inequality (LMI) $M(x)$ depends affinely on x: $M(x) = M_0 + \sum_{i=1}^m x_i M_i$ and $M_0 = M_0^T$, $M_i = M_i^T$. In that case finding a feasible x for LMIs is convex. The condition $\exists T : \overline{\sigma}(TAT^{-1}) < \gamma$ is equivalent to the LMI $\exists X > 0 : A^*XA - \gamma^2 X < 0$ or $\lambda_{max}(A^*XA - \gamma^2 X) < 0$.

Concerning stability of a nonlinear system $x_{k+1} = f(x_k)$, where $\overline{x}$ is an equilibrium point and $S(\overline{x}, R)$ denotes a spherical region in state space with center at $\overline{x}$ and radius R, some basic definitions are reviewed here. An equilibrium point $\overline{x}$ is *stable* if there is an $R_0 > 0$ for which the following is true: for every $R < R_0$, there is an r, $0 < r < R$, such that if $x(0)$ is inside $S(\overline{x}, R)$, then x_k is inside $S(\overline{x}, R)$ for all $k > 0$. An equilibrium point $\overline{x}$ is *asymptotically stable* whenever it is stable and in addition there is an $\overline{R_0} > 0$ such that whenever the state is initiated inside $S(\overline{x}, \overline{R_0})$, it tends to $\overline{x}$ as time increases. An equilibrium point $\overline{x}$ is *globally asymptotically stable* (also called asymptotically stable in the large) if it is asymptotically stable and, in addition, when initiated at any point in the entire state space the system state tends to $\overline{x}$. An equilibrium point $\overline{x}$ is *marginally stable* if it is stable but not asymptotically stable. A function V defined on a region Ω of the state space and containing $\overline{x}$ is a Lyapunov function for the nonlinear system if it satisfies the following three requirements: (1) V is continuous, (2) $V(x)$ has a unique minimum at $\overline{x}$ with respect to all other points in Ω, (3) The function $\Delta V(x) = V(f(x)) - V(x)$ satisfies $\Delta V(x) \le 0$ for all x in Ω. The following three conditions are sufficient (not necessary) then

in order to prove global asymptotical stability of the equilibrium point $\overline{x}$: (1) V is defined on the entire state space, (2) $\Delta V(x) < 0$ for all $x \neq \overline{x}$, (3) $V(x) \to \infty$ as any component of x gets arbitrarily large in magnitude (radially unbounded).
The following signals and systems are defined throughout this paper

Systems:

$\mathcal{M}_i$	Neural state space model ($i \in \{0, 1, ..., 3\}$)
$\mathcal{C}_j$	Neural state space controller ($j \in \{0, 1, ..., 5\}$)
Ξ_j^i	Member of the family of neural control problems $\{\Xi_j^i \vert i \in \{0, 1, ..., 3\}, j \in \{0, 1, ..., 5\}\}$, related to model $\mathcal{M}_i$ and controller $\mathcal{C}_j$
$\mathcal{S}_i$	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
LFT	Linear Fractional Transformation

Signals:

w_k	exogenous input
d_k	reference input
ϵ_k	white noise innovations input
e_k	regulated output
u_k	actuator input
y_k	sensed output
$\hat{x}_k$	state of model $\mathcal{M}_i$
z_k	state of controller $\mathcal{C}_j$
p_k	state of NL_q system for standard plant configuration

2 Introduction

Up till now among the properties of multilayer neural networks that are regarded as most important for their application in control are the universal approximation ability, the parallel network structure and the availability of algorithms to adapt the interconnection weights (Hunt *et al.* (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. From the control viewpoint parametrizations by neural networks may result into an improved tracking performance and are useful in order to solve optimal control problems.

However despite the many reported successes, there still remains scepticism mainly because of the lack of general stability results for control systems, containing neural networks. 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 will be precisely to cope with these last doubts about neural nets and to present sufficient conditions for global asymptotic stability, conditions for dissipativity of the systems and robust performance of the closed loop systems. The proposed framework in this paper is model based like the emulator approach proposed in Nguyen & Widrow (1990). Nonlinear state space models and nonlinear dynamic output feedback controllers are parametrized by feedforward neural nets (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 into so called NL_q models for which stability criteria will be given. It turns out that many problems appearing in classical, modern and neural control theory and neural networks are related to NL_qs

- all neural state space control problems
- discrete time Lur'e problem
- Linear Fractional Transformations with real diagonal Δ uncertainty block
- Hopfield network, multilayer cellular neural networks
- Locally recurrent globally feedforward neural networks

With respect to dissipativity of NL_qs it will be shown that the presented theory is a straightforward extension of linear H_∞ control theory and μ-theory towards general nonlinear systems.

Also numerical algorithms for control design are proposed. Results of convex optimization and linear matrix inequalities (LMIs) for which very efficient polynomial-time algorithms exist are used as much as possible. Like in many problems of systems and control (see Boyd *et al.* (1994)) LMIs play also here a crucial role. This leads finally to the formulation of a modified dynamic backpropagation algorithm that ensures global

asymptotic stability or dissipativity of the closed loop system.

This paper is organized as follows. In Section 3 prediction error learning algorithms for neural state space models are discussed using Narendra's sensitivity model approach, together with aspects of model validation. Section 4 introduces neural state space controllers, the control scheme, the standard plant representation and a complete family of neural state space control problems. In Section 5 the new NL_q system concept is proposed. In Section 6 global asymptotic stability criteria for NL_qs are given. The relationship with the discrete time Lur'e problem is given. In Section 7 input/output properties of NL_qs are investigated. Main Theorems on dissipativity with finite l_2-gain smaller than one are presented. The connection with linear H_∞ control theory is given. In Section 8 the influence of parametric uncertainties on NL_qs is studied. These perturbed NL_qs are shown to be NL_{q+1}s. The state space upper bound test in μ-theory can be interpreted under certain conditions as a special case of the general theory. In Section 9 learning algorithms for the neural state space controller are proposed. This includes tracking of a specific reference input according to Narendra's sensitivity model approach in dynamic backpropagation, the analysis problem where checking of a learned controller corresponds to solving a set of LMIs and the synthesis problem for which a modified dynamic backpropagation algorithm is introduced that can ensure global asymptotic stability or dissipativity with finite l_2-gain of the closed loop system. Finally in Section 10 it is illustrated that the NL_q concept goes far beyond problems in control theory. Neural network architectures such as locally recurrent globally feedforward and multilayer cellular neural networks are interpreted as NL_qs. In order to keep the text readable details and proofs are transferred to Appendices. Appendix A contains the formulas for sensitivity models of neural state space models. Proofs of Lemmas and Theorems can be found in Appendix B. In Appendix C a complete list is given of the family of neural state space control problems written as NL_qs.

3 Nonlinear system identification using neural state space models

Given a nonlinear plant, for system identification a true system in state space form

$$\left\{ \begin{array}{rcl} x_{k+1} & = & f_0(x_k, u_k) + \varphi_k \\ y_k & = & g_0(x_k, u_k) + \psi_k \end{array} \right. \tag{1}$$

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

$$E\{ \left[\begin{array}{c} \varphi_k \\ \psi_k \end{array} \right] [\varphi_s^t \ \psi_s^t] \} = \left[\begin{array}{cc} Q & S \\ S & R \end{array} \right] \delta_{ks}.$$

For such systems predictors are discussed in Suykens *et al.* (1994b), both for the case of deterministic identification ($\varphi_k = 0$, $\psi_k = 0$) as the stochastic case ($\varphi_k \neq 0$, $\psi_k \neq 0$). The main ideas will be reviewed here.

3.1 Neural state space models

The following predictor in innovations form was proposed for (1) in Goodwin & Sin (1984):

$$\mathcal{M}(\theta): \left\{ \begin{array}{rcl} \hat{x}_{k+1} & = & f(\hat{x}_k, u_k; \theta) + K(\theta)\epsilon_k \ ; \ \hat{x}_0 = x_0 \text{ given} \\ \hat{y}_k & = & g(\hat{x}_k, u_k; \theta) \end{array} \right. \tag{2}$$

with prediction error $\epsilon_k(\theta) = y_k - \hat{y}_k(\theta)$. Given N input/output data Z^N a prediction error algorithm aims then at minimizing the cost function

$$V_N(\theta, Z^N) = \frac{1}{N} \sum_{k=1}^{N} l(\epsilon_k(\theta)) \tag{3}$$

with solution

$$\hat{\theta}_N = \arg \min_{\theta \in D_{\mathcal{M}}} V_N(\theta, Z^N).$$

A typical choice for the scalar valued positive function $l(\epsilon_k)$ is $\frac{1}{2}\epsilon_k^t \epsilon_k$. In (2) $\mathcal{M}$ denotes the model structure and $\mathcal{M}^* = \{\mathcal{M}(\theta) \mid \theta \in D_{\mathcal{M}}\}$. The predictor makes use of a direct parametrization of the Kalman gain $K(\theta)$ instead of by a complicated dependency through a Riccati equation as for extended Kalman filtering. More details on these aspects can be found in Suykens *et al.* (1994b) and Goodwin & Sin (1984).

For the model structure (2) parametrizations by feedforward neural nets can be considered. Such parametrizations make sense because any continuous nonlinear function can be approximated arbitrarily well on a compact interval by a multilayer feedforward neural

network with one or more hidden layers (Cybenko (1989), Funahashi (1989), Hornik *et al.* (1989), Leshno *et al.* (1993)). The resulting models will be called here *neural state space models*[1]. The reason for this specific choice of parametrizations is that 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 f and g in the predictor (2) are made and called $\mathcal{M}_i$ ($i \in \{0,..,3\}$) depending on the specific parametrization:

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

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

$$\mathcal{M}_2(\theta): \begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_Bu_k + \beta_{AB}) + K\epsilon_k \\ y_k &= W_{CD}\tanh(V_C\hat{x}_k + V_Du_k + \beta_{CD}) + \epsilon_k \end{cases} \quad (4)$$

$$\mathcal{M}_3(\theta): \begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_{AB}\tanh(V_A\hat{x}_k + V_Bu_k + \beta_{ABv}) + \beta_{ABw}) + K\epsilon_k \\ y_k &= W_{CD}\tanh(V_{CD}\tanh(V_C\hat{x}_k + V_Du_k + \beta_{CDv}) + \beta_{CDw}) + \epsilon_k. \end{cases}$$

Remark. The linear model in innovations form $\mathcal{M}_0$ is the celebrated *Kalman filter* (see Ljung (1987)(1979)) 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 (4) the process is linear in ϵ_k. The case of deterministic identification corresponds to a zero Kalman gain K.

3.2 Prediction error learning algorithm

The prediction error algorithm related to a neural state space model $\mathcal{M}_i$ ($i \in \{0,1,...,3\}$), results into the following off-line optimization problem

$$\min_{\theta \in D_{\mathcal{M}}} V_N(\theta, Z^N) = \frac{1}{N}\sum_{k=1}^{N} l(\epsilon_k(\theta)) \quad (5)$$

subject to the dynamical model $\mathcal{M}_i(\theta)$ for $k = 0,...,N$. The latter is a nonlinear least squares problem. In general this problem has many local minima.

If one uses local optimization methods for minimizing the cost function (5) such as steepest descent, conjugate gradient, quasi-Newton or Levenberg-Marquardt (see e.g. van

[1] In neural network terminology *neural state space models* are called *recurrent neural networks*, but in our opinion the first term is better suited with respect to the field of system identification, because it reveals the fact that a predictor in state space form is parametrized by the neural net architecture. The term 'recurrent neural net' is more general, because it can be related to state space models as well as to input/output models (such as e.g. NARX or NARMAX in Billings *et al.* (1992)).

der Smagt (1994), Gill *et al.* (1981), Fletcher (1987)), one needs to know the gradient of V_N with respect to θ which is for $l(\epsilon) = \frac{1}{2}\epsilon^t\epsilon$ equal to

$$\frac{\partial V_N}{\partial \theta} = \frac{1}{N}\sum_{k=1}^{N}\epsilon^t\frac{\partial \epsilon_k}{\partial \theta} = -\frac{1}{N}\sum_{k=1}^{N}\epsilon^t\frac{\partial \hat{y}_k}{\partial \theta}. \quad (6)$$

In Suykens *et al.* (1994b) it is shown how Narendra's sensitivity model approach (Narendra & Parthasarathy (1991)) can be applied to neural state space models in order to generate the gradient of the cost function: given a predictor

$$\begin{cases} \hat{x}_{k+1} &= \Phi(\hat{x}_k, u_k; \alpha)\ ;\ \hat{x}_0 = x_0 \text{ given} \\ \hat{y}_k &= \Psi(\hat{x}_k, u_k; \beta) \end{cases} \quad (7)$$

such as $\mathcal{M}_i(\theta)$ ($i \in \{0,1,...,3\}$) with α, β elements of the parameter vector $\theta \in \mathbb{R}^p$, a sensitivity model is the model that generates the gradient $\frac{\partial \hat{y}_k}{\partial \alpha}$ and $\frac{\partial \hat{y}_k}{\partial \beta}$ and is obtained by taking the derivatives of (7) with respect to α and β

$$\begin{cases} \frac{\partial \hat{x}_{k+1}}{\partial \alpha} &= \frac{\partial \Phi}{\partial \hat{x}_k}.\frac{\partial \hat{x}_k}{\partial \alpha} + \frac{\partial \Phi}{\partial \alpha} \\ \frac{\partial \hat{y}_k}{\partial \alpha} &= \frac{\partial \Psi}{\partial \hat{x}_k}.\frac{\partial \hat{x}_k}{\partial \alpha} \\ \frac{\partial \hat{y}_k}{\partial \beta} &= \frac{\partial \Psi}{\partial \beta}, \end{cases} \quad (8)$$

which is a dynamical model with state vector $\frac{\partial \hat{x}_k}{\partial \alpha}$ driven by the input vector consisting of $\frac{\partial \Phi}{\partial \alpha}$, $\frac{\partial \Psi}{\partial \beta}$ and at the output $\frac{\partial \hat{y}_k}{\partial \alpha}$, $\frac{\partial \hat{y}_k}{\partial \beta}$ are generated. The expressions of $\frac{\partial \Phi}{\partial \alpha}$, $\frac{\partial \Psi}{\partial \beta}$, $\frac{\partial \Phi}{\partial \hat{x}_k}$ and $\frac{\partial \Psi}{\partial \hat{x}_k}$ for each of the neural state space models $\mathcal{M}_i$ can be found in Appendix A.

Remarks.

- The optimization method that makes use of such a sensitivity model for dynamic systems containing neural networks is called *dynamic backpropagation* in Narendra & Parthasarathy (1991). An important message of Narendra in that paper is that one has to be careful with taking the derivatives in the case of dynamic systems containing neural networks because the original (static) *backpropagation* algorithm (Rumelhart *et al.*(1986)) cannot be used in that case. Conceptually the predictor and its sensitivity model can be considered as one augmented system generating $\hat{y}_k$ as well as $\frac{\partial \hat{y}_k}{\partial \theta}$ at the output.

- For the linear model $\mathcal{M}_0$ the sensitivity model is also linear with constant matrices $\frac{\partial \Phi}{\partial \hat{x}_k}$ and $\frac{\partial \Psi}{\partial \hat{x}_k}$ equal to A and C respectively. System identification with $\mathcal{M}_0$ is an intensively studied topic in system identification theory. The present framework can be seen as a straightforward extension towards nonlinear systems using parametrizations by neural nets.

- *Parallel* training of the neural state space models is possible. Because at each iteration one has for each of the p parameters α, β of the parameter vector θ a simulation of the sensitivity model over N samples, these p simulations can be distributed over the available number of processors.

- For weakly nonlinear systems, linear models can be used as starting points for the nonlinear optimization problems (5) (see Suykens *et al.* (1994b)).

3.3 Model validation

Model validation for results of nonlinear system identification was studied e.g. in Billings *et al.* (1992). For adequate modeling the residual ϵ_k should be uncorrelated with all linear and nonlinear combinations of past inputs and outputs. The following conditions should hold:

$$\begin{array}{rcl} \phi_{\epsilon\epsilon}(\tau) & = & E[\epsilon_{k-\tau}\epsilon_k] = \delta_\tau \\ \phi_{u\epsilon}(\tau) & = & E[u_{k-\tau}\epsilon_k] = 0 \ (\forall\tau) \\ \phi_{u^{2'}\epsilon}(\tau) & = & E[(u_{k-\tau}^2 - \overline{u_k^2})\epsilon_k] = 0 \ (\forall\tau) \\ \phi_{u^{2'}\epsilon^2}(\tau) & = & E[(u_{k-\tau}^2 - \overline{u_k^2})\epsilon_k^2] = 0 \ (\forall\tau) \\ \phi_{\epsilon(\epsilon u)}(\tau) & = & E[\epsilon_k\epsilon_{k-1-\tau}u_{k-1-\tau}] = 0 \ (\tau \geq 0). \end{array} \tag{9}$$

Here $u^{2'} = u_k^2 - \overline{u_k^2}$, where $\overline{u_k^2}$ denotes the mean of u_k^2. In practice normalized correlations are computed. The sampled correlation function between two sequences α_k and β_k is given by $\hat{\phi}_{\alpha\beta}(\tau) = (\sum_{k=1}^{N-\tau}\alpha_k\beta_{k+\tau})/[\sum_{k=1}^{N}\alpha_k^2\sum_{k=1}^{N}\beta_k^2]^{1/2}$. This normalization ensures that $-1 \leq \hat{\phi}_{\alpha\beta} \leq 1$. 95% confidence bands are defined as $1.96/\sqrt{N}$ (N is the data length). The conditions can be applied both to the training data and the test data.

Another method (which is common in the field of neural networks) is to look simply for a model with the best generalization error, performing best on fresh data (test set) that were not used for the identification. One seeks normally for a minimal fitting error together with a minimal number of parameters of the model (as suggested by Akaike's criterion (see e.g. Ljung (1987))): such aspects are studied e.g. in Reed (1993) and Sjöberg & Ljung (1992).

4 A framework for control design

4.1 General aspects

The framework for control design that will be discussed in this paper is *indirect* in the sense that it is assumed that given input/output data on a plant, first nonlinear system identification is done according to Section 3 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 model. A difference between 'direct' and 'indirect' methods is normally made in the context of adaptive control. The present framework however as

it will be described here is non-adaptive. A second feature of the method is that both the model and the controller are dynamic systems, parametrized by multilayer feedforward neural nets, respectively called *neural state space models* and *neural state space controllers*.

4.2 Standard plant representation

In Fig.1 a two degree of freedom control scheme is given with a neural state space model and a neural state space controller. The neural state space model is given by

$$\left\{ \begin{array}{rcl} \hat{x}_{k+1} & = & f(\hat{x}_k, u_k, \epsilon_k; \theta_m) \\ y_k & = & g(\hat{x}_k, u_k, \epsilon_k; \theta_m) \end{array} \right. \tag{10}$$

parametrized by $\theta_m \in \mathbb{R}^{p_m}$, corresponding to one of the models $\mathcal{M}_i$ ($i \in \{0,..,3\}$) in (4) and a *dynamic output feedback* controller

$$\left\{ \begin{array}{rcl} z_{k+1} & = & h(z_k, y_k, d_k; \theta_c) \\ u_k & = & s(z_k, y_k, d_k; \theta_c) \end{array} \right. \tag{11}$$

parametrized by $\theta_c \in \mathbb{R}^{p_c}$. The state of the controller is $z_k \in \mathbb{R}^{n_z}$ and the reference input $d_k \in \mathbb{R}^l$. The following parametrizations for θ_c by neural nets are made.

Definition 2 [Neural state space controllers]. The following parametrizations by feedforward neural nets for h and s are made in (11) and are called Neural state space controllers $\mathcal{C}_i$ ($i \in \{0,..,5\}$) depending on the specific parametrization:

$$\mathcal{C}_0(\theta_c): \left\{ \begin{array}{rcl} z_{k+1} & = & Ez_k + Fy_k + F_2d_k \\ u_k & = & Gz_k + Hy_k + H_2d_k \end{array} \right.$$

$$\mathcal{C}_1(\theta_c): \left\{ \begin{array}{rcl} z_{k+1} & = & Ez_k + Fy_k + F_2d_k \\ u_k & = & \tanh(Gz_k + Hy_k + H_2d_k) \end{array} \right.$$

$$\mathcal{C}_2(\theta_c): \left\{ \begin{array}{rcl} z_{k+1} & = & W_{EF}\tanh(V_Ez_k + V_Fy_k + V_{F_2}d_k + \beta_{EF}) \\ u_k & = & W_{GH}\tanh(V_Gz_k + V_Hy_k + V_{H_2}d_k + \beta_{GH}) \end{array} \right.$$

$$\mathcal{C}_3(\theta_c): \left\{ \begin{array}{rcl} z_{k+1} & = & W_{EF}\tanh(V_Ez_k + V_Fy_k + V_{F_2}d_k + \beta_{EF}) \\ u_k & = & \tanh(W_{GH}\tanh(V_Gz_k + V_Hy_k + V_{H_2}d_k + \beta_{GH})) \end{array} \right.$$

$$\mathcal{C}_4(\theta_c): \left\{ \begin{array}{rcl} z_{k+1} & = & W_{EF}\tanh(V_{EF}\tanh(V_Ez_k + V_Fy_k + V_{F_2}d_k + \beta_{EF_v}) + \beta_{EF_w}) \\ u_k & = & W_{GH}\tanh(V_{GH}\tanh(V_Gz_k + V_Hy_k + V_{H_2}d_k + \beta_{GH_v}) + \beta_{GH_w}) \end{array} \right.$$

$$\mathcal{C}_5(\theta_c): \left\{ \begin{array}{rcl} z_{k+1} & = & W_{EF}\tanh(V_{EF}\tanh(V_Ez_k + V_Fy_k + V_{F_2}d_k + \beta_{EF_v}) + \beta_{EF_w}) \\ u_k & = & \tanh(W_{GH}\tanh(V_{GH}\tanh(V_Gz_k + V_Hy_k + V_{H_2}d_k + \beta_{GH_v}) + \beta_{GH_w})). \end{array} \right. \tag{12}$$

Remarks: The controllers $\mathcal{C}_1$, $\mathcal{C}_3$, $\mathcal{C}_5$ correspond respectively to the controllers $\mathcal{C}_0$, $\mathcal{C}_2$, $\mathcal{C}_4$ but with saturated output u_k. $\mathcal{C}_0$ is a classical linear dynamic output feedback controller (e.g. Boyd & Barratt (1991), Maciejowski (1989)).

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 $d_k - \hat{y}_k$ and possibly also other 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.

4.3 Family of neural state space control problems

Depending on the specific neural state space model $\mathcal{M}_i$ and neural state space controller $\mathcal{C}_j$ that is chosen in the control scheme of Fig.1, a neural control problem Ξ_j^i is selected out of a complete family of problems.

Definition 3 [Neural state space control problem Ξ_j^i]. A neural control problem Ξ_j^i is the control problem related to the control scheme of Fig.1 with a neural state space controller $\mathcal{C}_j$ ($j \in \{0,1,...,5\}$) and a neural state space model $\mathcal{M}_i$ ($i \in \{0,1,...,3\}$).

The family of problems given in the following table is considered throughout this paper. A problem Ξ_j^i corresponds to the ij-th entry in the table

	$\mathcal{C}_0$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$
$\mathcal{M}_0$	Ξ_0^0	Ξ_1^0	-	-	-	-
$\mathcal{M}_1$	Ξ_0^1	Ξ_1^1	Ξ_2^1	Ξ_3^1	-	-
$\mathcal{M}_2$	Ξ_0^2	Ξ_1^2	Ξ_2^2	Ξ_3^2	-	-
$\mathcal{M}_3$	Ξ_0^3	Ξ_1^3	Ξ_2^3	Ξ_3^3	Ξ_4^3	Ξ_5^3

5 NL$_q$ systems

An essential concept in the further derivation of global asymptotic stability criteria for the family of neural control problems Ξ_j^i is that of NL$_q$ systems.

Definition 4 [NL$_q$ system]. The following nonlinear state space model is called an NL$_q$ system

$$\begin{cases} x_{k+1} &= \Gamma_1(x_k,v_k)(\,V_1\,\Gamma_2(x_k,v_k)(\,V_2\,...\Gamma_q(x_k,v_k)(\,V_q\,x_k + B_q v_k) + B_{q-1}v_k) + ...) + B_1 v_k) \\ y_k &= \Lambda_1(x_k,v_k)(W_1\Lambda_2(x_k,v_k)(W_2...\Lambda_q(x_k,v_k)(W_q x_k + D_q v_k) + D_{q-1}v_k) + ...) + D_1 v_k) \end{cases} \tag{13}$$

where Γ_i, Λ_i ($i = 1,...,q$) are diagonal matrices with diagonal elements $\gamma_j(x_k,v_k)$, $\lambda_j(x_k,v_k) \in [0,1]$, depending continuously on the variables x_k, v_k.

Remark. The term 'NL$_q$' system refers to the alternating sequence of nonlinear and linear operators acting on the input arguments x_k, v_k: the diagonal matrices Γ_i, Λ_i depend on x_k, v_k and refer to 'N' in 'NL$_q$', the constant matrices V_i, W_i refer to 'L' in 'NL$_q$'. The index q refers to the number of nonlinear operations 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. Special cases for $q = 1, 2$ are of the form

$$\begin{aligned} q=1: & \begin{cases} p_{k+1} &= \Gamma_1(\,V_1\,p_k + B_1 w_k) \\ e_k &= \Lambda_1(W_1 p_k + D_1 w_k) \end{cases} \\ q=2: & \begin{cases} p_{k+1} &= \Gamma_1(\,V_1\,\Gamma_2(\,V_2\,p_k + B_2 w_k) + B_1 w_k) \\ e_k &= \Lambda_1(W_1\Lambda_2(W_2 p_k + D_2 w_k) + D_1 w_k) \end{cases} \end{aligned} \tag{14}$$

with $p_k \in \mathbb{R}^{n_p}$, $w \in \mathbb{R}^{n_w}$, $e_k \in \mathbb{R}^{n_e}$. The motivation for the definition of NL$_q$s is the following. Consider a discrete time Hopfield network (with synchronous updating)

$$x_{k+1} = \tanh(W x_k). \tag{15}$$

In Suykens *et al.* (1993)(1994b) it is shown that (15) can be written as

$$x_{k+1} = \Gamma(x_k) W x_k \tag{16}$$

with $\Gamma = \mathrm{diag}\{\gamma_i\}$ and $\gamma_i = \tanh(w_i^T x_k)/(w_i^T x_k)$. This is easily obtained by using an elementwise notation. Eqn.(15) becomes then

$$\begin{aligned} x^i &:= \tanh(\textstyle\sum_j w_j^i x^j) \\ &:= \frac{\tanh(\sum_j w_j^i x^j)}{\sum_j w_j^i x^j} \cdot \textstyle\sum_j w_j^i x^j \\ &:= \gamma_i^i \textstyle\sum_j w_j^i x^j. \end{aligned} \tag{17}$$

The time index can be omitted here because of the assignment operator ' :='. The notation γ_i^i means that this corresponds to the diagonal matrix $\Gamma(x_k)$. In Suykens *et al.* (1993) this became evident after using the Einstein summation convention for (15).
Also if an additional layer is considered

$$x_{k+1} = \tanh(V \tanh(W x_k)), \tag{18}$$

this can be written as

$$x_{k+1} = \Gamma_1(x_k) V \Gamma_2(x_k) W x_k \tag{19}$$

where $\Gamma_1 = \mathrm{diag}\{\gamma_{1_i}\}$ with $\gamma_{1_i} = \tanh(v_i^T \tanh(W x_k))/(v_i^T \tanh(W x_k))$ and $\Gamma_2 = \mathrm{diag}\{\gamma_{2_i}\}$ with $\gamma_{2_i} = \tanh(w_i^T x_k)/(w_i^T x_k)$. Again this becomes clear by writing (18) elementwise:

$$\begin{aligned} x^i &:= \tanh(\textstyle\sum_j v_j^i \tanh(\sum_l w_l^j x^l) \\ &:= \tanh(\textstyle\sum_j v_j^i \gamma_{2_j}^j \sum_l w_l^j x^l) \\ &:= \gamma_{1_i}^i \textstyle\sum_j v_j^i \gamma_{2_j}^j \sum_l w_l^j x^l. \end{aligned} \tag{20}$$

All systems treated in this paper, will be converted into NL_qs.
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_qs]. Neural state space models $\mathcal{M}_i$ ($i \in \{0,1,...,3\}$) 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 $i=3$ by taking the input $v_k=[u_k;\epsilon_k;1]$, output y_k and state x_k for the NL_q.

Proof: See Appendix B. By taking an elementwise notation, multiplying and dividing by the arguments of the activation functions, the alternating sequence of nonlinear and linear operators in (13) is obtained. The elements γ_i, λ_i are in $[0,1]$ because the activation function tanh is a nonlinearity belonging to sector [0,1].

□

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

Lemma 2 [Ξ_j^i's are NL_qs]. All members of the family of neural control problems Ξ_j^i can be written as NL_q systems.

Proof: An exhaustive list is presented in Appendix C. A trick of state augmentation is used in order to achieve the result.

□

Lemma 2 is illustrated with Ξ_1^0 and Ξ_2^2 below

Example 1. Given the Kalman filter ($\mathcal{M}_0$) and a linear dynamic output feedback controller with saturation ($\mathcal{C}_1$)

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

$$\mathcal{C}_1: \begin{cases} z_{k+1} &= Ez_k + Fy_k + F_2 d_k \\ u_k &= \tanh(Gz_k) \end{cases}$$

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

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

This state augmentation leads to the following equivalent system

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

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 $\Gamma_1 = \mathrm{diag}\{I, I, \Gamma_G(\xi_k, z_k, \hat{x}_k, d_k, \epsilon_k)\}$, where Γ_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 $\mathcal{C}_2$

$$\mathcal{M}_4: \begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_B u_k + \beta_{AB}) + K\epsilon_k \\ y_k &= W_C\tanh(V_C\hat{x}_k) + \epsilon_k \end{cases}$$

$$\mathcal{C}_2: \begin{cases} z_{k+1} &= W_{EF}\tanh(V_E z_k + V_F y_k + V_{F_2} d_k + \beta_{EF}) \\ u_k &= W_G\tanh(V_G z_k) \end{cases}$$

the state equation for the closed loop system is

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

State augmentation is done after defining

$$\begin{aligned} \xi_k &= \tanh(V_C\hat{x}_k) \\ \eta_k &= \tanh(V_G z_k), \end{aligned}$$

leading to the equivalent system

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

and can be written as an 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

$$V_1 = \begin{bmatrix} W_{AB} & & & \\ & W_{EF} & & \\ & & V_C W_{AB} & \\ & & & V_G W_{EF} \end{bmatrix}, V_2 = \begin{bmatrix} V_A & 0 & 0 & V_B W_G \\ 0 & V_E & V_F W_C & 0 \\ V_A & 0 & 0 & V_B W_G \\ 0 & V_E & V_F W_C & 0 \end{bmatrix}$$

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

If all bias terms are zero, it is possible that an NL_q with a lower q value may be obtained by applying another state augmentation. Other examples of NL_qs arising in the field of neural networks are presented in Section 10.

6 Global asymptotic stability criteria for NL_qs

In this section the autonomous case of NL_q systems for the standard plant configuration

$$\begin{cases} p_{k+1} &= \Gamma_1(p_k, w_k)(V_1 \Gamma_2(p_k, w_k)(V_2 ... \Gamma_q(p_k, w_k)(V_q p_k + B_q w_k) + B_{q-1} w_k) + ...) + B_1 w_k) \\ e_k &= \Lambda_1(p_k, w_k)(W_1 \Lambda_2(p_k, w_k)(W_2 ... \Lambda_q(p_k, w_k)(W_q p_k + D_q w_k) + D_{q-1} w_k) + ...) + D_1 w_k) \end{cases} \tag{21}$$

is considered ($w_k = 0$). Conditions for global asymptotic stability are derived. It it explained how the well known discrete time Lur'e problem from classical nonlinear control theory fits into the current framework.

6.1 Stability criteria

Theorem 1 [Diagonal scaling]. A sufficient condition for global asymptotic stability of the autonomous NL_q system ($w_k = 0$)

$$p_{k+1} = \prod_{i=1}^{q} \Gamma_i(p_k) V_i \, p_k \tag{22}$$

is to find diagonal matrices D_i such that

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

where

$$V_{tot} = \begin{bmatrix} 0 & V_2 & & & 0 \\ & 0 & V_3 & & \\ & & \ddots & & \\ & & & 0 & V_q \\ V_1 & & & & 0 \end{bmatrix}, V_i \in \mathbb{R}^{n_{h_i} \times n_{h_{i+1}}}, n_{h_1} = n_{h_{q+1}} = n_p$$

- $D_{tot} = \mathrm{diag}\{D_2, D_3, ..., D_q, D_1\}$
- $D_i \in \mathbb{R}^{n_{h_i} \times n_{h_i}}$ diagonal matrices with nonzero diagonal elements.

Proof: See Appendix B. 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. □

Remarks.

- The equilibrium point $p = 0$ in (22) is unique because the Lyapunov function is radially unbounded.
- The minimum over D_{tot} yields a maximal contraction rate of the flow for the NL_q system and shows how such a feasible point D_{tot} is found in practice

$$\min_{D_{tot}} \beta_D. \tag{24}$$

- Given a constant matrix V_{tot} the criterion (24) 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 (23) in the field of modern robust control theory (e.g. Boyd & Yang (1989), Doyle (1982), 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 8.

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.

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:

$$q_{ii} > \delta_Q . \sum_{j=1\,(j \neq i)}^{n} |q_{ij}| \quad , \quad \forall i = 1, ..., n. \tag{25}$$

Remark. This definition is consistent with the original definition of diagonal dominance (see e.g. 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 NL_q system ($w_k = 0$)

$$p_{k+1} = \prod_{i=1}^{q} \Gamma_i(p_k) V_i \, p_k \tag{26}$$

is to find matrices P_i, N_i such that

$$\prod_{i=1}^{q} (1+\alpha_i)^{1/2} \ \|P_{tot} V_{tot} P_{tot}^{-1}\|_2^q = c_\alpha \beta_P < 1 \tag{27}$$

where

- $c_\alpha = \prod_{i=1}^{q}(1+\alpha_i)^{1/2}$, $\beta_P = \|P_{tot} V_{tot} P_{tot}^{-1}\|_2^q$
- $$V_{tot} = \begin{bmatrix} 0 & V_2 & & & 0 \\ & 0 & V_3 & & \\ & & \ddots & & \\ & & & 0 & V_q \\ V_1 & & & & 0 \end{bmatrix}, V_i \in \mathbb{R}^{n_{h_i} \times n_{h_{i+1}}}, n_{h_1} = n_{h_{q+1}} = n_p$$
- $P_{tot} = \text{blockdiag}\{P_2, P_3, ..., P_q, P_1\}$, $P_i \in \mathbb{R}^{n_{h_i} \times n_{h_i}}$ of full rank.
- $Q_i = P_i^T P_i N_i$ are diagonal dominant matrices with $\delta_{Q_i} = (1+\alpha_i)/\alpha_i \geq 1$ ($\alpha_i \geq 0$) and N_i diagonal with positive diagonal elements.

Proof: See Appendix B. 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 δ_{Q_i} are derived from Gerschgorin's Theorem (see Wilkinson (1965)).

□

Lemma 3. Given a matrix $Q \in \mathbb{R}^{n \times n}$, the condition diagonal dominance of level δ_Q (25)

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

is equivalent to the condition

$$\|X_Q\|_\infty < 1/\delta_Q \tag{28}$$

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 B. It follows immediately from the definition of $\|X_Q\|_\infty$.

□

Remarks.

- Again the equilibrium point $p = 0$ in (26) is unique because the Lyapunov function is radially unbounded.

- The solution to optimization problem

$$\min_{P_i, N_i} c_\alpha \beta_P \tag{29}$$

yields a maximal contraction rate of the flow of the autonomous NL_q system with respect to the condition (27).

- 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 (27) can be interpreted as follows: search 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 α_i for $q = 1$ in Fig.6. There exists a trade-off between δ_Q and c_α: the lower the level of diagonal dominance on the 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_α is infeasible.

- Elimination of α_i in (27), 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 $l \in \{1, 2, ..., q\}$ for which $\Gamma_l = I$ there is no condition of diagonal dominance on Q_l, but only on the remaining Q_i's with $i \neq l$. This follows immediately from the proof of the Theorem. Examples of this are the neural state space control problems Ξ_0^1 and Ξ_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

$$\min_{P \in \mathbb{R}^{n \times n}} \|PAP^{-1}\|_2 < 1 \tag{30}$$

because $\Gamma_1 = I$ and by taking the Lyapunov function $V(p) = \|Pp\|_2$. 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 (30) is also necessary.

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

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

to hold $\forall p$ is $\delta_Q = 1$, taking $N_1 = I$. Hence for that specific activation function (27) becomes for $q = 1$

$$\|PVP^{-1}\|_2 < 1 \tag{31}$$

subject to

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

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

6.2 Discrete time Lur'e problem

It will be shown now how the classical discrete time Lur'e problem can be interpreted within the current neural state space model framework as an NL_1 system.
The discrete time Lur'e problem can be stated as follows (Fig.7): consider a stable linear system connected by feedback to a memoryless nonlinearity $f(\sigma)$, that is single valued, monotone increasing and belongs to a sector $[0, K]$, which means that

$$0 < \frac{f(\sigma)}{\sigma} < K \ (\sigma \neq 0), \qquad f(0) = 0. \tag{32}$$

Early work on this problem was done around 1960 e.g. by Tsypkin, Szegö, Pearson and Gibson, Jury and Lee. For a survey on this subject see e.g. Brockett (1966) and Vidal (1969). The following Theorem by Tsypkin can be seen as the closest analog for discrete time systems to Popov's result for the corresponding continuous time case.

Theorem 3 [Tsypkin]. Let $n(z)$ and $d(z)$ denote polynomials without common factors and let

$$\left\{ \begin{array}{rcl} x_{k+1} & = & Ax_k + bu_k \\ y_k & = & c^t x_k \end{array} \right.$$

be an irreducible representation of $G(z) = n(z)/d(z)$. If the zeros of $d(z)$ lie inside the disk $|z| = 1$, then the null solution of

$$x_{k+1} = Ax_k - bf(c^t x_k)$$

is asymptotically stable in the large, if $0 \leq f(y)/y \leq K$ and

$$Re[G(z)] + \frac{1}{K} \geq 0 \ \forall z : |z| = 1 \tag{33}$$

where $G(z) = c^T(A - zI)^{-1}b$.

Proof: See Jury (1974).

□

Within the neural state space model framework the Lur'e problem can be formulated as follows:

Theorem 4 [Lur'e problem as NL_1 system]. Given the discrete time Lur'e problem for a multivariable system with m inputs and m outputs, this can be formulated within the neural state space model framework as

$$\begin{array}{rl} \mathcal{M}_0: & \left\{ \begin{array}{rcl} x_{k+1} & = & Ax_k + Bu_k \\ y_k & = & Cx_k \end{array} \right. \\ \mathcal{C}: & u_k = -f(y_k) \end{array} \tag{34}$$

with memoryless, single valued, monotone increasing nonlinearity $f(\sigma)$, belonging to a sector $[0, K]$. One of the following conditions is sufficient for global asymptotic stability of the system (34)

- (Diagonal scaling)

$$K < 1/\beta_D \tag{35}$$

with

$$\beta_D = \min_D \|DV_{tot}D^{-1}\|_2$$

- (Diagonal dominance)

$$K < 1/(c_\alpha \beta_P) \tag{36}$$

with

$$c_\alpha \beta_P = \min_{P,N} (\delta_Q/(\delta_Q - 1))^{1/2} \|PV_{tot}P^{-1}\|_2$$

where

$$V_{tot} = \left[\begin{array}{cc} A & -B \\ CA & -CB \end{array} \right].$$

D is a diagonal matrix with nonzero diagonal elements and $Q = P^T PN$ with P a full rank matrix. The level of diagonal dominance of Q is $\delta_Q \geq 1$ and N is a diagonal matrix with positive diagonal elements.

Proof: See Appendix B. The proof is similar to the proof of Theorem 2.

□

Remark. For the special case of $\delta_Q \to \infty$ condition (36) reduces to (35).

7 Input/output properties - l_2 theory

In this section the non-autonomous NL_q system (21) will be studied. First some equivalent representations for NL_qs are formulated. Then an l_2 theory is presented with respect to these representations. The link between input/output stability and internal stability will be clarified through the concept of dissipativity.

7.1 Equivalent representations for NL_qs

Lemma 4 [Equivalent representations for NL_qs]. Equivalent representations for the NL_q system (21) are

$$\left\{ \begin{array}{rcl} p_{k+1} & = & \prod_{i=1}^{q} \Gamma_{i,e}(p_k, w_k) M_i \left[\begin{array}{c} p_k \\ w_k \end{array} \right] \\ \\ e_k & = & \prod_{i=1}^{q} \Lambda_{i,e}(p_k, w_k) N_i \left[\begin{array}{c} p_k \\ w_k \end{array} \right] \end{array} \right. \tag{37}$$

and

$$\left[\begin{array}{c} p_{k+1} \\ e_k^{ext} \end{array} \right] = \prod_{i=1}^{q} \Omega_i(p_k, w_k) R_i \left[\begin{array}{c} p_k \\ w_k \end{array} \right] \tag{38}$$

where for (37)

$$\Gamma_{1,e} = \Gamma_1, \quad \Gamma_{i,e} = \left[\begin{array}{cc} \Gamma_i & 0 \\ 0 & I \end{array} \right], \quad M_1 = [V_1 \ B_1], \quad M_i = \left[\begin{array}{cc} V_i & B_i \\ 0 & I \end{array} \right] \ (i = 2, ..., q)$$

$$\Lambda_{1,e} = \Lambda_1, \quad \Lambda_{i,e} = \left[\begin{array}{cc} \Lambda_i & 0 \\ 0 & I \end{array} \right], \quad N_1 = [W_1 \ D_1], \quad N_i = \left[\begin{array}{cc} W_i & D_i \\ 0 & I \end{array} \right] \ (i = 2, ..., q)$$

and for (38)

$$\Omega_i = \left[\begin{array}{ccc} \Gamma_{i,e} & 0 & 0 \\ 0 & \Lambda_{i,e} & 0 \\ 0 & 0 & 0 \end{array} \right], R_i = \left[\begin{array}{ccc} M_i & 0 & 0 \\ 0 & N_i & 0 \\ 0 & 0 & 0 \end{array} \right], R_q = \left[\begin{array}{c} M_q \\ N_q \\ 0 \end{array} \right] \ (i = 1, ..., q-1)$$

and 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 $\prod_{i=1}^{q} R_i$ square.

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:

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

with

$$\left\{ \begin{array}{rcl} p_{k+1} & = & \Gamma_1 [V_1 \ B_1] \left[\begin{array}{cc} \Gamma_2 & 0 \\ 0 & I \end{array} \right] \left[\begin{array}{cc} V_2 & B_2 \\ 0 & I \end{array} \right] \left[\begin{array}{c} p_k \\ w_k \end{array} \right] \\ \\ e_k & = & \Lambda_1 [W_1 \ D_1] \left[\begin{array}{cc} \Lambda_2 & 0 \\ 0 & I \end{array} \right] \left[\begin{array}{cc} W_2 & D_2 \\ 0 & I \end{array} \right] \left[\begin{array}{c} p_k \\ w_k \end{array} \right] \end{array} \right.$$

and

$$\left[\begin{array}{c} p_{k+1} \\ e_k^{ext} \end{array} \right] = \left[\begin{array}{ccc} \Gamma_1 & \Lambda_1 & 0 \end{array} \right] \left[\begin{array}{ccc} V_1 & B_1 & \\ & & W_1 \ D_1 \\ & & 0 \end{array} \right] \left[\begin{array}{ccc} \Gamma_2 & & \\ & \Lambda_2 & \\ & & 0 \end{array} \right] \left[\begin{array}{cc} V_2 & B_2 \\ & I \\ W_2 & D_2 \\ & I \\ 0 & 0 \end{array} \right] \left[\begin{array}{c} p_k \\ w_k \end{array} \right].$$

7.2 Main Theorems

The following Theorems are formulated for NL_qs with respect to their equivalent representations.

Theorem 5 [l_2 theory - Diagonal scaling]. Given the representation (38)

$$\left[\begin{array}{c} p_{k+1} \\ e_k^{ext} \end{array} \right] = \prod_{i=1}^{q} \Omega_i(p_k, w_k) R_i \left[\begin{array}{c} p_k \\ w_k \end{array} \right]$$

for an NL_q system, if there exist matrices D_i such that

$$\|D_{tot} R_{tot} D_{tot}^{-1}\|_2^q = \beta_D < 1, \tag{39}$$

then the following holds

1. For finite N:
$$\sum_{k=0}^{N-1} \|e_k\|_2^2 \leq \beta_D^2 \sum_{k=0}^{N-1} \|w_k\|_2^2 + r_0^2 \tag{40}$$
where $r_0 = \|D_1 p_0\|_2$.

2. For $N \to \infty$:
There exist constants c_1, c_2 such that
$$c_2(1 - \beta_D^2)\|p\|_2^2 + \|e\|_2^2 \leq \beta_D^2 \|w\|_2^2 + c_1 \|p_0\|_2^2 \tag{41}$$

provided that $\{w_k\}_{k=0}^{\infty} \in l_2$. Hence also

$$\|e\|_2^2 \leq \beta_D^2 \|w\|_2^2 + c_1 \|p_0\|_2^2$$

holds in that case.

Here

- $$R_{tot} = \begin{bmatrix} 0 & R_2 & & 0 \\ & 0 & R_3 & \\ & & \ddots & \\ & & & 0 \quad R_q \\ R_1 & & & 0 \end{bmatrix}, R_i \in \mathbb{R}^{n_{r_i} \times n_{r_{i+1}}}, n_{r_1} = n_{r_{q+1}} = n_p + n_w$$
- $D_{tot} = \text{diag}\{D_2, D_3, ..., D_q, D_{S_1}\}$
- $D_{S_1} = \text{diag}\{D_1, I_{n_w}\}$
- $D_1 \in \mathbb{R}^{n_p \times n_p}$, $D_i \in \mathbb{R}^{n_{r_i} \times n_{r_i}}$ diagonal matrices with nonzero diagonal elements.

Proof: See Appendix B. The proof is similar to the proof given in Packard & Doyle (1993) related to the state space upper bound test for the robust performance problem in μ-theory.

□

Theorem 6 [l_2 theory - Diagonal dominance]. Given the representation (38) in Theorem 5, if there exist matrices P_i, N_i such that

$$\prod_{i=1}^{q} (1+\alpha_i)^{1/2} \|P_{tot} R_{tot} P_{tot}^{-1}\|_2^q = c_\alpha \beta_P < 1 \tag{42}$$

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

1. For finite N:
$$\sum_{k=0}^{N-1} \|e_k\|_2^2 \leq c_\alpha^2 \beta_P^2 \sum_{k=0}^{N-1} \|w_k\|_2^2 + r_0^2 \tag{43}$$
where $r_0 = \|D_1 p_0\|_2$.

2. For $N \to \infty$:
There exist constants c_1, c_2 such that
$$c_2(1 - c_\alpha^2 \beta_P^2)\|p\|_2^2 + \|e\|_2^2 \leq c_\alpha^2 \beta_P^2 \|w\|_2^2 + c_1 \|p_0\|_2^2 \tag{44}$$
provided that $\{w_k\}_{k=0}^{\infty} \in l_2$. Hence also
$$\|e\|_2^2 \leq c_\alpha^2 \beta_P^2 \|w\|_2^2 + c_1 \|p_0\|_2^2$$
holds in that case.

where

- $P_{tot} = \text{blockdiag}\{P_2, P_3, ..., P_q, P_{S_1}\}$
- $P_{S_1} = \text{blockdiag}\{P_1, I_{n_w}\}$
- $Q_i = P_i^T P_i N_i$ are diagonal dominant matrices with $\delta_{Q_i} = (1+\alpha_i)/\alpha_i \geq 1$, $(\alpha_i \geq 0)$, N_i are diagonal matrices with positive diagonal elements and P_i are of full rank.

Proof: See Appendix B. The proof is based on the proofs of Theorems 2 and 3.

□

Remarks.

- The case $N \to \infty$ only applies to systems without bias vectors, because the bias vectors correspond to a constant input 1 in the exogenous input w_k. This constant input belongs not to the class l_2. The finite N case is still applicable.
- For a linear control system Ξ_0^0 Theorem 6 is related to the H_∞ control problem with dynamic output feedback (see e.g. Stoorvogel (1992)). In this case is $\Omega_1 = I$ and then it follows immediately from the proof of the Theorem that condition (42) becomes
$$\left\| \begin{bmatrix} P_1 & 0 \\ 0 & I \end{bmatrix} R_1 \begin{bmatrix} P_1 & 0 \\ 0 & I \end{bmatrix}^{-1} \right\|_2 < 1. \tag{45}$$
Now it is precisely a condition like (45) that can be found in literature in the context of the H_∞ norm for a linear dynamical system (see e.g. Packard & Doyle (1993)). In that paper it is stated that for a linear system
$$\begin{cases} p_{k+1} &= A_R p_k + B_R w_k \\ e_k &= C_R p_k + D_R w_k \end{cases}$$
the following five conditions are equivalent to condition (45), assuming $R_1 = [A_R \; B_R; C_R \; D_R]$ and the system is stable ($\rho(A_R) < 1$):
  1. $\|D_R + C_R(zI - A_R)^{-1} B_R\|_\infty < 1$
  2. S has no eigenvalues on the unit circle and $\|D_R + C_R(I - A_R)^{-1} B_R\|_2 < 1$
  3. $\exists X \geq 0$ solution to the Riccati equation
  $$E^T X E - X - E^T X G (I + XG)^{-1} X E + Q = 0$$
  with $I - D_R^T D_R - B_R^T X B_R > 0$ and $(I + GX)^{-1} E$ stable.
  4. $\exists X > 0$ such that
  $$E^T X E - X - E^T X G (I + XG)^{-1} X E + Q < 0$$
  and $I - D_R^T D_R - B_R^T X B_R > 0$.

5. $\exists X > 0$ such that

$$R_1^T \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix} R_1 - \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix} < 0$$

where by definition

$$\begin{aligned} E &= A_R + B_R(I - D_R^T D_R)^{-1} D_R^T C_R \\ G &= -B_R(I - D_R^T D_R)^{-1} B_R^T \\ Q &= C_R^T(I - D_R^T D_R)^{-1} C_R. \end{aligned}$$

It is supposed that E is nonsingular and a symplectic matrix S is defined as

$$\begin{bmatrix} E + GE^{-T}Q & -GE^{-T} \\ -E^{-T} & E^{-T} \end{bmatrix}.$$

A comparison between Theorem 1 and 5 and between Theorem 2 and 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 & Moylan (1980) and van der Schaft (1992) (but in a continuous time framework). The following definition related to discrete time systems is taken from Byrnes & Lin (1994).

Definition 6 [Dissipativity]. The dynamic system (21) 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 $\forall k \geq 0$:

$$V(p_{k+1}) - V(p_k) \leq W(e_k, w_k) \tag{46}$$

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

Lemma 5. The NL_q system (21) is dissipative under the condition of Theorem 5, 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 6 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 $c_\alpha \beta_P < 1$.

Proof: See Appendix B. □

Remark. Several kinds of dissipativity exist in general, depending on the supply rate. A system with supply rate $W = e_k^T w_k$ is called passive, while a supply rate $W = \kappa \|w_k\|_2^2 - \|e_k\|_2^2$ stands for finite l_2-gain κ (see Hill & Moylan (1976)(1977)(1980)).

8 Robust performance problem

Conditions for robust global asymptotic stability and robust performance with respect to parametric uncertainties on the interconnection weights of the neural state space model and/or controller will be derived now. It will be shown in what sense μ-theory (a modern theory in linear robust control) can be considered as a special case of the neural state space control theory.

8.1 Perturbed NL_qs

The conditions for global asymptotic stability and finite l_2-gain of NL_q systems will be extended now with respect to parametric uncertainties on the interconnection weights of the neural state space models. 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 on R_l with $R_l = R_l^{nom} + \Delta R_l$ on the NL_q (38)

$$\begin{bmatrix} p_{k+1} \\ e_k^{ext} \end{bmatrix} = \prod_{i=1}^{l-1} \Omega_i R_i \, \Omega_l (R_l^{nom} + \Delta R_l) \prod_{j=l+1}^{q} \Omega_j R_j \begin{bmatrix} p_k \\ w_k \end{bmatrix}. \tag{47}$$

Suppose this perturbation on R_l is through one of the interconnection matrices $X \in \mathbb{R}^{n \times m}$ and that the elements x_{ij} belong to bounded intervals: $x_{ij} \in [\underline{x}_{ij}, \overline{x}_{ij}]$. According to Steinbuch *et al.* (1992), by defining

$$\begin{aligned} x_{ij}^{nom} &= (\overline{x}_{ij} + \underline{x}_{ij})/2 \\ r_{ij} &= (\overline{x}_{ij} - \underline{x}_{ij})/2 \end{aligned} \tag{48}$$

one obtains

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

The matrix $X = X^{nom} + \Delta X$ can be written with

$$\Delta X = S_x \Delta_x R_x \tag{49}$$

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

$$\begin{aligned} \Delta_x &= \mathrm{diag}\{\delta_{11}, ..., \delta_{1m}, \delta_{21}, ..., \delta_{2m}, \delta_{n1}, ..., \delta_{nm}\} \\ S_x &= \mathrm{blockdiag}\{1_{1\times m}, ..., 1_{1\times m}\} \\ R_x &= [\mathrm{diag}\{r_{1i}\}; \mathrm{diag}\{r_{2i}\}; ...; \mathrm{diag}\{r_{ni}\}], \ (i = 1, ..., m). \end{aligned} \tag{50}$$

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

$$R_l(X) = R_l^{nom} + U_{R_l} \Delta_x V_{R_l}. \tag{51}$$

The perturbed system (47) can then be written as

$$\left\{ \begin{array}{rcl} \left[\begin{array}{c} p_{k+1} \\ e_k^{ext} \\ s_k \end{array} \right] & = & \prod_{i=1}^{q} \Upsilon_i T_i \left[\begin{array}{c} p_k \\ w_k \\ r_k \end{array} \right] \\ r_k & = & \Delta_x s_k, \qquad \|\Delta_x\| \leq 1 \end{array} \right. \tag{52}$$

with

$$\Upsilon_i = \left[\begin{array}{cc} \Omega_i & \\ & I \end{array} \right], T_i = \left[\begin{array}{cc} R_i & \\ & I \end{array} \right] (i \neq l), T_l = \left[\begin{array}{cc} R_l^{nom} & U_{R_l} \\ V_{R_l} & 0 \end{array} \right]$$

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

Lemma 6 [Perturbed NL_qs as NL_{q+1}s]. The perturbed NL_q (52) 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

$$\left[\begin{array}{c} p_{k+1} \\ e_k^{ext} \\ s_k \end{array} \right] = \prod_{i=1}^{q} \Upsilon_i T_i \left[\begin{array}{ccc} I & & \\ & I & \\ & & \Delta_x \end{array} \right] I \left[\begin{array}{c} p_k \\ w_k \\ s_k \end{array} \right] \tag{53}$$

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

Proof: Eliminate r_k in (52).

□

First a simple example is given here in order to illustrate (48)-(50).

Example 1. For a 2×2 matrix X in (49) is given by

$$\left[\begin{array}{cc} x_{11} & x_{12} \\ x_{21} & x_{22} \end{array} \right] = \left[\begin{array}{cc} x_{11}^{nom} & x_{12}^{nom} \\ x_{21}^{nom} & x_{22}^{nom} \end{array} \right] + \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array} \right] \left[\begin{array}{cccc} \delta_{11} & & & \\ & \delta_{12} & & \\ & & \delta_{21} & \\ & & & \delta_{22} \end{array} \right] \left[\begin{array}{cc} r_{11} & 0 \\ 0 & r_{12} \\ r_{21} & 0 \\ 0 & r_{22} \end{array} \right].$$

The following example illustrates the perturbed NL_q (47).

Example 2. The Ξ_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. According to Appendix C this can be written as

$$\left[\begin{array}{c} x_{k+1} \\ z_{k+1} \\ \xi_{k+1} \end{array} \right] = \Gamma_1 \left[\begin{array}{ccc} A & 0 & B \\ F(C + \Delta C) & E & FD \\ GF(C + \Delta C) & GE & GFD \end{array} \right] \left[\begin{array}{c} x_k \\ z_k \\ \xi_k \end{array} \right]$$

or in the form (52) as

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

with

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

This is easily verified by straightforward calculation.

The following Theorem can then be proven with respect to the perturbed NL_q system.

Theorem 7 [l_2 theory - perturbed NL_q system]. Given the perturbed NL_q system (52), if there exist matrices P_i, N_i such that

$$\prod_{i=1}^{q} (1 + \alpha_i)^{1/2} \|P_{tot} T_{tot} P_{tot}^{-1}\|_2^q = c_\alpha \beta_P < 1 \tag{54}$$

with $c_\alpha = \prod_{i=1}^{q} (1 + \alpha_i)^{1/2}$ and $\beta_P = \|P_{tot} T_{tot} P_{tot}^{-1}\|_2^q$ then

1. For finite N:
$$\sum_{k=0}^{N-1} \|e_k\|_2^2 \leq c_\alpha^2 \beta_P^2 \sum_{k=0}^{N-1} \|w_k\|_2^2 + r_0^2 \tag{55}$$
where $r_0 = \|D_1 p_0\|_2$.

2. For $N \to \infty$:
There exist constants c_1, c_2 such that
$$c_2 (1 - c_\alpha^2 \beta_P^2) \|p\|_2^2 + \|e\|_2^2 \leq c_\alpha^2 \beta_P^2 \|w\|_2^2 + c_1 \|p_0\|_2^2 \tag{56}$$
provided that $\{w_k\}_{k=0}^{\infty} \in l_2$. Hence also
$$\|e\|_2^2 \leq c_\alpha^2 \beta_P^2 \|w\|_2^2 + c_1 \|p_0\|_2^2$$
holds in that case.

for all real diagonal matrices Δ_x that satisfy the property $\|\Delta_x\| \leq 1$, where

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

- $P_{tot} = \text{blockdiag}\{P_2, P_3, ..., P_q, P_{S_1}\}$
- $P_{S_1} = \text{blockdiag}\{P_1, I, D_1\}$, D_1 diagonal with nonzero diagonal elements.
- $Q_i = P_i^T P_i N_i$ diagonal dominant matrices with $\delta_{Q_i} = (1+\alpha_i)/\alpha_i \geq 1 \quad (\alpha_i \geq 0)$, N_i diagonal matrices with positive diagonal elements and P_i of full rank.

Proof: See Appendix B. The proof is similar to the proof of Theorem 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.

8.2 Connections with μ-theory

In μ-theory, a modern approach in linear robust control, one analyses the influence of parametric uncertainties, unmodeled 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 LFTs (Linear Fractional Transformations). Related to the standard plant a state space formulation for an LFT looks like

$$\begin{cases} \begin{bmatrix} x_{k+1} \\ e_k^{ext} \\ s_k \end{bmatrix} = M \begin{bmatrix} x_k \\ w_k \\ r_k \end{bmatrix} \\ r_k = \Delta\, s_k \end{cases} \tag{57}$$

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 $\Delta \in \mathbf{\Delta} \subset \mathbb{C}^{n_{tot} \times n_{tot}}$ (see Doyle (1982), Packard & Doyle (1993))

$$\mathbf{\Delta} = \{\text{diag}[\delta_1 I_{r_1}, ..., \delta_s I_{r_s}, \Delta_{s+1}, ..., \Delta_{s+f}] : \delta_i \in \mathbb{C}, \Delta_{s+j} \in \mathbb{C}^{m_j \times m_j}; i = 1, ..., s; j = 1, ..., f\} \tag{58}$$

where $\sum_{i=1}^{s} r_i + \sum_{j=1}^{f} m_j = n_{tot}$ for consistency. Remark 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 lead to complex values. The uncertainties can be of time-invariant, time-variant or nonlinear nature. For time-invariant uncertainties Δ

a necessary and sufficient condition for robust performance (finite l_2-gain smaller than one, regardless of any Δ belonging to $\mathbf{B}_\Delta = \{\Delta \in \mathbf{\Delta} : \|\Delta\|_2 \leq 1\}$):

$$\|e\|_2 \leq \beta \|w\|_2 \qquad \beta \in [0, 1) \tag{59}$$

for zero-initial-state-response is

$$\mu_{\Delta_S}(M) < 1 \tag{60}$$

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

$$\rho(M) \leq \max_{Q \in \mathcal{Q}} \rho(QM) \leq \max_{\Delta \in \mathbf{B}_\Delta} \rho(\Delta M) = \mu_\Delta(M) \leq \min_{D \in \mathcal{D}} \|DMD^{-1}\|_2 \leq \|M\|_2 \tag{61}$$

where $\mathcal{Q} = \{Q \in \Delta : Q^*Q = I_n\}$, $\mathcal{D} = \{\text{diag}[D_1, ..., D_s, d_{s+1} I_{m_1}, ..., d_{s+f} I_{m_f}] : D_i \in \mathbb{C}^{r_i \times r_i}, D_i = D_i^* > 0, d_{s+j} \in \mathbb{R}, d_{s+j} > 0\}$.

The upper bound of μ corresponds to diagonal scaling of M, which is similar to the diagonal scaling in Theorem 1 or 5. Working with this upper bound instead of μ, guarantees also 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 (57), if

$$\min_{D_S} \|D_S M D_S^{-1}\|_2 = \beta < 1 \tag{62}$$

with $D_S = \text{diag}\{D_1, d_2 I, D\}$ and D_1, D diagonal matrices with positive elements, then the following holds for zero initial state

$$\|e\|_2^2 \leq \beta^2 \|w\|_2^2 \tag{63}$$

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

Proof: See Packard & Doyle (1993).

□

The following Theorem states that for a real diagonal Δ uncertainty block the Theorem by Packard & Doyle is a special case of Theorem 7.

Theorem 9 [Connection between Theorem 7-8]. The LFT (57) is a special case of the perturbed NL_q system (52) 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 Δ is real

5. Diagonal Δ block: $r_i = 0(\forall i)$, $m_j = 1(\forall j)$, $s = 0$, $f = n_{tot}$ in (58).

Proof: Readily checked by comparing (52)(54) with (57)(58)(62).

□

9 Control design

9.1 General aspects

In the previous sections conditions for global asymptotic stability and finite l_2-gain were derived for control systems, parametrized by neural state space models and controllers. The final goal however is to design a neural state space controller, given an identified neural state space model. In modern *linear* control theory control 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 to search within this class for a controller with optimal performance. Many problems in linear control theory also lead to *convex* optimization problems, which have a unique minimum that can be found in polynomial time (see Boyd *et al.* (1990)(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 learned based on this emulator. In many cases the neural controller is learned then with respect to a specific reference input. The disadvantage of this method is that many open problems exist around stability of the closed loop system or how the controller would generalize to other reference inputs.

The present *neural state space control theory* is a compromise between those two extremes: the neural controller can be learned 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 (like this is done in H_∞ 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.

9.2 Tracking of a specific reference input

Like for the learning algorithm related to the neural state space models $\mathcal{M}_i$ in Section 3, Narendra's sensitivity model approach can also be applied in order to generate the gradient of the cost function in the following tracking problem

$$\min_{\theta_c} J(\theta_c) = \frac{1}{2N} \sum_{k=1}^{N} \{[d_k - \hat{y}_k(\theta_c)]^T [d_k - \hat{y}_k(\theta_c)] + \lambda . u_k(\theta_c)^T u_k(\theta_c)\} \tag{64}$$

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 $\theta_c \in \mathbb{R}^{p_c}$ contains the interconnection weights of the neural controller. λ is a constant and N is the time horizon. Suppose the closed loop system for a given Ξ_j^i problem of Fig.1 is expressed as

$$\begin{cases} p_{k+1} &= \zeta(p_k, d_k; \alpha) \\ \hat{y}_k &= \chi(p_k, d_k; \beta) \\ u_k &= \vartheta(p_k, d_k; \gamma) \end{cases} \tag{65}$$

instead of in the standard plant form of Fig.3. Here α, β, γ are elements of θ_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 (65) 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 makes then use of the gradient

$$\frac{\partial J}{\partial \theta_c} = \frac{1}{N} \sum_{k=1}^{N} \{[d_k - \hat{y}_k(\theta_c)]^T (-\frac{\partial \hat{y}_k}{\partial \theta_c}) + \lambda . u_k(\theta_c)^T \frac{\partial u_k}{\partial \theta_c}\} \tag{66}$$

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

$$\begin{cases} \frac{\partial p_{k+1}}{\partial \alpha} &= \frac{\partial \zeta}{\partial p_k} . \frac{\partial p_k}{\partial \alpha} + \frac{\partial \zeta}{\partial \alpha} \\ \frac{\partial \hat{y}_k}{\partial \alpha} &= \frac{\partial \chi}{\partial p_k} . \frac{\partial p_k}{\partial \alpha} \\ \frac{\partial \hat{y}_k}{\partial \beta} &= \frac{\partial \chi}{\partial \beta} \\ \frac{\partial u_k}{\partial \gamma} &= \frac{\partial \vartheta}{\partial \gamma}, \end{cases} \tag{67}$$

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

9.3 Analysis problem

9.3.1 Problem statement

Suppose a neural controller in Ξ_j^i has learned to track a specific reference input according to Section 9.2, its properties can then be analyzed 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

$$\|D_{tot} Z_{tot} D_{tot}^{-1}\|_2^q < 1 \tag{68}$$

- *Diagonal dominance:* find matrices P_i, N_i such that

$$c_\alpha \|P_{tot} Z_{tot} P_{tot}^{-1}\|_2^q < 1 \tag{69}$$

and

$$\delta_{Q_i} = \frac{1+\alpha_i}{\alpha_i}, \quad \alpha_i > 0, \quad Q_i = P_i^T P_i N_i.$$

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

$$\begin{array}{rcll} Z_{tot} & = & V_{tot} & : \text{ global asymptotic stability} \\ & = & R_{tot} & : \text{ finite } l_2\text{-gain} < 1 \\ & = & T_{tot} & : \text{ robust performance} \end{array}$$

and $D_{tot} = \text{diag}\{D_2, D_3, ..., D_q, D_{S_1}\}$ with

$$\begin{array}{rcll} D_{S_1} & = & D_1 & : \text{ global asymptotic stability} \\ & = & \text{diag}\{D_1, I\} & : \text{ finite } l_2\text{-gain} < 1 \\ & = & \text{diag}\{D_{1_a}, I, D_{1_b}\} & : \text{ robust performance} \end{array}$$

and $P_{tot} = \text{blockdiag}\{P_2, P_3, ..., P_q, P_{S_1}\}$ with

$$\begin{array}{rcll} P_{S_1} & = & P_1 & : \text{ global asymptotic stability} \\ & = & \text{blockdiag}\{P_1, I\} & : \text{ finite } l_2\text{-gain} < 1 \\ & = & \text{blockdiag}\{P_1, I, D_1\} & : \text{ robust performance} \end{array}$$

9.3.2 Formulation as LMI problem

Both (68) and (69) correspond to nondifferentiable 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 $A^T A - \gamma I < 0$. For the special case of linear matrix inequalities (LMIs) this matrix M depends affinely on x: $M(x) = M_0 + \sum_{i=1}^m x_i M_i$ with $M_0 = M_0^T$, $M_i = M_i^T$. A set of LMIs $M_l(x) < 0$ $(l = 1, .., , L)$ is equivalent to a single LMI by considering $M = \text{blockdiag}\{M_1, ..., M_L\} < 0$ (see Boyd *et al.* (1994)). Finding a

feasible x for such an LMI is then a convex problem, which has a unique minimum. LMIs occur frequently in systems and control problems. The recent book by Boyd *et al.* (1994) gives an excellent overview on this subject. From a computational point of view LMIs are very attractive because efficient algorithms exist for solving it. A general theory of interior-point polynomial-time methods for convex programming is presented in Nesterov & Nemirovskii (1994)[2] : this area has been intensively developed since Karmarkar's famous paper in 1984 on linear programming (Karmarkar (1984)). Other related work is Boyd & El Ghaoui (1993), Vandenberghe & Boyd (1993), Overton (1988).
The problems (68) and (69) will be discussed now in more detail

- **Diagonal scaling:** The LMI which corresponds to (68) is

$$Z_{tot}^T D_{tot}^2 Z_{tot} < D_{tot}^2 \tag{70}$$

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

- **Diagonal dominance:** A set of LMIs will be associated now to the diagonal dominance case (69). In the sequel we have to assume that $N_i = I$ for $i = 1, ..., q$. Condition (69) is formulated then as a convex feasibility problem in the unknown symmetric matrices Q_i, Q_{S_1}

$$Z_{tot}^T Q_{tot} Z_{tot} < \gamma_1 Q_{tot} \tag{71}$$

with $Q_{tot} = P_{tot}^T P_{tot} = \text{blockdiag}\{Q_2, ..., Q_q, Q_{S_1}\}$ such that

$$\delta_{Q_i} > \gamma_{2_i} \ (\forall i \mid \Gamma_i \neq I), \tag{72}$$

where $\gamma_1 \leq 1$, $\gamma_{2_i} \geq 1$ are user defined constants. In order to obtain a set of LMIs the constraint for diagonal dominance (72) is replaced by

$$\left\{ \begin{array}{rcl} Q_i - \gamma_{3_i} I & > & 0 \ (\forall i \mid \Gamma_i \neq I) \\ \|Q_i\|_2 & < & \gamma_{4_i} \ (\forall i \mid \Gamma_i \neq I) \\ Q_i & > & 0, \ (\forall i) \end{array} \right. \tag{73}$$

and the Schur complement form is used for $\|Q_i\|_2 < \gamma_{4_i}$ which is

$$\left[\begin{array}{cc} \gamma_{4_i} I & Q_i \\ Q_i^T & \gamma_{4_i} I \end{array} \right] > 0 \tag{74}$$

A qualitative explanation why the set of LMIs (73) enforces the matrices Q_i to be diagonal dominant can be given based on Gerschgorin's Theorem and Lemma 3. Let

[2]Software for solving LMIs is available in Matlab's LMI lab (The MathWorks Inc. (1994)) and implements one of the algorithms proposed by Nesterov & Nemirovskii.

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 $\sum_{j(j\neq i)} |q_{ij}|$. Now it is clear that for $\gamma_3 > 0$ and if $Q - \gamma_3 I > 0$ holds an increasing value γ_3 enforces Q to be more diagonal dominant. The introduction of an upper bound on $\|Q\|_2$ follows from Lemma 3. Indeed because $Q = D_Q(I + X_Q)$ and $\|X_Q\|_2 \leq \sqrt{n}\|X_Q\|_\infty$ (assuming $Q \in \mathbb{R}^{n\times n}$) and $\|X_Q\|_\infty < 1/\delta_Q$ from (28) one obtains

$$\|Q\|_2 < \|D_Q\|_2(1 + \sqrt{n}/\delta_Q) \tag{75}$$

which means that some upper bound γ_4 on $\|Q\|_2$ is needed. The choice of γ_3 in (73) influences $\|D_Q\|_2$. One may use the following rule of thumb: choose first γ_3 and secondly γ_4 proportionally to γ_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}.

9.4 Synthesis problem

9.4.1 Problem statement

More difficult than the analysis problem is the synthesis problem where the controller parameter vector θ_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

1. **Type 1:**

   - *Diagonal scaling:* feasibility problem in θ_c, D_i such that

   $$Z_{tot}(\theta_c)^T D_{tot}^2 Z_{tot}(\theta_c) < D_{tot}^2 \tag{76}$$

   - *Diagonal dominance:* feasibility problem in θ_c, $Q_i = Q_i^T$ such that

   $$\begin{cases} Z_{tot}(\theta_c)^T Q_{tot} Z_{tot}(\theta_c) < \gamma_1 Q_{tot} \\ Q_i - \gamma_{3_i} I > 0 \\ Q_i > 0 \\ \begin{bmatrix} \gamma_{4_i} I & Q_i \\ Q_i^T & \gamma_{4_i} I \end{bmatrix} > 0 \end{cases} \tag{77}$$

   where $\gamma_1, \gamma_{3_i}, \gamma_{4_i}$ are user defined constants and in (77) $Q_i > 0$ must hold $\forall i$ and the other conditions $\forall i$ for which $\Gamma_i \neq I$ (except Q).

2. **Type 2:** feasible controller θ_c with optimal tracking of a given reference input

   $$\min_{\theta_c, Q_i \text{ or } D_i} J(\theta_c) \tag{78}$$

   such that the type 1 condition (diagonal scaling (76) or diagonal dominance (77)) holds. The cost function J corresponds to (64).

9.4.2 Non-convex nondifferentiable optimization

A framework for solving non-convex nondifferentiable optimization problems with singular value inequality constraints is proposed in Polak & Wardi (1982) in the context of semi-infinite optimization and applies to the two types of problems (76)-(78). Also ellipsoid algorithms have been applied to non-convex problems, although they are normally intended for convex problems (see e.g. Ecker & Kupferschmid (1985), Shor (1985), Boyd & Barratt (1991)). A disadvantage of these algorithms is that they are quite slow and more work remains to be done on quadratically convergent algorithms for such problems.

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

$$\min_{x\in\mathbb{R}^n} \lambda_{max}[M(x)], \quad M = M^T > 0 \tag{79}$$

where λ_{max} is the maximal eigenvalue of M. M denotes the matrix inequality related to (76) or (77). 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

$$[\nabla\lambda_{max}[M(x)]]_i = u_1(x)^T \frac{\partial M(x)}{\partial x_i} u_1(x) \tag{80}$$

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

$$\partial\lambda_{max}[M(x)] = co\{v \in \mathbb{C}^n \mid v_i = (Uz)^* \frac{\partial M(x)}{\partial x_i}(Uz), z \in \mathbb{C}^{k(x)}, \|z\| = 1\} \tag{81}$$

where co denotes the convex hull of the set in $\{.\}$ and $k(x)$ is such that $\lambda_1(x) = \lambda_2(x) = ... = \lambda_{k(x)}(x) > \lambda_{k(x)+1}(x)$. One should also note that the use of a generalized gradient doesn't lead necessarily to a descent direction like in differentiable problems. The algorithms described in Polak & Wardi (1982), Polak *et al.* (1984), Polak & Mayne (1985) 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 & Salcudean (1989).

9.4.3 A modified dynamic backpropagation algorithm

Instead of applying methods of non-convex nondifferentiable optimization to the general problem in θ_c and Q_i or D_i it is possible to exploit the fact that the Type 1 or 2 problem contains a *convex subproblem*. Without loss of generality let us consider the Type 2 problem with diagonal scaling:

$$\min_{\theta_c, D_i} J(\theta_c) \quad \text{such that} \quad M(\theta_c, D_i) < 0. \tag{82}$$

This can be thought of as two nested optimization problems: a non-convex and a convex one. For a given value of θ_c, called θ_c^0, the 'inner' feasibility problem

$$M(\theta_c^0, D_i) < 0 \tag{83}$$

is convex. The 'outer' optimization problem is then

$$\min_{\theta_c} J(\theta_c) \quad \text{such that} \quad \lambda_{max}[M(\theta_c)] < 0. \tag{84}$$

The gradient of $J(\theta_c)$ is given by (66) and the gradient for the constraint at a given point θ_c^0 is given by

$$[\nabla \lambda_{max}[M(\theta_c^0)]]_i = u_1(\theta_c^0)^T \frac{\partial M(\theta_c^0, D_i^{(*)})}{\partial \theta_{c_i}^0} u_1(\theta_c^0) \tag{85}$$

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 (83).

Remarks.

- It may happen that it is impossible to find a feasible point for (83). In that case one may not conclude that the system is not globally asymptotically stable because the condition is only sufficient. One has to rely then on the classical dynamic backpropagation algorithm.
- It is also possible to consider a modified dynamic backpropagation algorithm for the system identification problem (Section 3) like for the tracking problem (82). In that case the objective (5) of the prediction error algorithm is modified with an LMI constraint that expresses global asymptotic stability of the model

$$\min_{\theta_m, D_i} V_N(\theta_m, Z^N) \quad \text{such that} \quad M(\theta_m, D_i) < 0. \tag{86}$$

9.5 Example

A simple example is given now 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 $\mathcal{C}_0$

and a reference model $\mathcal{L}$:

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

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

$$\mathcal{C}_0: \; \begin{cases} z_{k+1} &= Ez_k + Fy_k + F_2 d_k \\ u_k &= Gz_k \end{cases}$$

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

First nonlinear system identification was done according to Section 3.2. 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 (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 7 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 to (5) (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 $V_{fit} = 8.3731\,e-05$ and a generalization error of $V_{gen} = 7.9801\,e-05$ and was acceptable according to correlation tests (9).

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 $w_k = [r_k; \epsilon_k; 1]$

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

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

$$\mathrm{V}_1 = \begin{bmatrix} W_{AB} & & & \\ & I & & \\ & & V_C W_{AB} & \\ & & & 1 \end{bmatrix}, \mathrm{V}_2 = \begin{bmatrix} V_A & V_B G & 0 & 0 \\ 0 & E & FW_{CD} & F_2 \\ V_A & V_B G & 0 & 0 \\ 0 & 0 & 0 & 0.6 \end{bmatrix}.$$

The original dynamic backpropagation algorithm (64) as well as the modified dynamic backpropagation algorithm (82) 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 an SQP method (Sequential Quadratic Programming) 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.10. A solution to the unconstrained problem is plotted in Fig.a. In Fig.b 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 (64) and (82): J = 0.0066 (Fig.a), J = 0.0073 (Fig.b), J = 0.0124 (Fig.c). Hence the tracking performance for the specific reference input degrades by ensuring a higher contraction rate of the flow for the autonomous control system. The minimum of Fig.a was not globally asymptotically stable according the the diagonal scaling condition. Although this may depend on the initial controller and the neural state space model, the example is quite representative.

10 More examples on NL_qs

In this section we will illustrate that the NL_q concept goes beyond the domain of control theory and may also serve as a generalized architecture in neural network theory. It will be shown how LRGFs (Locally recurrent globally feedforward neural nets) and GCNNs (Generalized cellular neural networks) can be written as NL_qs.

10.1 LRGFs as NL_qs

In Tsoi & Back (1994) a general LRGF architecture is described, that aims at unifying work on locally recurrent globally feedforward neural networks. The general LRGF includes the local synapse feedback architecture as well as the local output feedback architecture. This architecture proposed by Tsoi is shown in Fig.11. Assuming the transfer functions G_i $(i = 1,..,n)$ (which may have both poles and zeros) have a state space realization (A_i, B_i, C_i) this architectures can be written in state space form as

$$\begin{cases} \xi_{k+1}^{(i)} &= A^{(i)}\xi_k^{(i)} + B^{(i)}u_k^{(i)}, \qquad i = 1,...,n-1 \\ z_k^{(i)} &= C^{(i)}\xi_k^{(i)} \\ \\ \xi_{k+1}^{(n)} &= A^{(n)}\xi_k^{(n)} + B^{(n)}f(\sum_{j=1}^n z_k^{(j)}) \\ z_k^{(n)} &= C^{(n)}\xi_k^{(n)} \\ y_k &= f(\sum_{j=1}^n z_k^{(j)}). \end{cases}$$

Using the trick of state augmentation by defining $\eta_k = f(\sum_{j=1}^n z_k^{(j)})$ an NL_1 system is obtained with $p_k = [\xi_k^{(1)}; \xi_k^{(2)}; ...; \xi_k^{(n-1)}; \xi_k^{(n)}; \eta_k]$ and $w_k = [u_k^{(1)}; ...; u_k^{(n-1)}]$ in (21) with matrices

$$V_1 = \begin{bmatrix} A^{(1)} & & & & \\ & A^{(2)} & & & \\ & & \ddots & & \\ & & & A^{(n-1)} & \\ C^{(1)}{}_0 A^{(1)} & C^{(2)}{}_0 A^{(2)} & \cdots & C^{(n-1)}{}_0 A^{(n-1)} & C^{(n)}{}_{A^{(n)}} \end{bmatrix}, B_1 = \begin{bmatrix} B^{(1)} & & & \\ & B^{(2)} & & \\ & & \ddots & \\ & & & B^{(n-1)} \\ C^{(1)}{}_0 B^{(1)} & C^{(2)}{}_0 B^{(2)} & \cdots & C^{(n-1)}{}_0 B^{(n-1)} \end{bmatrix}.$$

The latter holds provided that the nonlinearity $f(.)$ belongs to sector $[0, 1]$. As a consequence the properties of the LRGF architecture (global asymptotic stability, finite l_2-gain) can be studied as a function of the choice of the transfer functions G_i by means of the stability and dissipativity Theorems, which was one of the open issues formulated in Tsoi & Back (1994).

10.2 GCNNs as NL_qs

Generalized cellular neural networks consist of multiple layers of cellular neural networks, characterized by a certain inter- and intra-layer connectivity. In cellular neural networks (CNNs), introduced by Chua (Chua & Yang (1988)), a given cell is connected only to its neighboring cells, in order to make a hardware implementation more feasible. A generalized CNN (GCNN) is obtained by connecting several CNNs in a feedforward, cascade or recurrent way and are studied in continuous time as well as discrete time (see e.g. Guzelis & Chua (1993), Harrer (1993)).
A GCNN is considered here that consists of four CNNs interconnected according to Fig.12 with a feedforward, cascade as well as a feedback interconnection. The i_l-th cell of layer l (l-th CNN) is described by

$$\begin{cases} \xi_{k+1}^{(i_l)} &= A^{(i_l)}\xi_k^{(i_l)} + B^{(i_l)}u_k^{(i_l)} \\ z_k^{(i_l)} &= f(C^{(i_l)}\xi_k^{(i_l)}), \end{cases}$$

where $\xi_k^{(i_l)}$ is the internal state, $u_k^{(i_l)}$ is the input and $z_k^{(i_l)}$ is the output of the cell and the nonlinearity $f(.)$ belongs to sector $[0, 1]$. For the indices we have $l = 1, 2, 3, 4$ and $i_l = 1, 2, ..., N_l$ where N_l is the number of cells of the l-th CNN.
The intra-layer connectivity of a given CNN will be unimportant in order to show that this system can be written as an NL_q because the cellular nature of the network is revealed only by the sparseness of the matrices $A^{(i_l)}$, $B^{(i_l)}$, $C^{(i_l)}$. The inter-layer connectivity of the GCNN is characterized by

$$\begin{aligned} u_k^1 &= E^1 z_k^1 + F^1 z_k^2 + G^1 v_k^1 \\ u_k^2 &= E^2 z_k^2 + F^2 z_k^1 \\ u_k^3 &= E^3 z_k^3 + F^3 v_k^3 \\ u_k^4 &= E^4 z_k^4 + F^4 z_k^1 + G^4 z_k^3 \end{aligned}$$

where v_k^1 and v_k^3 are externals inputs to layer 1 and 3 respectively. The upper index on u_k, z_k and v_k is the layer index ('i' is omitted because the intra-layer connectivity layer is

irrelevant here).
After applying the state augmentation $\eta_k^l = f(C^l \xi_k^l)$ $(l = 1, 2, 3, 4)$ it is straightforward calculation to obtain an NL_1 system with state vector $p_k = [\xi_k^1; \xi_k^2; \xi_k^3; \xi_k^4; \eta_k^1; \eta_k^2; \eta_k^3; \eta_k^4]$ and input $w_k = [v_k^1; v_k^3]$ in (21) and

$$\mathrm{V}_1 = \begin{bmatrix} A^1 & 0 & 0 & 0 & B^1E^1 & B^1F^1 & 0 & 0 \\ 0 & A^2 & 0 & 0 & B^2F^2 & B^2E^2 & 0 & 0 \\ 0 & 0 & A^3 & 0 & 0 & 0 & B^3E^3 & 0 \\ 0 & 0 & 0 & A^4 & B^4F^4 & 0 & B^4G^4 & B^4E^4 \\ C^1A^1 & 0 & 0 & 0 & C^1B^1E^1 & C^1B^1F^1 & 0 & 0 \\ 0 & C^2A^2 & 0 & 0 & C^2B^2F^2 & C^2B^2E^2 & 0 & 0 \\ 0 & 0 & C^3A^3 & 0 & 0 & 0 & C^3B^3E^3 & 0 \\ 0 & 0 & 0 & C^4A^4 & C^4B^4F^4 & 0 & C^4B^4G^4 & C^4B^4E^4 \end{bmatrix}, \mathrm{B}_1 = \begin{bmatrix} B^1G^1 & 0 \\ 0 & 0 \\ 0 & B^3F^3 \\ 0 & 0 \\ C^1B^1G^1 & 0 \\ 0 & 0 \\ 0 & C^3B^3F^3 \\ 0 & 0 \end{bmatrix}.$$

Conditions for global asymptotic and input/output stability for continuous time GCNNs were derived in Guzelis & Chua (1993). The Theorems in this paper apply to the discrete time case.

11 Conclusions

From the results of this paper it follows that the new concept of NL_qs plays a unifying role for many problems arising in classical, modern and neural control theory and neural networks. It was shown that the following systems could be interpreted as NL_qs: all neural state space control problems Ξ_j^is, the discrete time Lur'e problem, LFTs with real diagonal Δ block, Hopfield nets, multilayer CNNs, locally recurrent globally feedforward neural nets. A common aspect is that nonlinearities belonging to a sector $[0, K]$ are involved. The problems related to classical and modern control theory correspond to $q = 1$. Because of this unifying nature, theorems proven for NL_qs may have a simultaneous impact upon the different fields. In this paper such theorems were derived for global asymptotic stability and dissipativity with finite l_2-gain of NL_qs and the links were revealed with H_∞ control theory and μ-theory. 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). Another feature is that all systems considered in this paper are in state space form: this made it easy to define state augmentations in order to convert a given system into an NL_q. Vector norms as Lyapunov functions and properties of induced norms served as tools to derive the main results.

For control design efficient algorithms of convex optimization can be used for checking stability and dissipativity of the learned controller (analysis problem). In the synthesis problem a modified dynamic backpropagation algorithm was proposed such that global stability or dissipativity for the closed loop system can be ensured. Linear matrix inequalities are the central issue in these problems. An open problem here is e.g. whether it is possible to find a parametrization of all stabilizing neural controllers for a specific Ξ_j^i problem. Such a parametrization could simplify the synthesis problem.

We hope that in addition to the universal approximation ability, the parallel network structure and the learning ability also stability and dissipativity Theorems will be regarded

in the future as belonging to the attractive features of dynamical systems, containing multilayer neural networks and that this work may have delivered a significant contribution towards this.

References

Billings S., Jamaluddin H., Chen S. (1992). Properties of neural networks with applications to modelling non-linear dynamical systems, *International Journal of Control*, Vol.55, No.1, pp.193-224.

Boyd S., Barratt C. (1991). *Linear controller design, limits of performance*, Prentice-Hall.

Boyd S., Barratt C., Norman S. (1990). Linear controller design: limits of performance via convex optimization, *Proceedings of the IEEE*, Vol.78, No.3, pp.529-574.

Boyd S., Yang Q. (1989). Structured and simultaneous Lyapunov functions for system stability problems, *International Journal of Control*, Vol.49, No.6, pp.2215-2240.

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.

Boyd S., El Ghaoui L. (1993). Method of centers for minimizing generalized eigenvalues, *Linear Algebra and Applications, special issue on Numerical Linear Algebra Methods in Control, Signals and Systems*, 188, July.

Brockett R.W. (1966). The status of stability theory for deterministic systems, *IEEE Transactions on Automatic Control*, pp.597-605, July.

Byrnes C.I., Lin W. (1994). Losslessness, feedback equivalence and the global stabilization of discrete-time nonlinear systems, *IEEE Transactions on Automatic Control*, Vol.39, No.1, pp.83-98.

Chen S., Billings S., Grant P. (1990a). Nonlinear system identification using neural networks, *International Journal of Control*, Vol.51, No.6, pp. 1191-1214.

Chen S., Cowan C., Billings S., Grant P. (1990b). Parallel recursive prediction error algorithm for training layered neural networks, *International Journal of Control*, Vol.51, No.6, pp. 1215-1228.

Chua L.O., Yang L. (1988). Cellular neural networks: Theory, *IEEE Transactions on Circuits and Systems*, Vol. 35, No.10, pp.1257-1272.

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

Desoer C.A., Vidyasagar M. (1975). *Feedback systems: input-output properties*, New York: Academic Press.

Doyle J. (1982). Analysis of feedback systems with structured uncertainties, *IEE Proceedings-D*, Vol.129, No.6, pp.243-250.

Doyle J., Packard A., Zhou K. (1991). Review of LFTs, LMIs, and μ, *Proceedings of the 30th Conference on Decision and Control*, Brighton, England, pp.1227-1232, Dec.

Ecker J.G., Kupferschmid M. (1985). A computational comparison of the ellipsoid algorithm with several nonlinear programming algorithms, *SIAM Journal on Control and Optimization*, Vol.23, No.5, pp. 657-674.

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*, Vol.2, pp.183-192.

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

Golub G.H., Van Loan C.F. (1989). *Matrix Computations*, Baltimore MD: Johns Hopkins University Press.

Goodwin G., Sin K. (1984). *Adaptive filtering, prediction and control*, Prentice-Hall.

Grossberg S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures, *Neural Networks*, Vol.1, pp.17-61.

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

Harrer H. (1993). Multiple layer discrete-time cellular neural networks using time-variant templates, *IEEE Transactions on Circuits and Systems - II*, Vol. 40, No.3, pp.191-199.

Hill D.J., Moylan P.J. (1980). Connections between finite-gain and asymptotic stability, *IEEE Transactions on Automatic Control*, Vol.AC-25, No.5, pp.931-936.

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

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

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

Hunt K.J., Sbarbaro D., Zbikowski R., Gawthrop P.J. (1992). Neural networks for control systems - a survey, *Automatica*, Vol. 28., No. 6, pp. 1083-1112.

Jury E.I. (1974). *Inners and stability of dynamic systems*, John Wiley & Sons.

Karmarkar N. (1984). A new polynomial-time algorithm for linear programming, *Combinatorica*, 4, pp.373-395.

Kaszkurewicz E., Bhaya A. (1993). Robust stability and diagonal Liapunov functions, *SIAM Journal on Matrix Analysis and Applications*, Vol.14, No.2, pp.508-520.

Kato T. (1976). *Perturbation theory for linear operators*, Springer-Verlag.

Leshno M., Lin V.Y., Pinkus A., Schocken S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, *Neural Networks*, Vol.6, pp.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*, Vol.39, No.10, pp. 798-807.

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

Ljung L. (1979). Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems, *IEEE Transactions on Automatic Control*, Vol.AC-24, No.1, pp.36-50.

Luenberger D.G. (1979). *Introduction to dynamic systems: theory, models and applications*, John Wiley & Sons.

Maciejowski J.M. (1989). *Multivariable feedback design*, 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*, Vol.1, No.1, pp. 4-27.

Narendra K.S., Parthasarathy K. (1991). Gradient methods for the optimization of dynamical systems containing neural networks, *IEEE Transactions on Neural Networks*, Vol.2, No.2, pp.252-262.

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

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

Overton M.L. (1988). On minimizing the maximum eigenvalue of a symmetric matrix, *SIAM Journal on Matrix Analysis and Applications*, Vol.9, No.2, pp.256-268.

Packard A., Doyle J. (1993). The complex structured singular value, *Automatica*, Vol.29, No.1, pp.71-109.

Polak E., Mayne D. (1985). Algorithm models for nondifferentiable optimization, *SIAM Journal on Control and Optimization*, Vol.23, No.3, pp.477-491.

Polak E., Mayne D., Stimler D. (1984). Control system design via semi-infinite optimization: a review, *Proceedings of the IEEE*, Vol.72, No.12, pp. 1777-1794.

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*, Vol.34, No.3, pp.268-276.

Polak E., Wardi Y. (1982). Nondifferentiable optimization algorithm for designing control systems having singular value inequalities, *Automatica*, Vol.18, No. 3, pp.267-283.

Reed R. (1993). Pruning algorithms - a survey, *IEEE Transactions on Neural Networks*, Vol.4, No.5, pp.740-747.

Rumelhart D.E., Hinton G.E., Williams R.J. (1986). Learning representations by back-propagating errors, *Nature*, Vol.323, pp.533-536.

Shor N.Z. (1985). *Minimization methods for non-differentiable functions*, Berlin, Heidelberg: Springer-Verlag.

Siljak D.D. (1989). Parameter space methods for robust control design, *IEEE Transactions on Automatic Control*, Vol.34, No.7, pp.674-688.

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, pp.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*, Vol.139, No.3, pp.301-307.

Stoorvogel A.A. (1992). The discrete time H_∞ control problem with measurement feedback, *SIAM Journal on Control and Optimization*, Vol.30, No.1, pp.182-202.

Suykens J.A.K., De Moor B., Vandewalle J. (1994a). Static and dynamic stabilizing neural controllers, applicable to transition between equilibrium points, *Neural Networks*, Vol.7, No.5, pp.819-831.

Suykens J.A.K., De Moor B., Vandewalle J. (1994b). Nonlinear system identification using neural state space models, applicable to robust control design, *International Journal of Control*, to appear.

Suykens J.A.K., De Moor B., Vandewalle J. (1993). Neural network models as linear systems with bounded uncertainty, applicable to robust controller design, *NOLTA 93 International Symposium on Nonlinear Theory and its Applications, Honolulu, Hawaii*, Vol.2, pp.419-422, Dec.

The MathWorks Inc. (1994). Matlab, *Optimization Toolbox* (Version 4.1), User's guide, Nov. 1992 & *Robust Control Toolbox* (Version 4.1), User's guide, Aug. 1992 & *LMI lab* (Version 2.0), User's guide, 1993.

Tsoi A.C., Back A.D. (1994). Locally recurrent globally feedforward networks: a critical review of architectures, *IEEE Transactions on Neural Networks*, Vol.5, No.2, pp. 229-239.

Vandenberghe L., Boyd S. (1993). A primal-dual potential reduction method for problems involving matrix inequalities, *Mathematical Programming Series B, special issue on nondifferentiable and large scale optimization*, to appear.

van der Schaft A.J. (1992). L_2-gain analysis of nonlinear systems and nonlinear state feedback $\mathcal{H}_\infty$ control, *IEEE Transactions on Automatic Control*, Vol.AC-37, pp.770-784.

van der Schaft A.J. (1993). Nonlinear state space $\mathcal{H}_\infty$ control theory, in *Essays on Control: perspectives in the theory and its applications*, Eds. H.L. Trentelman, J.C. Willems, Birkhäuser.

van der Smagt P.P. (1994). Minimisation methods for training feedforward neural networks, *Neural Networks*, Vol.7, No.1, pp.1-11.

Vidal P. (1969). *Non-linear sampled-data systems*, Gordon and Breach Science Publishers Inc.

Vidyasagar M. (1993). *Nonlinear systems analysis*, Prentice-Hall.

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

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

A Sensitivity models for neural state space models

Given the predictor (7), referring to one of the neural state space models $\mathcal{M}_i$ ($i = 0, ..., 3$) and defining

$$\begin{aligned}
\varphi^l &= \textstyle\sum_r v_{A_r}^l \hat{x}^r + \sum_r v_{B_r}^l u^r + \beta_{AB}{}^l \\
\psi^l &= \textstyle\sum_r v_{C_r}^l \hat{x}^r + \sum_r v_{D_r}^l u^r + \beta_{CD}{}^l \\
\rho^l &= \textstyle\sum_r \kappa_r^l \epsilon^r \\
\eta^l &= \textstyle\sum_r v_{AB_r}^l \tanh(\varphi^r) + \beta_{AB_w}{}^l \\
\omega^l &= \textstyle\sum_r v_{CD_r}^l \tanh(\psi^r) + \beta_{CD_w}{}^l
\end{aligned}$$

the derivatives for the sensitivity model are given below. No distinction is made between β_{AB} and β_{AB_v}, β_{CD} and β_{CD_v}. δ_j^i denotes the Kronecker delta ($\delta_j^i = 1$ if $i = j$ and $\delta_j^i = 0$ if $i \neq j$).

For $\frac{\partial \Phi}{\partial \hat{x}}$:

$$\begin{aligned}
\tfrac{\partial \Phi^i}{\partial \hat{x}^r} &= a_r^i, \quad (\mathcal{M}_0) \\
\tfrac{\partial \Phi^i}{\partial \hat{x}^r} &= \textstyle\sum_j w_{AB_j^i}(1 - \tanh^2(\varphi^j)) v_{A_r^j}, \quad (\mathcal{M}_{1,2}) \\
\tfrac{\partial \Phi^i}{\partial \hat{x}^r} &= \textstyle\sum_j w_{AB_j^i}(1 - \tanh^2(\eta^j)) \sum_l v_{AB_l^j}(1 - \tanh^2(\varphi^l)) v_{A_r^l}, \quad (\mathcal{M}_3)
\end{aligned}$$

For $\frac{\partial \Psi}{\partial \hat{x}}$:

$$\begin{aligned}
\tfrac{\partial \Psi^i}{\partial \hat{x}^r} &= c_r^i, \quad (\mathcal{M}_{0,1}) \\
\tfrac{\partial \Psi^i}{\partial \hat{x}^r} &= \textstyle\sum_j w_{CD_j^i}(1 - \tanh^2(\psi^j)) v_{C_r^j}, \quad (\mathcal{M}_2) \\
\tfrac{\partial \Psi^i}{\partial \hat{x}^r} &= \textstyle\sum_j w_{CD_j^i}(1 - \tanh^2(\omega^j)) \sum_l v_{CD_l^j}(1 - \tanh^2(\psi^l)) v_{C_r^l}, \quad (\mathcal{M}_3)
\end{aligned}$$

For $\frac{\partial \Phi}{\partial \alpha}$:

$$\frac{\partial \Phi^i}{\partial a_l^j} = \delta_j^i \hat{x}^l, \quad (\mathcal{M}_0)$$

$$\frac{\partial \Phi^i}{\partial b_l^j} = \delta_j^i u^l, \quad (\mathcal{M}_0)$$

$$\frac{\partial \Phi^i}{\partial \kappa_l^j} = \delta_j^i \epsilon^l, \quad (\mathcal{M}_{0,1,2,3})$$

$$\frac{\partial \Phi^i}{\partial w_{AB_l^j}} = \begin{cases} \delta_j^i \tanh(\varphi^l), & (\mathcal{M}_{1,2}) \\ \delta_j^i \tanh(\eta^l), & (\mathcal{M}_3) \end{cases}$$

$$\frac{\partial \Phi^i}{\partial v_{A_l^j}} = \begin{cases} w_{AB}{}_j^i (1 - \tanh^2(\varphi^j))\hat{x}^l, & (\mathcal{M}_{1,2}) \\ \sum_r w_{AB}{}_r^i (1 - \tanh^2(\eta^r)) v_{AB}{}_j^r (1 - \tanh^2(\varphi^j))\hat{x}^l, & (\mathcal{M}_3) \end{cases}$$

$$\frac{\partial \Phi^i}{\partial v_{B_l^j}} = \begin{cases} w_{AB}{}_j^i (1 - \tanh^2(\varphi^j)) u^l, & (\mathcal{M}_{1,2}) \\ \sum_r w_{AB}{}_r^i (1 - \tanh^2(\eta^r)) v_{AB}{}_j^r (1 - \tanh^2(\varphi^j)) u^l, & (\mathcal{M}_3) \end{cases}$$

$$\frac{\partial \Phi^i}{\partial v_{AB_l^j}} = w_{AB}{}_j^i (1 - \tanh^2(\eta^j)) \tanh(\varphi^l), \quad (\mathcal{M}_3)$$

$$\frac{\partial \Phi^i}{\partial \beta_{AB^j}} = w_{AB}{}_j^i (1 - \tanh^2(\varphi^j)), \quad (\mathcal{M}_{1,2})$$

$$\frac{\partial \Phi^i}{\partial \beta_{AB_w{}^j}} = w_{AB}{}_j^i (1 - \tanh^2(\eta^j)), \quad (\mathcal{M}_3)$$

$$\frac{\partial \Phi^i}{\partial \beta_{AB_v{}^j}} = \sum_r w_{AB}{}_r^i (1 - \tanh^2(\eta^r)) v_{AB}{}_j^r (1 - \tanh^2(\varphi^j)), \quad (\mathcal{M}_3)$$

For $\frac{\partial \Psi}{\partial \beta}$:

$$\frac{\partial \Psi^i}{\partial c_l^j} = \delta_j^i \hat{x}^l, \quad (\mathcal{M}_{0,1})$$

$$\frac{\partial \Psi^i}{\partial d_l^j} = \delta_j^i u^l, \quad (\mathcal{M}_{0,1})$$

$$\frac{\partial \Psi^i}{\partial w_{CD_l^j}} = \begin{cases} \delta_j^i \tanh(\psi^l), & (\mathcal{M}_2) \\ \delta_j^i \tanh(\omega^l), & (\mathcal{M}_3) \end{cases}$$

$$\frac{\partial \Psi^i}{\partial v_{C_j^l}} = \begin{cases} w_{CD}{}_j^i (1 - \tanh^2(\psi^j))\hat{x}^l, & (\mathcal{M}_2) \\ \sum_r w_{CD}{}_r^i (1 - \tanh^2(\omega^r)) v_{CD}{}_j^r (1 - \tanh^2(\psi^j))\hat{x}^l, & (\mathcal{M}_3) \end{cases}$$

$$\frac{\partial \Psi^i}{\partial v_{D_l^j}} = \begin{cases} w_{CD}{}_j^i (1 - \tanh^2(\psi^j)) u^l, & (\mathcal{M}_2) \\ \sum_r w_{CD}{}_r^i (1 - \tanh^2(\omega^r)) v_{CD}{}_j^r (1 - \tanh^2(\psi^j)) u^l, & (\mathcal{M}_3) \end{cases}$$

$$\frac{\partial \Psi^i}{\partial v_{CD_l^j}} = w_{CD}{}_j^i (1 - \tanh^2(\omega^j)) \tanh(\psi^l), \quad (\mathcal{M}_3)$$

$$\frac{\partial \Psi^i}{\partial \beta_{CD^j}} = w_{CD}{}_j^i (1 - \tanh^2(\psi^j)), \quad (\mathcal{M}_2)$$

$$\frac{\partial \Psi^i}{\partial \beta_{CD_w{}^j}} = w_{CD}{}_j^i (1 - \tanh^2(\omega^j)), \quad (\mathcal{M}_3)$$

$$\frac{\partial \Psi^i}{\partial \beta_{CD_v{}^j}} = \sum_r w_{CD}{}_r^i (1 - \tanh^2(\omega^r)) v_{CD}{}_j^r (1 - \tanh^2(\psi^j)), \quad (\mathcal{M}_3)$$

B Proof of Theorems

Proof of Lemma 1:

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

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 $\Gamma_1 = I, \Gamma_2 = \Gamma_{AB}$, $V_1 = W_{AB}$, $V_2 = V_A$, $B_2 = [V_B\ 0\ \beta_{AB}]$, $B_1 = [0\ K\ 0]$, $\Lambda_1 = I$, $\Gamma_2 = I$, $W_1 = I$, $W_2 = C$, $D_2 = [D\ I\ 0]$, $D_1 = I$. Γ_{AB} is a diagonal matrix with diagonal elements $\in [0, 1]$, like this is shown in (15)-(20).

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 $V_{k+1} = \|D_1\Gamma_1 V_1 \Gamma_2 V_2 ... \Gamma_q V_q p_k\|_2$. The following procedure is followed then: (1) insert $D_i^{-1} D_i$ after Γ_i where $D_i \in \mathbb{R}^{n_{h_i} \times n_{h_i}} (i = 1, ..., q-1)$ and insert $D_1^{-1} D_1$ after V_q; (2) the diagonal matrices Γ_i and D_i^{-1} commute; (3) use the properties of induced norms and the fact that $\|\Gamma_i\|_2 \leq 1$. Hence

$$\begin{aligned} V_{k+1} &= \|D_1\Gamma_1 D_1^{-1} D_1 V_1 ... \Gamma_q D_q^{-1} D_q V_q D_1^{-1} D_1 p_k\|_2 \\ &\leq \|\Gamma_1\|_2 \|D_1 V_1 D_2^{-1}\|_2 \|\Gamma_2\|_2 \|D_2 V_2 D_3^{-1}\|_2 ... \|\Gamma_q\|_2 \|D_q V_q D_1^{-1}\|_2 V_k \\ &\leq \|D_1 V_1 D_2^{-1}\|_2 \|D_2 V_2 D_3^{-1}\|_2 ... \|D_q V_q D_1^{-1}\|_2 V_k. \end{aligned}$$

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

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

or if

$$\overline{\sigma}\left(\begin{bmatrix} D_1 V_1 D_2^{-1} & & & \\ & D_2 V_2 D_3^{-1} & & \\ & & \ddots & \\ & & & D_q V_q D_1^{-1} \end{bmatrix} \right) \leq \beta_D^{1/q} < 1$$

or if

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

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

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

where $D_{tot} = \text{diag}\{D_2, D_3, ..., D_q, D_1\}$ and V_{tot} is defined as in (23).

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 $V_{k+1} = \|P_1\Gamma_1 V_1 \Gamma_2 V_2 ... \Gamma_q V_q p_k\|_2$. Defining $q_k = V_1\Gamma_2 V_2 ... \Gamma_q V_q p_k$, we will study first under what condition there exist a matrix P_1 and a positive scalar α_1 such that the following is satisfied $\forall q$ and $\forall \Gamma_1$ ($\gamma_{1_i} \in [0, 1]$):

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

or $(1 + \alpha)S - \Gamma S \Gamma > 0$, where $S = P_1^T P_1$, $\Gamma = \Gamma_1$ (index 1 is omitted for notational reasons). The condition

$$(1 + \alpha)S - \Gamma S \Gamma > 0$$

is equivalent to

$$(1 + \alpha)U - \Gamma U \Gamma > 0$$

applying a congruence transformation, because Γ, N^{-1} and N commute if one defines $U = N^{-1} S N$, with N a diagonal matrix with positive diagonal elements.
Defining then $R = (1+\alpha)U - \Gamma U \Gamma$ 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 λ_l of R are lying in at least one of the circular disks with centers r_{ii} and radii $\sum_{j(j\neq i)} |r_{ij}|$:

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

or $|\lambda_l - (1 - \gamma_i^2 + \alpha) s_{ii}| \leq n_i^{-1} \sum_{j(j\neq i)} |(1 - \gamma_i\gamma_j + \alpha) n_j s_{ij}|$. Because $\gamma_i \in [0, 1]$ and in order to have $R > 0$:

$$\alpha n_i s_{ii} > (1 + \alpha) \sum_{j(j\neq i)} |n_j s_{ij}|, \quad \forall i.$$

Defining $Q = SN = P^T P N$ this corresponds to

$$q_{ii} > \delta_Q \sum_{j(j\neq i)} |q_{ij}|, \quad \forall i$$

with $\delta_Q = (1+\alpha)/\alpha$.
Hence under that condition

$$\begin{aligned} V_{k+1} &= \|P_1\Gamma_1 q_k\|_2 \\ &\leq (1+\alpha_1)^{1/2}\|P_1 q_k\|_2 \\ &\leq (1+\alpha_1)^{1/2}\|P_1 V_1 P_2^{-1}\|_2\|P_2\Gamma_2 r_k\|_2 \end{aligned}$$

where $r_k = \|P_2\Gamma_2 V_2 ... \Gamma_q V_q p_k\|_2$. 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

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

with $P_{q+1} = P_1$ and diagonal dominant matrices $Q_i = P_i^T P_i N_i$ with level of diagonal dominance equal to $\delta_{Q_i} = (1+\alpha_i)/\alpha_i$.
Defining $\beta_P = \prod_{i=1(\mathrm{mod}q)}^{q}\|P_i V_i P_{i+1}^{-1}\|_2$ and $c_\alpha = \prod_{i=1}^{q}(1+\alpha_i)^{1/2}$, global asymptotic stability of the NL_q system is obtained for

$$\beta_P < 1/c_\alpha$$

or if

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

or if

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

where $P_{tot} = \mathrm{blockdiag}\{P_2, P_3, ..., P_q, P_1\}$ and V_{tot} as in (27).

Proof of Lemma 3

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

Proof of Theorem 4

Defining the additional state variable $\xi_k = f(Cx_k)$ the closed loop system is of the type NL_1 with $p_k = [x_k; \xi_k]$ in $p_{k+1} = \Gamma(p_k)V_{tot}p_k$ where $V_{tot} = [A \ -B; CA \ -CB]$ and $\Gamma = \mathrm{diag}\{I, \Gamma_C\}$. Here $\|\Gamma_C\| < K$ because $\gamma_{C_i} = f(\varphi_k)/\varphi_k \leq K$ $(\varphi_k \neq 0)$ and the nonlinearity $f(.)$ belongs to a sector $[0, K]$ and $\xi_{k+1} = f(\varphi_k) = \Gamma_C(\varphi_k)\varphi_k$ where by definition $\varphi_k = CAx_k - CB\xi_k$.
The case of diagonal scaling is then easily proven by proposing the Lyapunov function $V(p) = \|Dp\|_2$ where D is a diagonal matrix. Hence

$$V_{k+1} = \|D\Gamma(p_k)V_{tot}p_k\|_2 \leq K\|DV_{tot}D^{-1}\|_2 V_k$$

which proves the Theorem.
For the other case a Lyapunov function $V(p) = \|Pp\|_2$ is proposed. Like in Theorem 2 a condition on P and α is sought such that the following holds

$$V_{k+1} = \|P\Gamma V_{tot}p_k\|_2 \leq (1+\alpha)^{1/2}\|PV_{tot}P^{-1}\|_2 V_k.$$

Gerschgorin's Theorem is applied to $R = (1+\alpha)U - \Gamma U\Gamma$, with $U = N^{-1}SN$ $S = P^TP$, but now $\|\Gamma\| \leq K$ $(\gamma_i \in [0, K])$. Hence

$$(1 - K^2 + \alpha)n_i s_{ii} \geq (1+\alpha)\sum_{j(j\neq i)}|n_j s_{ij}|.$$

Defining $Q = SN$, then

$$q_{ii} \geq \delta_Q \sum_{j(j\neq i)}|q_{ij}|, \qquad \delta_Q = (1+\alpha)/(1-K^2+\alpha)$$

and $(1+\alpha)^{1/2} = K(\delta_Q/(\delta_Q - 1))^{1/2}$.

Proof of Theorem 5

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

$$\begin{aligned} \|D_{S_1}\begin{bmatrix} p_{k+1} \\ e_k^{ext} \end{bmatrix}\|_2^2 &= \|D_1 p_{k+1}\|_2^2 + \|e_k\|_2^2 \\ &\leq \|D_{S_1}R_1 D_2^{-1}\|_2^2\|D_2 R_2 D_3^{-1}\|_2^2 ... \|D_q R_q D_{S_1}^{-1}\|_2^2(\|D_1 p_k\|_2^2 + \|w_k\|_2^2) \end{aligned}$$

according to the proof of Theorem 1, because $\|\Omega_i\|_2 \leq 1$ and with D_i $(i = 2, ..., q)$ diagonal with nonzero diagonal elements.
Defining $\beta_D = \|D_{S_1}R_1 D_2^{-1}\|_2 \prod_{i=2}^{q-1}\|D_i R_i D_{i+1}^{-1}\|_2\|D_q R_q D_{S_1}^{-1}\|_2$ 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 $D_{tot} = \mathrm{diag}\{D_2, D_3, ..., D_q, D_{S_1}\}$.
Defining $r_k = \|D_1 p_k\|_2$, the following holds, provided that $\beta_D < 1$:

$$\begin{array}{lrcl} & r_{k+1}^2 + \|e_k\|_2^2 & \leq & \beta_D^2 r_k^2 + \beta_D^2\|w_k\|_2^2 \\ \Rightarrow & \sum_{k=0}^{N-1} r_{k+1}^2 + \sum_{k=0}^{N-1}\|e_k\|_2^2 & \leq & \beta_D^2\sum_{k=0}^{N-1} r_k^2 + \beta_D^2\sum_{k=0}^{N-1}\|w_k\|_2^2 \\ \Rightarrow & r_N^2 + (1-\beta_D^2)\sum_{k=0}^{N-1} r_k^2 + \sum_{k=0}^{N-1}\|e_k\|_2^2 & \leq & r_0^2 + \beta_D^2\sum_{k=0}^{N-1}\|w_k\|_2^2 \\ \Rightarrow & \sum_{k=0}^{N-1}\|e_k\|_2^2 & \leq & r_0^2 + \beta_D^2\sum_{k=0}^{N-1}\|w_k\|_2^2. \end{array}$$

Furthermore for $N \to \infty$: $(1-\beta_D^2)\|r\|_2^2 + \|e\|_2^2 \leq r_0^2 + \beta_D^2\|w\|_2^2$. Defining constants $c_1 = \overline{\sigma}_{D_1}^2$, $c_2 = \underline{\sigma}_{D_1}^2$ one obtains (40)(41).

Proof of Theorem 6

Defining $P_{S_1} = \text{blockdiag}\{P_1, I_{n_w}\}$ and according to the proofs in Theorem 2 and 5

$$\begin{aligned} \|P_{S_1}\begin{bmatrix} p_{k+1} \\ e_k^{ext} \end{bmatrix}\|_2^2 &= \|P_1 p_{k+1}\|_2^2 + \|e_k\|_2^2 \\ &\leq \prod_{i=1}^{q}(1+\alpha_i)\|P_{S_1}R_1P_2^{-1}\|_2^2 ... \|P_iR_iP_{i+1}^{-1}\|_2^2 ... \|P_qR_qP_{S_1}^{-1}\|_2^2(\|P_1p_k\|_2^2 + \|w_k\|_2^2) \end{aligned}$$

with diagonal dominant matrices $Q_i = P_i^T P_i N_i$, $\delta_{Q_i} = (1+\alpha_i)/\alpha_i$. Defining $c_\alpha = \prod_{i=1}^{q}(1+\alpha_i)^{1/2}$, $\beta_P = \|P_{S_1}R_1P_2^{-1}\|_2 \prod_{i=2}^{q-1}\|P_iR_iP_{i+1}^{-1}\|_2 \|P_qR_qP_{S_1}^{-1}\|_2$ 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 $P_{tot} = \text{blockdiag}\{P_2, P_3, ..., P_q, P_{S_1}\}$. Putting $r_k = \|P_1 p_k\|_2$ then

$$\begin{array}{rcll} & r_{k+1}^2 + \|e_k\|_2^2 & \leq & c_\alpha^2\beta_P^2 r_k^2 + c_\alpha^2\beta_P^2\|w_k\|_2^2 \\ \Rightarrow & r_N^2 + (1-c_\alpha^2\beta_P^2)\sum_{k=0}^{N-1} r_k^2 + \sum_{k=0}^{N-1}\|e_k\|_2^2 & \leq & r_0^2 + c_\alpha^2\beta_P^2\sum_{k=0}^{N-1}\|w_k\|_2^2 \\ \Rightarrow & \sum_{k=0}^{N-1}\|e_k\|_2^2 & \leq & r_0^2 + c_\alpha^2\beta_P^2\sum_{k=0}^{N-1}\|w_k\|_2^2. \end{array}$$

Defining constants $c_1 = \overline{\sigma}_{P_1}^2$, $c_2 = \underline{\sigma}_{P_1}^2$ and letting $N \to \infty$ one obtains (43)(44).

Proof of Lemma 5

This follows immediately from the proof of Theorem 5 and 6, because if $c_\alpha\beta_P < 1$ then $r_{k+1}^2 + \|e_k\|_2^2 \leq c_\alpha^2\beta_P^2 r_k^2 + c_\alpha^2\beta_P^2\|w_k\|_2^2 \leq r_k^2 + c_\alpha^2\beta_P^2\|w_k\|_2^2$ or $r_{k+1}^2 - r_k^2 \leq c_\alpha^2\beta_P^2\|w_k\|_2^2 - \|e_k\|_2^2$ which proves the Lemma.

Proof of Theorem 7

Consider the perturbed NL_q written as an NL_{q+1} system (53). Defining $P_{S_1} = \text{blockdiag}\{P_1, I, D_1\}$ and $\Upsilon_{q+1} = \text{diag}\{I, I, \Delta_x\}$ one obtains like in the proof of Theorem 6:

$$\begin{aligned} \|P_{S_1}\begin{bmatrix} p_{k+1} \\ e_k^{ext} \\ s_k \end{bmatrix}\|_2^2 &= \|P_1 p_{k+1}\|_2^2 + \|e_k\|_2^2 + \|D_1 s_k\|_2^2 \\ &\leq \prod_{i=1}^{q}(1+\alpha_i)\|P_{S_1}T_1P_2^{-1}\|_2^2 ... \|P_iT_iP_{i+1}^{-1}\|_2^2 ... \|P_qT_qP_{S_1}^{-1}P_{S_1}\Upsilon_{q+1}\begin{bmatrix} p_k \\ w_k \\ s_k \end{bmatrix}\|_2^2 \\ &\leq \prod_{i=1}^{q}(1+\alpha_i)\|P_{S_1}T_1P_2^{-1}\|_2^2 ... \|P_iT_iP_{i+1}^{-1}\|_2^2 ... \|P_qT_qP_{S_1}^{-1}\|_2^2 \\ &\quad (\|P_1p_k\|_2^2 + \|w_k\|_2^2 + \|D_1 s_k\|_2^2) \end{aligned}$$

because Υ_{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 $c_\alpha = \prod_{i=1}^{q}(1+\alpha_i)^{1/2}$,
$\beta_P = \|P_{S_1}T_1P_2^{-1}\|_2 \prod_{i=2}^{q-1}\|P_iT_iP_{i+1}^{-1}\|_2 \|P_qT_qP_{S_1}^{-1}\|_2$ and according to Theorem 2, $c_\alpha\beta_P < 1$

holds if $\|P_{tot}T_{tot}P_{tot}^{-1}\|_2^q < 1/c_\alpha$ where $P_{tot} = \text{blockdiag}\{P_2, P_3, ..., P_q, P_{S_1}\}$. Putting $r_k = \|P_1 p_k\|_2$ then

$$\begin{array}{rcll} & r_{k+1}^2 + \|e_k\|_2^2 + \|D_1 s_k\|_2^2 & \leq & c_\alpha^2\beta_P^2(r_k^2 + \|w_k\|_2^2 + \|D_1 s_k\|_2^2) \\ & & \leq & c_\alpha^2\beta_P^2 r_k^2 + c_\alpha^2\beta_P^2\|w_k\|_2^2 + \|D_1 s_k\|_2^2 \\ \Rightarrow & r_{k+1}^2 + \|e_k\|_2^2 & \leq & r_k^2 + \|w_k\|_2^2 \end{array}$$

because $c_\alpha\beta_P < 1$, yielding (55)(56).

C Family of Neural Control problems Ξ_j^i

In this proof of Lemma 2, an exhaustive list is presented for the neural control problems Ξ_j^i written as NL_qs given a neural state space model $\mathcal{M}_i$ ($i \in \{0,1,2,3\}$) and a neural state space controller $\mathcal{C}_j$ ($j \in \{0,1,...,5\}$). The following information is given

1. Assumptions on the interconnection matrices
2. State augmentation
3. Closed loop system after state augmentation
4. Matrices V_i, B_i of the NL_q system, related to the standard plant representation

C.1 Problem Ξ_0^0

1. Assumption: $H = 0$
2. State augmentation: none
3. Closed loop system:

$$\begin{cases} \hat{x}_{k+1} &= A\hat{x}_k + BGz_k + BH_2d_k + K\epsilon_k \\ z_{k+1} &= Ez_k + FC\hat{x}_k + FDGz_k + FDH_2d_k + F\epsilon_k + F_2d_k \end{cases}$$

4. NL_q: $q = 1$, $p_k = [\hat{x}_k; z_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} A & BG \\ FC & E+FDG \end{bmatrix}, B_1 = \begin{bmatrix} BH_2 & K & 0 \\ FDH_2+F_2 & F & 0 \end{bmatrix}$$

C.2 Problem Ξ_1^0

1. Assumption: $H = 0$, $H_2 = 0$
2. State augmentation:

$$\xi_k = \tanh(Gz_k)$$

3. Closed loop system:

$$\begin{cases} \hat{x}_{k+1} &= A\hat{x}_k + B\xi_k + K\epsilon_k \\ z_{k+1} &= Ez_k + FC\hat{x}_k + FD\xi_k + F\epsilon_k + F_2d_k \\ \xi_{k+1} &= \tanh(GEz_k + GFC\hat{x}_k + GFD\xi_k + GF\epsilon_k + GF_2d_k) \end{cases}$$

4. NL_q: $q = 1$, $p_k = [\hat{x}_k; z_k; \xi_k]$, $w_k = [d_k; \epsilon_k; 1]$

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

C.3 Problem Ξ_0^1

1. Assumption: $H = 0$, $H_2 = 0$
2. State augmentation: none
3. Closed loop system:

$$\begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_BGz_k + \beta_{AB}) + K\epsilon_k \\ z_{k+1} &= Ez_k + FC\hat{x}_k + FDGz_k + F\epsilon_k + F_2d_k \end{cases}$$

4. NL_q: $q = 2$, $p_k = [\hat{x}_k; z_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} W_{AB} & \\ & I \end{bmatrix}, V_2 = \begin{bmatrix} V_A & V_BG \\ FC & E+FDG \end{bmatrix}, B_2 = \begin{bmatrix} 0 & 0 & \beta_{AB} \\ F_2 & F & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

C.4 Problem Ξ_1^1

1. Assumption: $H = 0$, $H_2 = 0$
2. State augmentation:

$$\xi_k = \tanh(Gz_k)$$

3. Closed loop system:

$$\begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_B\xi_k + \beta_{AB}) + K\epsilon_k \\ z_{k+1} &= Ez_k + FC\hat{x}_k + FD\xi_k + F\epsilon_k + F_2d_k \\ \xi_{k+1} &= \tanh(GEz_k + GFC\hat{x}_k + GFD\xi_k + GF\epsilon_k + GF_2d_k) \end{cases}$$

4. NL_q: $q = 2$, $p_k = [\hat{x}_k; z_k; \xi_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} W_{AB} & & \\ & I & \\ & & I \end{bmatrix}, V_2 = \begin{bmatrix} V_A & 0 & V_B \\ FC & E & FD \\ GFC & GE & GFD \end{bmatrix}, B_2 = \begin{bmatrix} 0 & 0 & \beta_{AB} \\ F_2 & F & 0 \\ GF_2 & GF & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

C.5 Problem Ξ_2^1

1. Assumption: $V_H = 0$, $V_{H_2} = 0$, $\beta_{GH} = 0$
2. State augmentation:

$$\xi_k = \tanh(V_Gz_k)$$

3. Closed loop system:

$$\begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_BW_{GH}\xi_k + \beta_{AB}) + K\epsilon_k \\ z_{k+1} &= W_{EF}\tanh(V_Ez_k + V_FC\hat{x}_k + V_FDW_{GH}\xi_k + V_F\epsilon_k + V_{F_2}d_k + \beta_{EF}) \\ \xi_{k+1} &= \tanh(V_GW_{EF}\tanh(V_Ez_k + V_FC\hat{x}_k + V_FDW_{GH}\xi_k + V_F\epsilon_k + V_{F_2}d_k + \beta_{EF})) \end{cases}$$

4. NL_q: $q = 2$, $p_k = [\hat{x}_k; z_k; \xi_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} W_{AB} & & \\ & W_{EF} & \\ & & V_GW_{EF} \end{bmatrix}, V_2 = \begin{bmatrix} V_A & 0 & V_BW_{GH} \\ V_FC & V_E & V_FDW_{GH} \\ V_FC & V_E & V_FDW_{GH} \end{bmatrix}$$

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

C.6 Problem Ξ_3^1

1. Assumption: $V_H = 0$, $V_{H_2} = 0$, $\beta_{GH} = 0$
2. State augmentation:
$$\xi_k = \tanh(W_{GH}\tanh(V_G z_k))$$
3. Closed loop system:
$$\begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_B\xi_k + \beta_{AB}) + K\epsilon_k \\ z_{k+1} &= W_{EF}\tanh(V_E z_k + V_F C\hat{x}_k + V_F D\xi_k + V_F\epsilon_k + V_{F_2}d_k + \beta_{EF}) \\ \xi_{k+1} &= \tanh(W_{GH}\tanh(V_G W_{EF}\tanh(V_E z_k + V_F C\hat{x}_k + V_F D\xi_k + V_F\epsilon_k + V_{F_2}d_k + \beta_{EF}))) \end{cases}$$
4. NL_q: $q = 3$, $p_k = [\hat{x}_k; z_k; \xi_k]$, $w_k = [d_k; \epsilon_k; 1]$
$$V_1 = \begin{bmatrix} I & & \\ & I & \\ & & W_{GH} \end{bmatrix}, V_2 = \begin{bmatrix} W_{AB} & & \\ & W_{EF} & \\ & & V_G W_{EF} \end{bmatrix}, V_3 = \begin{bmatrix} V_A & 0 & V_B \\ V_F C & V_E & V_F D \\ V_F C & V_E & V_F D \end{bmatrix}$$
$$B_3 = \begin{bmatrix} 0 & 0 & \beta_{AB} \\ V_{F_2} & V_F & \beta_{EF} \\ V_{F_2} & V_F & \beta_{EF} \end{bmatrix}, B_2 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{GH} \end{bmatrix}, B_1 = 0$$

C.7 Problem Ξ_0^2

1. Assumption: $H = 0$, $H_2 = 0$, $\beta_{CD} = 0$
2. State augmentation:
$$\xi_k = \tanh(V_C\hat{x}_k)$$
3. Closed loop system:
$$\begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_B G z_k + \beta_{AB}) + K\epsilon_k \\ z_{k+1} &= E z_k + F W_C\xi_k + F\epsilon_k + F_2 d_k \\ \xi_{k+1} &= \tanh(V_C W_{AB}\tanh(V_A\hat{x}_k + V_B G z_k + \beta_{AB}) + V_C K\epsilon_k) \end{cases}$$
4. NL_q: $q = 2$, $p_k = [\hat{x}_k; z_k; \xi_k]$, $w_k = [d_k; \epsilon_k; 1]$
$$V_1 = \begin{bmatrix} W_{AB} & & \\ & I & \\ & & V_C W_{AB} \end{bmatrix}, V_2 = \begin{bmatrix} V_A & V_B G & 0 \\ 0 & E & F W_C \\ V_A & V_B G & 0 \end{bmatrix}$$
$$B_2 = \begin{bmatrix} 0 & 0 & \beta_{AB} \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{AB} \end{bmatrix}, B_2 = \begin{bmatrix} 0 & K & 0 \\ F_2 & F & 0 \\ 0 & V_C K & 0 \end{bmatrix}$$

C.8 Problem Ξ_1^2

1. Assumption: $V_D = 0$, $H = 0$, $H_2 = 0$, $\beta_C = 0$
2. State augmentation:
$$\begin{aligned} \xi_k &= \tanh(V_C\hat{x}_k) \\ \eta_k &= \tanh(G z_k) \end{aligned}$$

3. Closed loop system:
$$\begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_B\eta_k + \beta_{AB}) + K\epsilon_k \\ z_{k+1} &= E z_k + F W_C\xi_k + F\epsilon_k + F_2 d_k \\ \xi_{k+1} &= \tanh(V_C W_{AB}\tanh(V_A\hat{x}_k + V_B\eta_k + \beta_{AB}) + V_C K\epsilon_k) \\ \eta_{k+1} &= \tanh(GE z_k + GF W_C\xi_k + GF\epsilon_k + GF_2 d_k) \end{cases}$$
4. NL_q: $q = 2$, $p_k = [\hat{x}_k; z_k; \xi_k; \eta_k]$, $w_k = [d_k; \epsilon_k; 1]$
$$V_1 = \begin{bmatrix} W_{AB} & & & \\ & I & & \\ & & V_C W_{AB} & \\ & & & I \end{bmatrix}, V_2 = \begin{bmatrix} V_A & 0 & 0 & V_B \\ 0 & E & F W_C & 0 \\ V_A & 0 & 0 & V_B \\ 0 & GE & GF W_C & 0 \end{bmatrix}$$
$$B_2 = \begin{bmatrix} 0 & 0 & \beta_{AB} \\ F_2 & F & 0 \\ 0 & 0 & \beta_{AB} \\ GF_2 & GF & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & V_C K & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

C.9 Problem Ξ_2^2

1. Assumption: $V_D = 0$, $V_H = 0$, $V_{H_2} = 0$, $\beta_C = 0$, $\beta_G = 0$
2. State augmentation:
$$\begin{aligned} \xi_k &= \tanh(V_C\hat{x}_k) \\ \eta_k &= \tanh(V_G z_k) \end{aligned}$$
3. Closed loop system:
$$\begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_A\hat{x}_k + V_B W_G\eta_k + \beta_{AB}) + K\epsilon_k \\ z_{k+1} &= W_{EF}\tanh(V_E z_k + V_F W_C\xi_k + V_F\epsilon_k + V_{F_2}d_k + \beta_{EF}) \\ \xi_{k+1} &= \tanh(V_C W_{AB}\tanh(V_A\hat{x}_k + V_B W_G\eta_k + \beta_{AB}) + V_C K\epsilon_k) \\ \eta_{k+1} &= \tanh(V_G W_{EF}\tanh(V_E z_k + V_F W_C\xi_k + V_F\epsilon_k + V_{F_2}d_k + \beta_{EF})) \end{cases}$$
4. NL_q: $q = 2$, $p_k = [\hat{x}_k; z_k; \xi_k; \eta_k]$, $w_k = [d_k; \epsilon_k; 1]$
$$V_1 = \begin{bmatrix} W_{AB} & & & \\ & W_{EF} & & \\ & & V_C W_{AB} & \\ & & & V_G W_{EF} \end{bmatrix}, V_2 = \begin{bmatrix} V_A & 0 & 0 & V_B W_G \\ 0 & V_E & V_F W_C & 0 \\ V_A & 0 & 0 & V_B W_G \\ 0 & V_E & V_F W_C & 0 \end{bmatrix}$$
$$B_2 = \begin{bmatrix} 0 & 0 & \beta_{AB} \\ V_{F_2} & V_F & \beta_{EF} \\ 0 & 0 & \beta_{AB} \\ V_{F_2} & V_F & \beta_{EF} \end{bmatrix}, B_1 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & V_C K & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

C.10 Problem Ξ_3^2

1. Assumption: $V_D = 0$, $V_H = 0$, $V_{H_2} = 0$, $\beta_C = 0$, $\beta_G = 0$
2. State augmentation:
$$\begin{aligned} \xi_k &= \tanh(V_C\hat{x}_k) \\ \eta_k &= \tanh(W_G\tanh(V_G z_k)) \end{aligned}$$

3. Closed loop system:

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

4. NL_q: $q = 3$, $p_k = [\hat{x}_k; z_k; \xi_k; \eta_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} I & & & \\ & I & & \\ & & I & \\ & & & W_G \end{bmatrix}, V_2 = \begin{bmatrix} W_{AB} & & & \\ & W_{EF} & & \\ & & V_C W_{AB} & \\ & & & V_G W_{EF} \end{bmatrix}, V_3 = \begin{bmatrix} V_A & 0 & 0 & V_B \\ 0 & V_E & V_F W_C & 0 \\ V_A & 0 & 0 & V_B \\ 0 & V_E & V_F W_C & 0 \end{bmatrix}$$

$$B_3 = \begin{bmatrix} 0 & 0 & \beta_{AB} \\ V_{F_2} & V_F & \beta_{EF} \\ 0 & 0 & \beta_{AB} \\ V_{F_2} & V_F & \beta_{EF} \end{bmatrix}, B_2 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & V_C K & 0 \\ 0 & 0 & 0 \end{bmatrix}, B_1 = 0$$

C.11 Problem Ξ_0^3

1. Assumption: $H = 0$, $H_2 = 0$, $\beta_{CD_v} = 0$, $\beta_{CD_w} = 0$
2. State augmentation:

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

3. Closed loop system:

$$\begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_{AB}\tanh(V_A\hat{x}_k + V_B G z_k + \beta_{AB_v}) + \beta_{AB_w}) + K\epsilon_k \\ z_{k+1} &= Ez_k + FW_{CD}\xi_k + F\epsilon_k + F_2 d_k \\ \xi_{k+1} &= \tanh(V_{CD}\tanh(V_C W_{AB}\tanh(V_{AB}\tanh(V_A\hat{x}_k + V_B G z_k + \beta_{AB_v}) + \beta_{AB_w}) + V_C K\epsilon_k)) \end{cases}$$

4. NL_q: $q = 4$, $p_k = [\hat{x}_k; z_k; \xi_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} I & & & \\ & I & & \\ & & V_{CD} & \\ & & & I \end{bmatrix}, V_2 = \begin{bmatrix} W_{AB} & & & \\ & I & & \\ & & V_C W_{AB} & \\ & & & I \end{bmatrix}, V_3 = \begin{bmatrix} V_{AB} & & & \\ & I & & \\ & & V_{AB} & \\ & & & I \end{bmatrix}$$

$$V_4 = \begin{bmatrix} V_A & 0 & 0 & V_B \\ 0 & E & FW_{CD} & 0 \\ V_A & 0 & 0 & V_B \\ 0 & GE & GFW_{CD} & 0 \end{bmatrix}, B_4 = \begin{bmatrix} 0 & 0 & \beta_{AB_v} \\ F_2 & F & 0 \\ 0 & 0 & \beta_{AB_v} \\ GF_2 & GF & 0 \end{bmatrix}, B_3 = \begin{bmatrix} 0 & 0 & \beta_{AB_w} \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{AB_w} \\ 0 & 0 & 0 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & V_C K & 0 \\ 0 & 0 & 0 \end{bmatrix}, B_1 = 0$$

C.12 Problem Ξ_1^3

1. Assumption: $V_D = 0$, $V_H = 0$, $V_{H_2} = 0$, $\beta_{CD_v} = 0$, $\beta_{CD_w} = 0$
2. State augmentation:

$$\begin{aligned} \xi_k &= \tanh(V_{CD}\tanh(V_C\hat{x}_k)) \\ \eta_k &= \tanh(Gz_k) \end{aligned}$$

3. Closed loop system:

$$\begin{cases} \hat{x}_{k+1} &= W_{AB}\tanh(V_{AB}\tanh(V_A\hat{x}_k + V_B\eta_k + \beta_{AB_v}) + \beta_{AB_w}) + K\epsilon_k \\ z_{k+1} &= Ez_k + FW_{CD}\xi_k + F\epsilon_k + F_2 d_k \\ \xi_{k+1} &= \tanh(V_{CD}\tanh(V_C W_{AB}\tanh(V_{AB}\tanh(V_A\hat{x}_k + V_B\eta_k + \beta_{AB_v}) + \\ & \quad \beta_{AB_w}) + V_C K\epsilon_k)) \\ \eta_{k+1} &= \tanh(GEz_k + GFW_{CD}\xi_k + GF\epsilon_k + GF_2 d_k) \end{cases}$$

4. NL_q: $q = 4$, $p_k = [\hat{x}_k; z_k; \xi_k; \eta_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} I & & & \\ & I & & \\ & & V_{CD} & \\ & & & I \end{bmatrix}, V_2 = \begin{bmatrix} W_{AB} & & & \\ & I & & \\ & & V_C W_{AB} & \\ & & & I \end{bmatrix}, V_3 = \begin{bmatrix} V_{AB} & & & \\ & I & & \\ & & V_{AB} & \\ & & & I \end{bmatrix}$$

$$V_4 = \begin{bmatrix} V_A & 0 & 0 & V_B \\ 0 & E & FW_{CD} & 0 \\ V_A & 0 & 0 & V_B \\ 0 & GE & GFW_{CD} & 0 \end{bmatrix}, B_4 = \begin{bmatrix} 0 & 0 & \beta_{AB_v} \\ F_2 & F & 0 \\ 0 & 0 & \beta_{AB_v} \\ GF_2 & GF & 0 \end{bmatrix}, B_3 = \begin{bmatrix} 0 & 0 & \beta_{AB_w} \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{AB_w} \\ 0 & 0 & 0 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & V_C K & 0 \\ 0 & 0 & 0 \end{bmatrix}, B_1 = 0$$

C.13 Problem Ξ_2^3

1. Assumption: $V_D = 0$, $V_H = 0$, $V_{H_2} = 0$, $\beta_{CD_v} = 0$, $\beta_{CD_w} = 0$, $\beta_{GH} = 0$
2. State augmentation:

$$\begin{aligned} \xi_k &= \tanh(V_{CD}\tanh(V_C\hat{x}_k)) \\ \eta_k &= \tanh(V_G z_k) \end{aligned}$$

3. Closed loop system:

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

4. NL_q: $q = 4$, $p_k = [\hat{x}_k; z_k; \xi_k; \eta_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} I & & & \\ & I & & \\ & & V_{CD} & \\ & & & I \end{bmatrix}, V_2 = \begin{bmatrix} W_{AB} & & & \\ & I & & \\ & & V_C W_{AB} & \\ & & & I \end{bmatrix}, V_3 = \begin{bmatrix} V_{AB} & & & \\ & W_{EF} & & \\ & & V_{AB} & \\ & & & V_G W_{EF} \end{bmatrix}$$

$$V_4 = \begin{bmatrix} V_A & 0 & 0 & V_B W_{GH} \\ 0 & V_E & V_F W_{CD} & 0 \\ V_A & 0 & 0 & V_B W_{GH} \\ 0 & V_E & V_F W_{CD} & 0 \end{bmatrix}, B_4 = \begin{bmatrix} 0 & 0 & \beta_{AB_v} \\ V_{F_2} & V_F & \beta_{EF} \\ 0 & 0 & \beta_{AB_v} \\ V_{F_2} & V_F & \beta_{EF} \end{bmatrix}, B_3 = \begin{bmatrix} 0 & 0 & \beta_{AB_w} \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{AB_w} \\ 0 & 0 & 0 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & V_C K & 0 \\ 0 & 0 & 0 \end{bmatrix}, B_1 = 0$$

C.14 Problem Ξ_3^3

1. Assumption: $V_D = 0$, $V_H = 0$, $V_{H_2} = 0$, $\beta_{CD_v} = 0$, $\beta_{CD_w} = 0$, $\beta_{GH} = 0$
2. State augmentation:

$$\begin{aligned} \xi_k &= \tanh(V_{CD} \tanh(V_C \hat{x}_k)) \\ \eta_k &= \tanh(W_{GH} \tanh(V_G z_k)) \end{aligned}$$

3. Closed loop system:

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

4. NL_q: $q = 4$, $p_k = [\hat{x}_k; z_k; \xi_k; \eta_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} I & & & \\ & I & & \\ & & V_{CD} & \\ & & & I \end{bmatrix}, V_2 = \begin{bmatrix} W_{AB} & & & \\ & I & & \\ & & V_C W_{AB} & \\ & & & W_{GH} \end{bmatrix}, V_3 = \begin{bmatrix} V_{AB} & & & \\ & W_{EF} & & \\ & & V_{AB} & \\ & & & V_G W_{EF} \end{bmatrix}$$

$$V_4 = \begin{bmatrix} V_A & 0 & 0 & V_B \\ 0 & V_E & V_F W_{CD} & 0 \\ V_A & 0 & 0 & V_B \\ 0 & V_E & V_F W_{CD} & 0 \end{bmatrix}, B_4 = \begin{bmatrix} 0 & 0 & \beta_{AB_v} \\ V_{F_2} & V_F & \beta_{EF} \\ 0 & 0 & \beta_{AB_v} \\ V_{F_2} & V_F & \beta_{EF} \end{bmatrix}, B_3 = \begin{bmatrix} 0 & 0 & \beta_{AB_w} \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{AB_w} \\ 0 & 0 & 0 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & V_C K & 0 \\ 0 & 0 & 0 \end{bmatrix}, B_1 = 0$$

C.15 Problem Ξ_4^3

1. Assumption: $V_D = 0$, $V_H = 0$, $V_{H_2} = 0$, $\beta_{CD_v} = 0$, $\beta_{CD_w} = 0$, $\beta_{GH_v} = 0$, $\beta_{GH_w} = 0$

2. State augmentation:

$$\begin{aligned} \xi_k &= \tanh(V_{CD} \tanh(V_C \hat{x}_k)) \\ \eta_k &= \tanh(V_{GH} \tanh(V_G z_k)) \end{aligned}$$

3. Closed loop system:

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

4. NL_q: $q = 4$, $p_k = [\hat{x}_k; z_k; \xi_k; \eta_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} I & & & \\ & I & & \\ & & V_{CD} & \\ & & & V_{GH} \end{bmatrix}, V_2 = \begin{bmatrix} W_{AB} & & & \\ & W_{EF} & & \\ & & V_C W_{AB} & \\ & & & V_G W_{EF} \end{bmatrix}$$

$$V_3 = \begin{bmatrix} V_{AB} & & & \\ & V_{EF} & & \\ & & V_{AB} & \\ & & & V_{EF} \end{bmatrix}, V_4 = \begin{bmatrix} V_A & 0 & 0 & V_B W_{GH} \\ 0 & V_E & V_F W_{CD} & 0 \\ V_A & 0 & 0 & V_B W_{GH} \\ 0 & V_E & V_F W_{CD} & 0 \end{bmatrix}$$

$$B_4 = \begin{bmatrix} 0 & 0 & \beta_{AB_v} \\ V_{F_2} & V_F & \beta_{EF_v} \\ 0 & 0 & \beta_{AB_v} \\ V_{F_2} & V_F & \beta_{EF_v} \end{bmatrix}, B_3 = \begin{bmatrix} 0 & 0 & \beta_{AB_w} \\ 0 & 0 & \beta_{EF_w} \\ 0 & 0 & \beta_{AB_w} \\ 0 & 0 & \beta_{EF_w} \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & V_C K & 0 \\ 0 & 0 & 0 \end{bmatrix}, B_1 = 0$$

C.16 Problem Ξ_5^3

1. Assumption: $V_D = 0$, $V_H = 0$, $V_{H_2} = 0$, $\beta_{CD_v} = 0$, $\beta_{CD_w} = 0$, $\beta_{GH_v} = 0$, $\beta_{GH_w} = 0$
2. State augmentation:

$$\begin{aligned} \xi_k &= \tanh(V_{CD} \tanh(V_C \hat{x}_k)) \\ \eta_k &= \tanh(W_{GH} \tanh(V_{GH} \tanh(V_G z_k))) \end{aligned}$$

3. Closed loop system:

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

4. NL_q: $q = 5$, $p_k = [\hat{x}_k; z_k; \xi_k; \eta_k]$, $w_k = [d_k; \epsilon_k; 1]$

$$V_1 = \begin{bmatrix} I & & & \\ & I & & \\ & & I & \\ & & & W_{GH} \end{bmatrix}, V_2 = \begin{bmatrix} I & & & \\ & I & & \\ & & V_{CD} & \\ & & & V_{GH} \end{bmatrix}$$

$$V_3 = \begin{bmatrix} W_{AB} & & & \\ & W_{EF} & & \\ & & V_C W_{AB} & \\ & & & V_G W_{EF} \end{bmatrix}, V_4 = \begin{bmatrix} V_{AB} & & & \\ & V_{EF} & & \\ & & V_{AB} & \\ & & & V_{EF} \end{bmatrix}$$

$$V_5 = \begin{bmatrix} V_A & 0 & 0 & V_B \\ 0 & V_E & V_F W_{CD} & 0 \\ V_A & 0 & 0 & V_B \\ 0 & V_E & V_F W_{CD} & 0 \end{bmatrix}, B_5 = \begin{bmatrix} 0 & 0 & \beta_{AB_v} \\ V_{F_2} & V_F & \beta_{EF_v} \\ 0 & 0 & \beta_{AB_v} \\ V_{F_2} & V_F & \beta_{EF_v} \end{bmatrix}, B_4 = \begin{bmatrix} 0 & 0 & \beta_{AB_w} \\ 0 & 0 & \beta_{EF_w} \\ 0 & 0 & \beta_{AB_w} \\ 0 & 0 & \beta_{EF_w} \end{bmatrix}$$

$$B_3 = \begin{bmatrix} 0 & K & 0 \\ 0 & 0 & 0 \\ 0 & V_C K & 0 \\ 0 & 0 & 0 \end{bmatrix}, B_2 = 0, B_1 = 0$$

List of Captions

Figure 1. Control scheme with neural state space model (emulator) $\mathcal{M}_i$ and a neural state space controller $\mathcal{C}_j$. The model $\mathcal{M}_i$ is described by means of a nonlinear state space model of the form $\hat{x}_{k+1} = f(\hat{x}_k, u_k, \epsilon_k; \theta_m)$, $y_k = g(\hat{x}_k, u_k, \epsilon_k; \theta_m)$ and $\mathcal{C}_j$ 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 θ_m for f, g and θ_c for h, s respectively. Several parametrizations are considered, resulting in a certain neural control problem Ξ_j^i related to a model $\mathcal{M}_i$ and a controller $\mathcal{C}_j$. The signal y_k represents the output of the plant, d_k the reference input, u_k the control signal and ϵ_k a white noise innovations signal.

Figure 2. Specific case of Figure 1 with neural state space model $\mathcal{M}_2$ and neural state space controller $\mathcal{C}_2$.

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. $\mathcal{S}_i$ is called the *augmented plant model.*

Figure 4. System identification: a nonlinear dynamic model and its corresponding Narendra's sensitivity model for generating the gradient of the cost function with respect to the parameter vector θ in a prediction error learning algorithm ($\alpha, \beta \in \theta$).

Figure 5. 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 Ω_i on the input arguments $[p_k; w_k]$. The matrices Ω_i are diagonal, the diagonal elements depend on p_k, w_k and satisfy the condition $\|\Omega_i\| \leq 1$. All Ξ_j^i neural control problems are formulated as NL_qs.

Figure 6. Interpretation of Theorem 2 for $q = 1$: Fig.a/ upper bounds $1/c_\alpha$ (full line) and $1/\delta_Q$ (dashed line) as a function of α. Fig.b/ trade-off curve between c_α and δ_Q. The curve shows that a high correction factor c_α 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_α. The limiting case $\delta_Q \to \infty$ corresponds to diagonal scaling.

Figure 7. Fig.a/ Discrete time Lur'e problem: a classical problem in nonlinear control theory. The problem can be formulated as a special case of NL_qs with $q = 1$. Fig.b/ Nonlinearity $f(\sigma)$ belonging to sector $[0, K]$. The nonlinearity $\tanh(\sigma)$ belongs to sector $[0, 1]$.

Figure 8. Perturbed NL_qs: 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.

Figure 9. LFT (Linear Fractional Transformation) representation as it occurs in modern robust control theory (μ-theory). M is the augmented plant (nominal system) and the uncertainty on M is represented through the feedback interconnection Δ.

Figure 10. Example of the modified dynamic backpropagation algorithm with respect to classical dynamic backpropagation: Fig.a/ classical dynamic backpropagation: best tracking performance on a specific reference input; Fig.b/ and c/ modified dynamic backpropagation with guaranteed global asymptotic stabiblity of the closed loop system according 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 Fig.b and c respectively. (Full line: output of reference model, dashed line: output of controlled plant).

Figure 11. General LRGF architecture (Locally Recurrent Globally Feedforward neural network), proposed by Tsoi & Back. The architecture includes the local synapse feedback architecture as well as the local output feedback architecture. Assuming that the transfer functions $G_1, ..., G_{n-1}$ and G_n (which may have both poles and zeros) have a state space realization (A_i, B_i, C_i) and the nonlinearity $f(.)$ belongs to sector $[0, 1]$, this architecture can be written as an NL_1 system.

Figure 12. Example of a generalized cellular neural network (GCNN) consisting of four CNNs with feedforward, cascade and feedback interconnection. The network can be written as an NL_1 system.

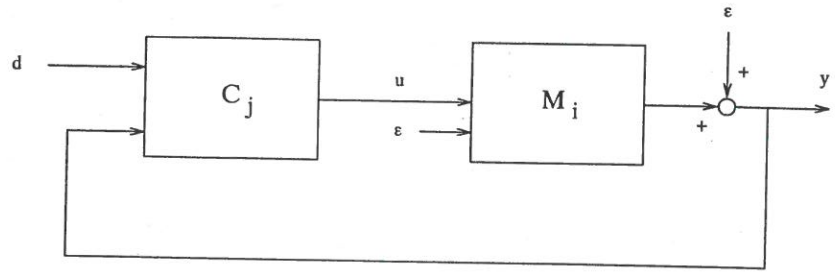


Figure 1:

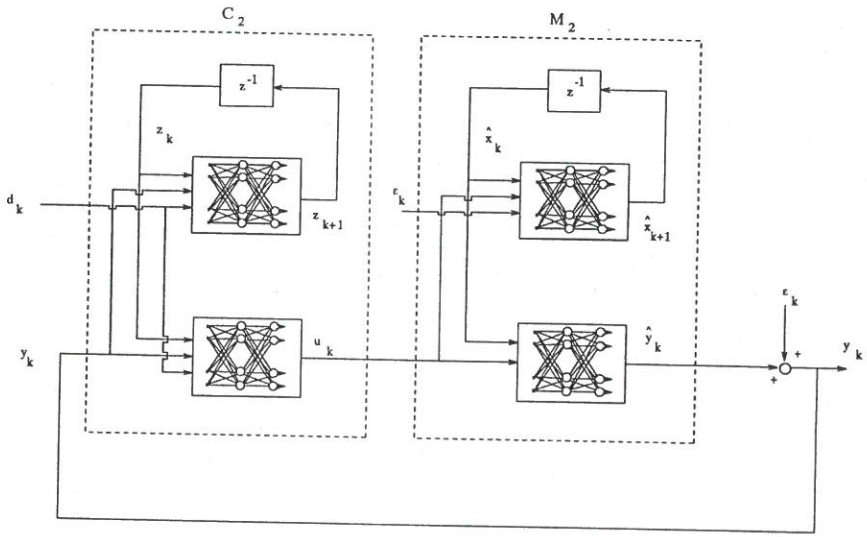


Figure 2:

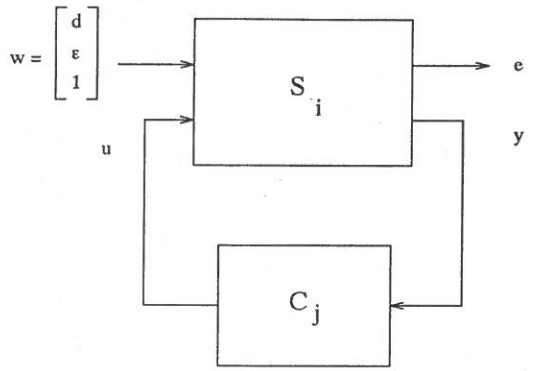


Figure 3:

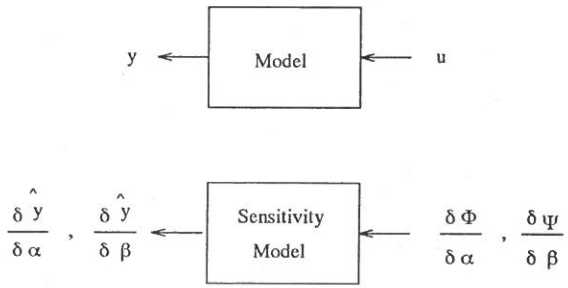


Figure 4:

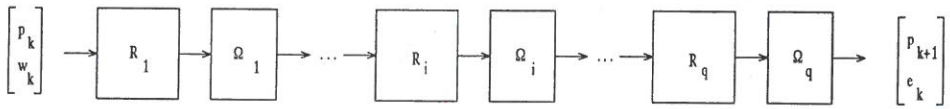


Figure 5:

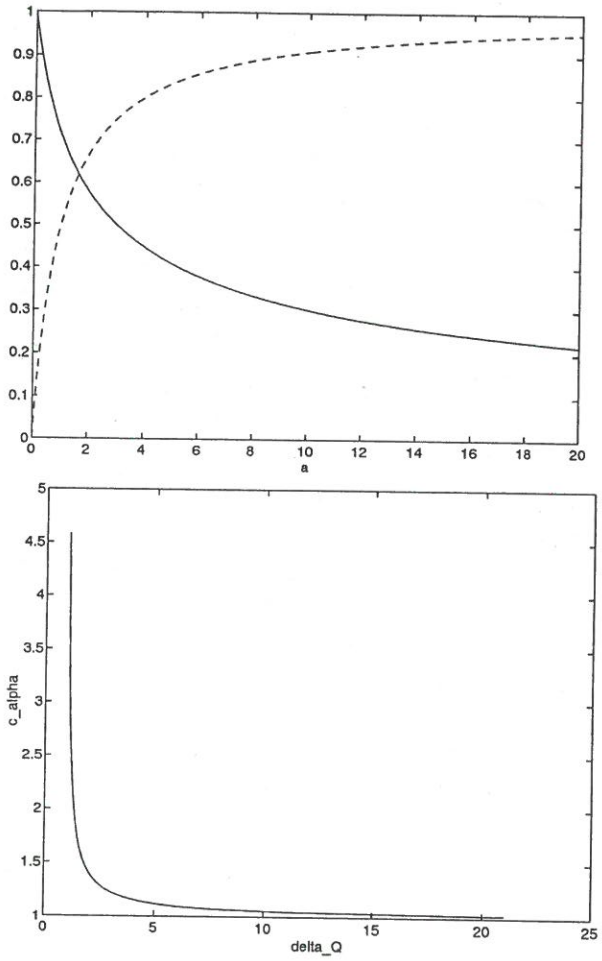
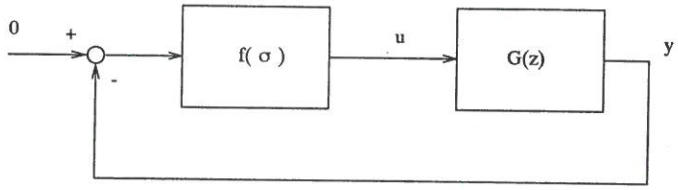


Figure 6:

a/



b/

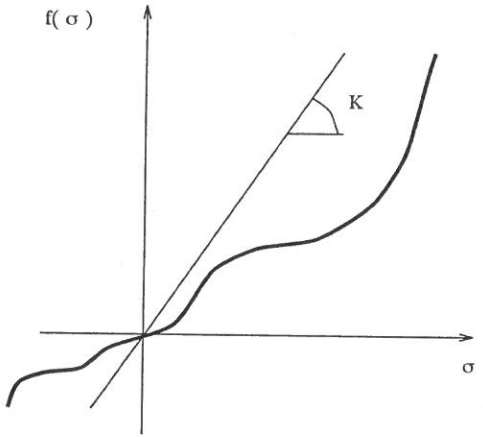


Figure 7:

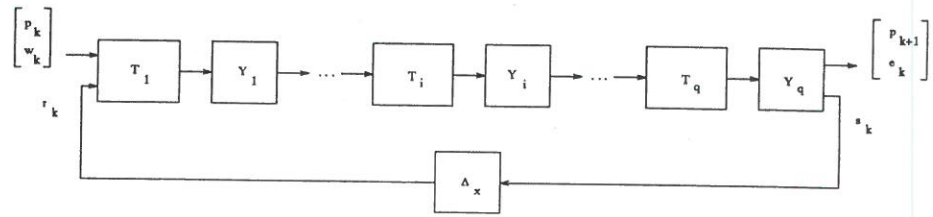


Figure 8:

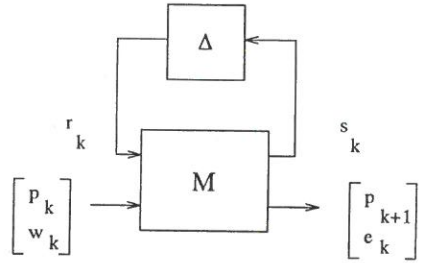


Figure 9:

a/

b/

c/

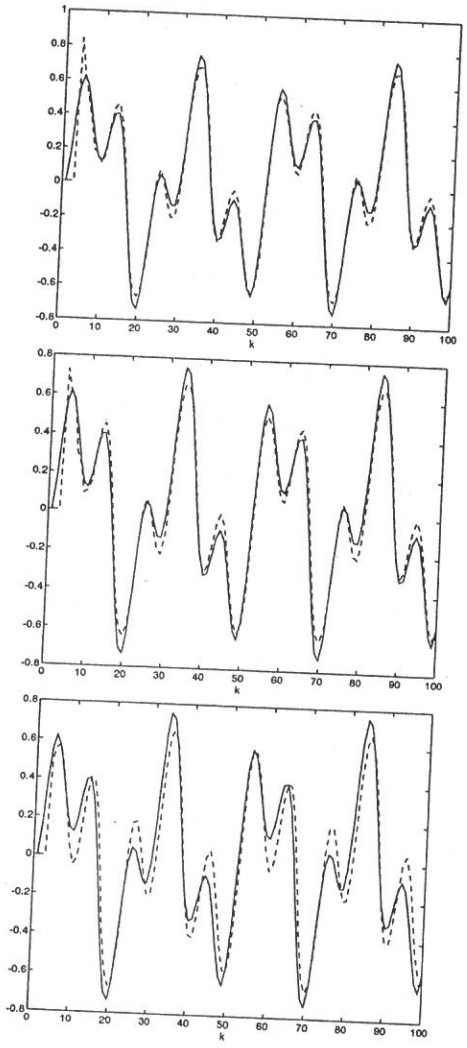


Figure 10:

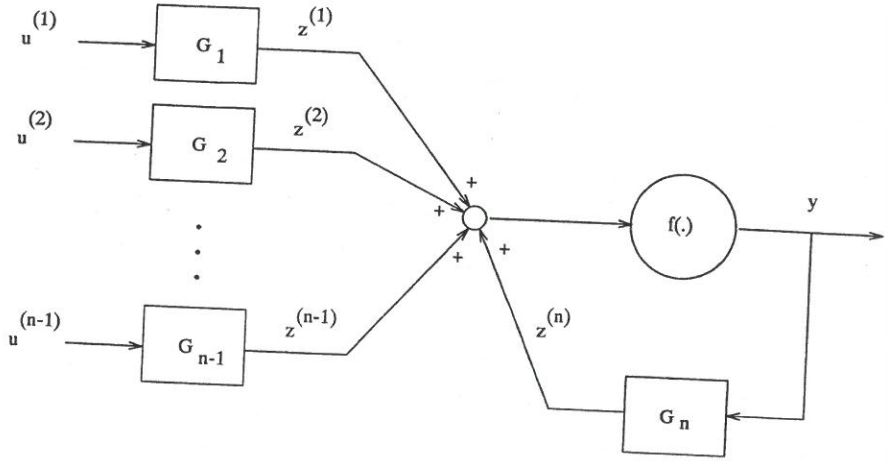


Figure 11:

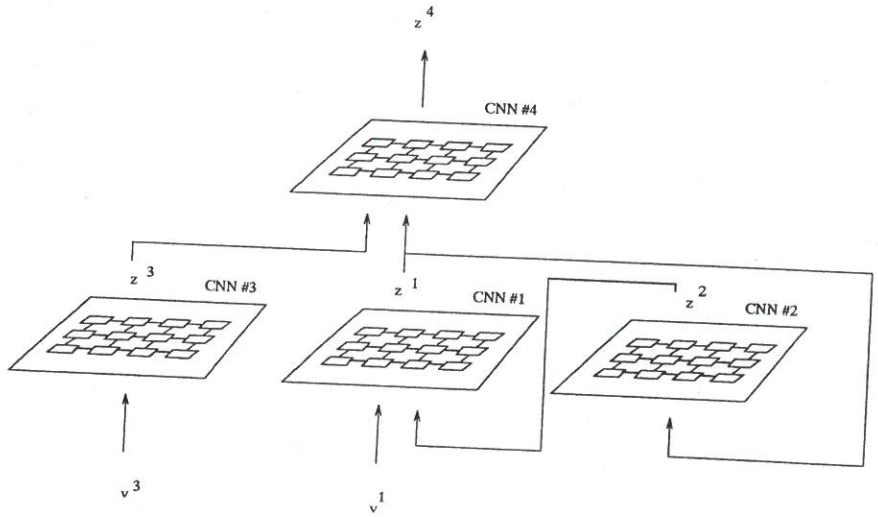


Figure 12: