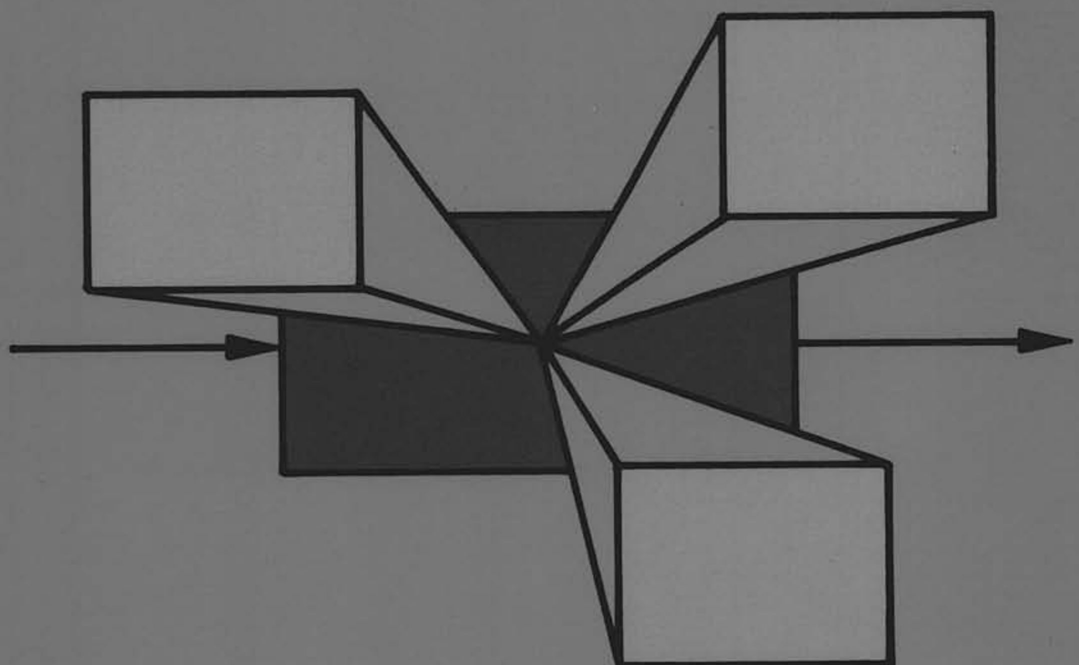
SELECTED TOPICS IN

Identification Modelling and Control

Volume 1, 1990

Edited by O.H. Bosgra and P.M.J. Van den Hof



Delft University Press

IDENTIFICATION, MODELLING AND CONTROL

SELECTED TOPICS IN IDENTIFICATION, MODELLING AND CONTROL

Progress Report on Research Activities in the Mechanical Engineering Systems and Control Group

Edited by O.H. Bosgra and P.M.J. Van den Hof

Volume 1, April 1990

Laboratory for Measurement and Control
Department of Mechanical Engineering and Marine Technology
Delft University of Technology

Delft University Press/1990

Published and Distributed by

Delft University Press
Stevinweg 1
2628 CN Delft
The Netherlands
Tel.: (0)15-783254
Telefax: (0)15-781661

By order of

Laboratory for Measurement and Control
Department of Mechanical Engineering and Marine Technology
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
Tel.: +31-15-786400; Telefax: +31-15-784717

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Selected

Selected topics in modelling, identification and control: progress report on research activities in the mechanical engineering systems and control group. – Delft: Laboratory for Measurement and Control, Department of Mechanical Engineering and Marine Technology, Delft University of Technology, Vol. 1 – ed. by O.H. Bosgra and P.M.J. Van den Hof. – ill. Met lit.opg.
ISBN 90-6275-591-7
SISO 656 UDC 531.7+681.5 NUGI 841

Cover design by Ruud Schrama

Contents

Volume 1, April 1990

Editorial

This is the first issue of a bi–annual publication in which recent research contributions of the Mechanical Engineering Systems and Control Group at Delft University are presented. The aim of the publication is to provide a means for fast publication of recent results of current research projects. It also serves to encourage post–graduate students and research associates to contribute to the written literature in an early stage of their research projects, and to get acquainted with the mechanisms of writing papers and dealing with reviews of their papers. This publication involves an account of some of the projects that are currently under study in our group, without aiming at completeness. Next issues therefore will amplify the picture of our group. We hope that this publication will contribute to creating fruitful communications with other groups and researchers on subjects on common research interests.
The research in our group aims at theory and applications of dynamic modelling, system identification and control system design. The applications involved include electromechanical servo systems (robots, electrical drives, wind power systems), and multivariable process control (power systems, chemical separation processes). In these projects a certain merging of system theory research and application–oriented projects takes place. We try to be involved only in those applications in which the achievements of recent theoretical results in model reduction, system identification and robust control will contribute to relevant engineering results.
The present issue especially contains results of projects oriented towards theoretical results. The very stimulating educational climate of the *Dutch Graduate Program on Systems and Control* certainly has contributed significantly to some of the results presented here, and consequently these efforts are gratefully acknowledged.
This issue also contains some contributions which have resulted from collaborative research projects performed in cooperation with industrial research partners. Such cooperative projects are experienced to be of paramount importance to our group.
The next issue will settle the balance between theory and applications by providing a number of contributions from various applied projects.

Okko Bosgra
Paul Van den Hof
Editors

Selected Topics in Identification,
Modelling and Control; Vol. 1, 1990

A family of reduced order models, based on open-loop balancing

Peter Heuberger

Lab. Measurement and Control, Dept. Mechanical Engineering and Marine Tech., Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. In this paper we generalize a well-known model reduction method, called balanced truncation, to a whole family of reduced order models, which are all based on the semi-canonical form of a balanced realization. This family will be parametrized by one real valued parameter, i.e. $\{G^{\alpha}, \alpha \in \mathbb{R}\}$. Several members of this family , as balanced truncation and singular perturbations, are already well known in the literature. The generalized approach presented covers both the continuous and discrete time case. Further conditions are given under which one can guarantee stability and minimality of the reduced order models, and a bound is given for the L_∞-norm of the error transfer function. It is shown that this frequency error can be much smaller than obtained with the standard methods.

Keywords. Model reduction; balanced realizations; singular perturbations; frequency error;

1 INTRODUCTION

Model reduction methods based on balanced realizations play an important role in various fields of system and control techniques. The first contribution in this area is due to Moore(1981), who introduced the truncation of balanced realizations of continuous time systems, which under weak conditions results in a balanced realization for the reduced order model, that is again stable and minimal. The same goes for the discrete time case, but the reduced model is not balanced any more. For these model reduction methods there is also a bound on the frequency error available.
Fernando and Nicholson (1982,1983), Al-Saggaf and Franklin (1988) and Liu and Anderson(1989) introduced the singular perturbation approach to reduce balanced models, which lead to reduced order models with the same nice properties. We will generalize these methods to a one parameter family of reduced order models and give the conditions under which these are stable and minimal. Further, we will give a bound for the frequency error and show by means of some examples that the generalized method we propose can lead to much smaller frequency errors than the 'known' methods..
The article is outlined as follows: In section 2 we will briefly repeat the main notions of balanced realizations and their relation with the Hankel singular values. Sections 3 and 4 deal with the currently existing methods in continuous and discrete time respectively. In section 5 we extend these methods to a generalized form, and the main properties of this method are given in theorem 5.4. We conclude with some examples in section 6.
Throughout this paper we only consider finite dimensional linear time invariant asymptotically systems, -which in the sequel will be abbreviated with -FDLTS systems- ,with state space realizations:

In continuous time:

$$\dot{x}(t) = Ax(t) + Bu(t) \qquad (1.1a)$$
$$y(t) = Cx(t) + Du(t) \qquad (1.1b)$$

In discrete time:

$$x_{k+1} = Ax_k + Bu_k \qquad (1.1c)$$
$$y_k = Cx_k + Du_k \qquad (1.1d)$$

The quadruple [A,B,C,D] is called a realization of the transfer function

$$G(p) = C[pI-A]^{-1}B+D \qquad (1.2)$$

where p is a complex variable. We use G(s) (p=s) for continuous time systems and G(z) (p=z) for discrete time systems.
We will make an extensive use of the ω-transformation to switch between continuous time and discrete time. This is the bilinear

transformation that maps the imaginary axis into the unit circle by ω: $s \to z=\frac{s+1}{1-s}$.
This transformation preserves stability and Hankel singular values. A thorough treatment is given in Glover (1984). We use the term ω-transformation for the transformation $s \to z$ as well as for $z \to s$; it will be clear from the context which one is used.

2 BALANCING TRANSFORMATIONS

In this section we explain the notion of balanced realizations, which was introduced by Moore (1981). Since this is a well known concept in the literature we will treat it only very briefly, giving the most important definitions and properties. In words one may say that a balanced realization of a system has the property that the amount of controllability of a certain element of the state vector is equal to the amount of observability of this element. As shown in for instance (Enns, 1984; Glover, 1984) we can consider the Gramians of a system as a tool to measure the controllability and observability of a realization. This is used in the balanced realization approach.

For a realization [A,B,C,D] of a FDLTS system G(p) the controllability and observability Gramian are defined as follows:

Continuous time:

$$P = \int_0^\infty e^{At}BB^Te^{A^Tt}dt \tag{2.1a}$$

$$Q = \int_0^\infty e^{A^Tt}C^TCe^{At}dt \tag{2.1b}$$

Discrete time:

$$P = \sum_{i=0}^{\infty} A^iBB^TA^{T^i} \tag{2.1.c}$$

$$Q = \sum_{i=0}^{\infty} A^{T^i}C^TCA^i \tag{2.1.d}$$

It is well known that these Gramians satisfy the following Lyaponov equations:

Continuous time:

$$AP + PA^T + BB^T = 0 \tag{2.2a}$$

$$A^TQ + QA + C^TC = 0 \tag{2.2b}$$

Discrete time:

$$APA^T + BB^T = P \tag{2.2c}$$

$$A^TQA + C^TC = Q \tag{2.2d}$$

A minimal realization [A,B,C,D] of a FDLTS system G(p) is called (*internally*) *balanced* w.r.t. Σ if

$$P = Q = \Sigma = \mathrm{diag}\{\sigma_1,\sigma_2,\cdots,\sigma_n\} \tag{2.3}$$

with $\sigma_i \geq \sigma_{i+1}$, $i=1,2,\cdots,n-1$ and $\sigma_n>0$.

The set $\{\sigma_i\}$ is the set of the non-zero Hankel singular values of the system G(p), which are the singular values of the Hankel operator of G(p) (Glover, 1984). In the sequel we will consider reduced order models of McMillan degree k<n and we will use the following partitioning of [A,B,C,D] and Σ, conformable with k and n:

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},\quad B=\begin{bmatrix}B_1\\ B_2\end{bmatrix},$$
$$C=[C_1 \;\; C_2],\quad \Sigma=\begin{bmatrix}\Sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix} \tag{2.4}$$

where $A_{11} \in \mathbb{R}^{k\times k}$ etc.

An interpretation of the Gramians of a realization is given in (Glover, 1984) and applied on a balanced realization this interpretation shows that the amount of energy to reach a state $x_\infty=x(\infty)$ from $x(0)=0$ is equal to $x_\infty^T\Sigma^{-1}x_\infty$. Thus if the i[th] singular value σ_i is very small it will take a large amount of energy to reach the state $x_o=e_i$, the i[th] unit vector, and therefore this state is almost unreachable. The interpretation of the observability Gramian shows that, with $u(t)\equiv 0$ $t\geq 0$, the amount of energy in the output on the interval $[0,\infty)$ is given by $x(0)^T\Sigma x(0)$. Hence initial states $x(0)=e_i$, with small σ_i make a small contribution to the output and are therefore almost unobservable. The equality of the energies leads to the term 'balancing'.

These realizations were introduced by Moore (1981) in the context of model reduction and they are of major importance in various applications. Laub (1980) gave an algorithm to calculate these realizations. It has also been shown (Gray and Verriest, 1987; Mullis and Roberts, 1976; Prabhakara, 1989) that these realizations are numerically superior to others, both with respect to parameter sensitivity and roundoff errors in simulation.

3 CONTINUOUS TIME MODEL REDUCTION

Based on the concept of balancing, Moore (1981) proposed a model reduction method for continuous time systems, which eliminates the states that are weakly observable and controllable. The singular values of the system provide a measure for determining how observable and controllable a certain state is, resulting in neglecting the states that correspond to the smallest singular values. This results in the following model reduction procedure.

DEFINITION 3.1. Let G(s) be a FDLTS system and [A,B,C,D] a balanced realization of G w.r.t. Σ, partitioned according to (2.4). Then $\hat{G}(s)=\mathcal{CB}_k(G)$, the *Continuous Balanced Reduced Model of order k*, is defined as

$$\hat{G}(s) = D + C_1[sI-A_{11}]^{-1}B_1 \tag{3.1}$$

◇

The rationale behind this procedure is to replace σ_i by 0, for i=k+1,..,n, and to retain the resulting system. This will generally lead to satisfactory results if the discarded singular values are relatively small. The next proposition gives the condition to retain stability and minimality.

PROPOSITION 3.2. [Moore, 1981; Pernebo and Silverman, 1983]. If $\sigma_k > \sigma_{k+1}$, then $[A_{11},B_1,C_1,D]$ is balanced w.r.t. Σ_1 and is a stable, minimal realization. ◇

One would like to have an exact measure of the error created by this procedure, but there is no such measure known. One can however bound the L_∞ norm of the error.

PROPOSITION 3.3. [Glover, 1984; Enns, 1984]. Under the conditions of definition 3.1 and proposition 3.2, the error of the approximation is bounded in the L_∞– norm:

$$\|G(s)-\mathcal{CB}_k(G(s))\|_\infty \leq 2\cdot(\sigma_{k+1}+\cdots+\sigma_n) \qquad (3.2)$$

and for k=n–1, this bound is tight. ◇

In general this model reduction method produces very good results, and is numerically efficient and stable. Only if the poles of the original system G(s) are close to the imaginary axis, then the balancing procedure tends to have numerical problems. A favorable feature of the method is the stability and minimality of the approximations.

A problem we have not discussed so far is the nonuniqueness of the balanced realizations. In (Ober, 1987; Ober and McFarlane, 1988) canonical forms are derived for balanced realizations.

Another favorable property of this method is the consistency, which means that

$\mathcal{CB}_r(\mathcal{CB}_k(G(s)) = \mathcal{CB}_r(G(s))$, if $r \leq k$, in other words once we have a k^{th} order reduced model, we can use this model to construct lower order approximations. This is a situation which will often occur in practical applications, where one is searching the lowest order approximation that would fulfil the designers specifications.

Fernando and Nicholson (1982), Al–Saggaf and Franklin (1988) and Liu and Anderson (1989) combined the balanced model reduction method with the method of singular perturbational approximations, resulting in the following model reduction method.

DEFINITION 3.4. Let G(s) be a FDLTS system and [A,B,C,D] a balanced realization of G w.r.t. Σ, partitioned according to (2.4). We define $\hat{G}(s)=\mathcal{CSB}_k(G)$, the *Continuous Singular Perturbational Balanced Reduced Model of order k*,

by $$\hat{G} = \hat{D} + \hat{C}[sI-\hat{A}]^{-1}\hat{B} \qquad (3.3a)$$

where, $$\hat{A} = A_{11} - A_{12}A_{22}^{-1}A_{21} \qquad (3.3b)$$

$$\hat{B} = B_1 - A_{12}A_{22}^{-1}B_2 \qquad (3.3c)$$

$$\hat{C} = C_1 - C_2A_{22}^{-1}A_{21} \qquad (3.3d)$$

$$\hat{D} = D - C_2A_{22}^{-1}B_2 \qquad (3.3e)$$

◇

The rationale behind this approximation method is as follows: Let x(t),u(t) and y(t) be respectively the state–, input– and output vector of the realization [A,B,C,D] and let x(t) be partitioned conformably as $x(t)=\begin{bmatrix}x_1(t)\\x_2(t)\end{bmatrix}$. The state space equations are:

$$\dot{x}_1(t) = A_{11}x_1(t) + A_{12}x_2(t) + B_1u(t) \qquad (3.4a)$$

$$\dot{x}_2(t) = A_{21}x_1(t) + A_{22}x_2(t) + B_2u(t) \qquad (3.4b)$$

$$y(t) = C_1x_1(t) + C_2x_2(t) + Du(t) \qquad (3.4c)$$

Assume that x_2 is a very fast stable state, such that (3.4b) can be approximated by $\dot{x}_2=0$:
This results in an algebraic state equation, which can be transformed to:

$$x_2(t) = -A_{22}^{-1}\left[A_{21}x_1(t) + B_2u(t)\right] \qquad (3.5a)$$

Substitution of (3.5a) in (3.4a,c) leads to:

$$\dot{x}_1(t) = \hat{A}x_1(t) + \hat{B}u(t) \qquad (3.5b)$$

$$y(t) = \hat{C}x_1(t) + \hat{D}u(t) \qquad (3.5.c)$$

The feasibility of this method is shown by the following two propositions, which have similar counterparts in $\mathcal{CB}$.

PROPOSITION 3.5. [Fernando and Nicholson, 1982; Liu and Anderson, 1989]. $[\hat{A},\hat{B},\hat{C},\hat{D}]$ given by (3.3) is balanced w.r.t. Σ_1 and is a stable, minimal realization. ◇

PROPOSITION 3.6. [Al–Saggaf and Franklin, 1988; Liu and Anderson, 1989]. The error of the approximation (3.3) is bounded in the L_∞ – norm:

$$\|G(s)-CSB_k(G(s))\|_\infty \leq 2(\sigma_{k+1}+\cdots+\sigma_n) \qquad (3.6)$$

and for k=n–1, this bound is tight. ◇

REMARK 3.7. Note that in definition 3.4 we explicitly pose the condition $\sigma_k \neq \sigma_{k+1}$. This is necessary to guarantee the stability of A_{22} and thus the existence of A_{22}^{-1} (Pernebo and Silverman, 1983).
This method replaces the 'fast' dynamical equations with algebraic ones, causing the static gain of $\mathcal{CSB}_k(G)$ to be equal to the static gain of G. Again this method is consistent, so if $r \leq k$, then $\mathcal{CSB}_r(\mathcal{CSB}_k(G(s)) = \mathcal{CSB}_r(G(s))$.
It should be stressed here that proposition 3.5 and 3.6. are valid without any condition on $x_2(t)$. This shows that $\mathcal{CSB}$ will be a good reduction method if the discarded singular values are small.
An important difference with $\mathcal{CB}$ is the better approximation of the low frequency components of the original system. ◇

4 DISCRETE TIME MODEL REDUCTION

So far we only dealt with model reduction of continuous time systems, based on balanced realizations. In this section we deal with the discrete time version, where we make a distinction between the balanced truncation, as proposed by Pernebo and Silverman (1983) and the result of combining $\mathcal{CB}$ and the ω-transformation, proposed by Al-Saggaf and Franklin (1988).

The discrete truncation is created in the same way as $\mathcal{CB}$:

Definition 4.1. Let G(z) be a FDLTS system and [A,B,C,D] a balanced realization of G w.r.t. Σ, partitioned according to (2.4). Then $\hat{G}(s)=\mathcal{DTB}_k(G)$, the *Discrete Truncated Balanced Reduced Model of order k*, is defined by:

$$\hat{G}(z) = D + C_1[zI-A_{11}]^{-1}B_1. \tag{4.1}$$

◇

Pernebo and Silverman (1983) show that this approximation is again minimal and stable, but contrary to the continuous time case this does not apply for the other subsystem $[A_{22},B_2,C_2,D]$. Also the given approximation will generally not be balanced, nor have $\{\sigma_1,\cdots,\sigma_k\}$ as its singular values. Nevertheless the same bound for the L_∞-norm of the approximation error holds true.

Proposition 4.2. [Al-Saggaf and Franklin, 1987]. The error of the approximation (4.1) is bounded in the L_∞-norm:

$$\|G(s)-CSB_k(G(s))\|_\infty \le 2(\sigma_{k+1}+\cdots+\sigma_n) \tag{4.2}$$

with strict inequality if $\sigma_k \neq \sigma_n$. ◇

Notice that this proposition implies that if $\sigma_k>\sigma_{k+1}$ we have a strict bound in (4.2)., contrary to (3.2) and (3.6).

Al-Saggaf and Franklin (1987) propose a method, that is slightly different from the above, but which is consistent with the continuous time method by applying the ω-transformation. We know that under this transformation Gramians are invariant (Glover, 1984), which shows that the transformation of a continuous realization, which is balanced with respect to Σ, is a discrete realization, balanced with respect to Σ. Since we implicitly assume stability we are assured that the ω-transformation is well defined. The reduction method they propose thus consists of the following steps:

Procedure 4.3.

1 Given a G(z) create a realization $[A_d,B_d,C_d,D_d]$ of G, balanced w.r.t. Σ.

2 Transform this realization with the ω-transformation to $[A_c,B_c,C_c,D_c]$.

3 Retrieve $[\hat{A}_c,\hat{B}_c,\hat{C}_c,\hat{D}_c]$ with definition 3.1.

4 Transform this realization with the ω-transformation to $[\hat{A}_d,\hat{B}_d,\hat{C}_d,\hat{D}_d]$. ◇

Clearly this procedure guarantees that the properties of $\mathcal{CB}$ are valid for this method, so the approximation is minimal, stable and balanced w.r.t. Σ_1 (2.4), if $\sigma_k>\sigma_{k+1}$, and the method is consistent. In calculating this procedure we do not have to go through all these steps. The following proposition shows how the calculation can be done without actually using the ω-transformation.

Proposition 4.4. [Al-Saggaf and Franklin, 1987] Let G(z) be a FDLTS system with realization $[A_d,B_d,C_d,D_d]$, partitioned according to (2.4) and balanced w.r.t. Σ, with $\sigma_k>\sigma_{k+1}$. Further, let $[\hat{A}_d,\hat{B}_d,\hat{C}_d,\hat{D}_d]$ be the k^{th} order approximation, calculated with procedure 4.3. Then:

$$\hat{A}_d = A_{11}-A_{12}[I+A_{22}]^{-1}A_{21} \tag{4.3a}$$

$$\hat{B}_d = B_1 - A_{12}[I+A_{22}]^{-1}B_2 \tag{4.3b}$$

$$\hat{C}_d = C_1 - C_2[I+A_{22}]^{-1}A_{21} \tag{4.3c}$$

$$\hat{D}_d = D - C_2[I+A_{22}]^{-1}B_2 \tag{4.3d}$$

◇

We will refer to this procedure as discrete balanced model reduction:

Definition 4.5. Let G(z) be a FDLTS system and [A,B,C,D] a balanced realization of G w.r.t. Σ, partitioned according to (2,4). Then $\hat{G}(s)=\mathcal{DB}_k(G)$, the *Discrete Balanced Reduced Model of order k*, is defined as

$$\hat{G}(z) = \hat{D} + \hat{C}[zI-\hat{A}]\hat{B} \tag{4.4}$$

with $[\hat{A},\hat{B},\hat{C},\hat{D}]$ defined by (4.3). ◇

The discrete analog of $\mathcal{CSB}$ has been reported by [Fernando and Nicholson '83], however without the adaptation of the D-matrix, which was added in (Al-Saggaf and Franklin, 1988; Liu and Anderson, 1989). It is in fact the result of the previous procedure if $\mathcal{CB}$ is replaced by $\mathcal{CSB}$. It is again a combination of balancing and singular perturbational model reduction.

Definition 4.6. Let G(z) be a FDLTS system and [A,B,C,D] a balanced realization of G w.r.t. Σ, partitioned according to (2.4). We define $\hat{G}(z)=\mathcal{DSB}_k(G)$, the *Discrete Singular Perturbational Balanced Reduced Model of order k*, by

$$\hat{G}(z) = \hat{D} + \hat{C}[zI-\hat{A}]^{-1}\hat{B} \tag{4.5a}$$

where,

$$\hat{A} = A_{11} + A_{12}[I-A_{22}]^{-1}A_{21} \tag{4.5b}$$

$$\hat{B} = B_1 + A_{12}[I-A_{22}]^{-1}B_2 \tag{4.5c}$$

$$\hat{C} = C_1 + C_2[I-A_{22}]^{-1}A_{21} \tag{4.5d}$$

$$\hat{D} = D + C_2[I-A_{22}]^{-1}B_2. \tag{4.5e}$$

◇

Again all the properties of $\mathcal{CSB}$ carry over to $\mathcal{DSB}$ which is stated in the following corollary.

Corollary 4.7. Let [A,B,C,D] be a balanced realization w.r.t. Σ of a FDLTS system G(z), with $\sigma_k > \sigma_{k+1}$. Let $\hat{G}(z) = \mathcal{DB}_k(G)$ or $\hat{G}(z) = \mathcal{DSB}_k(G)$ with realization given by (4.3) or (4.5). Then this realization is stable, minimal and balanced w.r.t. Σ_1. Furthermore the approximation error is bounded in the L_∞–norm by

$$\|G(z) - \hat{G}(z)\|_\infty \leq 2 \cdot (\sigma_{k+1} + \cdots + \sigma_n) \qquad (4.6)$$

and if k=n–1 the bound is achieved. ◇

Liu and Anderson (1989) propose to use combinations of the standard methods to get better results on frequency error and DC–error (static gain). Such a combination consists of two or more steps, for instance using $\mathcal{CB}$ to reduce from order n to k_1 and $\mathcal{CSB}$ to reduce further to order k_2. In the next section we propose a generalized method, which can make these errors considerably smaller and can be accomplished in only one step.

5 A FAMILY OF MODEL REDUCTION METHODS BASED ON BALANCING

In this paragraph we combine the results of the previous two paragraphs and define a generalized model reduction method, that has the five methods –$\mathcal{CB}$, $\mathcal{CSB}$, $\mathcal{DB}$, $\mathcal{DTB}$, $\mathcal{DSB}$– as special cases. First we will give the rationale of the method that we propose, after which we will formally define it.

The idea behind this framework is, among others, due to Santiago and Jamshidi (1986) and is based on a general partitioning of a transfer function matrix.
Let G(p) be a finite dimensional linear time invariant system (not necessarily stable) with a realization [A,B,C,D] , $G(p) = D + C[pI-A]^{-1}B$. Let 0<k<n and let A,B,C be partitioned conformably as in (2.4). We can rewrite G(p) in the following partitioning:

$$G(p) = \bar{D}(p) + \bar{C}(p)[pI-\bar{A}(p)]^{-1}\bar{B}(p) \qquad (5.1a)$$

with

$$\bar{A}(p) = A_{11} + A_{12}\,[pI-A_{22}]^{-1}A_{21} \qquad (5.1b)$$

$$\bar{B}(p) = B_1 + A_{12}\,[pI-A_{22}]^{-1}B_2 \qquad (5.1c)$$

$$\bar{C}(p) = C_1 + C_2\,[pI-A_{22}]^{-1}A_{21} \qquad (5.1d)$$

$$\bar{D}(p) = D + C_2\,[pI-A_{22}]^{-1}B_2 \qquad (5.1e)$$

We use no specific time domain here, implying that we can either use p=s or p=z. All model reduction methods we considered so far can in fact directly be obtained from this partitioning by approximating $[\bar{A}(p),\bar{B}(p),\bar{C}(p),\bar{D}(p)]$ by $[\bar{A}(p_o),\bar{B}(p_o),\bar{C}(p_o),\bar{D}(p_o)]$ with p_o a fixed parameter. Take $p_o=\infty$ and p=s then we have $\mathcal{CB}$; $p_o=1$ and p=z results in $\mathcal{DSB}$ etc.

The approach presented here is to define the family of reduced order models by letting p_o vary over $\mathbb{R}$ and to find the restrictions, that have to be satisfied in order to guarantee stable and minimal reduced order models.
Note that from the above partitioning of G(p) one would expect that p_o should be chosen on the imaginary axis or the unit circle, which in general would lead to complex valued reduced order systems. However we will show that it does make sense to choose p_o real.
Santiago and Jamshidi (1986) propose this idea to define a model reduction method for systems with unstable poles, which in continuous time comes down to:

1. find a p_o such that $A-p_oI$ is stable
2. apply $\mathcal{CB}$ on $[A-p_oI,B,C,D]$
3. shift the resulting $\hat{A}$ back to $\hat{A}+p_oI$.

It will be clear that the result of this procedure depends highly on the choice of p_o and can change the number of unstable poles, which in applications as control design is not advisable.
They also indicate that different values of p_o might lead to better results for systems with different time scales. In the next definition we formalize this reduction method.

Definition 5.1. Let G(p) be a FDLTS system and [A,B,C,D] a balanced realization of G w.r.t. Σ with $\sigma_k > \sigma_{k+1}$, partitioned according to (2.4). Let $\alpha \in \mathbb{R}$ such that $\alpha \notin \sigma(A_{22})$. We define $\hat{G}(p) = \mathcal{GB}_k^\alpha(G)$, the *General Balanced Reduced Model with order k and reduction parameter α.* as

$$\hat{G}(p) = \hat{D} + \hat{C}[pI-\hat{A}]^{-1}\hat{B} \qquad (5.2a)$$

where

$$\hat{A} = A_{11} + A_{12}[\alpha I-A_{22}]^{-1}A_{21} \qquad (5.2b)$$

$$\hat{B} = B_1 + A_{12}[\alpha I-A_{22}]^{-1}B_2 \qquad (5.2c)$$

$$\hat{C} = C_1 + C_2[\alpha I-A_{22}]^{-1}A_{21} \qquad (5.2d)$$

$$\hat{D} = D + C_2[\alpha I-A_{22}]^{-1}B_2. \qquad (5.2e)$$

◇

As stated before,we defined no time domain, writing G(p) where p can be both p=s or p=z. The following proposition shows how definition 5.1 covers the model reduction methods, defined previously.

Proposition 5.2. Let G(p) be a FDLTS system. If

p=s: $\mathcal{CB}_k = \mathcal{GB}_k^\infty$, $\mathcal{CSB}_k = \mathcal{GB}_k^0$,

p=z: $\mathcal{DB}_k = \mathcal{GB}_k^{-1}$, $\mathcal{DSB}_k = \mathcal{GB}_k^1 \Leftrightarrow \mathcal{DTB}_k = \mathcal{GB}_k^\infty$. ◇

Proof: Follows directly from substitution of the values of α in definition 5.1 and comparing the result with the definitions of the 'standard' model reduction methods. □

The next lemma shows the effect of the ω–transformation on the different reduction methods.

LEMMA 5.3. [Heuberger, 1990] Let G(p) be a FDLTS system.

1. If p=s and $G_d(z) = \omega(G(s))$ then $\omega(\mathcal{GB}_k^\alpha(G))=\mathcal{GB}_k^\beta(G_d)$ with $\beta = \frac{1+\alpha}{1-\alpha}$.
2. If p=z and $G_c(s) = \omega(G(z))$ then $\omega(\mathcal{GB}_k^\alpha(G))=\mathcal{GB}_k^\beta(G_c)$ with $\beta = \frac{\alpha-1}{\alpha+1}$. ◇

The next theorem is the main result of this paper. It gives the conditions under which $\mathcal{GB}$ will lead to stable and minimal reduced order models and gives a bound for the approximation error.

THEOREM 5.4.
Consider the situation as formulated in definition 5.1. Let $AR \subset \mathbb{R}$, the admissible region, be given by:

$$AR=[0,\infty] \text{ if G continuous,} \tag{5.3a}$$
$$AR=[-\infty,-1]\cup[1,\infty] \text{ if G is discrete.} \tag{5.3b}$$

Then

1 $[\hat{A},\hat{B},\hat{C},\hat{D}]$ is stable and minimal for $\alpha \in AR$.
2 The error of the approximation is bounded: $\|G-\hat{G}\|_\infty \le 2(\sigma_{k+1}+\cdots\sigma_n)$ for $\alpha \in AR$ with strict inequality if α is in the interior of AR. ◇

PROOF: Appendix A.

REMARK 5.5:
1. In this section we used a real valued parameter α, which in fact indexes the family of reduced order models. It is straightforward to show that one can get a similar result if α is allowed to be complex. In this case the admissible region AR, as defined in (5.3) is $\{\alpha\in\mathbb{C},\ \mathrm{real}(\alpha)\ge 0\}$ for continuous time systems and $\{\alpha\in\mathbb{C},\ |\alpha|\ge 1\}$ for discrete time systems. Note that in general this leads to complex valued reduced order systems, which is the reason we did not focus on this.
2. The reduction parameter α connects the standard methods in a continuous way. This is understood best if we consider the continuous time case, where $\alpha=\infty$ coincides with $\mathcal{CB}$ and $\alpha=0$ with $\mathcal{CSB}$. Variation of α from 0 to ∞ gives a continuous transition from a match on the very low to the very high frequencies, with the result that in the interval $(0,\infty)$ these two goals are more or less weighted against each other with weight factors depending on the choice of α.
Hence the freedom in the choice of α can be used to optimize the frequency characteristics of the approximant according to the designers specifications, in the bandwidth which is of importance. This is a major advantage over the standard methods that only leave the choice between matching either the very high or very low frequency behavior.
3. From practical experiments we have the very strong impression that there exists only one value of α for which the L_∞-norm of the error transfer functions reaches a minimum. If one would define a function $f(\alpha) = \|G(p)-GB_k^\alpha(G)\|_\infty$ then this function will have only one global minimum $f(\alpha_{min})$ and no local minima. If we consider the continuous time then $f(\alpha)$ will reach 2 maxima on the boundary of the admissible region, i.e. $\alpha=0$ and $\alpha=\infty$, and have no other local maxima. However we have not yet succeeded in finding a value for α_{min} and $f(\alpha_{min})$.
4. Liu and Anderson (1989) propose to combine the standard methods in order to improve the frequency characteristics of the reduced order model. They use for instance the combination of $\mathcal{CB}$ and $\mathcal{CSB}$ and show through some examples how the error bound improves. We believe that a 'good' choice of α can do an even better job in just one reduction step without using several 'one at a step' reductions. As mentioned before we have not yet succeeded in finding rules for the optimal value of α, but the improvement can be quite impressive, as will be shown in the next section. ◇

6 EXAMPLES

EXAMPLE 1

As a first example of the influence of the parameter α, we consider a simple 3rd order system, which was used in (Enns, 1984). The transfer function is:

$$G(s) = \frac{(s+0.8)(s+2)}{(s+1.5)(s^2+1.4s+1)}$$

The singular values of this system are

$$\{\sigma_1,\sigma_2,\sigma_3\} = \{0.6985,\ 0.1599,\ 0.0053\}.$$

We approximate G(s) with 1st order reduced models, applying different values of α. As to be expected the result shows that for $\alpha=0$ ($\mathcal{CSB}$) the approximation has the same static gain as G(s), while for $\alpha=\infty$ ($\mathcal{CB}$) the high frequency behavior is matched. This is shown in Fig. 1 and Fig 2. Figure 1 shows the Bode plot of the original model and the approximations with $\alpha=0,1,\infty$. In Fig. 2 the frequency errors are shown for the same values of α. It is clear that the response for $\alpha=1$ is more or less in between the responses of the approximations with $\alpha=0$ and $\alpha=\infty$.
Figure 3 depicts the L_∞-norm of the error transfer function as a function of α, to be precise it is a plot of the function $f(\alpha) = \|G(s)-\mathcal{GB}_1^\alpha(G)\|_\infty$.

The form of this function is typical for what we found with all kind of different systems, which lead to the impression mentioned in remark 5.5–3.

EXAMPLE 2

We consider the example used by Liu and Anderson (1989) and create 2nd order approximations of

$$G(s) = \frac{(s+4)}{(s+1)(s+3)(s+5)(s+10)}$$

with singular values

$$\{\sigma_1,\sigma_2,\sigma_3,\sigma_4\} = \{1.5938x10^{-2}, 2.7243x10^{-3}, 1.272x10^{-4}, 8.006x10^{-6}\}$$

The theoretical bound is $2(\sigma_3+\sigma_4) = 2.7024x10^{-4}$. Liu and Anderson use a mixture of one at a step standard reductions (*CB* and *CSB*) to compare the frequency errors and the errors at DC (s=0). This means they first reduce to order 3 and then from order 3 to 2. This is denoted by *CB/CSB* if the first method used is *CB* and the second method is *CSB*. We calculated the optimal α with respect to the frequency error to be α=11.83. This results in a far better frequency error, as can be seen in Table 1.

TABLE 1. Frequency Errors of the Reductions x 10^{-4}

	CB	*CSB*	*CB/CSB*	*CSB/CB*	*GB*
$\lVert G-\hat{G}\rVert_\infty$	2.4802	2.3692	2.5248	2.6402	1.3415
DC–err	2.384	0.0	0.1601	2.5441	0.9810

While the DC–error is still acceptable,the frequency error is almost half of what can be achieved by the other methods.
In Fig. 4 the frequency errors of the approximation are shown on the whole frequency scale, and it shows that *GB* makes a trade–off between matching high and low frequencies. It should be pointed out however that the frequency error of *CSB* is only large for the high frequencies, which may be of no interest. In Fig. 5 we depicted the frequency error as a function of α, and it shows a similar curve as Fig. 3 , with only one global minimum.
It is also interesting to consider the Hankel norm of the approximation error, where the theoretical bound is σ_3=1.272x10^{-4}. This is given in Table 2, from which we conclude that for this example *GB* is also superior to the other methods in the Hankel norm.

TABLE 2. Hankel norm of Reduction Errors x 10^{-4}

	CB	*CSB*	*CB/CSB*	*CSB/CB*	*GB*
$\lVert G-\hat{G}\rVert_H$	2.4291	1.8646	2.5874	1.9722	1.3177

For Table 2 we used again α=11.83 for *GB*, but this is not the optimal value of α for the Hankel norm. In Fig. 6 the Hankel norm of the reduction error is shown as a function of the reduction parameter and it reaches a minimum 1.2931x10^{-4} in α=13.28, which is near the theoretical underbound. This shows that, for the optimal value of α, *GB* results in a very good approximation with respect to the Hankel norm. In Fig. 6 we see again that there is only one global minimum.

7 CONCLUSIONS

It has been shown how the standard model reduction techniques, based on internally balanced realizations fit in naturally within a general framework of a one parameter family of reduced order models. For this family we have given conditions under which stability and minimality of the resulting approximations are assured and we have given a bound for the L_∞–norm of the error transfer functions, which is never worse than the bounds, that are known for the standard methods. This general framework leads to an extra freedom –the so called reduction parameter– to design reduced order models, which makes a considerable reduction possible of the frequency error in the bandwidth one wishes to consider. Optimal values of the reduction parameter are not yet known, but practical experience indicates that such optima always exist, both one with much better frequency behavior than the results of the standard methods as well as with lower Hankel norm of the error transfer function. Further research on this subject is therefore highly recommended.

ACKNOWLEDGEMENT

This research was sponsored by Shell Research/KSLA, Grant 85010 MSE. We wish to thank explicitly Dr.W.Schinkel and Dr.G.A.Van Zee for helpful discussions and their cooperation.

REFERENCES

Al–Saggaf and G.F. Franklin (1987). An error bound for a discrete reduced of a linear multivariable system. *IEEE Trans. Automat. Contr. AC–32*, 815–819.

Al–Saggaf and G.F. Franklin (1988). Model reduction via balanced realizations: an extension and frequency weighting techniques. *IEEE Trans. Automat. Contr. AC–33*, 687–692.

Enns D.F. (1984), *Model Reduction for Control System Design.* PhD. diss., Dept. of Aeronautics and Astronautics, Standford University.

Fernando K.V. and H. Nicholson (1982). Singular perturbational model reduction of balanced systems. *IEEE Trans. Automat. Contr., AC–27*, 466–468

Fernando K.V. and H. Nicholson (1983). Singular perturbational approximations for discrete time balanced systems. *IEEE Trans. Automat. Contr. AC-28,* 240–242

Glover K. (1984). All optimal Hankel norm approximations of linear multivariable systems and their L_∞ - error bound. *Int.J.Control 39,* 1115–1193.

Gray W.S. and E.I. Verriest (1987). Optimality properties of balanced realizations: minimum sensitivity. In: *Proc. 26th Conf. Decision and Control,* Los Angeles, 124–128.

Heuberger P.S.C. (1990). *On Model Reduction, Complete Orthonormal Functions and System Identification.* PhD. thesis, Delft University of Technology. To appear.

Kailath T. (1980). *Linear Systems,* Prentice Hall, New Jersey.

Laub A.J. (1980). Computation of balancing transformations. *Proc. JACC ,FA8–E,* San Francisco, Ca.

Liu Y. and B.D.O. Anderson (1989). Singular perturbation approximation of balanced systems. In: *Proc. 28th Conf. Decision and Control,* Tampa, Florida, 1355–1360.

Moore B.C. (1981). Principal component analysis in linear systems: controllability, observability and model reduction. *IEEE Trans. Automat. Contr., AC–26,* 17–32.

Mullis C.T. and R.A. Roberts (1976). Synthesis of minimum roundoff noise fixed point digital filters. *IEEE Trans. Circuits Syst., CAS–23,* 551–561.

Ober R.J. (1987). Balanced realizations: canonical form, parametrization, model reduction. *Int.J.Control 46,* 643–670.

Ober R. and D. McFarlane (1988). *Balanced Canonical Forms for Minimal Systems: a Normalized Coprime Factor Approach.* Report CUED/F–INFENG/TR11, University of Cambridge, England.

Pernebo L. and L.M. Silverman (1983). Model reduction via balanced state space representation. *IEEE Trans. Automat. Contr., AC–27,* 382–387.

Prabhakara Rao C.V.K. and G. Maskarinec (1989). Implementation of arbitrary realizations in orthogonal systems with applications to principle axis realizations. In: M.A. Kaashoek et al.(Ed.), *Progress in Systems and Control Theory,* Birkhauser Boston Inc., 1990. Proc. Int. Symp. MTNS, Amsterdam, 1989.

Santiago J.M and M. Jamshidi M. (1986) On the extensions of balanced approach for model reduction. *Contr. Theory and Adv. Techn. Vol.2, no.2,* 207–226.

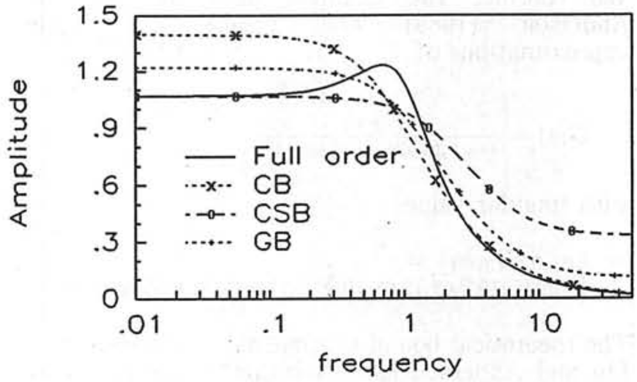


Fig. 1. Bode amplitude plot of approximations (example 1).
$\mathcal{CB}$ ($\alpha=\infty$), $\mathcal{CSB}$ ($\alpha=0$) and $\mathcal{GB}$ ($\alpha=1$)

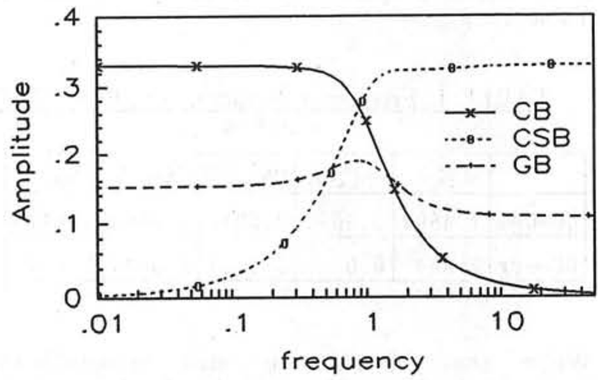


Fig. 2. Frequency errors of approximations (example 1).
$\mathcal{CB}$ ($\alpha=\infty$), $\mathcal{CSB}$ ($\alpha=0$) and $\mathcal{GB}$ ($\alpha=1$)

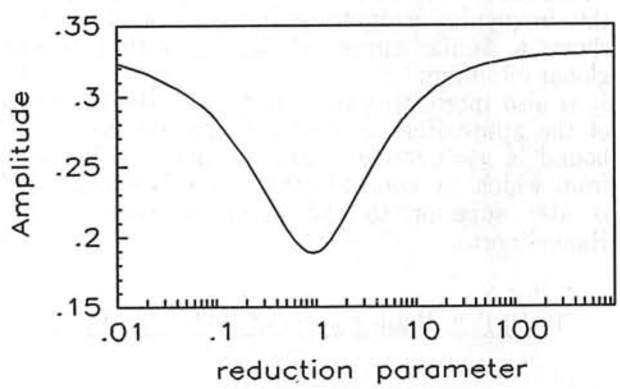


Fig. 3. $f(\alpha) = \|G(j\omega) - GB_1^\alpha(G(j\omega)\|_\infty$ (example 1)

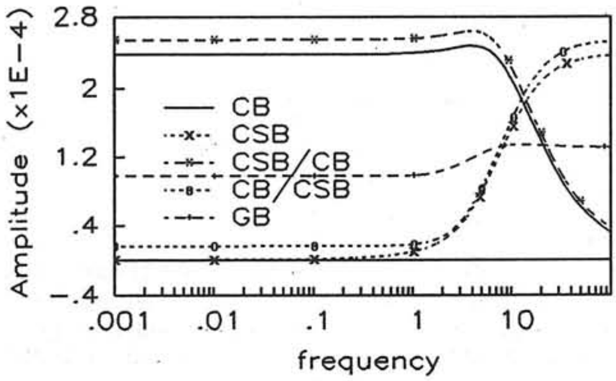


Fig. 4. Frequency errors of approximations (example 2).
$\mathcal{CB}$ ($\alpha=\infty$), $\mathcal{CSB}$ ($\alpha=0$) and $\mathcal{GB}$ ($\alpha=11.83$)

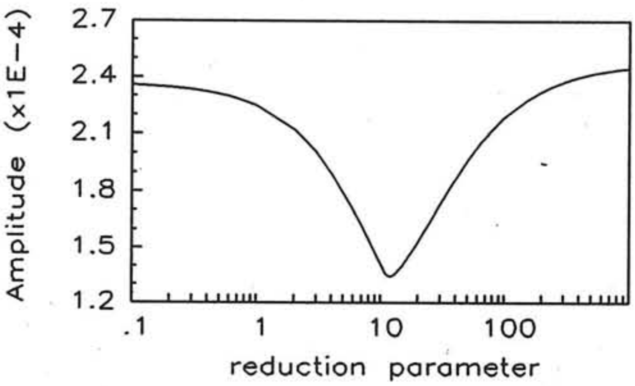


Fig. 5. $f(\alpha) = \|G(j\omega) - GB_2^{\alpha}(G(j\omega)\|_\infty$ (ex. 2)

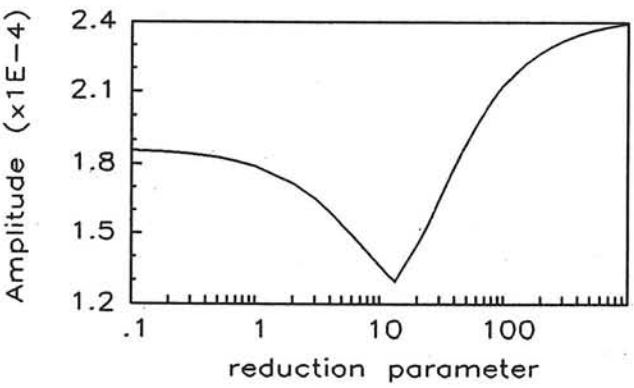


Fig. 6. $g(\alpha) = \|G(s) - GB_2^{\alpha}(G(s)\|_H$ (ex. 2)

APPENDIX A Proof of theorem 5.4.

Part 1 – Stability

Let G be discrete.
Proposition 5.2 shows that the case $|\alpha|=1$ is one of the standard methods for which stability was already proved (see corollary. 4.7).

So let $|\alpha|>1$. From Pernebo and Silverman (1983) we know that A_{22} is stable, so $\alpha\notin\sigma(A_{22})$ and thus $\hat{A}$ is well defined. Now suppose that $\hat{G}$ is not stable, so:

$$\exists x\in\mathbb{R},\ \lambda\in\mathbb{C}\ ,x\neq 0,\ |\lambda|\geq 1 \text{ with } \hat{A}x=\lambda x. \qquad \text{(A1)}$$

We will show that this leads to a contradiction

$$\text{(A1)} \Rightarrow \left[A_{11} + A_{12}[\alpha I - A_{22}]^{-1}A_{21}\right]x = \lambda x$$

$$\Rightarrow \begin{bmatrix}A_{11} & A_{12}\end{bmatrix}\begin{bmatrix} I \\ [\alpha I - A_{22}]^{-1}A_{21}\end{bmatrix}x = \lambda x \qquad \text{(A2)}$$

$$\begin{bmatrix}A_{21} & A_{22}\end{bmatrix}\begin{bmatrix} I \\ [\alpha I-A_{22}]^{-1}A_{21}\end{bmatrix} = \left[I+A_{22}[\alpha I-A_{22}]^{-1}\right]A_{21} = \alpha[\alpha I - A_{22}]^{-1}A_{21} \qquad \text{(A3)}$$

Combining (A2) and (A3) gives:

$$A\begin{bmatrix} I \\ [\alpha I-A_{22}]^{-1}A_{21}\end{bmatrix}x = \begin{bmatrix} \lambda I \\ \alpha[\alpha I-A_{22}]^{-1}A_{21}\end{bmatrix}x \qquad \text{(A4)}$$

$$\text{Let } y=\begin{bmatrix} y_1 \\ y_2\end{bmatrix}=\begin{bmatrix} I \\ [\alpha I-A_{22}]^{-1}A_{21}\end{bmatrix}x, \qquad \text{(A5)}$$

$$\text{then (A4) becomes:} \quad Ay = \begin{bmatrix}\lambda & 0 \\ 0 & \alpha\end{bmatrix}y. \qquad \text{(A6)}$$

Note that $y_1\neq 0\neq y_2$ since $y_1=x\neq 0$ and if $y_2=0$ then (A2) shows $A_{11}x=\lambda x$ but A_{11} is stable (Pernebo and Silverman ,1983), so $y_2\neq 0$.

Now $\|Ay\|_2 \leq \|A\|_s\|y\|_2 \leq \|y\|_2$, where $\|\ \|_s$ denotes the spectral norm (Pernebo and Silverman, 1983).

and $$\left\| \begin{bmatrix}\lambda & 0 \\ 0 & \alpha\end{bmatrix}y \right\|_2 \geq \|y\|_2$$

with equality iff $|\lambda|=|\alpha|=1$, since $y_1\neq 0\neq y$ and $|\alpha|,|\lambda|\geq 1$.

Thus we can conclude that $|\alpha|=1$, which is in contradiction with the assumption $|\alpha|>1$.

This shows that $\hat{G}$ is stable for $|\alpha|\geq 1$.
The stability of the continuous time equivalent follows from lemma 5.3., because the function $\alpha \to \frac{\alpha-1}{\alpha+1}$ maps $[-\infty,-1]\cup[1,\infty]$ into $[0,\infty]$. This proves the stability of $\hat{G}$ for $\alpha\in AR$.

□

Part 1 – Minimality

Consider the continuous time case. The case $\alpha=0$ is covered in proposition 3.5. Pernebo and Silverman (1983) proved the minimality of $\{\hat{A},\hat{B},\hat{C}\}$ for the discrete time case, with $\alpha=\infty$, which with lemma 5.3 shows the correctness for the continuous time case for $\alpha=1$. The correctness for $\alpha=\infty$ (continuous time) is given in proposition 3.2.

Now let $0<\alpha<\infty$ and define:

$$[\tilde{A},\tilde{B},\tilde{C},\tilde{D}] \overset{\Delta}{=} [\alpha^{-1}A,\alpha^{-\frac{1}{2}}B,\alpha^{-\frac{1}{2}}C,D]. \tag{A7}$$

It is easy to see that this realization is still balanced w.r.t. Σ and stable. We just showed that the reduction of such a system with $\alpha=1$ gives a stable minimal approximation $[\hat{\tilde{A}},\hat{\tilde{B}},\hat{\tilde{C}},\hat{\tilde{D}}]$ with:

$$\begin{aligned}
\hat{\tilde{A}} &= \tilde{A}_{11} + \tilde{A}_{12}\,[I-\tilde{A}_{22}]^{-1}\tilde{A}_{21} \\
&= \alpha^{-1}\Big[A_{11} + A_{12}\,[\alpha I-A_{22}]^{-1}A_{21}\Big] = \alpha^{-1}\hat{A}.
\end{aligned}$$

$$\begin{aligned}
\hat{\tilde{B}} &= \tilde{B}_1 + \tilde{A}_{12}\,[I-\tilde{A}_{22}]^{-1}\tilde{B}_2 \\
&= \alpha^{-\frac{1}{2}}\Big[B_1 + A_{12}\,[\alpha I-A_{22}]^{-1}B_2\Big] = \alpha^{-\frac{1}{2}}\hat{B}
\end{aligned}$$

$$\begin{aligned}
\hat{\tilde{C}} &= \tilde{C}_1 + \tilde{C}_2\,[I-\tilde{A}_{22}]^{-1}\tilde{A}_{21} \\
&= \alpha^{-\frac{1}{2}}\Big[C_1 + C_2\,[\alpha I-A_{22}]^{-1}A_{21}\Big] = \alpha^{-\frac{1}{2}}\hat{C}
\end{aligned}$$

Because $\{\alpha^{-1}\hat{A},\alpha^{-\frac{1}{2}}\hat{B},\alpha^{-\frac{1}{2}}\hat{C}\}$ is minimal the Popov–Belevitch–Hautus test (Kailath, 1980) shows immediately the minimality of $[\hat{A},\hat{B},\hat{C},\hat{D}]$. Consequently we have proven the minimality for continuous time systems for $0\leq\alpha\leq\infty$.
The minimality of the discrete counterpart follows from lemma 5.3.

□

Part 2.
Let E(p) be the difference transfer function:

$$E(p)=G(p)-\hat{G}(p), \text{ with } \hat{G}(p) = \mathcal{GB}_k^{\alpha}(G).$$

Our aim is to proof that $\|E(p)\|_\infty \leq 2\sum_{k+1}^{n}\sigma_i$ with strict inequality if α is in the interior of AR. The cases with G discrete and $\alpha=-1,1,\infty$ are proven by Al–Saggaf and Franklin (1987,1988). The ω-transformation then gives the corresponding bounds for G continuous and $\alpha=0,1,\infty$.

Now let G(s) be a continuous time system, $0<\alpha<\infty$ and $\hat{G}(s)=\mathcal{GB}_k^{\alpha}(G)$. Define $\tilde{G}(s)=G(\alpha s)$ and $\tilde{\hat{G}}(s)=\hat{G}(\alpha s)$. Note that (A7) defines a stable realization of $\tilde{G}$, still balanced with respect to Σ and that $\tilde{\hat{G}}(s)$ has a realization $[\alpha^{-1}\hat{A},\alpha^{-\frac{1}{2}}\hat{B},\alpha^{-\frac{1}{2}}\hat{C},\hat{D}]$. It is straightforward that $\tilde{\hat{G}}(s) = GB_k^1(\tilde{G})$ and hence:

$$\begin{aligned}
\|G(s)-\hat{G}(s)\|_\infty &= \\
&= \|G(\alpha s)-\hat{G}(\alpha s)\|_\infty \\
&= \|\tilde{G}(s)-\tilde{\hat{G}}(s)\|_\infty \\
&< 2(\sigma_{k+1}+\cdots+\sigma_n).
\end{aligned}$$

This completes the proof for G continuous and $0<\alpha<\infty$ and thus also for $0\leq\alpha\leq\infty$. Lemma 5.3 and the properties of the ω-transformation now gives the proof for G discrete and $\alpha\leq-1$ or $\alpha\geq1$ and hence we have proven part 2.

□

This completes the proof of theorem 5.4.

□

Selected Topics in Identification,
Modelling and Control; Vol. 1, 1990

Pole-zero cancellations in the multivariable mixed sensitivity problem

Sven G. Smit

Lab. Measurement and Control, Dept. Mechanical Engineering and Marine Techn., Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
Mechanics and Systems & Control Group, Philips Research Laboratories, WY–p, P.O.Box 80.000, 5600 JA, Eindhoven, The Netherlands

Abstract–The Mixed Sensitivity Problem, often proposed in literature being a formulation for handling both performance and robustness in controller design, is shown to have poor robustness properties. If the controller is computed using the state–space formulae of Glover and Doyle, 1988, for H_2 ($\gamma \to \infty$) or H_∞ norm bounded design, it is shown that all plant poles are canceled by controller zeros. This result holds in the multivariable case, regardless of the weighting functions which are introduced in order to specify performance and robustness.

Key Words–Mixed Sensitivity; pole–zero cancellation; H_2 and H_∞ optimization; robustness.

INTRODUCTION

It is expected that H_∞ control theory could lead to robust controller design, due to the absolute bound the H_∞ norm gives on the singular values of a transfer function matrix. A method which has frequently been suggested in literature for both handling performance and robustness (Kwakernaak, 1983; Maciejowski, 1989; Verma and Jonkheere, 1984; Francis, 1988) is the Weighted Mixed Sensitivity Problem (WMSP). In this problem the performance of a controlled system is measured by its tracking properties involving the sensitivity matrix of the system. The robustness properties of the controlled system are measured by the singular values of the complementary sensitivity matrix thereby specifying how much multiplicative output uncertainty the controlled system can tolerate before instability occurs. In this paper the influence of weightings and plant dynamics on the controller dynamics will be investigated, giving insight in the (robustness) properties of controllers evolving from the WMSP.
In this paper the WMSP will be described in more detail, by describing the control set up, the H_∞ mixed sensitivity criterion and the evolving standard plant. Then the controller satisfying an H_∞ norm bound on the WMSP–criterion will be derived and analyzed with respect to the location of its poles and zeros, in relation to the poles and zeros of weightings and plant. Finally the conclusions of this study are presented.

THE WEIGHTED MIXED SENSITIVITY PROBLEM

In the following the control set up in figure 1 is used, in which the controller K(s) is in cascade with the plant G(s) and measures the tracking error $\epsilon = y - w$, where ϵ, w and y possibly are vectors. The following transfer function matrices are defined:

Sensitivity Matrix:
$$S(s) = (I + G(s)K(s))^{-1} \quad (1)$$
Complementary Sensitivity Matrix:
$$T(s) = (I + G(s)K(s))^{-1}G(s)K(s) \quad (2)$$
Control Sensitivity Matrix:
$$C(s) = K(s)(I + G(s)K(s))^{-1} \quad (3)$$

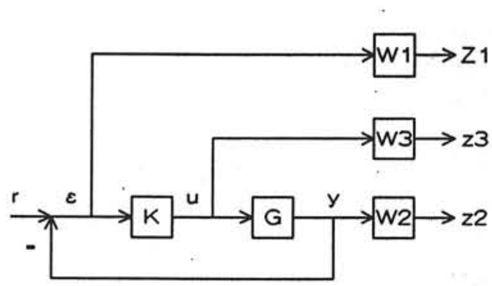


Figure 1 Control setup and cost functions

Typically the sensitivity matrix is used to measure the tracking and disturbance rejection properties of the controlled system, whereas the complementary sensitivity matrix is used for measuring the "singular value stability margin" of the controlled system for multiplicative output uncertainty (Doyle and Stein, 1981 and Safonov et al, 1981). The control sensitivity matrix is a measure for the control effort necessary to yield a certain performance. By using weighting functions the frequency dependence of the specifications on (1) – (3) for the set up in figure 1 can be handled yielding the H_∞ Weighted Mixed Sensitivity Problem of finding controllers such that:

$$\|T_{wz}\|_\infty < \gamma \Leftrightarrow \left\| \begin{matrix} W_1S \\ W_2T \\ W_3C \end{matrix} \right\|_\infty < \gamma \tag{4}$$

In (4) the controller is absorbed in T_{wz} so actually T_{wz} is a function of K(s). From this the following design problem can be stated:

Weighted Mixed Sensitivity Design Problem

Find stabilizing controllers K(s) such that:

$$\|T_{wz}(K)\|_\infty < \gamma$$

where the design parameter γ is chosen such that a stabilizing controller exists ■

DERIVATION OF THE CENTRAL H_∞ CONTROLLER

The formulae of Glover and Doyle [1988] for stabilizing controllers satisfying an H_∞ norm bound use a general standard plant in state space form as in (5) below:

$$SP \begin{cases} \dot{x} = Ax + B_1w + B_2u \\ z = C_1x + D_{11}w + D_{12}u \quad \text{and } u = Ky \\ y = C_2x + D_{21}w + D_{22}u \end{cases} \tag{5}$$

where $A \in \mathbb{R}^{n\times n}$, $w \in \mathbb{R}^{m1}$, $u \in \mathbb{R}^{m2}$, $z \in \mathbb{R}^{p1}$ and $y \in \mathbb{R}^{p2}$.

Now the A, B_i, C_j and D_{ij} matrices for the specific plant in (4) become (assuming $G(s) = C_g(sI-A_g)^{-1}B_g$ strictly proper):

$$A = \begin{bmatrix} A_g & 0 & 0 & 0 \\ -B_{w1}C_g & A_{w1} & 0 & 0 \\ B_{w2}C_g & 0 & A_{w2} & 0 \\ 0 & 0 & 0 & A_{w3} \end{bmatrix}$$

$$B_1 = \begin{bmatrix} 0 \\ B_{w1} \\ 0 \\ 0 \end{bmatrix} \quad B_2 = \begin{bmatrix} B_g \\ 0 \\ 0 \\ B_{w3} \end{bmatrix}$$

$$C_1 = \begin{bmatrix} -D_{w1}C_g & C_{w1} & 0 & 0 \\ D_{w2}C_g & 0 & C_{w2} & 0 \\ 0 & 0 & 0 & C_{w3} \end{bmatrix}$$

$$C_2 = [\, -C_g \; 0 \; 0 \; 0 \,]$$

$$D_{11} = \begin{bmatrix} D_{w1} \\ 0 \\ 0 \end{bmatrix} \quad D_{12} = \begin{bmatrix} 0 \\ 0 \\ D_{w3} \end{bmatrix}$$

$$D_{21} = [\, I \,] \quad D_{22} = [\, 0 \,] \tag{6}$$

The matrices A_{wi}, B_{wi}, C_{wi} and D_{wi} represent the weighting filters W_i. Below the assumptions made in Glover and Doyle (1988) are restated together with their specific implications for the WMSP:

A1 (A, B_2, C_2) is stabilizable and detectable

I1: The weightings W_1, W_3, and W_3 must be stable since they are not observable and the plant G(s) must be stabilizable and detectable

A2 rank D_{12} equals number of measurements y (p_2), rank D_{21} equals number of controls u (m2)

I2: D_{w3} must be of full rank m2

A3 A scaling of u and y, together with a unitary transformation of w and z, enables to assume without loss of generality that (by A2)

$$D_{12} = \begin{bmatrix} 0 \\ I \end{bmatrix}, \quad D_{21} = [\, 0 \;\; I \,] \text{ and}$$

$$D_{11} = \begin{array}{cc} \begin{bmatrix} D_{1111} & D_{1112} \\ D_{1121} & D_{1122} \end{bmatrix} & \begin{array}{l} m1\text{-}p2 \\ p2 \end{array} \\ \begin{array}{cc} p1\text{-}m2 & m2 \end{array} & \end{array}$$

I3: $Dw_3 = I$

A4 $D_{22} = 0$ (satisfied if G(s) is strictly proper)

A5 $\text{rank} \begin{bmatrix} A-j\omega I & B_2 \\ C_1 & D_{12} \end{bmatrix} = n + m_2 \;\forall\, \omega \in \mathbb{R}$

A6 $\text{rank} \begin{bmatrix} A-j\omega I & B_1 \\ C_2 & D_{21} \end{bmatrix} = n + p_2 \;\forall\, \omega \in \mathbb{R}$

A7 A_g is stable, this assumption facilitates the derivations below but is not essential.

The solution to an algebraic Ricatti equation (ARE) will be denoted via its Hamiltonian matrix, as

$$X = \text{Ric} \begin{bmatrix} A & -P \\ Q & -A^* \end{bmatrix}, \quad P = P^*, \; Q = Q^* \text{ where}$$

this implies that $X = X^*$ and

$$\begin{bmatrix} A & -P \\ Q & -A^* \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} I \\ X \end{bmatrix} [\, A - PX \,],$$

$\text{Re}\, \lambda_i[\, A-PX \,] < 0$

Now following the formulae in Glover, Doyle 1988 the controller satisfying the WMSP can be derived. Define:

$$D_{1\cdot} = [D_{11} \;\; D_{12}] = \begin{bmatrix} D_{w1} & 0 \\ 0 & 0 \\ 0 & I \end{bmatrix} \rightarrow$$

$$R = D_{1\cdot}^{*} D_{1\cdot} - \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} D_{w1}^{*} D_{w1} - \gamma^2 I & 0 \\ 0 & I \end{bmatrix}$$

$$D_{\cdot 1} = \begin{bmatrix} D_{11} \\ D_{21} \end{bmatrix} = \begin{bmatrix} D_{w1} \\ 0 \\ I \end{bmatrix} \rightarrow$$

$$\tilde{R} = D_{\cdot 1} D_{\cdot 1}^{*} - \begin{bmatrix} -\gamma^2 I & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} D_{w1} D_{w1}^{*} - \gamma^2 I & 0 \\ 0 & I \end{bmatrix} \tag{7}$$

Define X_∞ and Y_∞ as solutions to the following ARE's (assuming that solutions exist):

$$X_\infty =$$

$$\mathrm{Ric}\left\{ \begin{bmatrix} A & 0 \\ -C_1'C_1 & -A' \end{bmatrix} - \begin{bmatrix} B_1 & B_2 \\ -C_1' D_{11} & -C1' D_{12} \end{bmatrix} R^{-1} \begin{bmatrix} D_{11}'C_1 & B_1' \\ D_{12}'C_2 & B_2' \end{bmatrix} \right\}$$

$$\Leftrightarrow$$

$$\mathrm{Ric}\left\{ \begin{bmatrix} -A_{x\infty}' & -P_{x\infty} \\ Q_{x\infty} & A_{x\infty} \end{bmatrix} \right\}$$

$$Y_\infty =$$

$$\mathrm{Ric}\left\{ \begin{bmatrix} -A' & 0 \\ -B_1B_1' & A \end{bmatrix} - \begin{bmatrix} C_1' & C_2' \\ -B_1D_{11}' & -B_1D_{21}' \end{bmatrix} \tilde{R}^{-1} \begin{bmatrix} D_{11}B_1' & C_1 \\ D_{21}B_1' & C_2 \end{bmatrix} \right\}$$

$$\Leftrightarrow$$

$$\mathrm{Ric}\left\{ \begin{bmatrix} -A' + \begin{bmatrix} 0 & -C_g'B_{w1}' & 0 \\ 0 & 0 & 0 \end{bmatrix} & \gamma^{-2}C_1'C_1 - C_2'C_2 \\ 0 & A - \begin{bmatrix} 0 & 0 \\ -B_{w1}C_g & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix} \right\}$$

$$\Leftrightarrow$$

$$\mathrm{Ric}\left\{ \begin{bmatrix} -A_{y\infty}' & -P_{y\infty} \\ 0 & A_{y\infty} \end{bmatrix} \right\} \tag{8}$$

Here the (2,1) block is a zero matrix which implies that $Y_\infty = 0$ by lemma 3.1. General conditions for the occurrence of zero X_∞ and Y_∞ are given in this lemma.

Lemma 3.1
Y_∞ is zero if D_{21} is of full rank and $A_{y\infty}$ is stable and by duality X_∞ is zero if the D_1 block is of full rank and $A_{x\infty}$ is stable. ■

Proof See the appendix

The central H_∞ controller in state-space can easily be derived if Y_∞ is zero and equals:

$$K\{A_c, B_c, C_c, D_c\} = \left[\begin{array}{c|c} A_{hc} + B_{2c}C_{1c} & B_{1c} \\ \hline C_{1c} & 0 \end{array} \right] \tag{9}$$

where

$A_c =$

$$\begin{bmatrix} Ag - B_pXg & -B_pXw_1 & -B_pXw_2 & B_gCw_3 - B_pXw_3 \\ 0 & Aw_1 & 0 & 0 \\ Bw_2Cg & 0 & Aw_2 & 0 \\ BwXg & -BwXw_1 & -BwXw_2 & Aw_3 - Bw_3Cw_3 - BwXw_3 \end{bmatrix}$$

$$B_c = \begin{bmatrix} 0 \\ Bw_1 \\ 0 \\ 0 \end{bmatrix}$$

$C_c = [-B_gB_g' \;\; 0 \;\; 0 \;\; B_{w3}B_{w3}']$

$D_c = [\,0\,]$

where:

$X_\infty = [X_g \;\; X_{w1} \;\; X_{w2} \;\; X_{w3}]$

$B_P = [B_gB_g' \;\; 0 \;\; 0 \;\; B_gB_{w3}']$

$B_w = [B_{w3}B_g' \;\; 0 \;\; 0 \;\; B_{w3}B_{w3}']$

POLES AND ZEROS OF THE CENTRAL CONTROLLER

Now that the central H_∞ controller for the WMSP has been stated explicitly in section 3 the analysis of the influence of weighting functions and plant dynamics on the controller dynamics can be stated explicitly too. This is done in two lemmas regarding the poles and zeros of the central H_∞ controller for the WMSP.

Lemma 4.1 Poles of the central H_∞ controller
All the poles of the Sensitivity weight W_1 become poles of the central H_∞ controller ■

Proof
From (9) it is easy to verify that the eigenvalues of A_{w1} are eigenvalues of A_c and thus become poles of the central H_∞ controller. ■

Lemma 4.2 Zeros of the central H_∞ controller
If A_g is stable (by A7) all plant poles and the poles of the control weighting W_3 become zeros of the central H_∞ controller if the number of outputs of the controller does not exceed its number of inputs. ■

Proof
The (transmission) zeros of a system are defined by the $\lambda \in \mathbb{C}$ (if λ is not a pole of G(s)) for which:

$$\text{rank}\left[\begin{array}{c|c} \lambda I-A & B \\ \hline -C & D \end{array}\right] < n + \min(ni,no)$$

where ni is the number of controller inputs and no is the number of controller outputs.
The controller zeros then can be determined as the values λ for which:

$$\text{rank}\begin{bmatrix} \lambda I-A_{hc}-B_{2c}C_{1c} & B_{1c} \\ -C_{1c} & 0 \end{bmatrix} < n + \min(ni,no)$$

Since the rank remains unchanged by adding rows multiplied by constants to other rows the rank can also be evaluated from:

$$\text{rank}\begin{bmatrix} \lambda I-A_{hc} & B_{1c} \\ -C_{1c} & 0 \end{bmatrix} \Leftrightarrow$$

$$\text{rank}\begin{bmatrix} \lambda I-A_g & 0 & 0 & 0 & 0 \\ 0 & \lambda I-A_{w1} & 0 & 0 & B_{w1} \\ B_{w2}C_g & 0 & \lambda I-A_{w2} & 0 & 0 \\ 0 & 0 & 0 & \lambda I-A_{w3} & 0 \\ & & -C_{1c} & & 0 \end{bmatrix}$$

If now the rank of the matrix given above is evaluated by rows, noting that the number of controller outputs is assumed to be less or equal to the number of controller inputs, it is easily verified that the zeros of the controller equal the poles of the plant and the control weighting W_3. ■

Lemma 4.2 states that all stable plant poles become controller zeros. The following lemma strengthens this to pole zero cancellation.

lemma 4.3 Pole zero cancellation
All stable poles of the the plant to be controlled are canceled by controller zeros. ■

Proof
To determine the cancellation of all stable plant poles by controller zeros the transfer KG has to be regarded.

$$K(A_c,B_c,C_c,D_c)G(A_g,B_g,C_g,D_g) =$$

$$\left[\begin{array}{cc|c} A_g & B_gC_c & 0 \\ 0 & A_c & B_c \\ \hline C_g & 0 & 0 \end{array}\right] =$$

$$\left[\begin{array}{cc|c} A_g & B_gC_{1c} & 0 \\ 0 & \begin{matrix} A_g+B_gC_{1c} \\ A_w \end{matrix} & \begin{matrix} 0 \\ B_w \end{matrix} \\ \hline C_g & 0 & 0 \end{array}\right] =$$

$$\left[\begin{array}{cc|c} A_g & 0 & 0 \\ 0 & \begin{matrix} A_g+B_gC_{1c} \\ A_w \end{matrix} & \begin{matrix} 0 \\ B_w \end{matrix} \\ \hline C_g & C_g & 0 \end{array}\right]$$

where clearly the modes of the plant (eigenvalues of A_g) are uncontrollable, and thus are canceled by the controller zeros. ■

Remark 1
The assumption that the plant to be controlled is stable can be removed and then lemma 4.2 changes to: all the stable plant poles are canceled by controller zeros. ■

Remark 2
Note that for $\gamma \to \infty$ the H_∞ controller becomes the H_2 optimal controller for the WMSP and that the lemmas 3.1, 4.1 and 4.2 also hold for a H_2 solution to the WMSP. ■

CONCLUSIONS

By deriving the Central H_∞ controller following Glover and Doyle (1988) for the Weighted Mixed Sensitivity Problem, explicit relations between controller poles and zeros and the poles and zeros of plant and weightings have been stated. The most important result is that a H_∞ controller for the WMSP cancels all stable plant poles, regardless of the weightings which are introduced to specify performance and robustness. Therefore it can not be expected that controller designs which result from the H_∞ Mixed Sensitivity Problem have good robustness *and* performance properties in the face of varying system poles.

APPENDIX
ZERO SOLUTIONS TO H_∞ ARE's

Proof of lemma 3.1

Suppose D_{21} is of full rank then by A3 D_{21} can be assumed to be the identity, so:

$$D_{21} = I \tag{A.1}$$

The (2,1) block in the Hamiltonian for Y_∞ is:

$$
\begin{aligned}
&Hy_\infty(2,1)\\
&= -B_1B_1' + B_1[D_{.1}'\tilde{R}^{-1}D_{.1}]B_1'\\
&= -B_1[I - D_{.1}'\tilde{R}^{-1}D_{.1}]B_1'
\end{aligned} \tag{A.2}
$$

This obviously yields a zero block if

$$D_{.1}'\tilde{R}^{-1}D_{.1} = I \tag{A.3}$$

Now since $D_{21} = I$ the left hand side of (A.3) can be written as:

$$[D_{11}' \;\; I]\begin{bmatrix} D_{11}D_{11}'-\gamma^2 I & D_{11} \\ D_{11}' & I \end{bmatrix}^{-1}\begin{bmatrix} D_{11} \\ I \end{bmatrix} \tag{A.4}$$

Using the formulae for inversion of block matrices in Patel and Munro (1982) we obtain:

$$[D_{11}' \;\; I]\begin{bmatrix} \alpha^{-1}+\alpha^{-1}D_{11}XD_{11}'\alpha^{-1} & -\alpha^{-1}D_{11}X \\ -XD_{11}'\alpha^{-1} & X \end{bmatrix}\begin{bmatrix} D_{11} \\ I \end{bmatrix} \tag{A.5}$$

where $\alpha = D_{11}D_{11}'-\gamma^2 I$ and $X = (I - D_{11}'\alpha^{-1}D_{11})^{-1}$ equation (A.5) is equivalent to:

$$D_{11}'\alpha^{-1}D_{11} + (I - D_{11}'\alpha^{-1}D_{11})X(I - D_{11}'\alpha^{-1}D_{11}) \tag{A.6}$$

Now substituting X in (A.6) shows that the block (A.3) holds and thus $Hy_\infty(2,1)$ equals zero.
The ARE associated with the Hamiltonian Hy_∞ then takes the following form:

$$Y_\infty Ay_\infty + Ay_\infty' Y_\infty - XPX = 0$$

where Ay_∞ and P follow from (8). Since it is assumed that Ay_∞ is stable, Y_∞ obviously equals zero, which completes the proof.
The proof that X_∞ is zero if D_{12} is of full rank follows by duality ∎

REFERENCES

Chen, C. (1984), *Linear System Theory and Design*, Holt–Saunders International Editions.

Doyle, J.C. Glover, K. (1988), State space formulae for all stabilizing controllers that satisfy an H_∞–norm bound and relations to risk sensitivity, *Systems& Control letters* vol. 11 pp. 167–172.

Doyle, J.C. Stein G. (1981), Multivariable Feedback Design: Concepts for a classical modern synthesis, *IEEE Trans.Auto.Cont.*,vol. AC–26, pp.4–16

Francis, B.A. (1988), *A course in H_∞ Control Theory*, Lecture notes in Control and information sciences, Springer Verlag Berlin

Kwakernaak, H. (1983), Robustness optimization of linear feedback systems, *Proceedings of the IEEE Conference on Decision and Control*, pp.

Patel R.V. Munro. N. (1982), *Multivariable System Theory and Design*, Pergamon Press

Maciejowski, J.M. (1989), *Multivariable Feedback Design*, Addison Wesley

Safonov, M.G. Laub, A.J. Hartmann, G.L. (1981), Feedback properties of multivariable systems: the role and use of the return difference matrix, *IEEE Trans.Auto.Cont.*,vol. AC–26, pp.47–65

Verma, M. Jonkheere E. (1984) L_∞–compensation with mixed sensitivity as a broadband matching problem, *Systems& Control Letters*, vol. 4, pp. 125–130

Selected Topics in Identification,
Modelling and Control; Vol. 1, 1990

Modal reduction guided by Hankel singular value intervals

Pepijn Wortelboer and Henk van Oostveen

Lab. Measurement and Control, Dept. Mechanical Engineering and Marine Techn., Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. For extremely high–dimensional lightly damped systems, that are often represented in modal form, modal reduction is an efficient model reduction method. In these situations other methods such as balanced reduction, requiring additional computations, become very complicated. Modal realizations of lightly damped systems enable straightforward estimation of the system–invariant Hankel singular values (HSV's), that indicate the input–output importance of states in a balanced realization. In addition HSV *intervals* including the exact HSV's are determined based on such modal realizations; eigenvalue perturbation theory (Geršgorin, Weyl) is applied to the (non–diagonal) product of reachability and observability Gramian. An HSV–ordering of *sets* of modes is established and modal reduction is performed by truncating mode sets in the lower HSV intervals. In case sets of modes are too large, scaling and partially balancing transformations are used to split the associated HSV intervals.

Keywords. large–scale systems, model reduction, modal reduction, balanced reduction, Geršgorin eigenvalue regions, Hankel singular value intervals.

0 NOTATION

$\lvert z\rvert, \bar{z}$	*modulus, complex conjugate of* $z\in\mathbb{C}$
$\lVert z\rVert$	*Euclidean norm of* $z\in\mathbb{C}^n$
$Z\in\mathbb{C}^{n\times m}$	$n\times m$ *complex matrix* Z
Z^T	*transpose of* Z
Z^H	*Hermitian adjoint of* Z $(=[\bar{Z}]^T)$
P, Q	*reachability, observability Gramian*
σ_i,	*i'th Hankel singular value* (HSV)
Σ	$\Sigma=\mathrm{diag}(\sigma_i)$, *with* $\sigma_i \geq \sigma_{i+1}$
ϑ_k	*dominance measure of vibration mode* k
A^{-T}	*inverse transpose of* $A\in\mathbb{C}^{n\times n}$
$\lambda(A)$	*set of* n *eigenvalues of* $A\in\mathbb{C}^{n\times n}$
$A=A^H>0$	*Hermitian, and positive definite* $A\in\mathbb{C}^{n\times n}$
$\grave{A}, \mathring{A}$	*diagonal, off–diagonal part of matrix* $A\in\mathbb{C}^{n\times n}$; $A=\grave{A}+\mathring{A}$
$E_i(A)$	*i'th absolute row sum* (DEFINITION 1)
$Ш_i(A)$	*i'th absolute column sum* (DEFINITION 2)
$\mathcal{G}(A)$, $\mathcal{F}(D,F)$	*Geršgorin's eigenvalue inclusion regions,* (THEOREM 1, CORROLARY 1)
$\mathrm{diag}(A_i)$	*block diagonal matrix with* $A_i\in\mathbb{C}^{n_i\times n_i}$
$(\breve{A},\breve{B},\breve{C},D)$, $(A,B,C,D)\sim(\tilde{A},\tilde{B},\tilde{C},D)$	*balanced, similar state–space realizations*

1 INTRODUCTION

Controllers for extremely high–dimensional systems as encountered in large space structure applications, can only be designed after some model simplification. In most high–dimensional lightly damped systems vibration modes play a crucial role as they enable a physical interpretation and are all dynamically decoupled. This has made mode selection *(modal reduction)* one of the most important model–order reduction methods for extremely large systems. In the analysis of flexible mechanical structures for instance, one usually represents the infinite–dimensional system by a modal subsystem and all modes outside a certain frequency range are simply neglected. In this way responses to forces with known frequency contents can be computed efficiently. If we are mainly interested in the motion of specific points in the structure, selection of modes based on their input–output contribution seems more appropriate than mere truncation of modes outside a certain frequency range.

The residual system (the difference between original and reduced system) is completely defined by the truncated modes and is always of lower order than the original system; this may facilitate stability robustness analysis in robust control applications.

Model reduction methods that try to recover the input–output behaviour are numerous, but mostly involve additional computations (assumed we start out from a modal realization). Reduction methods that hinge on small norms of the residual system (being particularly attractive in robustness analysis of controlled systems) are optimal Hankel norm reduction (Glover, 1984) and balanced reduction (Enns, 1984).

However, the residual systems are of higher order and exact norm calculations become laborious. Besides the poles of the reduced-order model do not correspond to poles of the original model and available reduction procedures for balanced reduction and particularly for optimal Hankel norm reduction are computationally demanding compared to modal reduction procedures.

In this contribution modal reduction is discussed within a 'balancing' setting in order to obtain input–output dominance measures of modes or sets of modes. It is well-known that from lightly damped vibration modes accurate estimates of the HSV's can be obtained (Gregory, 1984). In addition we present methods to bound the exact HSV's (leading to *HSV intervals*) and to associate *sets* of modes with these HSV intervals. These sets have a definite *HSV–ordering* and modal reduction is achieved by truncating mode sets associated with the lower HSV intervals.

In section 2 balancing theory is reviewed, and similarity between truncated state-space realizations is discussed.

In section 3 a modal realization is analysed for its correspondence with a balanced realization by means of closed-form solutions of the reachability and observability Gramians. The diagonal elements of these Gramians provide HSV estimates that are used as a measure for the input–output importance of vibration modes. It is shown that if damping goes to zero these estimates converge to the exact HSV's. Besides, truncation of a modal and balanced realization based on these HSV's becomes identical, provided poles do not occur repeatedly.

For systems with non-vanishing damping a new procedure is introduced.

In section 4 eigenvalue perturbation theory is used to establish bounds on the HSV's based on the HSV estimates derived for each mode. Theorems of Geršgorin and Weyl are discussed in detail. HSV intervals can be found that cluster subsets of modes. Sets of modes with a definite HSV–ordering are treated as entities in the proposed truncation procedure. Although this method is not restricted to arbitrarily lightly damped systems, increased damping may lead to impractically large mode sets and eventually all ordering of mode sets will be lost.

In section 5 scaling and partially balancing transformations are explored that give better bounds on the HSV's (i.e. smaller and possibly more HSV intervals), thus providing additional ordering of modes.

In section 6 a general procedure is presented to find a sufficient number of HSV intervals, on which the reduction can be based. A characteristic example is given to illustrate the efficiency of the procedure.

2 BALANCED REDUCTION

Truncation of a balanced realization is now one of the most popular methods for model-order reduction. In a balanced realization states are equally reachable (from input) and observable (at output) and their input–output importance is measured by associated HSV's. Lines of thought that led to this concept can be found in Moore (1981).

For a minimal state-space realization of a time-invariant and stable system,

$$\dot{x} = A\,x + B\,u \qquad y = C\,x + D\,u \tag{1}$$

with state vector $x(t)\in\mathbb{R}^n$, input vector $u(t)\in\mathbb{R}^m$, output vector $y(t)\in\mathbb{R}^p$, and A,B,C, and D real constant matrices, the associated reachability and observability Gramians P and Q are defined as,

$$P = \int_0^\infty \exp(At)BB^T\exp(tA^T)\,dt \tag{2a}$$

$$Q = \int_0^\infty \exp(A^Tt)C^TC\exp(tA)\,dt \tag{2b}$$

The HSV's are fully stated by P and Q, and are system invariants:

$$\sigma_i \;= [\lambda_i(P\,Q)]^{\frac{1}{2}} \tag{3}$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$.

State-space realization (1) can be transformed into a balanced (sometimes called 'internally balanced') realization,

$$\dot{\breve{x}} = \breve{A}\,\breve{x} + \breve{B}\,u \qquad y = \breve{C}\,\breve{x} + D\,u \tag{4}$$

satisfying:

$$\begin{aligned} &\breve{P} = \breve{T}^{-1}P\breve{T}^{-T} = \breve{Q} = \breve{T}^TQ\breve{T} = \Sigma = \mathrm{diag}(\sigma_i) \\ &\breve{A} = \breve{T}^{-1}A\breve{T}, \quad \breve{B} = \breve{T}^{-1}B, \quad \breve{C} = C\breve{T} \end{aligned} \tag{5}$$

with $\breve{T}$ the balancing transformation matrix.

Partitioning $\Sigma=\mathrm{diag}(\Sigma_1,\Sigma_2)$ and $(\breve{A},\breve{B},\breve{C},D)$ conformably, the reduced-order model,

$$\dot{x} = \breve{A}_{11}x + \breve{B}_1u \qquad y = \breve{C}_1x + D\,u \tag{6}$$

is again stable and balanced with both reduced-order Gramians equal to Σ_1 (the partitioning should be chosen such that Σ_1 contains HSV's significantly larger those in Σ_2). Thus balanced reduction retains the input–output important part of the dynamics, but as opposed to modal reduction, the reduced system does not recover poles of the original system in general;

$\lambda(\breve{A}) \neq \{\lambda(\breve{A}_{11}), \lambda(\breve{A}_{22})\}$ for $\breve{A}_{12}$ and $\breve{A}_{21}$ non-zero.

The Gramians are usually solved from the (continuous time) reachability and observability Lyapunov equations,

$$A\,P + PA^T + BB^T = 0 \tag{7a}$$

$$A^TQ + Q\,A + C^TC = 0 \tag{7b}$$

For complex realizations similar to (A,B,C,D) in (1) $[\cdot]^T$ should be replaced by $[\cdot]^H$ in the formulas above. Calculation of the balancing transformation (Laub e.a., 1987) requires relatively large computation power which may cause problems for extremely high-dimensional systems.

In the sequel it is shown how full balancing transformation can be avoided if the Gramians are block–diagonal or nearly block–diagonal.

PROPOSITION 1

Let state–space system (A,B,C,D) *of order* n *be truncated to* (A_{11},B_1,C_1,D) *of order* n_1. *Similarity transformations* $T = \mathrm{diag}(T_1,T_2)$ *with* $T_1 \in \mathbb{C}^{n_1 \times n_1}$ *do not affect the truncation result.*

Proof:

Truncation after transformation yields $(T_1^{-1}A_{11}T_1,T_1^{-1}B_1,C_1T_1,D) \sim (A_{11},B_1,C_1,D)$ ■

PROPOSITION 2

Let $P=\mathrm{diag}(P_1,P_2)$ *and* $Q=\mathrm{diag}(Q_1,Q_2)$ *be associated with state–space system* (A,B,C,D) *and* $P_1,Q_1 \in \mathbb{C}^{n_1 \times n_1}$, $P_2,Q_2 \in \mathbb{C}^{n_2 \times n_2}$, *then a block–diagonal transformation matrix* $T=\mathrm{diag}(T_1,T_2)$ *with* $T_1 \in \mathbb{C}^{n_1 \times n_1}$, $T_2 \in \mathbb{C}^{n_2 \times n_2}$ *exists that balances* (A,B,C,D).
Moreover, if the HSV's related to P_1Q_1 *are all larger than those related to* P_2Q_2, *then direct and balanced truncation yield identical systems:* $(A_{11},B_1,C_1,D) \sim (\breve{A}_{11},\breve{B}_1,\breve{C}_1,D)$.
Realization (A,B,C,D) *will be called 'block–balanced'.*

Proof:

Since $\tilde{P}=T^{-1}PT^{-T}=\mathrm{diag}(T_1^{-1}P_1T_1^{-T},T_2^{-1}P_2T_2^{-T})$ and $\tilde{Q}=T^TQT=\mathrm{diag}(T_1^TQ_1T_1,T_2^TQ_2T_2)$, T_1 and T_2 can be found independently to make $\tilde{P}$ and $\tilde{Q}$ diagonal and equal. Proposition 1 says that this does not alter the truncation result. ■

As a direct consequence, each truncation of a realization with diagonal Gramians that satisfy $p_{ii} \cdot q_{ii} \geq p_{i+1,i+1} \cdot q_{i+1,i+1}$, yields identical reduced–order systems.

For given realizations with almost (block–) diagonal Gramians, direct truncation may be such close to balanced truncation that additional balancing transformations would complicate the reduction unnecessarily. For lightly damped systems in modal form, smallness of the off–diagonal elements of the Gramians is explained in the next section and quantified in section 4.

3 HSV ESTIMATES FROM MODAL REALIZATIONS

In literature it has been shown (Gregory, 1984; Jonckheere, 1984; Blelloch e.a., 1987) that differences between a particular modal and balanced realization vanish if damping approaches zero and poles do not occur repeatedly. Besides from each mode with non–zero damping a HSV can be estimated. In the sequel closed–form solutions of the Gramians P and Q are presented for realizations with diagonal state–space matrices. For damping going to zero the diagonal elements of P and Q tend to infinity whereas most off–diagonal elements remain finite (only repeated poles cause infinite off–diagonal elements). Systems that have non–diagonalizable state–space matrices are treated in the appendix.

Gramians of a modal realization

Given a modal realization of a strictly proper and oscillatory system,

$$\dot{\eta}_i = \lambda_i \eta_i + \beta_i u \quad y = \sum_{i=1}^{n} \gamma_i \eta_i \ \ (\text{n even}) \tag{8}$$

$$\lambda_{2k-1} = \rho_k + j\omega_k, \qquad \lambda_{2k} = \rho_k - j\omega_k \quad ,k \leq \tfrac{1}{2}n$$
$$\beta_{2k-1} = a_k + jb_k \in \mathbb{C}^{1\times m}, \quad \beta_{2k} = a_k - jb_k \in \mathbb{C}^{1\times m}$$
$$\gamma_{2k-1} = c_k + jd_k \in \mathbb{C}^{p\times 1}, \quad \gamma_{2k} = c_k - jd_k \in \mathbb{C}^{p\times 1}$$

with $\|\beta_i\| = \|\gamma_i\|$, then closed–form solutions of the Lyapunov equations (7) are:

$$p_{ij} = -\frac{\beta_i \ \beta_j^H}{\lambda_i + \bar{\lambda}_j}, \quad q_{ij} = -\frac{\gamma_i^H \ \gamma_j}{\bar{\lambda}_i + \lambda_j} \tag{9}$$

(indices 'i' and 'j' denote first–order modes, 'k' and 'l' denote vibration modes). Four types of denominators can be discerned,

$$\begin{aligned} \lambda_{2k-1} + \bar{\lambda}_{2l-1} &= \rho_k + \rho_l + j(\omega_k - \omega_l) \\ \lambda_{2k} + \bar{\lambda}_{2l} &= \rho_k + \rho_l - j(\omega_k - \omega_l) \\ \lambda_{2k-1} + \bar{\lambda}_{2l} &= \rho_k + \rho_l + j(\omega_k + \omega_l) \\ \lambda_{2k} + \bar{\lambda}_{2l-1} &= \rho_k + \rho_l - j(\omega_k + \omega_l). \end{aligned} \tag{10}$$

It can be shown that $\mathrm{diag}(p_{ii}) = \grave{P} = \grave{Q} = \mathrm{diag}(q_{ii})$ and we define

$$\begin{aligned} \vartheta_k &= p_{2k\,2k} = p_{2k-1\,2k-1} = \frac{\|\beta_{2k}\|^2}{2|\rho_k|} \\ &= q_{2k\,2k} = q_{2k-1\,2k-1} = \frac{\|\gamma_{2k}\|^2}{2|\rho_k|} \end{aligned} \tag{11}$$

as a measure of the input–output contribution of vibration mode k (HSV estimates). The off–diagonal elements of P and Q are generally complex.

The vanishing damping case

From (11) we conclude that if mode k becomes undamped ($\rho_k \to 0$) and remains reachable and observable, a pair of diagonal elements in P and in Q tends to infinity ($\vartheta_k \to \infty$). If all other elements remain finite, ϑ_k converges to HSV's σ_1 and σ_2, and mode k is clearly dominant. If the system has no repeated poles, arbitrarily small damping in any mode does not cause infinite off–diagonal elements (on the contrary if $\omega_k=\omega_l$ for $k\neq l$ and if $\rho_k,\rho_l\to 0$ then off–diagonal elements approach infinity too)

If all modes become undamped and $\omega_k\neq\omega_l$ for $k\neq l$, then p_{ij} and q_{ij} $(i\neq j)$ are negligible compared to p_{ii} and q_{ii}, and the balancing transformation relating both realizations tends to a permutation matrix times a diagonal sign matrix (a specific modal realization exists for which the balancing transformation tends to identity).

In the next section approximation errors are assessed for generally damped systems.

4 GENERALLY DAMPED SYSTEMS, HSV INTERVALS AND MODE SETS

For non–zero damping, the exact HSV's are only approximated by ϑ_k (11); for lightly damped systems these estimates will be 'better' than for well damped systems. This is made more precise in this section. Based on the Gramians of a given modal realization, intervals are derived that include the exact HSV's.
Deviations of the Gramians from diagonal structure as given by (11), are accounted for quantitatively. This goes beyond error analysis in literature: Gregory (1984) considers modal reduction of a modally damped system appropriate if for any two vibration modes the following quotient is 'small',

$$\frac{\max(\zeta_i,\zeta_j)\cdot\max(\omega_{0i},\omega_{0j})}{|\,\omega_{0i}-\omega_{0j}\,|}\ll 1 \qquad (12)$$

with $\omega_0=(\rho^2+\omega^2)^{\frac{1}{2}}$ the undamped frequency and $\zeta=|\rho|/\omega_0$ the modal damping ratio. This involves low frequencies and damping ratios, and a large frequency separation; no information concerning the input matrix B or output matrix C is taken into account. Blelloch (1987) found a similar condition for generally damped systems.

Our approach hinges on HSV intervals. We review two eigenvalue perturbation theories to establish these HSV intervals: a well–known theorem of Geršgorin to locate eigenvalues of complex matrices in disc–shaped regions and a theorem of Weyl to bound the real eigenvalues of Hermitian matrices individually. In these eigenvalue perturbation theories the matrix of interest is decomposed into a part with known or 'easy–to–find' eigenvalues and a 'small' residual part that is treated as a perturbation.

In balanced reduction the HSV's can be related to balanced states because the Gramians are both diagonal matrices. To link modal reduction to balanced reduction, both reachability and observability Gramian has to be sufficiently close to a diagonal or block–diagonal matrix. Several methods are introduced to evaluate the deviations from diagonal or block–diagonal form.

Eigenvalue perturbation theory of Geršgorin

In Geršgorin's theory a matrix is decomposed into a diagonal matrix (with known eigenvalues) and a off–diagonal perturbation matrix,

$$Z=\overset{\smallsetminus}{Z}+\mathring{Z},\qquad Z,\overset{\smallsetminus}{Z},\mathring{Z}\in\mathbb{C}^{n\times n} \qquad (13)$$

Definition 1

$E_i(Z)\equiv\sum_{j=1}^{n}|z_{ij}|$ is absolute row sum i of Z.

Definition 2

$\text{Ш}_j(Z)\equiv\sum_{i=1}^{n}|z_{ij}|$ is absolute column sum j of Z.

Theorem 1, Geršgorin.

All eigenvalues of $Z\in\mathbb{C}^{n\times n}$ *are located in the union of* n *discs*

$\bigcup_{i=1}^{n}\{\,x\in\mathbb{C}:\ |x-z_{ii}|\le E_i(\mathring{Z})\,\}\equiv\mathscr{G}_E(Z)$

A region of k *intersecting discs that is disjoint from all other discs contains exactly* k *eigenvalues of* A.

Proof:
This classic result can be found in most textbooks on matrix theory (see Horn and Johnson (1985) for a detailed discussion). ■

Geršgorin disks are centered at the diagonal elements of A and their radii are fully defined by the absolute values of the off–diagonal elements of A. Since the eigenvalues of A and A^H are the same, Geršgorin's theorem can be applied to rows as well as columns and an intersection yields better estimates in general,

$$\lambda(Z)\in\mathscr{G}_E(Z)\cap\mathscr{G}_{\text{Ш}}(Z) \qquad (14)$$

with $\mathscr{G}_{\text{Ш}}(Z)\equiv\bigcup_{i=1}^{n}\{x\in\mathbb{C}:\ |x-z_{ii}|\le\text{Ш}_i(\mathring{Z})\}$,
defining the column–based Geršgorin regions.

Since we know that the eigenvalues of PQ are real nonnegative, only the intersections of the discs with the real axis are of interest. The square roots of the interval bounds determine the HSV intervals. If the off–diagonal absolute row or column sums of PQ are sufficiently small theorem 1 provides accurate bounds on the exact HSV's, from which the feasibility of balanced truncation can be evaluated; we thus circumvent a complete eigenvalue solution (3).

The method discussed above to determine HSV bounds is based upon a decomposition (13) of the product of the Gramians that does not reflect our HSV estimates (11) obtained from separate modes. To ensure that the squared HSV estimates (ϑ_k^2) are included in the eigenvalue intervals of PQ, we decompose PQ as follows:

$$PQ = \grave{P}\grave{Q} + [\grave{P}\mathring{Q} + \mathring{P}\grave{Q} + \mathring{P}\mathring{Q}] \tag{15}$$

in which the first term represents the squared HSV estimates and the bracketed expression defines the perturbation matrix. Although the first matrix is diagonal, the second is full in general. The following corollary based on Geršgorin's theorem can be used in eigenvalue estimation problems with full perturbation matrices.

COROLLARY 1

Let $Z = D + F \in \mathbb{C}^{n\times n}$ *with* D *a diagonal matrix. The eigenvalues of* Z *are contained in the union of* n *discs*

$$\bigcup_{i=1}^{n} \{ x \in \mathbb{C}: |x-d_i| \leq \sum_{j=1}^{n} |f_{ij}| \} \equiv \mathcal{G}_E(D,F)$$

Proof:
Use Geršgorin with disc centers d_i+f_{ii}, then shift center to d_i while enlarging the radius by $|f_{ii}|$ to ensure inclusion of the original disc. ■

Again intersection of row-based and column-based discs yields sharper bounds on the eigenvalues:

$$\lambda(D+F) \in \mathcal{G}_E(D,F) \cap \mathcal{G}_{III}(D,F) \tag{16}$$

with $\mathcal{G}_{III}(D,F) \equiv \bigcup_{i=1}^{n} \{z \in \mathbb{C}: |z-d_i| \leq III_i(F)\}$, defining the modified column-based Geršgorin regions.

We mention that other eigenvalue inclusion regions similar to Geršgorin's have been derived in literature (see Horn and Johnson (1985) for Ostrowski's and Brauer's theorems). The elegant simplicity of Geršgorin's approach, however makes it well suited for the analysis of HSV's as will be demonstrated.

Eigenvalue perturbation theory of Weyl for Hermitian matrices

All eigenvalues of a Hermitian matrix are real and can be computed relatively easy. If the perturbation matrix is Hermitian too, eigenvalue intervals can be derived using a theorem of Weyl.

THEOREM 2, Weyl.

For $A = A^H$, $B = B^H \in \mathbb{C}^{n\times n}$, $C = A + B$, *and all eigenvalues arranged in increasing order, the eigenvalues of* C *can be bounded individually:*

$\lambda_k(A)+\lambda_1(B) \leq \lambda_k(C) \leq \lambda_k(A)+\lambda_n(B)$

All intervals have width $\lambda_n(B)-\lambda_1(B)$.

Proof: Horn and Johnson (1985) ■

This theorem can be useful for Hermitian matrices that are close to a Hermitian matrix of which the eigenvalues are known or easy to find. The eigenvalue bounds of P and Q, being Hermitian (and positive definite) can be obtained straightforward based on decomposition (13). However PQ, having real eigenvalues, is not Hermitian and we have to reformulate the eigenvalue problem in (3).

LEMMA 1

Let $P = p^Hp$ *and* $Q = q^Hq$ *be positive definite matrices. Then* $\lambda(PQ) = \lambda(pQp^H) = \lambda(qPq^H)$.

Proof:
This is an immediate result of $\lambda(AB) = \lambda(BA)$ for A and B square. ■

Now both pQp^H and qPq^H are Hermitian and $\grave{P}\grave{Q}$ is a Hermitian diagonal matrix with known eigenvalues (ϑ_k^2). The maximum and minimum eigenvalue of perturbation matrix $pQp^H-\grave{P}\grave{Q}$ or $qPq^H-\grave{P}\grave{Q}$ determine all eigenvalue intervals. The required factorization of P (or Q) however makes the eigenvalue estimation rather complicated.

If all eigenvalues of the perturbation matrix are available even sharper bounds on the individual eigenvalues of PQ can be obtained.

THEOREM 3, Weyl.

For $A = A^H$, $B = B^H \in \mathbb{C}^{n\times n}$, $C = A + B$, *and all eigenvalues arranged in increasing order, the eigenvalues of* C *satisfy the following bounds:*

$\lambda_j(A)+\lambda_{i-j+1}(B) \leq \lambda_i(C) \qquad 1\leq i\leq n,\ j=1,..,i$

$\lambda_i(C) \leq \lambda_j(A)+\lambda_{i-j+n}(B) \qquad 1\leq i\leq n,\ j=i,..,n$

Selection of the sharpest bounds yields:

$$\lambda_i(C) \in [\max_{j=1,i} \{\lambda_j(A)+\lambda_{i-j+1}(B)\},\ \min_{j=i,n} \{\lambda_j(A)+\lambda_{i-j+n}(B)\}]$$

Proof: Horn and Johnson (1985) ■

In general the simplicity in calculating HSV intervals is lost if a full eigensolution (Theorem 3) is required.

Linking HSV intervals to mode sets

Although we now have established HSV intervals containing a number of estimated HSV's that are coupled to separate modes, we cannot conclude that the underlying modal realization is close to a balanced realization. Situations may occur in which the product of two Hermitian matrices has relatively small off-diagonal elements, whereas the matrices itself have off-diagonal elements that are relative large. This can be shown by a simple example:

$$P = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix},\ Q = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} \text{ and } PQ = \begin{bmatrix} 3 & 1 \\ -1 & 8 \end{bmatrix}$$

This means that small HSV intervals associated with a modal realization do not allow conclusions on the approximately balancedness of this realization (i.e. off–diagonal elements in P and Q may still be large).

Eigenvalue perturbation analysis on P and Q separately can be used to check the smallness of the off–diagonal elements of P and Q. This procedure is heuristic since the eigenvalues of P and Q are not system invariants.

An alternate solution is the replacement of the off–diagonal elements of P and Q by their absolute values prior to multiplication. This results in a larger perturbation matrix ensuring inclusion of the original HSV intervals. By this modification each separate HSV interval can only be caused by separate eigenvalue intervals of P and Q.

As stated in section 2 a minimum requirement for a truncation to be equal to a balanced truncation is that P and Q are of the same block–diagonal structure. Although this requirement can never be met exactly for modal realizations of damped systems, Weyl's theorems can be used in evaluating the 'almost block–balancedness' (Proposition 2) of the modal realization. It is assumed that 'P and Q sufficiently close to block–diagonal matrices' leads to almost block–balancedness. Therefore let the off–diagonal blocks define the perturbation matrix and let the block–diagonal matrix be used for estimation of the eigenvalues of P and Q,

$$P=\mathrm{diag}(P_{11},P_{22})+\Delta P,\ Q=\mathrm{diag}(Q_{11},Q_{22})+\Delta Q$$
$$\text{with } \Delta P=\begin{bmatrix} 0 & P_{12} \\ P_{12}^{H} & 0 \end{bmatrix} \text{ and } \Delta Q=\begin{bmatrix} 0 & Q_{12} \\ Q_{12}^{H} & 0 \end{bmatrix}. \tag{17}$$

Then as a result of theorem 2, the following conditions ensure that eigenvalues of P (Q) estimated from P_{11} (Q_{11}) are also the largest eigenvalues of P (Q).

$$\begin{aligned} \lambda_{min}(P_{11})-\lambda_{max}(P_{22}) &> \lambda_{max}(\Delta P)-\lambda_{min}(\Delta P) \\ \lambda_{min}(Q_{11})-\lambda_{max}(Q_{22}) &> \lambda_{max}(\Delta Q)-\lambda_{min}(\Delta Q) \end{aligned} \tag{18}$$

If (18) is satisfied the truncation result (A_{11},B_1,C_1,D) will generally deviate little from $(\breve{A}_{11},\breve{B}_1,\breve{C}_1,D)$, the balanced and truncated system. In principle a similar analysis should be performed on PQ since there is no simple relationship between $\lambda(PQ)$ on the one hand and $\lambda(P)$ and $\lambda(Q)$ on the other. However as we have seen Weyl's theorem can only be applied to PQ after factorization of P or Q, and thus HSV interval analysis based on Weyl's theorems are less efficient than the approach based on Geršgorin's theorem.

5 SCALING AND PARTIALLY BALANCING TRANSFORMATIONS

In the previous section HSV intervals were derived based on a given modal realization. This realization is not unique and in this section it is shown how state transformations not destroying the modal structure can be utilized to contract and possibly split HSV intervals with more than one HSV.

In principle finding two HSV intervals, each associated with a parallel–connected subsystem, is sufficient for the truncation purpose (Proposition 1). Nor the retained part nor the truncated part has to be in modal form; they are only required to be dynamically uncoupled. This provides additional freedom of transformation. By exact balancing of sets of modes off–diagonal elements in the Gramians are zeroed and sharper HSV bounds can be achieved. If sets of modes that are transformed in this way, are retained or truncated as an entity, the reduction is still modal in essence. If, on the other hand, we allow truncation within transformed sets of modes, we have an intermediate form between modal and balanced truncation.

Scaling transformations

Since $T^{-1}ZT$ has the same eigenvalues as Z, simple transformations can be invoked to obtain sharper bounds on the eigenvalues. It is well known that Geršgorin's eigenvalue approximation may benefit from diagonal transformations. The advantage of diagonal transformations is that the approximate eigenvalues (the disk centers) do not alter while the disc radii can be manipulated. Formulas for eigenvalue regions based on Geršgorin's theory are quite simple.

COROLLARY 2, Geršgorin (Horn and Johnson, 1985)

All eigenvalues of $Z \in \mathbb{C}^{n\times n}$ *are located in the union of* n *discs*

$$\bigcup_{i=1}^{n} \{x\in\mathbb{C}: |x-z_{ii}| \le \frac{1}{s_i} \sum_{\substack{j=1\\ j\neq i}}^{n} s_j|z_{ij}|\} = \mathscr{G}_E(S^{-1}ZS)$$

with $s_1,s_2,..,s_n > 0$

Proof:
A scaling transformation does not change the eigenvalues nor the diagonal elements; only the off–diagonal elements vary and thus the absolute row sums. Because only absolute values of the off–diagonal are of interest, a positive scaling is not restrictive. ∎

LEMMA 2

Given Z = D+F, *with* D *a diagonal, and* F *a full complex matrix with column–radii* Ш(F), *then all eigenvalues of* Z *are located in the union of* n *discs*

$$\bigcup_{j=1}^{n} \{x\in\mathbb{C}:|x-d_j| \le s_j \sum_{i=1}^{n} \frac{1}{s_i}|f_{ij}|\} = \mathscr{G}_{Ш}(D,S^{-1}FS)$$

with $s_1,s_2,..,s_n > 0$.

Besides choosing one $s_j>1$ *and all other* $s_i=1$ $(i\neq j)$ *enlarges disc radius* j: $Ш_j(S^{-1}FS) > Ш_j(F)$ *and yields equal or reduced disc radii* i: $Ш_i(S^{-1}FS) \le Ш_i(F)$.

Note that the absolute row sum discs show precisely the opposite behaviour.

Proof:
The first part is just a generalization of the previous result as it does not take the diagonal elements as eigenvalue estimates; only the off–diagonal elements of F may change. The one–element scaling results in row j divided by s_j, and column j multiplied by s_j. Column sum j is thus enlarged, while all others can only diminish. ■

A systematic way to find scaling parameters d_i that yield some sharper eigenvalue bounds is not available to our best knowledge.

For matrices having real eigenvalues, we propose several procedures to find scalings likely to contract eigenvalue intervals of interest.

Proposition 3
Let $Z = D+F$, *with* D *a real diagonal matrix,* F *a full complex matrix, and* $\lambda(Z)$ *all real. Suppose* $\mathcal{G}_{III}(D,F)$ *is a set of eigenvalue intervals and interval* k *is denoted by* $\mathcal{G}_{III}(D,F)^k$, *then each 'disc'* k_j *in* $\mathcal{G}_{III}(D,F)^k$ *can be enlarged individually by an amount*

$$s_{k_j} = \frac{\min\{d_{k_j} - \bar{g}_{k-1}, \underline{g}_{k+1} - d_{k_j}\}}{\text{III}_{k_j}(F)} > 1$$

with $\bar{g}_{k-1}$, $\underline{g}_{k+1}$ *respectively maximum of* $\mathcal{G}_{III}(D,F)^{k-1}$, *and minimum of* $\mathcal{G}_{III}(D,F)^{k+1}$.

Proof:
As a consequence of the previous lemma, enlargement of one disc radius can never result in any other enlarged disc; even the other discs within interval k will shrink. This means all scalings can be computed individually; the minimum ensures interval k does not overlap interval k+1 nor k–1. ■

In most cases we are interested in tearing apart one specific eigenvalue interval, and mutually overlapping of all other intervals is of no concern because their mutual ordering remains valid. The following procedure exploits this additional freedom of scaling.

Proposition 4
Suppose $\mathcal{G}_{III}(D,F)^k$ *is an eigenvalue interval apart from all others and* $\underline{g}_k$, $\bar{g}_k$ *are its minimum and maximum value. Let* $\mathcal{k}$ *number all other intervals and* k_i, $\mathcal{k}_i$ *be the corresponding disc numbers. Then each disc* $\mathcal{k}_j$ *in* $\mathcal{G}_{III}(D,F)^{\mathcal{k}}$ *can be enlarged individually by*

$$s_{\mathcal{k}_j} = \frac{\max\{d_{\mathcal{k}_j} - \bar{g}_k, \underline{g}_k - d_{\mathcal{k}_j}\}}{\text{III}_{\mathcal{k}_j}(F)} > 1$$

Proof:
This again results from the fact that all discs k_i necessarily shrink if $s_{k_i}=1$ and all $s_{\mathcal{k}_i}>1$.

The maximum ensures a scaling greater than one is chosen. ■

In the previous methods no advantage is taken of the fact that enlargement of one disc allows a subsequent disc to be enlarged more. By updating the matrix and starting the procedure again most conservatism can be removed. In the next method a one–element–scaling is performed in each step.

Algorithm (iterative search for scalings)

Suppose $\mathcal{G}_{III}(D,F)^k$ is an eigenvalue interval apart from all others and $\underline{g}_k$, $\bar{g}_k$ are its minimum respectively maximum values. $\mathcal{G}_{III}(D,F)^{\mathcal{k}}$ represents the union of all other eigenvalue intervals.

1: Search disc $\mathcal{k}_j$ in $\mathcal{G}_{III}(D,F)^{\mathcal{k}}$ that can be enlarged most:
$\bar{s}_{\mathcal{k}} = \max_j \left[\text{III}_{\mathcal{k}_j}^{-1} \max\{ d_{\mathcal{k}_j} - \bar{g}_k, \underline{g}_k - d_{\mathcal{k}_j} \} \right]$

2: Scale matrix F:
$F := S^{-1}FS$, $S = \text{diag}(s_i)$, $s_i = 1$ for $i \neq \mathcal{k}_j$
$s_i = \bar{s}_{\mathcal{k}}$ for $i = \mathcal{k}_j$

3: Calculate eigenvalue intervals $\mathcal{G}_{III}(D,F)$.

4: Stop if $\mathcal{G}_{III}(D,F)^k$ has been split or if no significant contraction of intervals has been found.
Otherwise go to 1. ■

Because smaller column–based intervals are accompanied by larger row–based intervals, the row–based Geršgorin regions need not be recalculated if the scalings were based on column analysis. Scalings based completely on rows may improve the eigenvalue bounds by intersecting intervals from both analysis.

<u>Partially balancing transformations</u>

As mentioned earlier, balancing of a subsystem introduces zero off–diagonal elements in P and Q, which will reduce the HSV–interval sizes in most cases. This partially balancing can best be applied to sets of modes that are responsible for the largest off–diagonal elements in the Gramians. Truncation of the transformed realization is only similar to a modal truncation if all modes involved in the partially balancing transformation are retained or truncated (Proposition 1). For moderately damped high-dimensional systems with large sets of modes, interval splitting can only be achieved by partially balancing transformations involving many modes. Taking these modes together in modal set reduction may constrain the choice of the order reduction unacceptably. Dropping the requirement of modal-reduction-similar truncation we may design effective schemes for partly balanced, partly modal reduction.

Note that separate balancing of vibration modes prior to truncation does not affect the modal reduction. For systems with a realization as in (8) the balancing transformation will be (2×2)-block-diagonal and introduces zeros at the entries (2k–1,2k) and (2k,2k–1) in P and Q.

6 MODE SET SELECTION PROCEDURES

In this section it is shown how mode sets can be selected that are input-output most important. HSV intervals, scaling and partially balancing are used in a general procedure for selecting input-output important mode sets or parts of mode sets.

We start out from a modal realization with complex modal states that are scaled with respect to input and output contribution (8). Modes responsible for non-diagonalizable parts of the state-space matrix are treated as sets from the beginning (also see Appendix).

By means of methods presented in section 4, HSV intervals are calculated and scaling transformations (section 5) are applied to give maximum information on the HSV's (HSV intervals from different realizations should be intersected). Well spaced HSV intervals indicate the suitability of (balanced) order-reduction.

To make sure that the realization is close to a balanced realization the eigenvalue intervals of P and Q are evaluated. Based on these intervals together with the HSV intervals, an ordering of sets of modes is determined.

If order-reduction can be achieved by truncation of particular mode sets we can stop here. Otherwise additional ordering can be forced by partially balancing.

This may involve modes that are responsible for large off-diagonal contributions to P and Q, or mode sets that are sure to be retained or truncated. This latter procedure will not introduce 'couplings' with the mode set(s) in the medium HSV range of which an additional ordering is sought. However, selecting mode pairs because of their contribution to the off-diagonal matrices of P and Q generally couples the original mode sets, but is very effective in splitting HSV intervals. This can best be illustrated by means of a characteristic example.

EXAMPLE

A linear time-invariant two-input-two-output system is constructed that has ten complex poles, is non-minimum phase and typically lightly damped. The state-space matrix is diagonalizable and input and output matrices are scaled in order to satisfy (8). Vibration mode numbers are indicated by {k} and mode sets are denoted by capitals.

$$\begin{array}{ll} A = \mathrm{diag}(& -0.0041-0.3823j, -0.0041+0.3823j, \quad \{1\} \\ & -0.0022-0.7580j, -0.0022+0.7580j, \quad \{2\} \\ & -0.0026-1.0197j, -0.0026+1.0197j, \quad \{3\} \\ & -0.0084-1.8087j, -0.0084+1.8087j, \quad \{4\} \\ & -0.0072-1.8474j, -0.0072+1.8474j\ ) \quad \{5\} \end{array}$$

$$B = \begin{bmatrix} 0.00022 + 0.00475j & 0.00052 + 0.23242j \\ 0.00022 - 0.00475j & 0.00052 - 0.23242j \\ 0.00003 - 0.22813j & 0.00028 + 0.42735j \\ 0.00003 + 0.22813j & 0.00028 - 0.42735j \\ -0.04800 - 0.00020j & -0.00982 + 0.00144j \\ -0.04800 + 0.00020j & -0.00982 - 0.00144j \\ 0.16178 + 0.00564j & 0.11206 + 0.00737j \\ 0.16178 - 0.00564j & 0.11206 - 0.00737j \\ 0.35035 - 0.00694j & 0.43074 - 0.00542j \\ 0.35035 + 0.00694j & 0.43074 + 0.00542j \end{bmatrix} \begin{array}{c} \{1\} \\ \\ \{2\} \\ \\ \{3\} \\ \\ \{4\} \\ \\ \{5\} \\ \\ \end{array}$$

$$C^H = \begin{bmatrix} 0.00445 + 0.00021j & 0.00231 - 0.23242j \\ 0.00445 - 0.00021j & 0.00231 + 0.23242j \\ -0.20786 + 0.00011j & -0.00155 + 0.43756j \\ -0.20786 - 0.00011j & -0.00155 - 0.43756j \\ 0.00043 + 0.04852j & 0.00698 + 0.00031j \\ 0.00043 - 0.04852j & 0.00698 - 0.00031j \\ -0.00488 - 0.10270j & -0.16776 - 0.01000j \\ -0.00488 + 0.10270j & -0.16776 + 0.01000j \\ 0.00058 - 0.17478j & 0.52705 - 0.00548j \\ 0.00058 + 0.17478j & 0.52705 + 0.00548j \end{bmatrix} \begin{array}{c} \{1\} \\ \\ \{2\} \\ \\ \{3\} \\ \\ \{4\} \\ \\ \{5\} \\ \\ \end{array}$$

The HSV estimates associated with the vibration modes are (11):

$$\vartheta = \begin{bmatrix} 6.6621 \\ 54.1092 \\ 0.4559 \\ 2.3160 \\ 21.4060 \end{bmatrix} \begin{array}{c} \{1\} \\ \{2\} \\ \{3\} \\ \{4\} \\ \{5\} \end{array}$$

Closed-form solutions (9) are used to calculate both Gramians. Application of (16) with PQ-decomposition (15) gives a first indication of the HSV bounds (Fig. 1a, lower line; Table 1a). Figure 1b (Table 1b) shows that the eigenvalue intervals of P and Q support the division of modes into three sets (A,B,C): {3,4,1}, {5} and {2}. Moreover mode 1 is likely to be more important than modes 3 and 4; indeed the scaling transformation algorithm of section 4 is able to split mode set {3,4,1} within one iteration step into a least important mode set {3,4} and a moderately important mode(set) {1}. In Fig. 1a three steps of iteration are presented, all based on Geršgorin's absolute row sums. Ordering of modes 3 and 4 in set A could not be obtained by continuing the scaling transformation algorithm.

Analysis based on column information gave similar results.

Application of Weyl's Theorem 3 did not improve above results.

(1,2,3,4,5 modes, A,B,C,D mode sets)

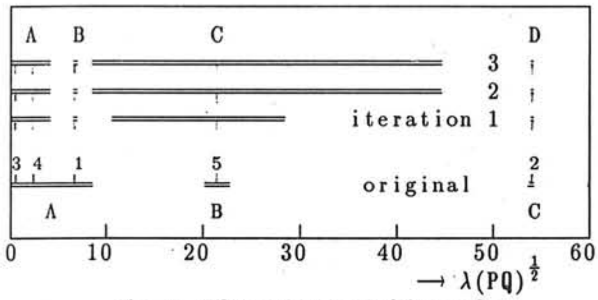


Fig. 1a. HSV estimates and intervals.

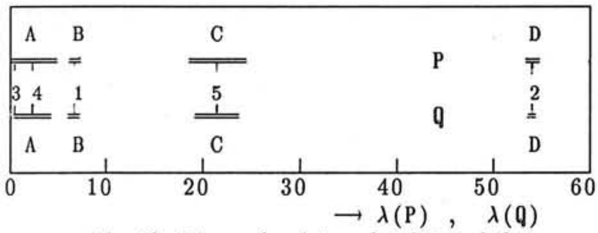


Fig. 1b. Eigenvalue intervals of P and Q.

We conclude that modal truncation of modes {3,4}, {3,4,1} or {3,4,1,5} will be close to an equivalent order-reduction by balanced truncation. For this simple example this is satisfactory, but for higher-dimensional systems HSV intervals and mode sets are generally much larger and then scaling transformations are not sufficient to split HSV intervals. To illustrate the procedure of partially balancing we try to split mode set {3,4}. A balancing transformation on modes {1,5,2} did not result in sufficiently smaller HSV intervals. Evaluation of the off-diagonal elements of P and Q revealed large couplings between modes 4 and 5 (with relatively close poles). Balancing the associated 4×4-block reduced all HSV and eigenvalue intervals dramatically (Fig. 2a/b; Table 2a/b). Note that the original modes 4 and 5 are now coupled and in order to preserve the reduction to be modal, modes {4,5} should be both truncated or both retained. We can now conclude that modal truncation of mode {3} or modes {3,4,1,5} will be close to an equivalent order-reduction by balanced truncation. If we do not strive towards pure modal truncation, truncation of the least important part of subsystem {4,5} can be considered.

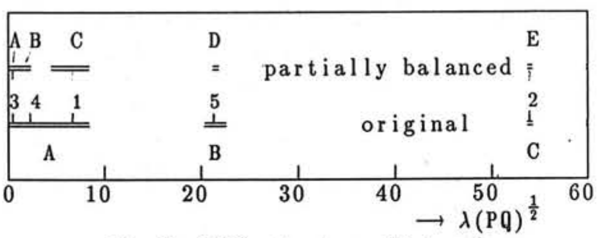


Fig. 2a. HSV estimates and intervals.

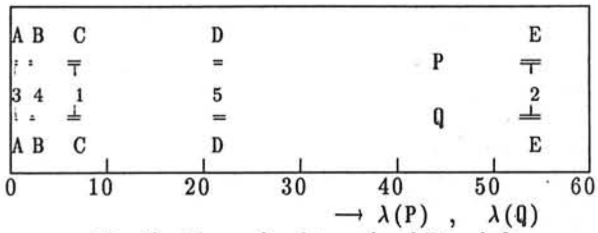


Fig. 2b. Eigenvalue intervals of P and Q, partially balanced.

TABLE 1a HSV intervals

	$\lambda(PQ)^{\frac{1}{2}}$		modes
original:	[0.0000	8.3355]	{1,3,4}
	[20.1727	22.5720]	{5}
	[53.8460	54.3711]	{2}
iteration 1:	[0.00	4.22]	{3,4}
	[6.57	6.76]	{1}
	[10.54	28.38]	{5}
	[54.07	54.14]	{2}
iteration 2:	[0.00	4.12]	{3,4}
	[6.57	6.75]	{1}
	[8.37	44.79]	{5}
	[54.08	54.14]	{2}
iteration 3:	[0.00	4.12]	{3,4}
	[6.57	6.75]	{1}
	[8.34	44.92]	{5}
	[54.08	54.14]	{2}

TABLE 1b Eigenvalue intervals

$\lambda(P)$		modes	$\lambda(Q)$	
[0.0000	4.9359]	{3,4}	[0.3776	4.2124]
[6.0917	7.2324]	{1}	[6.0423	7.2818]
[18.4903	24.3217]	{5}	[19.1836	23.6285]
[53.4265	54.7919]	{2}	[53.2043	55.0141]

TABLE 2a HSV intervals (partially balanced)

$\lambda(\tilde{P}\tilde{Q})^{\frac{1}{2}}$		modes
[0.0000	1.6282]	{3}
[1.6768	2.4116]	{4}/{5}
[4.4102	8.3256]	{1}
[21.4226	21.7329]	{5}/{4}
[53.8404	54.3767]	{2}

TABLE 2b Eigenvalue int. (part. balanced)

$\lambda(\tilde{P})$		modes	$\lambda(\tilde{Q})$	
[0.3670	0.5449]	{3}	[0.3753	0.5366]
[2.0135	2.1538]	{4}/{5}	[1.8735	2.3047]
[6.0947	7.2294]	{1}	[6.0561	7.2680]
[21.2261	21.9306]	{5}/{4}	[21.0485	22.1081]
[53.4162	54.8022]	{2}	[53.2012	55.0172]

For this simple example modal and balanced reduction gave almost indistinguishable results. Note however that our procedure is advised for high-dimensional systems that do not allow a thorough comparison with balancing results. More general model reduction techniques can be applied after a first modal reduction.

7 CONCLUSIONS

For generally lightly damped systems, modal realizations provide a good starting point for estimation of the HSV's. HSV intervals derived by means of Geršgorin's eigenvalue perturbation theory seem very effective in evaluating the reducibility of lightly damped systems, circumventing a (balancing) transformation of systems originally in modal form. Sets of modes naturally appear that have an input-output importance quantified by HSV intervals and they are truncated or retained as a whole. Modal reduction by truncation of mode sets avoids problems with the ordering of modes within these sets. Additional ordering information can be obtained by scaling transformations and by balancing of subsystems. Therefore a specific scaling procedure has been designed. By separately balancing (modal) subsystems, the advantages of modal and balanced truncation can be combined while avoiding full balancing transformations.

APPENDIX

For non-diagonalizable state-space matrices the closed-form solutions to the Lyapunov equations (7) are more complicated and it will be shown that the Gramians contain off-diagonal elements that reach infinity for damping coefficients going to zero.

Since A is now block-diagonal ($A=\mathrm{diag}(A_{ii})$), the Lyapunov equation can be solved per block. We point out the solution for the reachability Gramian only. Partitioning B and P conformably, we may write:

$$A_{ii}P_{ij} + P_{ij}A_{jj}^{H} + B_iB_j^{H} = 0$$

with $A_{ii} \in \mathbb{C}^{n_i \times n_i}$ Jordan blocks: $\begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \cdot & \cdot & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i \end{bmatrix}$.

To simplify the expressions we drop the block-matrix indices of P_{ij} and write BB for $B_iB_j^{H}$. Now P can be built up starting from the lower right element,

$$p_{n_in_j} = -\frac{bb_{n_in_j}}{\lambda_i + \bar{\lambda}_j},$$

following the arrows in $\left[\begin{array}{ccccccc} \cdot & & \cdot & & \cdot & \leftarrow & \cdot \\ & & & & \uparrow & & \uparrow \\ \cdot & & \cdot & \leftarrow & \cdot & \leftarrow & \cdot \\ & & \uparrow & & \uparrow & & \uparrow \\ \cdot & \leftarrow & \cdot & \leftarrow & \cdot & \leftarrow & x \end{array}\right]$.

Row n_i is found from,

$$(\lambda_i+\bar{\lambda}_j)\, p_{n_i,m} + p_{n_i,m+1} + bb_{n_i,m} = 0 \qquad m<n_j$$

and column n_j from,

$$(\lambda_i+\bar{\lambda}_j)\, p_{l,n_j} + p_{l+1,n_j} + bb_{l,n_j} = 0 \qquad l<n_i$$

All other elements are solved from

$$(\lambda_i+\bar{\lambda}_j)\, p_{l,m} + p_{l+1,m} + p_{l,m+1} + bb_{l,m} = 0$$

In all solutions we have a denominator term $\lambda_i+\bar{\lambda}_j$, that can only reach zero for vanishing damping if $\lambda_i=\lambda_j$. Thus blocks in P associated with a Jordan block or Jordan blocks in A with identical eigenvalues, contain elements approaching infinity for damping going to zero.

REFERENCES

Blelloch, P.A., D.L. Mingory and J.D. Wei (1987). Perturbation analysis of internal balancing for lightly damped mechanical systems with gyroscopic and circulatory forces. *AIAA Journal of Guidance, Control and Dynamics, vol.10,* 406–410.

Enns, D.F. (1984). Model reduction with balanced realizations: an error bound and a frequency weighted generalization. *Proc. IEEE Conf. on Decision and Control, (1984).* 127–132.

Glover, K. (1984). All optimal Hankel-norm approximations of linear multivariable systems and their L∞-error bounds. *Int. J. Control, vol.39,* 1115–1193.

Gregory, C.Z.Jr. (1984). Reduction of large flexible spacecraft models using internal balancing theory. *AIAA Journal of Guidance, Control and Dynamics, vol.7,* 725–732.

Horn, R.A. and C.R. Johnson (1985). *Matrix analysis.* Cambridge University Press, Cambridge.

Jonckheere, E.A. (1984). Principal component analysis of flexible systems – open-loop case. *IEEE Trans. Automat. Contr., AC–29,* 1095–1097.

Laub, A.J., M.T. Heath, C.C. Paige, and R.C. Ward (1987). Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms. *IEEE Trans. Automat. Contr., AC–32,* 115–122.

Moore, B.C. (1981). Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Automat. Contr., AC–26,* 17–32.

Selected Topics in Identification,
Modelling and Control; Vol. 1, 1990

Complete orthonormal sets based on linear systems and their application to system identification.

Peter Heuberger and Okko Bosgra

*Lab. Measurement and Control, Dept. Mechanical Engineering and Marine Techn.
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands*

Abstract. Orthogonal functions are of importance in various fields of system and control theory. In this paper it is shown that every finite dimensional time invariant linear discrete time system gives rise to two sets of orthonormal functions, which are complete in ℓ_2 and therefore can be considered as a basis for this space. Specific examples of these functions are the Laguerre polynomials and the discrete pulse functions. The derivation is based on the properties of discrete all–pass transfer functions. Through transformation of input and output signals of a system G in terms of these sets of orthonormal functions, new system descriptions are obtained and new possibilities arise for the construction of approximate identification methods.

Keywords. Discrete time systems; all–pass functions; orthonormal functions; Laguerre polynomials; system identification; system theory.

1 INTRODUCTION

Orthogonal functions and their application in system theory have been subject of research for many years, cf. the early work of Wiener (1949) and Lee (1933). In the past decade their use for problems like system analysis, optimal control and system identification has been investigated by many authors, cf. the work of King and Paraskevopoulos (1979), Paraskevopoulos (1985), Nurges and Yaaksoo (1981), Nurges (1987) and Wahlberg (1989) on Laguerre polynomials, the paper of Unbehauen and Rao (1988) on continuous time identification, and the references therein. There are many different sets of orthonormal functions and the choice of a specific set to attack a certain problem in all these papers is more or less arbitrary, and the choice is often more motivated by the nice properties of a certain set than by the problem at hand. For orthogonal polynomials , like Legendre , Chebychev and Laguerre polynomials, in general the most important property is the so called shift structure (Paraskevopoulos,1985).
It is to be expected that for a specific system and a specific problem there will be a 'best' choice from the whole family of orthogonal sets to solve the problem. We are merely interested in the problem of system identification and the question arises if linear systems give rise to orthogonal functions in a natural way, in order to find an answer to the question if there exists a natural coordinate basis to represent a specific system in terms of a small number of coefficients. The answer to this question is affirmative and in this paper we will give the basis for the theory involved.
We will show that every finite dimensional linear stable discrete time system gives rise to a complete set of orthonormal functions, based on input or output balanced realizations, or equivalently on the singular value decomposition of the Hankel matrix of the system. These functions are generalizations of the Laguerre polynomials. The theory is based on the properties of discrete all–pass functions, analogous to the continuous time results of Glover (1984).
These properties are given in section 2, and in section 3 we show how all–pass functions give rise to sets of orthonormal functions, which is extended to general transfer functions in section 4. In section 4 the completeness of these sets is proven and in section 5 some specific examples of these sets are presented. In analogy with the Laguerre polynomials we can use these functions to transform time_series and arbitrary linear systems to what we will call the orthogonal domain, which is explained in section 6. In section 7 two identification schemes are proposed based on these sets of functions. The application of known identification methods on transformed data changes the properties of the identified models, thus leading to new methods for approximate identification. These schemes can be seen as a search for the 'best' set of orthogonal functions for the identification problem.
In this paper we restrict ourselves to finite dimensional linear time invariant discrete time systems, abbreviated to FDLT systems and FDLTS systems if the system is asymptotically

stable. We will merely be dealing with state space descriptions:

$$x(t+1) = Ax(t)+Bu(t) \qquad (1.1a)$$
$$y(t) = Cx(t)+Du(t) \qquad (1.1b)$$

with $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{C}^{n \times m}$, $C \in \mathbb{C}^{p \times n}$, $D \in \mathbb{C}^{p \times m}$.

The corresponding transfer function is:

$$G(z)=C[zI-A]^{-1}B + D. \qquad (1.1c)$$

and [A,B,C,D] is called a realization of G. For a realization we define the controllability matrix M_c and the observability matrix M_o by:

$$M_c = [B \mid AB \mid A^2B \cdots] \qquad (1.2a)$$
$$M_o = [C^* \mid A^*C^* \mid A^{*2}C^* \cdots]^* \qquad (1.2b)$$

We denote by $\bar{A}$ the complex conjugate of A and by A^* the Hermitian transpose of A, so. $A^*=\bar{A}^T$. It is well known that for minimal realizations M_o and M_c have full rank n.
We assume that the reader is familiar with the notions of Gramians, Hankel singular values and the ω-transformation. A short treatment can be found in this issue (Heuberger, 1990a).

In this paper we use the notation ℓ_2 for square summable time sequences:

$$\ell_2 = \{x:\mathbb{N}^o \to \mathbb{C} \mid \sum_{i=0}^{\infty} x(i)x(i)^* < \infty\} \qquad (1.3)$$

When we deal with Kronecker products we use the operator Vec to transform a matrix into a vector:

If $X=(x_{ij}) \in \mathbb{C}^{n \times m}$, then $Vec(X) \in \mathbb{C}^{nm \times 1}$

$$Vec(X):=(x_{11},x_{12},\cdots,x_{1m},x_{21},\cdots,x_{nm})^T \qquad (1.4)$$

In section 6 we use the concept of the behavior of a system, which we define as follows.

Definition 1.1. Let G(z) be a FDLTS system. We define the behavior $\mathcal{B}(G)$ by

$\mathcal{B}(G)= \{(u(t),y(t)) \mid u(t) \in \ell_2$ and $\{u(t),y(t)\}$ is an input/output pair of G(z)) ◇

Note that in definition 1.1 $t \in \mathbb{N}^O$; we consider $\{u(t),y(t),t \geq 0\}$ to be an input/output pair if there exists a realization of G and an initial condition x(0), such that $\{u(t),y(t),x(0)\}$ obey the equations (1.1). Note that in this definition the stability of G(z) implies that also $y(t) \in \ell_2$.

2. PROPERTIES OF DISCRETE ALL-PASS FUNCTIONS

In this section we give a characterization of realizations of discrete all-pass functions. This is given in theorem 2.2, which is the discrete time version of theorem 5.1 in (Glover, 1984).
First we define all-pass transfer functions, following Glover (1984).

Definition 2.1. A discrete transfer function matrix E(z) of a FDLT system, with dimensions p×m is called an all-pass function if :

$$E(z)\bar{E}^T(\tfrac{1}{z})=I \quad p \leq m \qquad (2.1a)$$
$$\bar{E}^T(\tfrac{1}{z})E(z)=I \quad p \geq m \qquad (2.1b)◇$$

The next theorem shows that all Hankel singular values of a square all-pass function are equal to unity and it gives conditions for the existence of a state space realization.

Theorem 2.2. (Heuberger,1990b). Given a realization [A,B,C,0], (not necessarily stable) with $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{C}^{n \times m}$, $C \in \mathbb{C}^{m \times n}$, then

1. If {A,B,C} is completely controllable and completely observable the following two statements are equivalent:
   (a) $\exists$ $D \in \mathbb{C}^{m \times m}$ such that $G(z)\bar{G}^T(\frac{1}{z})=\sigma^2 I$, where $G(z) := D + C[zI-A]^{-1}B$.
   (b) $\exists$ $P,Q \in \mathbb{C}^{n \times n}$, such that
      (i) $P=P^*$, $Q=Q^*$ (2.2a)
      (ii) $A^*QA+C^*C=Q$ (2.2b)
      (iii) $APA^*+BB^*=P$ (2.2c)
      (iv) $PQ=\sigma^2 I$ (2.2d)
2. Without the condition on controllability or observability: Given that the conditions under (1.b) are satisfied then $\exists$ D satisfying
   (i) $D^*D+B^*QB=\sigma^2 I$ (2.3a)
   (ii) $DD^*+CPC^*=\sigma^2 I$ (2.3b)
   (iii) $C^*D+A^*QB=0$ (2.3c)
   (iv) $BD^*+APC^*=0.$ (2.3d)
   and any such D satisfies (1)(a). ◇

Remark 2.3.
1. Note that if A is not stable then P and Q cannot be seen as Gramians, since these are only defined for stable realizations.. Nevertheless P and Q are unique solutions of (2.2b,c) if A has no eigenvalues on the unit circle. If A does have eigenvalues on the unit circle there may be an infinite number of solutions to (2.2b,c), some of which will satisfy (2.2d) iff (2.2a) is satisfied.
2. If A is not stable, the condition $PQ=\sigma^2 I$ does not imply minimality of the realization. Take for example A=I, B=C=0, then P=Q=I but {A,B,C} is neither observable nor controllable. ◇

3 ORTHONORMAL FUNCTIONS GENERATED BY ALL-PASS FUNCTIONS

In this section we use theorem 2.2 to show that a square stable all-pass function gives rise to an infinite set of orthonormal functions. This derivation is based on the fact that the controllability Gramian P of a realization of a FDLTS system is equal to $P=M_cM_c^*$, where M_c is defined in (1.2a). Consider the rows of M_c as discrete time functions, then the entries of P are the inner products of these functions. So if P=I then these rows are mutually orthonormal in ℓ_2-sense. The next step is an embedding of an all-pass function with McMillan degree n in one with degree k×n, which has a controllability matrix with k×n rows. If we let $k\to\infty$ this leads to an infinite number of rows or orthonormal functions.

If $G(z)=C[zI-A]^{-1}B+D$ is a square stable all-pass function with McMillan degree n, then theorem 2.2 shows that P and Q, defined by (2.2) satisfy PQ=I. We can always find a minimal realization with P=Q=I, using well known balancing techniques (Laub, 1980; Moore, 1981; Enns, 1984).

So $$AA^*+BB^*=I \tag{3.1a}$$
and $$A^*A+C^*C=I \tag{3.1b}$$

Stability and minimality imply that the controllability and observability matrix (1.2a,b) of the realization have an orthonormality property:

$$M_cM_c^* = P = I \tag{3.2a}$$
$$M_o^*M_o = Q = I \tag{3.2b}$$

Hence we can consider the rows of M_c (and M_o^*) as n mutually orthonormal discrete time functions.

Remark 3.1. For such a realization we can show that (3.2a,b) gives us also the singular value decomposition of the Hankel matrix H corresponding to G. It is well know that $H=M_oM_c$ and because the Hankel singular values of G are all equal to unity this gives the singular value decomposition of H:

$$H=U\Sigma V^*,\quad U=M_o,\quad \Sigma=I,\quad V=M_c^*. \tag{3.3}$$ ◇

The next step is to embed G in an all-pass function with larger McMillan degree. If G(z) is a square all-pass function then it is clear that $G^k(z)$ is all-pass for $k\in\mathbb{N}$. The following lemma shows that we can easily find a realization of $G^2(z)$ with the property (3.1a,b).

Lemma 3.2. (Heuberger, 1990b).

Let $G(z)=C[zI-A]^{-1}B+D$ be a square stable all-pass function with $AA^*+BB^*=I_n$ and $A^*A+C^*C=I_n$. Let $G_2(z)=G^2(z)$ then G_2 has a stable, minimal realization $[A_2,B_2,C_2,D_2]$ with:

$$A_2=\begin{bmatrix}A & 0\\ BC & A\end{bmatrix}\ B_2=\begin{bmatrix}B\\ BD\end{bmatrix}\ C_2=[DC\ C]\ D_2=D^2 \tag{3.4a}$$
and
$$A_2A_2^* + B_2B_2^* = A_2^*A_2 + C_2^*C_2 = I_{2n} \tag{3.4b}$$ ◇

As a result of lemma 3.2 we can again consider the rows of the controllability and observability matrix of $[A_2,B_2,C_2,D_2]$ as 2n orthonormal functions. Note that the first n 'controllability' functions are the orthonormal 'controllability' functions of [A,B,C,D] and that the last n 'observability' functions are the orthonormal 'observability' functions of [A,B,C,D] . The next theorem extends this property to arbitrary powers of G(z).

Theorem 3.3. (Heuberger, 1990b).

Let $G(z)=C[zI-A]^{-1}B+D$ be a square stable all-pass function with $AA^*+BB^*=A^*A+C^*C=I_n$. Let $G_k(z)=G^k(z)$ with $k\in\mathbb{N}$, k>1. Then G_k has a stable, minimal realization $[A_k,B_k,C_k,D_k]$ with:

$$A_k = \begin{bmatrix}A & 0 & & & \cdot\\ BC & A & 0 & & \cdot\\ BDC & BC & A & 0 & \cdot\\ \cdot & \cdot & \cdot & \cdot & \\ BD^{k-2}C & \cdots & BDC & BC & A\end{bmatrix} \tag{3.5a}$$

$$B_k = \begin{bmatrix}B\\ BD\\ BD^2\\ \cdot\\ BD^{k-1}\end{bmatrix} \tag{3.5b}$$

$$C_k = [D^{k-1}C \cdots D^2C\ DC\ C] \tag{3.5c}$$

$$D_k = D^k \tag{3.5d}$$

and

$$A_kA_k^* + B_kB_k^* = I_{(k+1)n} \tag{3.6a}$$
$$A_k^*A_k + C_k^*C_k = I_{(k+1)n} \tag{3.6b}$$ ◇

By letting $k\to\infty$ theorem 3.3 actually shows the construction of two infinite sequences of orthonormal functions, represented by the controllability and observability matrices of $\{A_k,B_k,C_k\}$. Note that the 'controllability' functions induced by G_{k-1} are the first k×n functions induced by G_k and the 'observability' function of G_{k-1} are the last k×n 'observability' functions of G_k..

Remark 3.4. As mentioned in remark 3.1. the controllability and observability matrices of $[A_k,B_k,C_k,D_k]$ define the singular value decomposition of the Hankel matrix of G_k. The structure of the realization (3.5) with the decomposition (3.3) shows that we have actually extended the matrices U,V by adding extra columns, such that these extended matrices are still unitary. ◇

4 ORTHONORMAL FUNCTIONS FROM GENERAL TRANSFER FUNCTIONS

In this section we use the results of the previous section in order to define sets of orthonormal functions based on an arbitrary FDLTS system G with McMillan degree n. This will be accomplished by splitting of an all–pass function and to use the method described in section 2. The line of thought is best understood by considering the Hankel matrix H of G. The singular value decomposition of H is

$$H=U\Sigma V^* \tag{4.1a}$$

$$U^*U = VV^* = I \tag{4.1b}$$

and Σ is the diagonal matrix with singular values.

The unitarity of U and V implies that the columns of U and V can be seen as n orthonormal discrete time functions. We will extend one of these to an infinite number of orthonormal functions, such that we again have a recursive structure as in section 2. In general it is not possible in general to extend U and V simultaneously, for aming at this recursive structure, because the Hankel singular values are not equal. We will consider the extension of V.

If G(z) is an arbitrary FDLTS system then we can always construct a so called input balanced realization (Enns, 1984). This realization has the property $AA^* + BB^* = I$, $A^*\Sigma^2A + C^*C = \Sigma^2$, where Σ is the diagonal matrix with Hankel singular–values. Let M_c and M_o be as in (1.2) then $M_o^*M_o=\Sigma^2$ and $M_cM_c^*=I$. The Hankel matrix has a singular value decomposition (4.1) with

$U=M_o\Sigma^{-1}$ and $V=M_c$,

since $H=M_oM_c=(M_o\Sigma^{-1})\Sigma M_c$ and $U^*U=VV^*= I$.

We want to extend $V=M_c$ to a larger unitary matrix. This can be done with the theory in the previous section if we can consider it as the controllability matrix of a realization of an all–pass function. Thus we want to expand $\{A,B\}$ with new matrices $\{\tilde{C},\tilde{D}\}$ such that $\tilde{G}(z)=\tilde{C}[zI-A]^{-1}B+\tilde{D}$ is all–pass. Theorem 2.2 shows that it is sufficient to require that

$A^*A+\tilde{C}^*\tilde{C}=I$.

The following lemma shows that this is achieved through the singular value decomposition of A.

Lemma 4.1. (Heuberger,1990b). Let $A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{n\times m}$ with A stable, $\mathrm{rank}(B)=m\leq n$ and $AA^*+BB^*=I$. Let $A=U\Sigma V^*$ be the svd of A and define

$$F=UV^* \tag{4.2a}$$

$$\tilde{C}=B^*F. \tag{4.2b}$$

then 1. $A^*A+\tilde{C}^*\tilde{C}=I$. (4.3a)

2. $\exists\ \tilde{D}=\tilde{D}^*$ such that

$$\tilde{G}(z)=\tilde{C}[zI-A]^{-1}B+\tilde{D} \text{ is all–pass} \tag{4.3b}$$

$$B\tilde{D}=-FA^*B \tag{4.3c}$$

$$\tilde{D}\tilde{C}=-\tilde{C}A^*F \tag{4.3d}$$ ◇

Note that in lemma 4.1. we did not require that $\{A,B\}$ is part of an input balanced realization of a transfer function G, since $AA^*+BB^*=I$ does not imply $A^*A+C^*C=\Sigma^2$. However if we do require this it follows, as stated before, that $[B|AB|A^2B\cdots]$ is exactly the matrix with the right hand side singular vectors of the Hankel matrix of G.

The rank condition on B in lemma 4.1 is necessary to guarantee the existence of a Hermitian $\tilde{D}$ that obeys (4.3), which we will need for the proof of the next theorem.

Lemma 4.1 thus shows how we can 'split off' an all–pass function from a FDLTS system. If we now combine the results of theorem 3.3 and lemma 4.1, we can extend the unitary matrix V (4.1) in a recursive way to an infinitely large unitary matrix. Another way of putting this is that we can create an infinite set of orthonormal functions, based on transfer functions. The exact form of the extension is given in the following theorem.

Theorem 4.2. (Heuberger,1990b). Let $A\in\mathbb{C}^{n\times n}$, stable and $B\in\mathbb{C}^{n\times m}$ with $AA^*+BB^*=I$ and $\mathrm{rank}(B)=m\leq n$. Let $A = U\Sigma V^*$ be a singular value decomposition of A. Define:

$$F =UV^* \tag{4.4a}$$

$$P =-FA^* =-U\Sigma U^* \tag{4.4b}$$

$$X =I-A^*A =I-V\Sigma^2V^* \tag{4.4c}$$

$$A_e = \begin{bmatrix} A & 0 & & & \cdot\cdot \\ FX & A & 0 & & \cdot\cdot \\ PFX & FX & A & 0 & \cdot\cdot \\ P^2FX & PFX & FX & A & 0\cdot\cdot \\ \cdot & & \cdot & \cdot & \cdot \end{bmatrix} \tag{4.5a}$$

$$B_e = \begin{bmatrix} B \\ PB \\ P^2B \\ \vdots \end{bmatrix} \tag{4.5b}$$

Then $A_eA_e^* + B_eB_e^* = I$ ◇

Proof: Lemma 4.1 shows that there exist C and D such that $G(z)=C[zI-A]^{-1}B+D$ is all–pass and $BD=-FA^*B=PB$. Therefore $BD^k=P^kB$. Further $C=B^*F$, so $BC=BB^*F=[I-AA^*]F=F[I-A^*A]=FX$. Substitution of the expressions for BD^k and BC in theorem 3.3 gives A_e and B_e. □

Theorem 4.2 shows how a pair $\{A,B\}$, which obeys the conditions of the theorem, gives rise to an infinite set of orthonormal functions, which are the rows of the matrix $[B_e|A_eB_e|A_e^2B_e\ \cdots]$.

If an arbitrary pair $\{A,B\}$ is stable and reachable there exist a similarity transformation which transforms its Gramian into an identity matrix. The transformed pair then again gives rise to a set of orthonormal functions. Thus for any such pair $\{A,B\}$ we can define the set of orthonormal functions, which in the sequel we will denote by $\Psi_e\{A,B\}$. This is formalized in the following definition.

DEFINITION 4.3. EXTENSION PROCEDURE Let $A\in\mathbb{C}^{n\times n}$, stable, $B\in\mathbb{C}^{n\times m}$, $\operatorname{rank}(B)=m\leq n$, $\{A,B\}$ reachable and $P=P^*>0$ the solution of $APA^*+BB^*=P$. Let $W=\sqrt{P}$, $\tilde{A}=W^{-1}AW$ and $\tilde{B}=W^{-1}B$, leading to $\tilde{A}\tilde{A}^*+\tilde{B}\tilde{B}^*=I$. Create with $\{\tilde{A},\tilde{B}\}$ the matrices A_e and B_e as in theorem 4.2.
We define $\psi_{k-1}\{A,B\}$ as the k^{th} row of

$$[B_e|A_eB_e|A_e^2B_e\ \cdots] \tag{4.6a}$$

and denote the set of these functions by

$$\Psi_e\{A,B\}:=\{\psi_0^*,\psi_1^*,\cdots\}^* \tag{4.6b}$$ ◇

With a small abuse of notation we will also use Ψ_e to denote the matrix (4.6a).

We can interpret $\Psi_e\{A,B\}$ as responses of a system $G_e=[A_e,B_e,A_e,B_e]$ as follows: Let $B\in\mathbb{C}^{n\times m}$ and define the input vectors $u_i(t)=\delta_{it}$, i=1 to m. Apply this input to G_e, then the k^{th} output will be ψ_{k-1}. A more compact way of describing the functions in terms of signals, making full use of the structure, is presented in the following proposition.

PROPOSITION 4.4. (Heuberger,1990[b]). Let $\{A,B\}$ and F be as in theorem 4.2 and define for $k\in\mathbb{N}\cup\{0\}$ the transfer function

$$H_k(z) = \Big[[zI-A]^{-1}F[I-zA^*]\Big]^k z[zI-A]^{-1}B \tag{4.7a}$$

Let M_i for $i\in\mathbb{N}\cup\{0\}$ denote the Markov parameters of H_k and define the matrix

$$\mathcal{M}_k = [M_0|M_1|M_2\cdots] \tag{4.7b}$$

Then the rows of $\mathcal{M}_k$ are the elements of Ψ_e, number $k\times n+1$ to $(k+1)\times n$. ◇

The simplest example of proposition 4.4 is the case k=0, then $H_0(z)=z[zI-A]^{-1}B$, with $\mathcal{M}_0=[B|AB|A^2B\cdot\cdot]$ which are the first n functions. Note that if $B\in\mathbb{C}^n$, so only one input, then $\mathcal{M}_k$ gives the impulse responses of $H_k(z)$.
This property will be of use for transformation of a finite time-series in terms of the elements of Ψ_e, which will be covered in section 6.

REMARK 4.5. In this section we only dealt with the 'input side' of a transfer function. An analogous procedure can be carried out on the output side with output balanced realizations, taking the first n orthogonal functions from the left hand side singular vectors of the Hankel matrix of G. What we established in this section is thus that given a FDLTS G(z) with Hankel matrix $H=U\Sigma V^*$, we defined a method to extend the matrix V to an infinite matrix V_e by adding new columns or equivalently to extend U to U_e. ◇

Completeness

We have now defined a method to create an infinite sequence of orthonormal functions, based on a transfer function. Our goal is to use these functions to describe linear systems and to use them for system identification as is done for instance with Laguerre polynomials in (King and Paraskevopoulos, 1979; Nurges, 1987; Wahlberg, 1989; Heuberger, 1990[b]). A necessary condition will be that these functions form a basis for the function space we wish to consider, which in our case is ℓ_2 (1.3). In other words we have to show that, under appropriate conditions on $\{A,B\}$, $\Psi_e\{A,B\}$ forms a complete orthonormal basis for ℓ_2. This result is presented in the following theorem.

THEOREM 4.6. (Heuberger, 1990[b]).

Let $A\in\mathbb{C}^{n\times n}$, stable, $B\in\mathbb{C}^{n\times m}$, $\operatorname{rank}(B)=m\leq n$ and $\{A,B\}$ a reachable pair. Let $\Psi_e\{A,B\}$ be defined as in definition 4.3. Then this set of functions forms a complete orthonormal basis for ℓ_2, as defined by (1.3) ◇

The proof is based on $\Psi_e\Psi_e^*=\Psi_e^*\Psi_e=I_e$.
A simple example shows why the property that $\Psi_e\Psi_e^*=I_e$ is not sufficient for completeness and why we need $\Psi_e^*\Psi_e=I_e$. Consider the matrix Γ:

$$\Gamma = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \quad \text{then } \Gamma^*\Gamma=\begin{bmatrix} 0 & & \\ & 1 & \\ & & 1 & \\ & & & \ddots \end{bmatrix} \text{ and } \Gamma\Gamma^*=I,$$

and consider the rows of Γ as discrete time functions. It is clear that this constitutes an orthonormal set. All functions are also in ℓ_2, but we do not have a basis for ℓ_2 since the function $(1\ 0\ 0\ 0\cdots)$ cannot be written as a converging sum of the other functions. This can be translated to the fact that $\Gamma^*\Gamma\neq I$.

The rank condition on B in theorem 4.6 is necessary to omit situations like the one above. If B is for instance of the form $B=[B_1\ 0]$ then B_e will also be of this form, causing Ψ_e to have zero columns.

Theorem 4.6 shows that the set of orthonormal functions, based on a transfer function, that we introduced, forms an orthonormal basis for the space ℓ_2. This shows that any ℓ_2–time series can be written as a converging sum of these functions. In section 6 we will apply this to input/output pairs $\{u(t),y(t)\}$ of a linear system, with $u,y \in \ell_2$, and we will show how we can use these results in order to define an alternative description of a linear system. We first give some examples of the extension procedure.

5 EXAMPLES OF ORTHONORMAL SETS

In this section we will give 2 examples of well known orthogonal sets of functions, the Laguerre polynomials and the discrete pulse functions, and we will show that they can be derived using the extension procedure outlined in the previous paragraphs, by choosing a specific system as 'generator'.

1. Laguerre polynomials

Let G(z) be a first order stable SISO–system with an input balanced realization [A,B,C,D]. Let $A=\xi$ $|\xi|<1$ and $B=\sqrt{\eta}$ where $\eta:=1-\xi^2$. Now follow the procedure outlined in theorem 4.2. The singular value decomposition of A is $A=U\Sigma V^*$ with $U=V=1$ and $\Sigma=\xi$. Substitute this in (4.4), then we get:

$$F=1,\ \ P=-\xi,\ \ X=\eta \tag{5.1a}$$

and substitution in (4.5) results in:

$$A_e=\begin{bmatrix} \xi & 0 & & & \\ \eta & \xi & 0 & & \\ -\xi\eta & \eta & \xi & 0 & \\ \xi^2\eta & -\xi\eta & \eta & \xi & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad B_e=\begin{bmatrix} \sqrt{\eta} \\ -\xi\sqrt{\eta} \\ \xi^2\sqrt{\eta} \\ \vdots \end{bmatrix} \tag{5.1b}$$

These are exactly the matrices that constitute the finite difference Laguerre polynomials (Paraskevopoulos, 1985). If we look at the generating transfer functions, defined in proposition 4.4

$$H_k(z) = \left[[zI-A]^{-1}F[I-zA^*]\right]^k z[zI-A]^{-1}B$$

and substitute $A=\xi$, $B=\sqrt{\eta}$ then we get

$$H_k(z) = \sqrt{\eta}\cdot z[1-z\xi]^k[z-\xi]^{-k-2} \tag{5.1c}$$

which are the generating Laguerre transfer functions (Nurges and Yaaksoo, 1981). This shows that with the extension procedure we generalized the construction of the Laguerre polynomials.

2. Pulse functions

Let G(z) be a system with a finite impulse response. We can construct a realization [A,B,C,D] of G with A=0,B=I, which in general will not be minimal but fulfils the conditions of theorem 4.2. A singular value decomposition of A is $A=U\Sigma V^*$ with $U=V=I$ and $\Sigma=0$. Substitution in (4.4) and (4.5) results in:

$$F=X=I,\ P=0 \tag{5.2a}$$

$$A_e=\begin{bmatrix} 0 & & & \\ I & 0 & & \\ 0 & I & 0 & \\ 0 & 0 & I & 0 \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad B_e=\begin{bmatrix} I \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} \tag{5.2b}$$

$$\Psi_e=[B_e\,|\,A_eB_e\,|\,A_e^2B_e\cdots] = \begin{bmatrix} I & 0 & & \\ 0 & I & 0 & \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{5.2c}$$

So the extended set of functions are the pulse functions $\psi_i(t)=\delta_{it}$, which is in fact the usual basis for ℓ_2.

These examples show that the extension procedure 4.3 is quite natural and leads to a generalization of well–known orthonormal bases for ℓ_2.

6 TRANSFORMATIONS

In this section we use the orthonormal functions as a basis for ℓ_2 and expand time series in these functions. We will show that if this is applied to the input/output variables of a linear system, this leads to another system description in terms of the coefficients of the expansion. Let {A,B} be stable and reachable, $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$, rank(B)=m≤n and let $\Psi_e\{A,B\}$ be defined by definition 4.3.

1. Time series

The set of functions Ψ_e is complete in ℓ_2, so we can expand any $\mathcal{L}_2^p$–time series f(t) in these functions:

$$f(t) = \sum_{k=0}^{\infty} F_k\psi_k(t) \tag{6.1a}$$

$$F_k = \sum_{t=0}^{\infty} f(t)\psi_k(t) \tag{6.1b}$$

where $F_k \in \mathbb{C}^p$

In order to make full use of the structure we will group the orthogonal functions in groups of n functions and define :

$$\varphi_k(t) := [\psi^*_{kn+1},\psi^*_{kn+2},\cdots,\psi^*_{kn+n}]^*. \tag{6.2}$$

This leads to:

$$f(t) = \sum_{k=0}^{\infty} L_k \varphi_k(t) \tag{6.3a}$$

$$L_k = \sum_{t=0}^{\infty} f(t)\varphi_k^*(t) \tag{6.3b}$$

where $L_k \in \mathbb{C}^{p \times n}$

It is our goal to use this transformation for identification purposes in which case we will actually have to calculate the orthonormal coefficients $\{L_k\}$. In practical situations, considering f(t) to be a sequence of measured input and output signals, the number of points of f will be finite, $f=[f(0),f(1),\cdots,f(N)]$.
In (Heuberger, 1990b) it is shown that we can calculate the coefficients L_k by leading the inverse sequence $[f(N),f(N-1),\cdots,f(0)]$ through the generating transfer functions $H_k(z)$, defined in (4.7a), and that L_k will be the output of this filter upon the last entry (f(0)). Because of the simple structure of $\{H_k(z)\}$ this calculation of the coefficients can be done using a simple cascade like network (Heuberger, 1990b) as is the case with the Laguerre polynomials (King and Paraskevopoulos, 1979).

2. Systems

Now suppose we have at hand an arbitrary pxm FDLTS system G(z) and let $\{u(t),y(t)\}$ be an input/output pair of G, with $u \in L_2^m$. The stability ensures that $y \in L_2^p$ and thus we can transform these signals with any set $\Psi_e\{A,B\}$. We do not assume any connection between G and $\{A,B\}$, but we will assume that $B \in \mathbb{C}^{n \times 1}$. Let U_k, Y_k denote the orthogonal coefficients (6.3) of u(t) and y(t), $U_k \in \mathbb{C}^{m \times n}$, $Y_k \in \mathbb{C}^{p \times n}$. The next theorem shows that these coefficients are also connected through a linear system. We first define the transformation of a behavior.

Definition 6.1. Let $\{A,B\}$ be a stable, reachable pair, $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{C}^{n \times m}$ with $\mathrm{rank}(B)=m \leq n$ and let G(z) be a FDLTS system. Let $\mathcal{B}(G)$ and $\Psi_e\{A,B\}$ be defined according to definition 1.1 respectively definition 4.3. We define the transform $\Psi(\mathcal{B}(G))$ of the behavior $\mathcal{B}(G)$ by:

$$\Psi(\mathcal{B}(G)) = \{(\mathrm{Vec}(U_k),\mathrm{Vec}(Y_k)) \mid \exists\ (u(t),y(t)) \in \mathcal{B}(G) \text{ with } U_k \text{ and } Y_k \text{ the orthonormal coefficients of u and y as defined by (6.3).}\} \tag{6.4}$$ ◇

Note that the completeness of Ψ_e implies that this is a bijective transformation.

Theorem 6.2. (Heuberger, 1990b). Let $\{A,B\}$ be a stable, reachable pair, $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{C}^{n \times 1}$.
Further, let G(z) be a p×m FDLTS system with McMillan degree n_g and let $\Psi(\mathcal{B}(G))$ be defined according to definition 6.1.
Then there exists a FDLTS system G_o with dimension $(pn_g \times mn_g)$ and McMillan degree n_g, such that

a $\mathcal{B}(G_o)=\Psi(\mathcal{B}(G))$
b For every eigenvalue of A that is a pole of G, the system G_o will have a pole in z=0. ◇

Corollary 6.3. (Heuberger, 1990b). Given the conditions of theorem 6.2, if the eigenvalues of A coincide with the poles of G(z), then the system $G_o(z)$ will only have poles in z=0. These poles are not coupled and $G_o(z)$ will thus have only two non zero Markov parameters. ◇

The conditions of this corollary are for instance fulfilled if A is the system matrix of a realization of G(z). One might say that in this situation all dynamic behavior is covered by the transformation.

In Fig. 1 we visualize the bijective system transformation which is induced by the transformation of the time series.

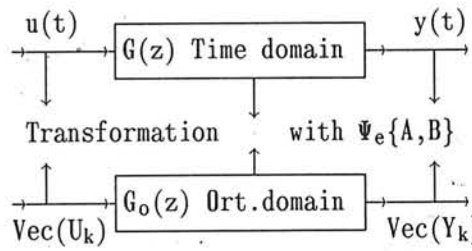


Fig. 1 Transformation of a system, applying the set $\Psi_e\{A,B\}$ of orthonormal functions.

Remark 6.4.
1. For the case that the orthonormal functions are the Laguerre polynomials then theorem 6.1 is given by Nurges and Yaaksoo (1981).
2. It is important to emphasize here that the input/output dimension of the transformed system is larger than the dimension of the original system.
3. In Heuberger (1990b) two conjectures are given which state that G and G_o have the same Hankel norm and the same L_∞ norm.

7. APPLICATION TO IDENTIFICATION

In analogy with the Laguerre polynomials (King and Paraskevopoulos, 1979;Nurges, 1987; Wahlberg, 1989; Heuberger, 1990b) we can use the generalized orthonormal functions in an identification setting. This approach can be of great use if we have some knowledge about the system at hand, for instance when eigenmodes are (partially) known or if we have an initial guess of the system from theoretical or experimental modeling. In that case we can create an $\{A,B\}$

pair which reflects our knowledge and use $\Psi_e\{A,B\}$ in an identification setting. In this section two identification methods are given, that use the generalized orthonormal functions. Both methods combine the use of the orthonormal functions with fairly simple estimation techniques, that lead to an easy to calculate solution. The motivation for this is to derive satisfactory results with simple techniques and to avoid the problems that arise with standard methods that use nonlinear optimization techniques.
If we use $\Psi_e\{A,B\}$ as a basis of the function space of inputs and outputs, then theorem 6.2 and corollary 6.3 show that a 'correct' $\{A,B\}$ pair will lead to a system with only 2 Markov parameters. This could be seen as a search for that set of orthonormal functions that minimizes the dynamic behavior. Method 1 is based on this idea.
Method 2 uses the generating transfer functions (4.7a) as a basis of the frequency domain, in other words it is based on an expansion of the transfer function of a system in the generating transfer functions. In the case that $\{A,B\}$ is 'correct' this would mean that only the first n elements of such an expansion will contain information. This method can also be considered as an approximation of the impulse response of a system in terms of the orthonormal functions Ψ_e.

1. Transformation and ARX

This method is based on the transformation of time series and systems as described in section 6 and is a generalization of identification methods, using Laguerre polynomials, proposed by King and Paraskevopoulos (1979) and Nurges (1987).
The estimation technique involved is referred to as ARX, which is a bit misleading because it is in fact a name for the following model structure:

$$y(t+n) + A_{n-1}y(t+n-1) + ... + A_0y(t) = B_nu(t+n) + ... + B_0u(t) + e(t) \qquad (7.1)$$

where u(t),y(t) and e(t) are respectively the input, output and disturbance of the model and A_i,B_i are constant matrices of appropriate dimensions. The parameters A_i,B_i in (7.1) can be estimated using a least squares algorithm (Ljung, 1987). We use the term ARX for this method.
The method we propose needs an orthonormal set to begin with. This can be the result of a priori knowledge or previous modeling. We often used Laguerre polynomials as a first choice. Now assume that a set Ψ_e is given and that we have recorded input and output sequences of a system G and we wish to find an estimate $\hat{G}$.
The procedure consists of the following steps:

1. transform input u(t) and output y(t) with Ψ_e into orthonormal coefficients (6.3b) U_k and Y_k.
2. Estimate in the 'orthonormal domain' a $\hat{G}_o$ with ARX from $Vec(U_k)$ and $Vec(Y_k)$.
3. Transform $\hat{G}_o$ back to a 'time domain' system $\hat{G}$.

This procedure might be done iteratively, by using the resulting estimate $\hat{G}$ to form a new set of functions $\Psi_e\{\hat{A},\hat{B}\}$ and repeating the procedure. This might be seen as a search for the 'best' basis for the decomposition of the signals.

2. Estimation of impulse response parameters

This method is in fact a generalization of the estimation of a finite number of Markov parameters of a system. As in the previous method we need an orthonormal set as initialization and we use the generating transfer functions $H_k(z)$, defined in (4.7a) and write:

$$G(z) = D + \frac{1}{z}\sum_{k=0}^{\infty} C_kH_k(z) + E(z) \qquad (7.2a)$$

where E(z) denotes the disturbance. The completeness of $\{H_k(z)\}$ for the frequency domain is a direct result of theorem 4.6, but we will not go into this here.
We approximate G(z) with a finite expansion

$$\hat{G}(z) = \hat{D} + \frac{1}{z}\sum_{k=0}^{N} \hat{C}_kH_k(z) + \hat{E}(z) \qquad (7.2b)$$

and estimate the $\hat{C}_k$ parameters, with a least squares algorithm. A well known example of this method is the case where the orthonormal functions are generated by A=0 and B=I. In section 5 it was shown how this leads to the pulse functions, with $H_k(z)=z^{-k+1}$. Hence in this case the C_k parameters are the Markov parameters of G, and the method is known a the estimation of a FIR (finite impulse response) model (Ljung, 1987). Note that if $\{A,B\}$ coincide with G, this leads to $C_k=0$, k>0. This procedure can be seen as a search for the 'best' basis to decompose the impulse response of a system and is a generalization of the algorithm of Zervos c.s (1985), using Laguerre polynomials.

3. Example

As an example of these methods, we have simulated a 4^{th} order SISO system, with a pseudo random binary signal as input and additive noise on the output, such that the signal to noise ratio on the output is 0 dB. The system has important high and low frequent behavior, which can be seen in Fig. 3 and Fig. 4, where the solid line depicts respectively the step response and the Bode amplitude of the system.

Method 1.
We compare the result of ARX in the time domain with the first orthonormal method, described above. First (in the time domain) an 8^{th}

order ARX system was estimated. For the orthogonal method we used a simple first order system (A=0.5) to generate the orthonormal functions. In Fig. 2 we show the deterministic output y(t) of the system, the additive noise and the orthonormal output Y_k, which is the transform of y(t)+noise. We transformed 1100 samples of y(t)+noise into 500 orthonormal coefficients Y_k. Coefficients Y_k with k>500 are negligible which shows that the transformation leads to a considerable data reduction. In Fig. 3 and 4 the step responses and Bode amplitudes are depicted of the original system and the approximations. As to be expected the ARX method gives an estimate which fits the first 8 true Markov parameters, (Swaanenburg and co-workers, 1985; Van den Hof and Janssen, 1987) which can be seen in Fig. 3. Figure 4 shows that the result of the ARX method is only satisfactory for the very high frequencies and that the orthogonal method gives a much better approximation over the whole frequency range.

Method 2.
For this method we used the same input and output data as for method 1. In Fig. 5 and 6 we compare the result of estimating Markov parameters (FIR) in the time domain with application of the second orthonormal method. The model that resulted from method 1, as described above, was used to generate the orthonormal functions. From the estimated Markov parameters a state space model was realized, using approximate realization, leading to a 13^{th} order model. This high order is the result of the large amount of noise, which leads to a large variance in the estimated parameters. Since the data are produced by an output error model it is to be expected that an output error method like FIR gives a better approximation then ARX. Comparison of Fig. 4 and Fig. 6 shows that this is indeed the case. The result of method 2 is clearly superior, it is a 5^{th} order model which is slightly better than the result of scheme 1.

CONCLUSIONS

We have shown that every finite dimensional stable linear discrete time system in a natural way gives rise to two sets of orthonormal functions, based on input and output balanced realizations, that are complete in ℓ_2. This is done by splitting of the all-pass part of the transfer function or, equivalently, by extending the matrices of singular vectors, corresponding with the Hankel matrix of the system. These functions, to be seen as a multivariable extension of the orthogonal polynomials, form a natural basis tc describe the system behavior. It has been shown that these functions give rise to new possibilities for the construction of approximate system identification methods.

ACKNOWLEDGEMENT

This research was sponsored by Shell Research/KSLA, Grant 85010 MSE. We wish to thank Dr. W.Schinkel and Ir. H.Swaanenburg for helpful discussion and their cooperation.

REFERENCES

Enns D.F. (1984), *Model Reduction for Control System Design.* PhD. diss., Dept. of Aeronautics and Astronautics, Standford University.

Glover, K. (1984). All optimal Hankel-norm approximations of linear multivariable systems and their L_∞-error bounds. *Int.J.Control, vol.39,* 1115–1193.

Heuberger, P.S.C. (1990a). A family of reduced order models, based on open loop balancing. *This issue.*

Heuberger, P.S.C. (1990b). *On Model Reduction, Complete Orthonormal Sets and System Identification.* Doctoral thesis, to be published. Delft University of Technology.

King, R.E. and P.N. Paraskevopoulos (1979). Parametric identification of discrete time SISO systems. *Int.J.Control, vol.30,* 1023–1029.

Laub, A.J. (1980). Computation of balancing transformations. *Proc. JACC, FA8-E:* San Francisco.

Lee, Y.W. (1933). Synthesis of electrical networks by means of the Fourier transforms of Laguerre functions. *J.Math.Physics., vol 11,* 83–113.

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

Moore B.C. (1981). Principal component analysis in linear systems: controllability, observability and model reduction. *IEEE Trans. Automat. Contr., AC-26,* 17–32.

Nurges Y. (1987). Laguerre models in problems of approximation and identification of discrete systems. *Autom. and Remote Contr. 48,* 346–352.

Nurges Y. and Y. Yaaksoo (1981). Laguerre state equations for a multivariable discrete system. *Autom. and Remote Contr., 42,* 1601–1603.

Paraskevopoulos, P.N. (1985). System analysis and synthesis via orthogonal polynomial series and Fourier series. *Mathematics and Computers in Simulation 27,* North-Holland, 453–469.

Swaanenburg, H.A.C., W.M.M. Schinkel, G.A. Van Zee and O.H. Bosgra (1985). Practical aspects of industrial multivariable process identification. In Barker H.A. and P.C Young (Eds.) (1985). *Identification and System Parameter Estimation,* IFAC Proc. Series 1985, York, UK, 201–206.

Unbehauen H. and G.P. Rao (1988). Continuous time approaches to system identification. In *Preprints of the 8th IFAC/IFORS Symp. on Identification and System Parameter Estimation,* Beijing, China, 60–68.

Van den Hof, P.M.J. and P.H.M. Janssen (1987). Some asymptotic properties of multivariable models identified by equation error techniques. *IEEE Trans. Automat. Contr. AC–32*, 89–92.
Wahlberg, B. (1989). System identification using high–order models, revisited. In *Proc. 28th Conf. on Decision and Control, Tampa, Florida*, 634–639.
Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. MIT Press, Cambridge, Ma.
Zervos C., P.R. Belanger and G.A. Dumont (1985). On PID controller tuning using orthonormal series identification. *IFAC Adapt. Contr. Chem. Proc. Frankfurt*, 13–20. Idem (1988). *Automatica, vol 24–2*, 165–175.

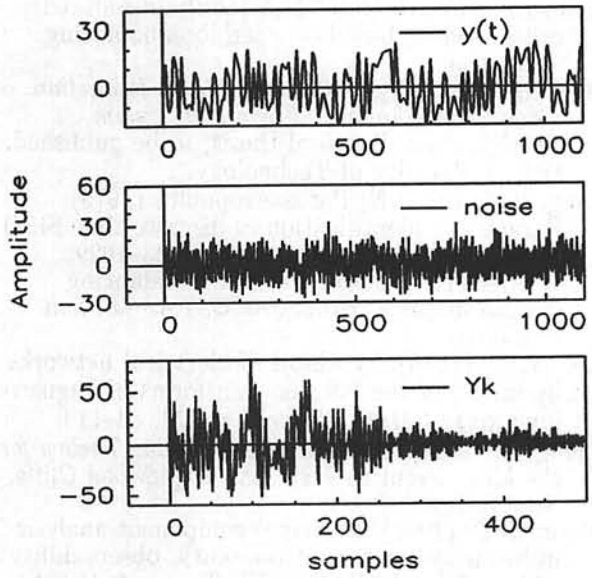


Fig. 2 Simulated output y(t) of the system, additive noise and transformed output Y_k.

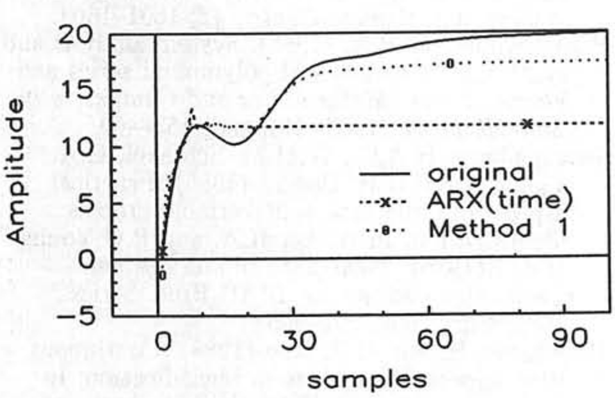


Fig. 3 Step responses of approximations (Method 1)

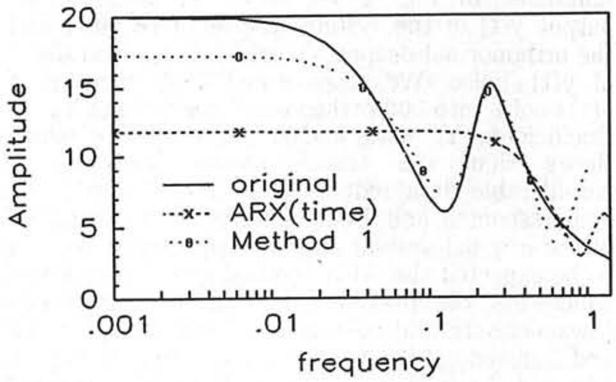


Fig. 4 Bode amplitudes of approximations (Method 1)

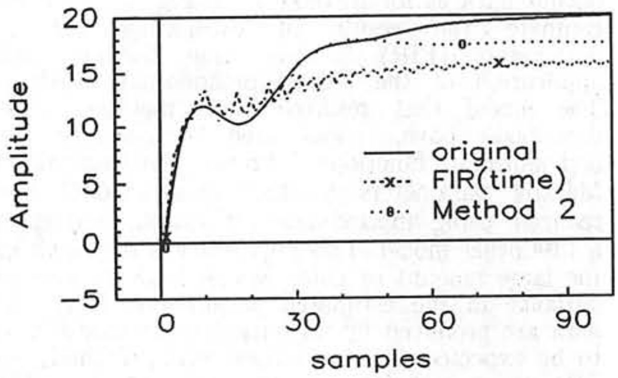


Fig. 5 Step responses of approximations (Method 2)

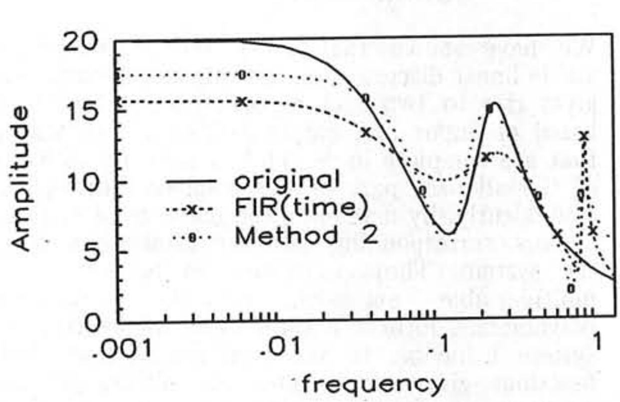


Fig. 6 Bode amplitudes of approximations (Method 2)

Selected Topics in Identification,
Modelling and Control; Vol. 1, 1990

Application of the fractional representation approach in identification: the noiseless case

Ruud J.P. Schrama

Lab. Measurement and Control, Dept. Mechanical Engineering and Marine Techn., Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands

Abstract. In approximate identification the actual purpose of the modeling procedure should be taken into account, in order to guarantee that the identified model is suited for its intended application. The fractional representation approach offers a setting that we claim to be suited to identify models, that can be used to design a controller for the system under consideration. In this paper we apply the algebraic systems theory to an uncorrupted linear feedback system with one input. Doing so, the closed loop identification problem is recasted into an open loop identification problem. The results presented are preliminary, but they are ready for generalization to a more general configuration.

Keywords. System identification, control design, algebraic systems theory, fractional representation, feedback system.

INTRODUCTION

In this paper we address the problem of identifying models, that have to be appropriate for control design. Let us first focus on this ultimate objective of the identification. Control design algorithms get intractable, if they are applied to models of high complexity. So in order to practise control design we have to come up with fairly simple models of complex systems. In fact these simple models have to reflect all characteristics of the plant, that are important in the closed loop, e.g. the feedback system of fig. 1. In robust control theory the ubiquitous approach is to approximate the plant by a nominal model and, in one way or another, to supply a supplementary model, that reflects the deficiency of the nominal model with respect to the plant (Doyle and Stein, 1981; Vidyasagar and Kimura, 1986; and many references in Dorato, 1987).

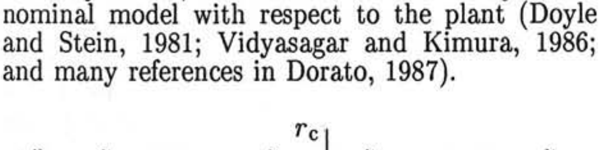


Fig. 1. Basic feedback system.

Often the nominal model is linear time-invariant and finite dimensional and the supplementary model consists of one or more bounded terms. The latter can be given e.g. by intervals, in which some parameters take their values, or e.g. a plant P can be modeled as $P_0+\Delta$, where P_0 has low complexity and Δ is bounded like $\|\Delta\|_\infty<\alpha$. Any such description will be called a supplementary model bound. In this way a model consists of *both* a nominal model and a supplementary model bound. Note that usually a model induces a class of input-ouput maps, since e.g. many Δ's satisfy the bound.

In view of the objective of control design a substantial difference between the nominal model and the plant may very well be acceptable or even required. In this regard we are obviously dealing with approximate identification. Now let us pay some attention to this aspect and recall several results from the literature: In Ljung and Van Overbeek (1978) it has been indicated, that to a large extent the outcome of an approximate identification is influenced by the specific conditions, that come into play while performing the procedure. Often several of these conditions can be chosen freely. One can think of the modelset, input signals etcetera. For some of these conditions, the consequences of a specific choice have been investigated in e.g. Ljung (1985, 1989) and Van den Hof (1989a, 1989b). A lot of attention has been paid to experiment design in approximate identification (Gevers and Ljung, 1986; Wahlberg and Ljung, 1986; Yuan and Ljung, 1985). The starting-point in these references is the observation, that in the ultimate application an approximate model will not perform as well as an exact model of the plant. Clearly the goal of experiment design is to minimize this performance degradation by choosing the right experimental conditions. Even so if an identified model P_0 is close to plant P, and thus $\|\Delta\|$ is small, then the model may be expected to result in a good performance, provided that in the identification the deficiency of the nominal model has been minimized in a proper sense. Apparently we may as well interpret approximate identification as defining the supplementary model and minimizing its bound.

In case the aim is control system design we might regard identification as obtaining a *good fit* of the nominal model as well as settling the supplementary model just there, where it affects the closed loop as little as possible. To our knowledge, the concepts of the pole placement controller and the minimum variance controller are the only control strategies, for which the performance degradation has been minimized analytically. This resulted in an optimal identification experiment design (Gevers and Ljung, 1986; Ljung, 1987); i.e. in its class the model identified under the prescribed conditions, is best suited to design the specific controller for the plant. Unfortunately such an explicit solution does not seem to be tractable for more complex control design methods.

The key to identification in behalf of control design is answering the question: what aspects of a system are important for control design? Clearly we would like to come up with a model, which *both* is close to the plant *and* gives rise to such a controller, that only small differences occur between the feedback system containing the plant and the feedback system containing the model. In this context we claim, that the fractional representation approach offers a proper setting for solving the identification problem. The incentive behind this claim is twofold.
First knowing that in approximate identification the resulting model depends in particular on the experimental conditions, intuition says that if the plant will operate ultimately in a closed loop, then the identification should be performed in some closed loop, that is very much alike. And secondly the set of all plants, that are stabilized by a known controller can be parametrized by means of the fractional representation (Hansen, 1989; Hansen *et al.*, 1988, 1989). So if we know a controller, that stabilizes the plant, then immediately the plant can be parametrized as a function of this controller. However the corresponding set is rather extensive: e.g. in case of a stable controller it contains also the zero system. Therefore we aim at shrinking this set by means of an identification procedure.

Our choice to use the fractional representation in identification with control design as an objective, can be solidified by recalling a couple of results from literature. First the fractional representation has been crucial in the development of control design techniques, that directly address the performance of the feedback system (Boyd, *et al.*, 1988; Gustafson and Desoer, 1983). And secondly in Hansen (1989) and Hansen *et al.* (1988, 1989) fractional representations have been used in the analysis of the exact identification of a plant in a noise corrupted feedback system. This resulted not only in an experiment design in terms of the loop inputs instead of the plant inputs, but also in an equivalent open loop identification problem. Though approximate identification of the nominal model has not been considered in these references, the results on recasting closed loop problems into open loop problems are quite promising towards this area.

In this paper we present a closer investigation into the application of fractional representations in (approximate) identification of a nominal model. More specific, as a start of a series of such investigations, we consider the identification of a plant in a noiseless environment. The results derived here can and will be generalized to a more general configuration.
We start with some notation and general preliminaries in the next section. Then given a controller C we use the fractional representation approach to parametrize all plants P, that make the feedback system of fig. 1 stable. We will indicate this feedback system by $H(P,C)$, which denotes the mapping from (r,r_c) into (u_c,u). Further we analyze the single variate control system $H_s(P,C)$, which equals $H(P,C)$ in case of $r_c\equiv 0$ (Desoer *et al.*, 1980). From this analysis we obtain the main result of the paper, i.e. a setting, that appears to be suited for the identification of models, that are appropriate for control design. In a discussion we outline some experiment design variables offered by this setting and we summarize a variety of aspects, that are worthwhile to be subjected to further investigations. We also pay some attention to the paradox between the aspect of control design and the paraphrase 'given a controller C'. Finally we end up with conclusions and future work.

NOTATION AND PRELIMINARIES

In this section we introduce some notation, we define the algebraic structure used in this paper and we summarize several results from the algebraic theory of fractional representations. For a proper introduction to this axiomatic theory we refer to Desoer *et al.* (1980, part II) and additionally Vidyasagar *et al.* (1982, parts I and II). A sufficient background on the standard algebraic terms can be found in Vidyasagar (1985, appendix A).

Algebraic structure. Let $\mathcal{R}$ be a principal ring and let $\mathcal{F}$ be the quotient field of $\mathcal{R}$, i.e. $\mathcal{F}:=\{a/b \mid a,b\in\mathcal{R},\ b\neq 0\}$. Furthermore let $\mathcal{J}$ be the group of units in $\mathcal{R}$: $\mathcal{J}:=\{a \mid a,a^{-1}\in\mathcal{R}\}$. Throughout the paper $\mathcal{R}$ will be considered as the set of all stable plants. As an example one could think of $\mathcal{R}$

to consist of all scalar rational plants, that have their poles in the open left half plane. Then $\mathcal{F}$ contains all rational not necessarily proper plants and every element of $\mathcal{J}$ is stable and stably invertible. However due to the generality of the algebraic setting the results also hold for discrete time systems and distributed systems. Our structure resembles the one used in Vidyasagar *et al.* (1982) and it differs from the algebraic structure built in Desoer *et al.* (1980), where, in terms of the example, only proper plants have been considered.
A plant P with m inputs and p outputs and with all its entries in the ring $\mathcal{H}$ is an element of $\mathcal{H}^{p\times m}$. However dimensions are not an issue in this paper and for the sake of conciseness we will denote $\mathcal{H}^{p\times m}$ as $\mathcal{H}$ and likewise for $\mathcal{F}$ and $\mathcal{J}$.

Algebraic theory. We recall several definitions and facts from the algebraic theory of fractional representations.
The factors $N,D\in\mathcal{H}$ are *right coprime* over the ring of stable plants if there exist $X,Y\in\mathcal{H}$ such that $XN+YD=I$. We will call the factors X,Y *right Bezout factors* of the pair (N,D). The pair (N,D) is said to be a *right coprime factorization* (rcf) of the plant $P\in\mathcal{F}$ if $\det(D)\neq 0$, $P=ND^{-1}$ and $N,D\in\mathcal{H}$ are right coprime. Analogously left coprimeness and a left coprime factorization (lcf) are defined with the pair $(\tilde{D},\tilde{N})$ such that $\tilde{N}\tilde{X}+\tilde{D}\tilde{Y}=I$ and $P=\tilde{D}^{-1}\tilde{N}$.
Some nice results with respect to the stability of the feedback system $H(P,C)$ of figure 1 have been based on these factorizations. In the sequel both plant P and controller C are considered to be in $\mathcal{F}$. The next lemma states a necessary and sufficient condition for a plant P and a controller C to make a stable feedback system $H(P,C)$.

Lemma 1 (Vidyasagar, *et al.*, 1982). Let (N_P,D_P) be a rcf of P and $(\tilde{D}_c,\tilde{N}_c)$ a lcf of C. Then the loop $H(P,C)$ is stable if and only if Λ, defined as

$$\Lambda = \tilde{D}_c D_P + \tilde{N}_c N_P, \tag{1}$$

is unimodular in $\mathcal{H}$, i.e. $\Lambda\in\mathcal{J}$. □

We like to recall, that in this lemma the notions of stability concerns the boundedness of the mapping $H(P,C)$ from the two outer loop signals r and r_c to the two inner loop signals u and u_c (fig.1). Clearly the stability condition holds irrespective of the fact whether the signals are deterministic or stochastic.
Since Λ in equation (1) is stably invertible, it can easily be shown, that for any rcf (N_P,D_P) of plant P, every stabilizing controller C has a $(\tilde{D}_c,\tilde{N}_c)$ such that

$$\tilde{D}_c D_P + \tilde{N}_c N_P = I \tag{2}$$

and thus $\tilde{D}_c,\tilde{N}_c$ are Bezout factors of N_P,D_P and vice versa.

ANALYSIS OF THE NOISELESS CASE

In this section we use the establishments of the fractional representation theory to analyze the single variate control system $H_s(P,C)$, which equals the feedback system $H(P,C)$ of fig. 1 in case $r_c\equiv 0$. Thereby we recast the closed loop identification problem into an open loop identification problem.

Like in Hansen *et al.* (1988, 1989) we model a plant P, that makes a stable feedback system $H(P,C)$, by means of the dual of the fractional representation approach to control design. In this modeling procedure it is pivotal to know a controller, that stabilizes the unknown plant. Later on we will use this model in the analysis of the single input feedback system $H_s(P,C)$.
Using the stability condition of lemma 1 together with just any plant P_0 that is stabilized by controller C, it is possible to derive the following necessary and sufficient condition for a plant P to make a stable feedback system $H(P,C)$. Though this is the dual of the well-established control design result (Desoer, *et al.*, 1980), we supply an alternative simple proof in appendix P.

Lemma 2. Given a controller C with rcf (N_c,D_c) and given a rcf (N_0,D_0) of just any plant P_0, such that $H(P_0,C)$ is stable, then $H(P,C)$ is stable, if and only if P admits a rcf[1] (N_P,D_P) with

$$N_P = (N_0+D_cR),\quad D_P = (D_0-N_cR), \tag{3}$$

and $R\in\mathcal{H}$ is such that $\det(D_0-N_cR)\neq 0$. □

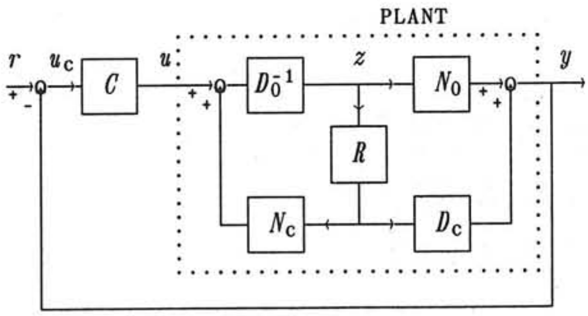


Fig. 2: R-parameterization of $H_s(P,C)$.

Apparently lemma 2 can be interpreted as follows: any $R\in\mathcal{H}$ with $\det(D_0-N_cR)\neq 0$ gives rise to a plant P such that $H(P,C)$ is stable. This we call the *R-parameterization* of the set of all plants, that are stabilized by controller C.
Since stability of the loop (P_0,C) is the only

1) As a consequence of the algebraic structure, only plants, that admit rcf's as well as lcf's over $\mathcal{H}$, are considered (see Desoer and Gundes, 1988; and Anantharam, 1985).

requirement on P_0 in lemma 2, we could as well use the Bezout factors $\tilde{X}_c, \tilde{Y}_c$ of $(\tilde{D}_c, \tilde{N}_c)$ as N_0 respectively D_0; i.e. $P_0 := \tilde{X}_c \tilde{Y}_c^{-1}$.

Corollary 1. Given controller C with rcf (N_c, D_c) and lcf $(\tilde{D}_c, \tilde{N}_c)$ satisfying $\tilde{N}_c \tilde{X}_c + \tilde{D}_c \tilde{Y}_c = I$, then $H(P,C)$ is stable, if and only if P admits a rcf (N_P, D_P) with

$$N_P = (\tilde{X}_c + D_c R), \quad D_P = (\tilde{Y}_c - N_c R), \tag{4}$$

where $R \in \mathcal{U}$ is such that $\det(\tilde{Y}_c - N_c R) \neq 0$. Moreover any such rcf constitutes Bezout factors of the lcf $(\tilde{D}_c, \tilde{N}_c)$ of the controller and vice versa. □

Proof. See appendix P.

Now we come to the key result of this paper. Getting ahead of the next section we state, that in view of control design it comes in useful to identify the plant P in terms of its right coprime factors N_P and D_P. In order to realize this we introduce the intermediate variable z as $z = D_P^{-1} u$, and with $y = Pu = N_P D_P^{-1} u$ this leads to

$$\begin{bmatrix} u \\ y \end{bmatrix} = \begin{bmatrix} D_P \\ N_P \end{bmatrix} z. \tag{5}$$

Now we focus on the single variate control system $H_S(P,C)$ of fig. 2. By equation (3) it is easy to verify, that z in equation (5) and z in fig. 2 are one and the same variable. Identification of the mapping from z to (u,y) would solve our problem. Referring to fig. 2 it is common to assume, that only u, y and r can possibly be measured. Therefore we propose the construction of the variable z from reference signal r, using the R-parameterization.

Proposition 1. Let the controller C with lcf $(\tilde{D}_c, \tilde{N}_c)$ stabilize both the unknown plant P and any plant P_0 with rcf (N_0, D_0)[2]. Then the intermediate variable z originating from the feedback system $H_S(P,C)$ of fig. 2, can be constructed by means of the stable mapping

$$z = \Lambda_0^{-1} \tilde{N}_c r, \tag{6}$$

with $\Lambda_0 = \tilde{D}_c D_0 + \tilde{N}_c N_0$.

Proof. See appendix P.

In case the signal r is not measurable, we still can construct the variable z by applying the next corollary, which follows from the proof of the proposition above.

Corollary 2. Under the conditions given in proposition 1, the variable z can be constructed by

$$z = \Lambda_0^{-1} (\tilde{D}_c u + \tilde{N}_c y). \tag{7}$$

Note that in equations (6) and (7) only factors of the controller and the known plant P_0 have been used. So no information on the plant P is needed to construct z from r. Further it is remarkable, that $\Lambda_0^{-1} \tilde{N}_c$ depends on the specific rcf (N_0, D_0) of P_0 and it does *not* depend on what lcf $(\tilde{D}_c, \tilde{N}_c)$ has been chosen for C: any other lcf of C can be written as $(A\tilde{D}_c, A\tilde{N}_c)$ leading to $A\Lambda$ in equation (1), and in the product $(A\Lambda)^{-1} A \tilde{N}_c$ the factor A is canceled. So without loss of generality, that is without affecting the mapping from r to z, we can choose a lcf for C, such that $\tilde{D}_c, \tilde{N}_c$ are right Bezout factors of (N_0, D_0) (see equation (2)). In this case $\Lambda_0 = I$ and $z = \tilde{N}_c r$.

We end up this section, making a remark with respect to the necessity of the condition in lemma 2. That result has been derived for the feedback system $H(P,C)$, which has two loop inputs, whereas we analyzed $H_S(P,C)$ with just one loop input. Indeed in the latter case the necessity of the condition does not hold as is shown by a counter example in appendix E. Nevertheless we restricted the investigation deliberately to only the set of plant given by the R-parametrization of lemma 2, in order that the setting is readily extendible to a configuration with a so-called two-input plant (Schrama, 1990).

A SUITABLE SETTING FOR IDENTIFICATION

The analysis of the previous section opens up several new possibilities in the identification for the purpose of control design. Here we like to mention a few of them and we have to admit, that the end of this section does not go without speculations.

Let us first return to proposition 1 and examine what happens, if the mapping in equation (5) is identified. Suppose the factorizations of C and P_0 have simple dynamics and suppose r is a white noise signal, then by equation (6) z will have a simple spectrum. If at the same time the plant P is very complex, then this complexity will be reflected in u and y, and thus it asserts itself in the identification. This 'simplicity' of z is not immediate from equation (7).

We can point out several variables, that influence the identification procedure. The controller and signal r share this property straight on: it is well-known, that the identification result can be manipulated via the signal spectra (Gevers and Ljung, 1986; Hansen *et al*, 1988, 1989; Yuan and Ljung, 1985) and these latter depend on both C and r. Further as indicated in the previous section the mapping $\Lambda_0^{-1} \tilde{N}_c$ of equation (6) depends on P_0 and its specific rcf (N_0, D_0). Therefore P_0 and its rcf can be seen as frequency weighting functions. We also mentioned, that without loss of generality the lcf $(\tilde{D}_c, \tilde{N}_c)$ of C can be chosen

[2] i.e. $H(P,C)$ is stable and thus lemma 2 is applicable.

such that $z=\tilde{N}_c r$. We emphasize that now every alteration of P_0 or its rcf immediately leads to a change of its right Bezout factors $\tilde{D}_c,\tilde{N}_c$, and thus $\tilde{N}_c$ cannot be chosen freely. Moreover if P_0 is used to stress e.g. low frequency dynamics, then this reasserts itself in $\tilde{N}_c$.
Since all these variables are at our disposal, they are actually experiment design variables. Though it is clear that they *do* affect the identification result, we do not know yet *how* it comes about. This aspect of experiment design definitely needs further investigation and most probably we can take advantage of the results of Hansen (1989) at this point.

Now we pay some attention to the aspect of approximation. If in equation (3) P_0 and (N_0,D_0) are such that R is small in any sense, then evidently the model P_0 can be said to be close to the plant P. At this stage we can clarify why we have chosen a rcf model of the plant instead of a lcf as in Hansen (1989) and Hansen, *et al.* (1988, 1989). In these references the experiment design problem for exact closed loop identification has been tackled by considering the identification of a term equivalent to R. Unfortunately in general this leads to an increase of the dimension of the problem in the sense of the order of the models involved. On the other hand if we use a rcf and equation (5), then we can restrain the order of the approximating model in a straightforward manner.

Next we consider control design. There is a strong relationship between the fractional representation and the graph topology, which is the weakest topology in which feedback stability is a robust property (Vidyasagar, *et al.*, 1982); simply stated if a sequence P_i converges to P in this topology, then the sequence of feedback systems $H(P_i,C)$ converges to $H(P,C)$. This topology is induced by the gap metric, which can be defined in terms of factorizations[3]. In fact if P_0 is close enough to P in this gap metric, then we can practice robust control design onto P_0, such that stabilization of P is guaranteed (Bongers and Bosgra, 1990; Glover and McFarlane, 1988, 1989). For more details on the gap metric we refer to Georgiou (1988).

An interesting question that arises, is how to parameterize the factorizations. Since we have not solved this problem yet, we can not supply an example at this moment. An even further reaching problem is the incorporation of the metric itself in the identification. That is, if the identification comes up with some model P_0, then given the data what can be said about R? The problem gets even more involved if some noise contributions are present. Since the latter will be the case in practice, first the setting has to be generalized to control systems with noise contributions and more inputs. This is currently performed (Schrama, 1990) based on Desoer and Gündes (1988) and Nett (1986). These latter references concern effectively the set of all proper linear systems, that have more than one input (and output) vector and that give rise to a stable feedback system.
Finally we address the paradox between the aspect of control design and the need for a known controller, that stabilizes the plant. We like to urge that everything hinges on the design of a *new* controller. This can be realized in two ways. First a new robust controller could be designed e.g. in relation to the gap metric as mentioned earlier. Secondly one could think of an iterative scheme, in which both identification and control design are performed consecutively and repeatedly. At this moment it is untransparent to what this iteration will lead, but we have the strong impression, that the knowledge obtained in the successive identification procedures should be turned to use. This might very well be done by means of the design variable P_0.

[3] At least for linear finite dimensional systems

CONCLUSIONS AND FUTURE WORK

We have put the single variate control system, i.e. the feedback system of fig. 1 with $r_c\equiv 0$, in a setting, that is suitable for approximate identification of the plant in terms of a right coprime factorization. Moreover the closed loop identification problem has been recasted into an open loop identification problem. Models, that will be identified in this setting, appear to be well suited to control design. We pointed out, that the identification result is affected by several variables, that we have at our disposal. The precise impact of these variables needs further investigation.
Furthermore we have mentioned, that this setting offers several possibilities in relation to approximate modeling and control. However it is not clear yet how to handle and to combine the different phenomena.
Finally the generalization of the setting to multiple input noisy feedback systems is currently under investigation (Schrama, 1990). And in the next future we will also examine the problem of parametrization and of pulling apart noise contributions from effects caused by unmodeled dynamics.

REFERENCES

Anantharam, V. (1985). On stabilization and the existence of coprime factorizations. *IEEE Trans. Automat. Contr., AC-30*, 1030-1031.
Bongers, P.M.M. and O.H. Bosgra (1990). Low order robust H_∞ controller synthesis. *Report Lab. Meas. Control, Delft Univ. of Techn.; Also submitted for publication.*

Boyd, S.P., V. Balakrishnan, C.H. Barratt, N.M. Khraishi, X. Li, D.G. Meyer and S.A. Norman (1988). A new CAD method and associated architectures for linear controllers. *IEEE Trans. Automat. Contr.*, *33*, 268-283.

Desoer, C.A., R.-W. Liu, J. Murray and R. Saeks (1980). Feedback system design: the fractional representation approach to analysis and synthesis. *IEEE Trans. Automat. Contr.*, *AC-25*, 399-412.

Desoer, C.A. and A.N. Gündes (1988). Algebraic theory of linear time-invariant feedback systems with two-input two-output plant and compensator. *Int. J. Control*, *47*, 33-51.

Dorato, P., ed. (1987). *Robust Control*, IEEE Press Selected Reprints, IEEE, Inc., New York.

Doyle, J.C. and G. Stein (1981). Multivariable feedback design: concepts for a classical/modern synthesis. *IEEE Trans. Automat. Contr.*, *AC-26*, 4-16.

Gevers, M. and L. Ljung (1986). Optimal experiment designs with respect to the intended model application. *Automatica*, *22*, 543-554.

Georgiou, T.T. (1988). On the computation of the gap metric. *Syst. Control Lett.*, *11*, 253-257.

Glover, K. and D. McFarlane (1988). Robust stabilization of normalized coprime factors: an explicit H∞ solution. *Proc. American Control Conf., Atlanta, GA*, 842-847.

Glover, K. and D. McFarlane (1988). Robust stabilization of normalized coprime factor plant descriptions with H∞-bounded uncertainty. *IEEE Trans. Automat. Contr.*, *34*, 821–830.

Gustafson, C.L. and C.A. Desoer (1983). Controller design for linear multivariable feedback systems with stable plants, using optimization with inequality constraints. *Int. J. Contr.*, *37*, 881-907.

Hansen, F. and G. Franklin (1988). On a fractional representation approach to closed-loop experiment design. *Proc. American Control Conf., Atlanta, GA*, 1319-1320.

Hansen, F. (1989). *A Fractional Representation Approach to Closed-loop System Identification and Experiment Design.* PhD thesis, Stanford University, Stanford, CA, March 1989.

Hansen, F., G. Franklin and R. Kosut (1989). Closed-loop identification via the fractional representation: experiment design. *Proc. American Control Conference, Pittsburgh, PA*, 1422-1427.

Ljung, L. (1985). Asymptotic variance expressions for identified black-box transfer function models. *IEEE Trans. Automat. Contr.*, *AC-30*, 834-844.

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

Ljung, L. (1989). Choices of prefilters and forgetting factors in noise-free identification. *Proc. American Control Conf., Pittsburgh, PA*, 1409-1415.

Ljung, L. and A.J.M. Van Overbeek (1978). Validation of approximate models obtained from prediction error identification. *Proc. 7th IFAC World Congress, june 12-16, Helsinki*, 1899-1906.

Nett, C.N. (1986). Algebraic aspects of linear control system stability. *IEEE Trans. Automat. Contr.*, *AC-31*, 941-949.

Schrama, R.J.P. (1990). Fractional representations in the identification of a two-input two-output plant in closed loop. *In preparation.*

Van den Hof, P.M.J. (1989a). A deterministic approach to approximate modelling of input-output data. *Proc.28th Conf. Decision Control, Tampa, FL*, pp.659-664.

Van den Hof, P.M.J. (1989b). *On residual-based parametrization and identification of multivariable systems.* Dr. Dissertation, Dept. Elect. Eng., Eindhoven University of Technology, The Netherlands, 1989.

Vidyasagar, M. (1985). *Control System Synthesis: A Factorization Approach.* M.I.T. Press, Cambridge, MA.

Vidyasagar, M., H. Schneider and B. Francis (1982). Algebraic and topological aspects of feedback stabilization. *IEEE Trans. Automat. Contr.*, *AC-27*, 880-894.

Vidyasagar, M. and H. Kimura (1986). Robust controllers for uncertain linear multivariable systems. *Automatica*, *22*, 85-94.

Wahlberg, B. and L. Ljung (1986). Design variables for bias distribution in transfer function estimation. *IEEE Trans. Automat. Contr.*, *AC-31*, 134-144.

Yuan, Z.-D. and L. Ljung (1985). Unprejudiced optimal open loop input design for identification of transfer functions. *Automatica*, *21*, 697-708.

APPENDIX E

By means of a counter example we show, that the necessity of the condition on P for $H(P,C)$ to be stable (lemma 2) does not hold for P to make a stable closed loop $H_S(P,C)$. Let P=s+1 and C=1/(s+1) then in $H_S(P,C)$ we have

$$e_1=\frac{1}{2(s+1)}\,r,\quad e_2=\tfrac{1}{2}r,$$

and thus all inner loop and output signals are bounded provided r is bounded. The factors $\tilde{N}_C, \tilde{D}_C, \tilde{X}_C, \tilde{Y}_C$ can be chosen as C,1,0,1 and R follows uniquely from equation (4): $R=\frac{1}{2}(s+1)$. This R is *not* an element of $\mathcal{X}$. Conclusively though $H_S(P,C)$ is stable, there is no R-parameterization of P. □

APPENDIX P

Proof of lemma 2.

If. Given $P = N_P D_P^{-1}$ with N_P and D_P as defined in equation (3) and a lcf $(\tilde{D}_c,\tilde{N}_c)$ of C. We show that the control system $H(P,C)$ is stable, and a fortiori that the pair (N_P,D_P) is right coprime. Irrespective of the coprimeness of (N_P,D_P) we substitute equation (3) in equation (1):

$$\begin{aligned}\Lambda &= \tilde{N}_c(N_0+D_cR) + \tilde{D}_c(D_0-N_cR)\\ &= \tilde{N}_cN_0 + \tilde{D}_cD_0 + (\tilde{N}_cD_c-\tilde{D}_cN_c)R.\end{aligned}$$

The factor $(\tilde{N}_cD_c-\tilde{D}_cN_c)$ equals $\tilde{D}_c(C-C)D_c$, and thus the term preceding R is zero. Furthermore (P_0,C) is stable, so $\Lambda\in\mathcal{J}$. Since by definition $\Lambda=\tilde{N}_cN_P+\tilde{D}_cD_P$, both the coprimeness of (N_P,D_P) and the stability of $H(P,C)$ are guaranteed.

Only if. Given $H(P,C)$ is stable then there exist a rcf (N_P,D_P) of P and lcf $(\tilde{D}_c,\tilde{N}_c)$ of C such that $\tilde{D}_cD_P+\tilde{N}_cN_P=I$. Next let P_0 be any plant, that makes a stable feedback system $H(P_0,C)$, and for the moment let (N_0,D_0) be a rcf that satisfies $\tilde{D}_cD_0+\tilde{N}_cN_0=I$. Then let R_x be given implicitly by $N_P=N_0+D_cR_x$, and thus $R_x=D_c^{-1}(N_P-N_0)$. In order to establish lemma 2 we have to prove consecutively, that a) D_P equals $D_0-N_cR_x$ and b) R_x is stable as in equation (3).

a) Denote $D_x=D_0-N_cR_x$, substitute R_x and $N_cD_c^{-1}=\tilde{D}_c^{-1}\tilde{N}_c$, then $D_x=D_0-\tilde{D}_c^{-1}\tilde{N}_c(N_P-N_0)$. Use $\tilde{D}_cD_0+\tilde{N}_cN_0=I$ in the rearrangement of this expression to $\tilde{D}_cD_x+\tilde{N}_cN_P=I$. Together with $\tilde{D}_cD_P+\tilde{N}_cN_P=I$ this shows that $D_x=D_P$.

b) Now $D_P=D_0-N_cR_x$ and while by definition $D_P\in\mathcal{H}$, we have $N_cR_x = D_0-D_P\in\mathcal{H}$. Furthermore $D_cR_x=(N_P-N_0)\in\mathcal{H}$. Since N_c,D_c are right coprime, there exist $X_c,Y_c\in\mathcal{H}$ such that $Y_cD_c+X_cN_c=I$. Now $X_c(N_cR_x)+Y_c(D_cR_x)=R_x$ and since $\mathcal{H}$ is a ring we have $R_x\in\mathcal{H}$. Finally extension of the proof to a rcf (N_0,D_0) of P_0 with $\tilde{D}_cD_0+\tilde{N}_cN_0=\Lambda_0$ and $I\neq\Lambda_0\in\mathcal{J}$ becomes self-evident by the choice of a rcf (N_P,D_P) of P, such that $\tilde{D}_cD_P+\tilde{N}_cN_P=\Lambda_0$. □

This proof is more concise than the proof in Desoer *et al.* (1980), which has been derived for proper P and C, that both have coprime factorizations.

Proof of corollary 1. Analogously to the proof of lemma 2 the equation $\Lambda=\tilde{N}_cN_P+\tilde{D}_cD_P$ can be reduced to $\Lambda=\tilde{N}_c\tilde{X}_c+\tilde{D}_c\tilde{Y}_c=I$. □

Proof of proposition 1.
Using the R-parameterization in equation (5) we obtain

$$\begin{aligned}u &= D_Pz = (D_0-N_cR)z\\ y &= N_Pz = (N_0+D_cR)z\end{aligned}$$

and from these equations

$$\begin{aligned}Rz &= D_c^{-1}(y-N_0z)\\ N_cRz &= D_0z-u\end{aligned}$$

(the corresponding variables appear in fig. 2). Substitute $N_cD_c^{-1} = C = \tilde{D}_c^{-1}\tilde{N}_c$ and eliminate Rz, then we get

$$(\tilde{D}_cD_0 + \tilde{N}_cN_0)\, z = \tilde{D}_cu + \tilde{N}_cy.$$

Rearrangement of the controller equation $u=\tilde{D}_c^{-1}\tilde{N}_c(r-y)$ shows, that the right hand term equals $\tilde{N}_cr$. Finally since the loop $H(P_0,C)$ is stable, the factor $(\tilde{D}_cD_0 + \tilde{N}_cN_0)$ is stably invertible. □

H_∞-norm computation using a Hamiltonian matrix

N.A. Bruinsma[1,2], M. Steinbuch[1]

[1]) *Philips Research Laboratories Eindhoven, P.O. Box 80.000, 5600 JA Eindhoven.*
[2]) *Lab. Measurement and Control, Dept. Mechanical Engineering and Marine Techn., Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands*

Abstract. This paper discusses various methods to compute the H_∞-norm of a transfer function matrix with use of a related Hamiltonian matrix. The underlying theory will be illustrated with some examples.

Keywords. H_∞-norm, L_∞-norm, H_∞-Control, Hamiltonian matrix, singular values

NOTATION

[A,B,C,D]	state-space representation of a transfer function matrix $G(s)= C[sI-A]^{-1}B+D$
$G'(s)$	transpose of $G(s)$
$G^{\sim}(s)$	$G'(-s)$
$G^*(s)$	$G'(\bar{s})$ ($\bar{s}$ = complex conjugate of s)
$\sigma_{\max}$	maximum singular value

1. INTRODUCTION

In the recent literature on robust analysis and control, see for instance (Doyle et al., 1989; Francis, 1987), the H_∞-norm of a transfer function matrix plays an important role. The computation of the H_∞-norm can be necessary either in analysis of a system, or in the synthesis of a controller, see for instance (Scherer, 1989).

Definition 1.1. Let a real-rational proper transfer function matrix G(s) be given by [A,B,C,D], and let all the eigenvalues of A have negative real part. Then the <u>H_∞-norm of G(s)</u> is defined as the supremum of the maximum singular value of G(s), evaluated over the right half plane:

$$\|G\|_\infty := \sup_{\mathrm{Re}(s)\geq 0} \sigma_{\max}(G(s)) = \\ = \sup_{\omega\in\mathbb{R}} \sigma_{\max}(G(j\omega)) \qquad (1)$$

The H_∞-norm is defined for systems that are analytical in the closed right half plane. Systems that have no poles on the imaginary axis have a L_∞-norm that is defined as the supremum of the maximum singular value of G(s) evaluated on the imaginary axis, so the last part of (1) also gives the L_∞-norm in the case of unstable G(s).

Until 1988 not much attention has been paid to the computation of the H_∞-norm. The 'computation' was done by a search over frequencies. The disadvantages of this approach are obvious: it cannot be used automatically within other algorithms, it takes a considerable amount of computer time, and no accuracy bound can be given. In 1988 a bisection algorithm was presented by Boyd, Balakrishnan and Kabamba (1988,1989) and Robel (1989), to compute the H_∞-norm with guaranteed accuracy, using the relation between the singular values of the transfer function matrix and the eigenvalues of a related Hamiltonian matrix. This bisection algorithm is much more efficient than a search over frequencies, but for repeated use as well as for very large systems, it is still not very fast.

Several attempts have been made to reduce the computing time. The use of derivatives of the Hamiltonian matrix for a search algorithm has been investigated by (Bruinsma, 1990). Boyd and Balakrishnan (1990) and independently Bruinsma and Steinbuch (1990) developed an algorithm approximating the H_∞-norm with a lower bound, to which we will refer as the 'two-step algorithm'. This algorithm is much faster than the other methods.

For an exact description and proofs of the algorithms we refer to the mentioned papers. Here we give a short description of the bisection algorithm, the algorithm using eigenvalue derivatives and the two-step algorithm. The role of the Hamiltonian matrix in these algorithms will be made clearer by giving some examples of how its eigenvalues behave. A comparison of the three algorithms will be given.

2. THEORETICAL BACKGROUND

2.1. Hamiltonian Eigenvalues and Singular Values

All algorithms described in this article are based on a relation between the singular values of a

transfer function G(s) and the eigenvalues of a related Hamiltonian matrix $H(\gamma)$.
Let system G(s) be given through

$$G(s) = [A,B,C,D] \qquad (2)$$

and *let A not have any eigenvalues on the imaginary axis.*
For $\gamma > 0$ *not equal to a singular value of D* we define the Hamiltonian matrix

$$H(\gamma) = \begin{bmatrix} A-BR^{-1}D'C & -\gamma BR^{-1}B' \\ \gamma C'S^{-1}C & -A'+C'DR^{-1}B' \end{bmatrix} \qquad (3)$$

where $R = (D'D - \gamma^2 I)$ and $S = (DD' - \gamma^2 I)$.

As stated in (Boyd et al., 1989), under the assumptions made, (2) and (3) are related by the following equivalence.

Proposition 2.1. For all $\omega_p \in \mathbb{R}$,

$$j\omega_p \text{ is an eigenvalue of } H(\gamma_1) \Leftrightarrow \gamma_1 \text{ is a singular value of } G(j\omega_p) \qquad (4)$$

This relation between the singular values of G(s) and the eigenvalues of $H(\gamma)$ has been proven in (Bruinsma and Steinbuch, 1990) via the fact that the transfer function matrix $[\gamma^2 I - G^{\sim}(s)G(s)]^{-1}$ has a realization with state matrix $H(\gamma)$. The proof follows by this fact, and by realizing that the singular values of G(s) are computed with
$\det[\gamma^2 I - G^*(s)G(s)] = 0$
and that
$G^*(s) = G^{\sim}(s)$ for $s = j\omega$, $\omega \in \mathbb{R}$.
From Prop. 2.1. follows the next corollary, important in both the bisection and the eigenvalue derivative algorithm.

Corollary 2.1. Let G(s) and $H(\gamma)$ be given by (2), (3) and let $\gamma > \sigma_{max}(D)$ then

$$\gamma > \|G\|_\infty \Leftrightarrow H(\gamma) \text{ has no imaginary eigenvalues} \qquad (5)$$

The proof follows directly from Prop. 2.1., as stated in (Boyd et al., 1989).

2.2. Behaviour of Hamiltonian eigenvalues

The consequences of the theory for the behaviour of the eigenvalues of the Hamiltonian matrix (3) as a function of γ will be discussed using some examples.

Example 1.
Consider the following system:

$$G(s) = \frac{1}{(\tau s+1)\cdot(s^2/\omega_0^2+2\beta s/\omega_0+1)} \qquad (6)$$

with τ=1s, ω_0=5rad/s, β=0.05.

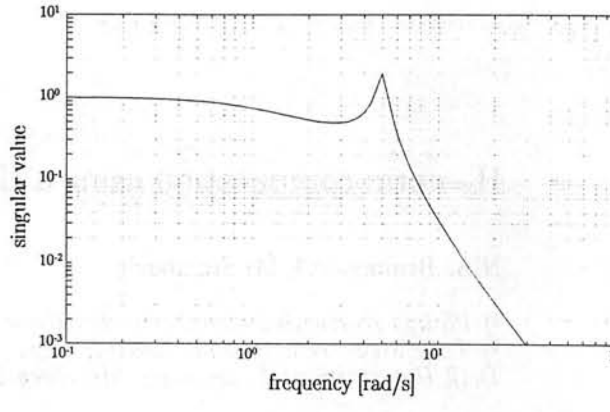


Fig. 1. Singular values of third-order system (6)

The singular value plot of this system (Fig. 1) shows that the H∞-norm is ≈ 2, and $\sigma_{max}(D) = 0$. We may expect (in accordance with Cor. 2.1) that for all $\gamma > \|G\|_\infty$ the Hamiltonian matrix $H(\gamma)$ will not have imaginary eigenvalues, and for all $0 < \gamma \leq \|G\|_\infty$ at least one of the loci of the eigenvalues of the Hamiltonian matrix will be on the imaginary axis.

The eigenvalues of the Hamiltonian matrix $H(\gamma)$ are computed for a number of values γ between 0.1 and 10 and plotted in the complex plane (Fig. 2a). The six eigenvalues lie symmetric with respect to both the real and the imaginary axis, as is inherent in the structure of the Hamiltonian matrix.
For $\gamma \to \infty$ the eigenvalues of $H(\gamma)$ will equal + and - the poles of G(s), as can be concluded from (3):

$$\lim_{\gamma \to \infty} H(\gamma) = \begin{bmatrix} A & 0 \\ 0 & -A' \end{bmatrix}$$

For $\gamma = 10$ the eigenvalues still are very close to + and - the poles of the system (-1 and $-0.25\pm4.99j$). For smaller γ the eigenvalues move towards the imaginary axis. In Fig. 2b the real part of the eigenvalues is plotted as a function of gamma. From this figure it follows that (in accordance with Cor. 2.1) below $\gamma \approx 2$ the Hamiltonian matrix has purely imaginary eigenvalues.

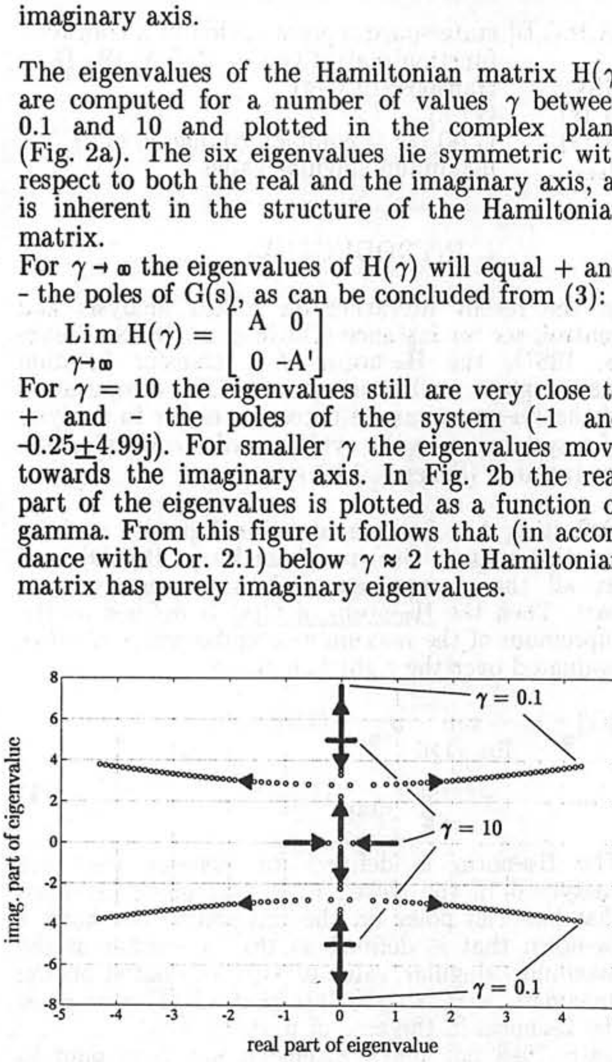


Fig. 2a. Eigenvalues of $H(\gamma)$ for system (6)

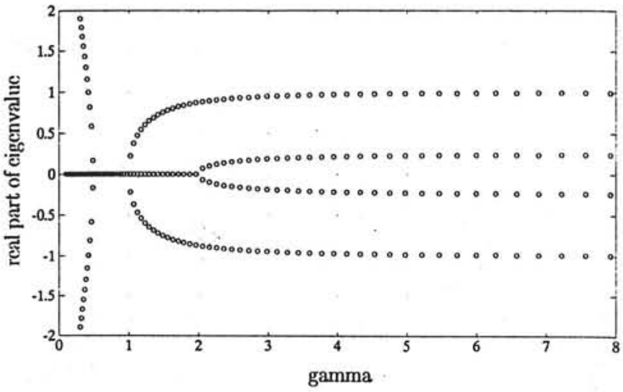


Fig. 2b. Real part of the eigenvalues as a function of γ

We will now relate the singular value plot (Fig. 1) to the eigenvalues of $H(\gamma)$ (Fig. 2), using Prop. 2.1. If a line in the singular value plot at some value γ would intersect the singular value at a number of frequencies ω_1 to ω_k,, then it follows from Prop. 2.1. that $H(\gamma)$ would have $2 \cdot k$ imaginary eigenvalues at $\pm j\omega_1$ to $j\omega_k$.

Relations between Fig. 1 and Fig. 2:
Global maximum. The singular value plot has a global maximum 2 at a frequency $\omega = 5$ rad/s $\Rightarrow$ for $\gamma = 2$ a quadruple of complex eigenvalues coincides at the imaginary axis at $\approx \pm 5j$, for smaller γ they split up in 4 imaginary eigenvalues.
Local maximum. The singular value plot has a local supremum 1 at a frequency $\omega = 0 \Rightarrow$ for $\gamma = 1$ a real pair of eigenvalues (not a quadruple because it is a maximum at $\omega = 0$) reaches the imaginary axis at the origin.
Number of imaginary eigenvalues. For a line in the singular value plot at value γ, the number of intersections with the singular value plot times 2 will give the number of imaginary eigenvalues of $H(\gamma)$, as can be verified by Fig. 2.

interval	number of intersections (Fig.1)	number of imag. eigenvalues (Fig.2)
$0<\gamma<\approx 0.5$	1	2
$\approx 0.5<\gamma<1$	3	6
$1<\gamma<2$	2	4
$2<\gamma$	0	0

Example 2.
For the second example we again take the third–order system given by (6), but with a higher damping factor β:

third-order system given by (6)
with τ=1s, ω_0=5rad/s, β=0.2 (7)

Because we increased the damping factor β to 0.2, the peak in the singular value plot (Fig. 3) caused by the second order term will be smaller than in the first example, and the maximum is achieved at $\omega = 0$. When decreasing γ from infinity, in this case the real eigenvalue pair will be the first to reach the imaginary axis (at $\gamma = 1$), as can be verified with Fig. 4a. and 4b. At $\gamma \approx 0.5$, which is the value of the local maximum at $\omega \approx 5$ rad/s, the quadruple of complex eigenvalues reaches the imaginary axis at imaginary value $\approx \pm 5j$.

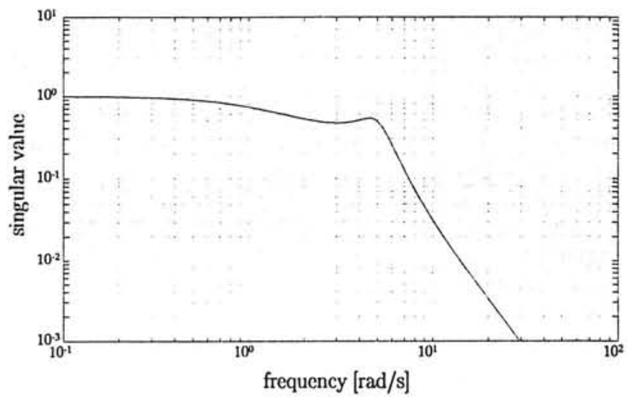


Fig. 3. Singular values of third-order system (7)

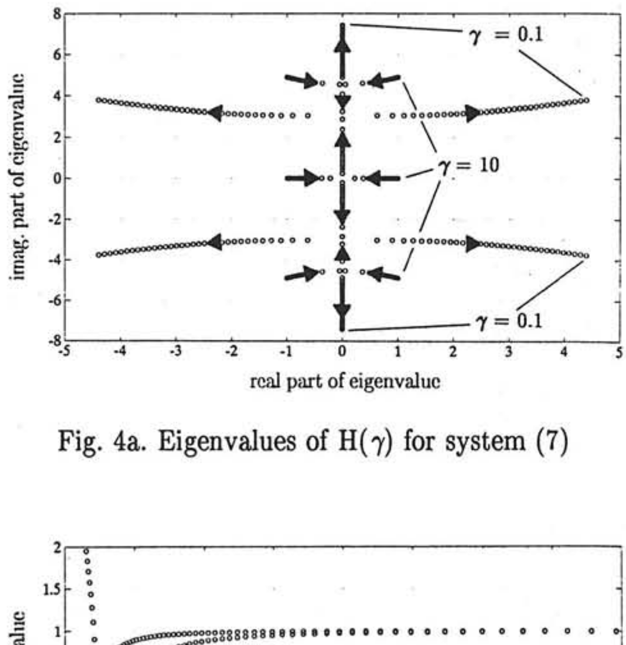


Fig. 4a. Eigenvalues of $H(\gamma)$ for system (7)

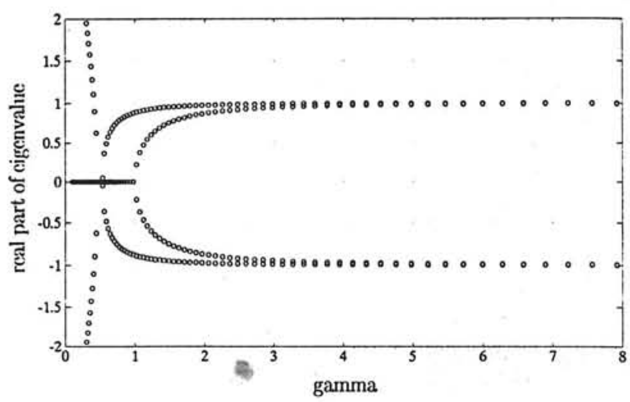


Fig. 4b. Real part of the eigenvalues as a function of γ.

Example 3.
As a third example we take the non-strictly proper system

$$G(s) = \begin{bmatrix} K_1(\tau_1 s+1)/(\tau_2 s+1) & 0 \\ 0 & K_2/(\tau_3 s+1) \end{bmatrix} \quad (8)$$

with K_1=5, K_2=0.5, τ_1= 1s, τ_2=5s, τ_{3_3}=1s

The singular value plot of this system (Fig. 5) demonstrates why in Cor. 2.1 in the right hand term γ must be larger than $\sigma_{max}(D)$. For a non strictly proper system not all singular values go to zero for $\omega \to \infty$. Because of this there may be values for $\gamma < \|G\|_\infty$ where the singular value plot is not intersected, and for which $H(\gamma)$ will not have imaginary eigenvalues. In Fig. 6 the real part of the eigenvalues as a function of γ is plotted, showing that for $0.5 < \gamma < 1$ there are no eigenvalues on the imaginary axis.

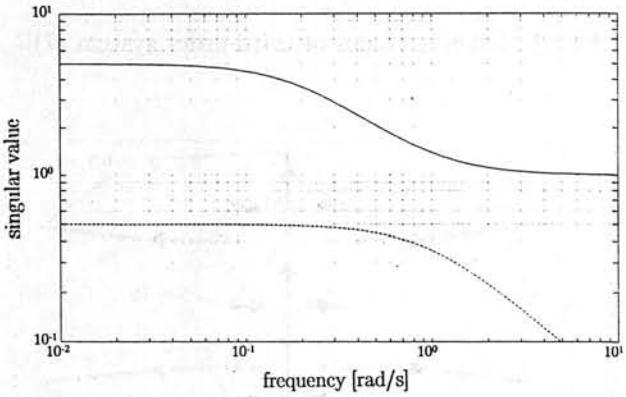


Fig. 5. Singular values for non strictly proper system (8)

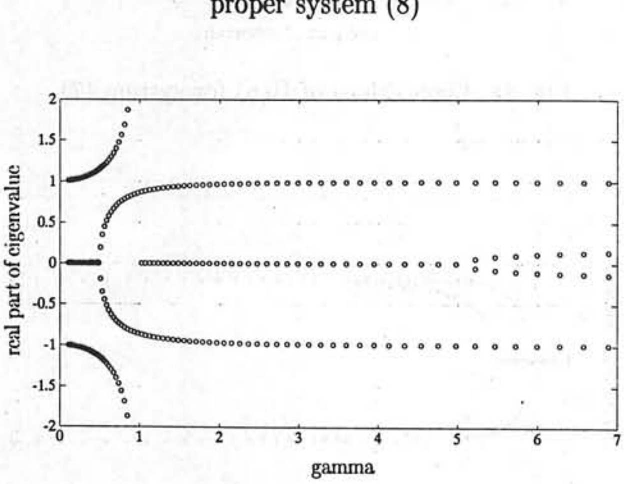


Fig. 6. Real part of the eigenvalues of $H(\gamma)$ as a function of γ for system (8).

Other examples have shown that the eigenvalues of $H(\gamma)$ do not necessarily all move towards the imaginary axis when decreasing γ from infinity. Some of them can move away from it, or for instance move towards the real axis.

3. THREE ALGORITHMS

3.1. Bisection algorithm

Using Cor. 2.1., the H_∞-norm of a system can be approximated with a simple bisection algorithm as described in (Boyd et al., 1988,1989) and (Robel, 1989). A starting interval $[\gamma_{lb}(0), \gamma_{ub}(0)]$ is determined (see § 3.4.) that certainly contains the H_∞-norm, and this interval is reduced by bisection until the required accuracy, specified by the maximum relative error ϵ, is achieved.

algorithm:
- compute lower and upper bound starting values γ_{lb} and γ_{ub}
- *repeat* until 'break'
  - $\gamma = 0.5\cdot(\gamma_{lb} + \gamma_{ub})$
  - compute the eigenvalues of $H(\gamma)$ (3)
  - *if* no imaginary eigenvalues
    $\gamma_{ub} = \gamma$
    else
    $\gamma_{lb} = \gamma$
  - *if* $\gamma_{lb} - \gamma_{ub} \leq 2\cdot\epsilon\cdot\gamma_{ub}$, break
- $\|G\|_\infty = 0.5\cdot(\gamma_{lb} + \gamma_{ub})$

3.2. Algorithm using eigenvalue derivatives

The derivatives of the Hamiltonian eigenvalues with respect to γ can be used to write an algorithm that converges in less steps than the bisection algorithm (Bruinsma, 1990). The eigenvalue derivatives can be computed with the next proposition (for derivation see (Rogers, 1970)).

Proposition 3.1. Let $A(\gamma)$ be a differentiable matrix function of γ with n distinct eigenvalues $\lambda_1(\gamma)$ to $\lambda_n(\gamma)$, then

$$\frac{d\lambda_i(\gamma)}{d\gamma} = y_i'(\gamma)\cdot\frac{dA(\gamma)}{d\gamma}\cdot x_i(\gamma) \quad (9)$$

where y_i is the 'left eigenvector' and x_i the 'right eigenvector' related to λ_i:

$y_i'A = \lambda_i y_i'$
$Ax_i = \lambda_i x_i$

with y_i scaled such that $y_i'\cdot x_i = 1$.

It follows from (3) that

$$\frac{dH(\gamma)}{d\gamma} = \quad (10)$$
$$= \begin{bmatrix} -2\gamma BR^{-2}D'C & -B(R^{-1}+2\gamma^2R^{-2})B' \\ C'(S^{-1}+2\gamma^2S^{-2})C & 2\gamma C'DR^{-2}B' \end{bmatrix}$$

with R and S as in (3).

Assuming that $H(\gamma)$ (3) has distinct eigenvalues, we can use Prop. 3.1. to compute the eigenvalue derivatives of $H(\gamma)$ for some upper bound $\gamma > \|G\|_\infty$. With the real part of the derivative we make an estimation of when the real part of the eigenvalue will become zero.

algorithm:

- compute upper and lower bound starting values γ_{ub} and γ_{lb}
- $\gamma = \gamma_{ub}$
- *repeat* until 'break'
  - compute the eigenvalues λ_i of $H(\gamma)$ and the eigenvalue derivatives der_i (using (9))
  - *if* no imaginary eigenvalues,
    $\gamma_{ub} = \gamma$
    $\text{step} = \min_i\{|\rho_1 \cdot \mathrm{Re}(\lambda_i) \cdot \mathrm{Re}(der_i)|\}$
    $\gamma = \gamma_{ub} - \text{step}$
    else
    $\gamma_{lb} = \max(\gamma_{lb}, \gamma)$
    $\gamma = \gamma_{ub} - \rho_2 \cdot \text{step}$
  - *if* $\gamma_{lb} - \gamma_{ub} \leq 2 \cdot \epsilon \cdot \gamma_{ub}$, break
- $\|G\|_\infty = 0.5 \cdot (\gamma_{lb} + \gamma_{ub})$

Experience has shown that appropriate choices for the multiplication factors ρ are

$\rho_1 = 0.6$
$\rho_2 = 0.8$

3.3. Two step algorithm on the lower bound

The algorithm described here approximates the H_∞-norm using only a lower bound. Some lower bound starting value is computed, and in an iteration loop the lower bound is increased until the required accuracy is achieved. Within each iteration two steps lead to the next γ_{lb}. In step 1 we use Prop. 2.1. to compute the frequencies corresponding to γ_{lb}.

For a description of the algorithm in detail we refer to (Bruinsma and Steinbuch, 1990). Local quadratic convergence of the algorithm has been proven by Boyd and Balakrishnan (1990). Here we only describe the main characteristic of the algorithm: the two steps to compute, given some lower bound $\gamma_{lb}(i)$, the next lower bound $\gamma_{lb}(i+1)$ (see Fig. 7.).

step 1:
Compute the frequencies ω_1 to ω_k corresponding to lower bound $\gamma_{lb}(i)$, using an eigenvalue computation of Hamiltonian matrix $H(\gamma_{lb}(i))$ (Prop. 2.1.)

step 2:
Take frequencies m_1 to m_{k-1} with $m_i = 0.5 \cdot (\omega_i + \omega_{i-1})$, compute the singular values of $G(jm_i)$ and take as new lower bound:
$\gamma_{lb}(i+1) = \max_i\{\sigma_{max}(G(jm_i))\}$.

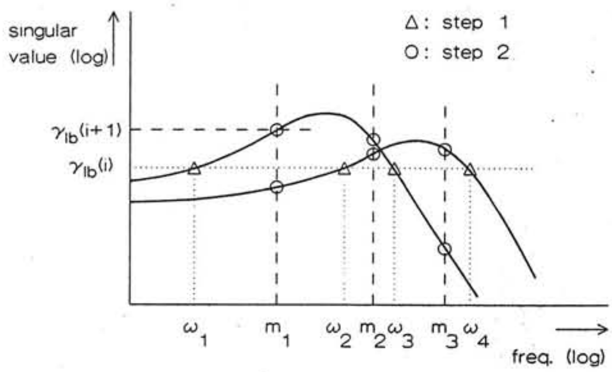


Fig. 7. Two steps to compute the next value $\gamma_{lb}(i+1)$

3.4. Comparison of the algorithms

The bisection algorithm and the algorithm using eigenvalue derivatives only use Cor. 2.1 to compute the H_∞-norm, using the Hamiltonian $H(\gamma)$ to search for the highest value γ for which there are imaginary eigenvalues. The two-step algorithm also uses Prop. 2.1, thus fully exploiting the relation between the imaginary eigenvalues of $H(\gamma)$ and the singular values.

The different algorithms will be compared by computing the H_∞-norm for some examples.
The lower and upper bound starting values γ_{lb} and γ_{ub} for the algorithms, used in these examples are:

$$\gamma_{ub} = 2 \cdot \Sigma\, \sigma_H\{G(s)\} + \sigma_{max}(D) \tag{11}$$

with σ_H = Hankel singular values

$$\gamma_{lb} = \max\{\sigma_{max}(G(0)), \sigma_{max}(D)\} \tag{12}$$

For a derivation of (11) see (Glover, 1984). For systems of high order n the computation of this upper bound takes relatively much time, because it requires the solution of two Lyapunov equations with dimension n.
Instead of lower bound (12) also other expressions are possible. A simple and effective lower bound is presented in (Bruinsma and Steinbuch, 1990). For the comparison of the three algorithms it is not relevant which lower bound we use, and (12) will do quite well.

As examples we take three systems:

- the first example from § 2.2,
- a 4th order system with 3 inputs and 2 outputs, with a random generated state-space matrix,
- a 13th order model of a wind energy conversion system, with 10 inputs and 10 outputs, extracted from (Steinbuch, 1989).

The H_∞-norms of these three systems have been computed with the three different algorithms. In

Table 1 the number of iteration steps are given, necessary to compute the H∞-norm with maximum relative error 10^{-6}, plus the computing time for the complete algorithm (on a 12 MHz AT).

TABLE 1 H∞-computation with rel. error 10^{-6}

		bisection	derivatives	two-step
example 1 3rd order 1x1	steps	14	9	2
	time [s]	12.9	15.0	3.2
random 4th order 2x3	steps	14	8	1
	time [s]	18.1	19.8	2.6
wind turb. 13th order 10x10	steps	16	11	4
	time [s]	331	421	108

The number of steps for the bisection algorithm is determined by the length of the starting interval and the maximum relative error.
The use of eigenvalue derivatives can substantially reduce the number of steps, but due to the more complicated computations within each step the complete algorithm is even slower. It needs to be said that the algorithm using derivatives can be improved with respect to numerical efficiency. Also, this algorithm should have a close upper bound starting value, otherwise the steps might be much too large initially (see for example Fig. 2b for large γ). The algorithm might become faster if the first steps are taken with the simpler bisection algorithm.
For all three examples the two-step algorithm is much faster than the other algorithms and needs only a few steps to achieve the required accuracy.

CONCLUSIONS

Three algorithms have been described, all based on the relationship between the H∞-norm of a transfer function and the eigenvalues of an Hamiltonian matrix. Both the bisection and the algorithm using eigenvalue derivatives use upper and lower bounds. The two-step algorithm uses only a lower bound, and is based on an alternating calculation of eigenvalues and singular values. Using numerical examples it has been shown that the two-step algorithm is superior, with respect to both the number of iterations and the calculation time.

ACKNOWLEDGEMENTS

The authors would like to thank O.H. Bosgra for his assistance and encouragement, P. van Dooren for his help on numerical aspects, and J. Ludlage for suggesting the problem.

REFERENCES

Boyd, S., V. Balakrishnan and P. Kabamba (1988). On computing the H∞-norm of a transfer matrix. *Proc. 1988 Amer. Contr. Conf.*, 396–397.
Boyd, S., V. Balakrishnan and P. Kabamba (1989). A bisection method for computing the H∞-norm of a transfer function matrix and related problems. *Mathematics of Control, Signals and Systems, 2*, 207–219.
Boyd, S. and V. Balakrishnan (1990). A regularity result for the singular values of a transfer function matrix and a quadratically convergent algorithm for computing the L∞-norm. *SIAM J. on Matrix Analysis and Applications*, (to appear).
Bruinsma, N.A. and M. Steinbuch (1990). A fast algorithm to compute the H∞-norm of a transfer function matrix. *Syst. Control Lett.*, (to appear).
Bruinsma, N.A. (1990). *On Hamiltonians, H∞-norms and optimal H∞-designs*, MSc. Thesis, Lab. Measurement and Control, Dept. Mechanical Engineering and Marine Techn., Delft University of Technology, (to appear).
Doyle, J., K. Glover, P. Khargonekar and B. Francis (1989). State-space solutions to standard H_2 and H∞ control problems. *IEEE Trans. Automat. Contr., 34*, 831–847.
Francis, B.A. (1987). *A Course in H∞ Control Theory*, Springer–Verlag.
Glover, K. (1984). All optimal Hankel-norm approximations of linear multivariable systems and their L∞-error bounds. *Int. J. of Control, 39*, 1115–1193.
Robel, G. (1989). On computing the infinity norm. *IEEE Trans. on Automat. Contr., 34*, 882–884.
Rogers, L.C. (1970). Derivatives of eigenvalues and eigenvectors. *AIAA Journal, 8*, 943–945.
Scherer, C. (1989). H∞-control by state feedback: An iterative algorithm and characterization of high-gain occurence. *Syst. Control Lett., 12*, 383–391.
Steinbuch, M. (1989) *Dynamic Modelling and Robust Control of a Wind Energy Conversion System*, Doctoral Dissertation, Delft University of Technology.

Selected Topics in Identification,
Modelling and Control; Vol. 1, 1990

Robust stability analysis of a flexible mechanism assuming real or complex parametric uncertainty

Jan C. Terlouw and Sven G. Smit

Lab. Measurement and Control, Dept. Mechanical Engineering and Marine Techn., Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands and Philips Research Laboratories, P.O.Box 80.000, 5600 JA, Eindhoven, The Netherlands

Abstract–For a controlled flexible mechanism the stability robustness is analyzed. A gain and spring stiffness variation are modelled as parametric uncertainty, which typically is a structured perturbation. Stability margins are then computed using singular value robustness analysis, complex structured singular value robustness analysis and real structured singular value robustness analysis. Comparing these results with a crude eigenvalue computation shows that only the real case leads to exact stability margins.

Key Words–Real and complex structured singular value; parametric uncertainty; stability robustness

INTRODUCTION

Robustness analysis and robust design of control systems has gained much attention in systems and control literature (Doyle, 1982; Maciejowski, 1989). Especially H_∞ methods have been propagated as a tool for robust controller design (Kwakernaak, 1983; Francis, 1988). H_∞ design yields stability margins for norm–bounded unstructured complex perturbations. However, in practice perturbations are often structured and real (i.e. parametric uncertainty). This may lead to very conservative designs. Therefore, in H_∞ design only the most important uncertainty can be taken into account (Smit, 1990), leaving a necessity for robustness analysis afterwards for a more realistic set of perturbations to compute the actual stability margins.

In this paper we investigate the robust stability of a simple flexible mechanism controlled by a H_∞ controller. In this H_∞ design the only and most important perturbation taken into account is a varying spring stiffness. However, since gain variations always occur in practice it is necessary to analyze the stability margin of the controlled system with respect to both spring stiffness and gain variations. This will be done using three robust stability analysis tests:

1. Singular Value Robustness Analysis (SVRA), (Doyle and Stein, 1981).
2. Complex Structured Singular Value Robustness Analysis (CSSVRA), (Doyle, 1982).
3. Real Structured Singular Value Robustness Analysis (RSSVRA), (Fan et al., 1990).

Our aim is to compare the stability margins obtained by these three methods. Because the combined perturbation of gain and spring stiffness is structured and real it can be expected that the system is best analyzed using RSSVRA, where "best" means the least conservative.

This paper is devided into three parts. In the first section the system is described and the uncertainty model is derived. The second section is devoted the stability analysis. A stability region for the uncertainty is determined by doing a crude closed loop pole computation of the system perturbed by the two varying parameters. Then SVRA, CSSVRA and RSSVRA are applied. The conservatism of the three methods is then evaluated by comparing the results with the computed stability region. Finally the conclusions are presented.

UNCERTAINTY MODELLING OF A FLEXIBLE MECHANISM

Robust stability analysis with structured singular values requires uncertainty modelling. The aim is to arrive at a specific representation of the perturbed closed loop system. This representation is called the interconnection structure (Doyle, 1982) and has all uncertainty collected in a block diagonal feedback matrix (see Fig. 2). In the next section the interconnection structure that will be derived in the following is needed to analyse our example system on its closed loop stability. The uncertainty model of the flexible mechanism has been derived using a parametric uncertainty modelling procedure on state space level described in (Steinbuch, 1989; Terlouw, 1990). Parametric uncertainty modelling is based on the following general uncertainty representation for a plant in state space

$$\begin{aligned} \dot{x} &= Ax + Bu + dAx + dBu \\ y &= Cx + Du + dCx + dDu \end{aligned} \tag{1}$$

In this equation A, B, C, and D are the nominal state space matrices, while dA, dB, dC and dD are perturbation matrices containing information on the variations in the entries of the state space matrices. In order to apply the robustness analysis methods it is necessary to rewrite these equations into a standard form (Doyle, 1982). The parametric variations occuring in (1) must be collected in a diagonal feedback perturbation matrix $\Delta = \mathrm{diag}(\delta_1, \delta_2, \ldots, \delta_n)$ replacing dA, dB, dC and dD. This requires a reformulation of equation (1):

$$\begin{aligned} \dot{x} &= Ax + Bu + B_2u_2 \\ y &= Cx + Du + D_{12}u_2 \\ y_2 &= C_2x + D_{21}u + D_{22}u_2 \\ u_2 &= \Delta y_2 \end{aligned} \tag{2}$$

The following equalities must be satisfied to guarantee the equivalence of (1) and (2):

$$dA = B_2(I - \Delta D_{22})^{-1}\Delta C_2 \tag{3}$$
$$dB = B_2(I - \Delta D_{22})^{-1}\Delta D_{21} \tag{4}$$
$$dC = D_{12}(I - \Delta D_{22})^{-1}\Delta C_2 \tag{5}$$
$$dD = D_{12}(I - \Delta D_{22})^{-1}\Delta D_{21} \tag{6}$$

Equations (3–6) determine the constraints on weighting matrices B_2, C_2, D_{12}, D_{21} and D_{22} and the perturbation Δ.

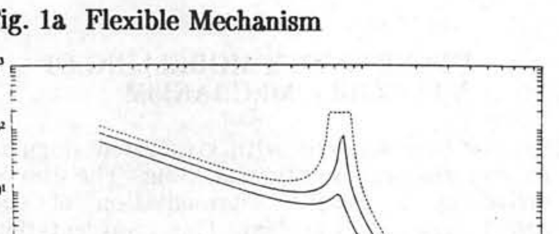


Fig. 1a Flexible Mechanism

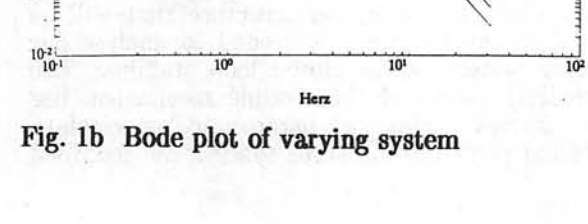


Fig. 1b Bode plot of varying system

This uncertainty modelling procedure will be used in the following to derive a model for a flexible mechanism. The system under consideration is a flexible shaft connected to two rotating masses (see Fig. 1a). One of the two (m_1) is a current driven motor, with J_1 and D_1 the inertia and damping of the motor. The relation between torque M and current I is modelled as

$$M = KI \tag{7}$$

with the real scalar K denoting the motor gain. The other mass (m_2) is beared, and modelled by the damping coefficient D_2 and inertia J_2. The flexible shaft in between has a spring stiffness C_s and a damping coefficient D_s and a neglectable mass. The aim is to control the rotational speed of the second mass (m_2). So the rotational speed is controlled by the DC–motor through the flexibility of the shaft. The goal is to achieve a closed loop bandwidth up to the resonance frequency of the shaft in spite of a varying spring stiffness C_s and a varying gain K (see Fig. 1b for the effects of these variations on the open loop behaviour). Define the state vector as $x = [\dot{\varphi}_1 \; \dot{\varphi}_2 \; \psi]'$, with $\dot{\varphi}_1$ the rotational speed of mass m_1, $\dot{\varphi}_2$ the rotational speed of mass m_2 and $\psi = \varphi_1 - \varphi_2$, then the state space matrices of the equations of motion yield:

$$A = \begin{bmatrix} -(D_1+D_s)/J_1 & D_s/J_1 & C_s/J_1 \\ D_s/J_2 & -(D_2+D_s)/J_2 & -C_s/J_2 \\ -1 & -1 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} K/J_1 \\ 0 \\ 0 \end{bmatrix}$$

$$C = [\,0 \quad 1 \quad 0\,]$$

$$D = [\,0\,]$$

We assume that the spring stiffness C_s and the motor gain K can vary:

$$C_s \in [C_{min}, C_{max}]$$
$$K \in [K_{min}, K_{max}]$$

Choose

$$\begin{aligned} C_{nom} &= (C_{min} + C_{max})/2 \\ K_{nom} &= (K_{min} + K_{max})/2 \\ \Delta C &= (C_{max} - C_{min})/2 \\ \Delta K &= (K_{max} - K_{min})/2 \end{aligned}$$

then

$$\begin{aligned} C_s &= C_{nom} + \Delta C \\ K &= K_{nom} + \Delta K \end{aligned}$$

Using description (1) to seperate the actual variations ΔC and ΔK from the nominal values C_{nom} and K_{nom}, the following perturbation matrices are obtained.

$$dA = \begin{bmatrix} 0 & 0 & -\Delta C/J_1 \\ 0 & 0 & \Delta C/J_2 \\ 0 & 0 & 0 \end{bmatrix}$$

$$dB = \begin{bmatrix} \Delta K/J_1 \\ 0 \\ 0 \end{bmatrix}$$

$$dC = [\ 0\ 0\ 0]$$

$$dD = [0]$$

In order to be able to apply the robustness tests in following section on a transformation with the constraints of (3–6) has to be carried out:

$$B_2 = \begin{bmatrix} -1 & 1 \\ J_1/J_2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$C_2 = \begin{bmatrix} 0 & 0 & 1/J_1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$D_{12} = [\ 0\ 0\ ]$$

$$D_{21} = \begin{bmatrix} 0 \\ 1/J_1 \end{bmatrix}$$

$$D_{22} = [0]$$

$$\Delta = \begin{bmatrix} \Delta C & 0 \\ 0 & \Delta K \end{bmatrix}$$

Note that the uncertainty matrix Δ has indeed a diagonal structure. In (Smit, 1990) a H_∞ controller has been designed to control the output rotational speed. In this design, which uses the uncertainty modelling described above, only the spring stiffness variation is taken into account (ΔK=0). The controller has the state–space realization:

$$\dot{p} = Ep + Fy \tag{8}$$
$$u = Gp$$

It is a fourth order controller designed accounting for a spring stiffness variation of 1/3 of its nominal value. Using this controller and the uncertainty model of the plant derived above, the interconnection structure (Doyle, 1982) can be derived and is shown in Fig. 2, with

$$M(s) = C_m(sI - A_m)^{-1}B_m \tag{9}$$

where

$$A_m = \begin{bmatrix} A & -BG \\ FC & E \end{bmatrix}$$

$$B_m = \begin{bmatrix} W_1 \\ 0 \end{bmatrix}$$

$$C_m = [\ V_1\ \ V_2G\ ]$$

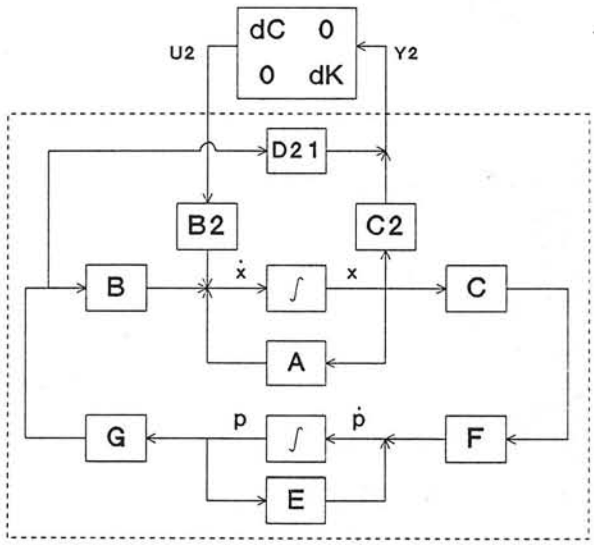


Fig. 2a Interconnection structure for ΔC and ΔK (Time domain)

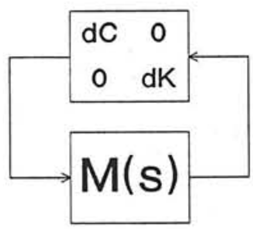


Fig. 2b Interconnection structure for ΔC and ΔK (Frequency domain)

ROBUSTNESS ANALYSIS

Before applying the three robustness analysis methods mentioned in the introduction, an exact reference for the robust stability problem is obtained by a crude eigenvalue computation. In the nominal case the closed loop matrix of the controlled system is equal to A_m in (9). Computing the eigenvalues of A_m for the varying gain and spring stiffness leads to the stability region $S = \{(C_s,K) |\ real(eig(A_m(C_s,K)))<0\}$. A part of the region S is shown in Fig. 3 below.

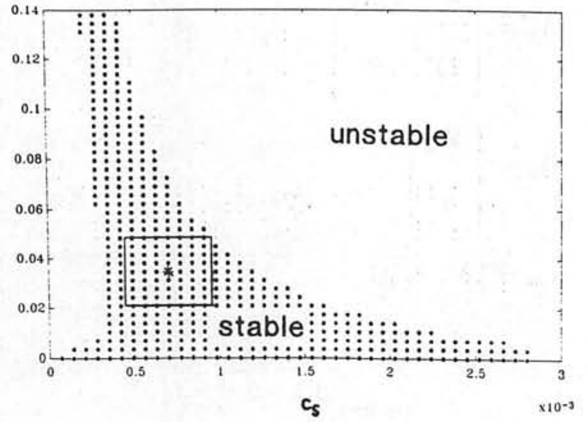


Fig. 3 Stability region S for for varying C_S and K

In (Smit, 1990) it is shown that the closed loop system has acceptable performance properties for a varying spring stiffness of about 1/3 of its nominal value and a nominal motor gain. Analysis afterwards learns that the motor gain may increase with about 40% of its nominal value before instability occurs. This can be seen by considering the box in Fig. 3 which represents all allowable K if ΔC is chosen as in the H_∞ design problem, since for one combination (C_S,K) the system becomes unstable (the box touches the instability region in the upper right corner). The cross in the middle of the box represents the nominal values of C_S and K. The maximum admissable ΔC and ΔK given by the box are absorbed in the interconnection structure M(s) so that the diagonal perturbation matrix in Fig. 2b is scaled to a 2 by 2 identity matrix who's elements can vary between −1 and +1.

Now the three robustness analysis tests can be applied to M(s). The theorems given below are stated by *computable* upperbounds and not by the exact definitions. They are based on the requirement that the loop gain of the interconnection structure (Fig. 2) remains smaller than one for all possible Δ, since ΔC and ΔK are scaled to one. Since it is the goal of this work to show computational *results*, theoretical background on the bounds and computational aspects will be ommited here. (Fan et al., 1990) provides an excellent explanatory text for the interested reader.

Theorem SVRA (Doyle and Stein, 1981)
Robust stability if

$$\overline{\sigma}(\omega) = \overline{\sigma}[M(j\omega)] < 1 \ \forall\omega$$ ■

Theorem CSSVRA (Doyle, 1982)
Robust stability if

$$\mu_c(\omega) = \min_{D}\{\overline{\sigma}[DM(j\omega)D^{-1}]\} < 1 \ \forall\omega$$

where D is a block–diagonal matrix according to the structure of Δ ■

Theorem RSSVRA (Fan et al., 1990)
Robust stability if

$$\mu_r(\omega) = \min_{D,G}\{\overline{\lambda}\{D^{-1}M(j\omega)^* D^2 M(j\omega)D^{-1} + j[GM(j\omega)^* - M(j\omega)G]\} < 1 \ \forall\omega$$

where D and G are block–diagonal matrices according to the structure of Δ. ■

with $\overline{\sigma}$ denoting the maximum singular value, $\overline{\lambda}$ denoting the largest eigenvalue, * denoting the complex conjugate transpose of a matrix.

The largest singular value denotes the largest gain of M. Since the largest gain of the perturbation Δ is less than one the loop gain $M\Delta$ of the interconnection structure should not exceed 1 and instability thus does not occur if $\overline{\sigma}[M(j\omega)] < 1 \ \forall\omega$. However if Δ is structured as in our example the largest singular value of M may be scaled and thereby minimized by a matrix D according to the structure of Δ.

If Δ is structured and real an additional scaling of the "imaginary part" of M may be applied resulting in a minimization over D, G as in theorem RSSVRA.

Remark 3.1.
For the assumed ΔC and ΔK the closed loop system can reach the edge of stability. Therefore the three tests above should be less than or equal to 1. If the peak value over all frequencies of $\overline{\sigma}(\omega)$, $\mu_c(\omega)$ and $\mu_r(\omega)$ is larger than 1 the tests state that the system is not robustly stable while it is and thus yield conservative results. ■

The first test is directly computable using standard software. The second test involves the minimization per frequency over a (block–) diagonal D and the third involves an optimization over a (block–) diagonal D and a (block–) diagonal G (an algorithm doing so has been programmed (Terlouw, 1990)). In Fig. 4 the results of the computations of the upperbounds given above are shown.

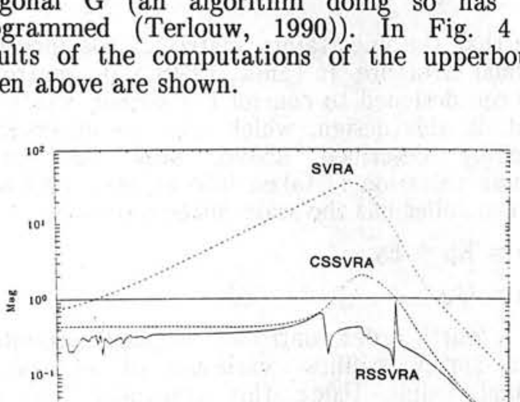


Fig. 4 Results for SVRA, CSSVRA and RSSVRA

Theorem SVRA holds for unstructured complex uncertainties and since the perturbations in this problem are structured *and* real the SVRA test is expected to be very conservative. This can be seen in Fig. 4 where $\overline{\sigma}[M(j\omega)]$ has a peak value of 35, implying that only an uncertainty 35 times smaller than the actual uncertainty would satisfy theorem SVRA. The CSSVRA takes the structure of the perturbations into account and therefore is less conservative: D–scaling reduces conservatism considerably (in this case the peak value equals 2). The RSSVRA–test computes an upper bound for structured *and* real perturbations. In Fig. 4 the computed upper bound equals 1 and hence is non–conservative.

Remark 3.2.
Our experience is that often at specific frequencies the real structured singular value $\mu_r(\omega)$ has peaks. An interpretation for scalar perturbations is the crossing of $M(j\omega)$ with the negative real axis in the complex plane (equivalent to the gain margin). The results of RSSVRA are not reliable if the computation is done with a frequency grid skipping such a crossing frequency. ■

CONCLUSIONS

In this paper the stability robustness of a flexible mechanism for gain and spring stiffness variations has been analysed. It has been shown that it is possible to isolate these variations in a diagonal feedback structure suitable for application of several robustness theorems.
Singular value robustness analysis, even for a simple 2 by 2 problem, can be extremely conservative. This has an implication for design too because this conservatism would yield very low performance H_∞ controllers.
Accounting for the structure of the perturbation reduces conservatism considerably, but still does not account for the real nature of perturbations. For the specific problem under investigation this conservatism can be completely removed by applying the computable upperbound of (Fan et al., 1990) for the real structured singular value.

REFERENCES

Doyle, J.C. (1982). Analysis of feedback systems with structured uncertainties, *IEE Proc. Pt.D.Control Theory and Applications*, 129 vol.6, pp. 242–250.
Doyle, J.C. and G. Stein (1981). Multivariable feedback design: Concepts for a classical/modern synthesis. *IEEE Trans. Automat. Contr.*, AC–26, pp. 4–16.
Fan, M.K.H., A.L. Tits and J.C. Doyle (1990). Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics. *IEEE Trans. Automat. Contr.*, to appear.
Francis, B. (1988). *A course in H_∞ Control Theory,* Lecture Notes Control and Information Sciences, Springer Verlag Berlin.
Kwakernaak, H. (1983). Robustness optimization of linear feedback systems. *Proc. IEEE Conf. on Decision and Contr.*, pp. 618–624.
Maciejowski, J.M. (1989). *Multivariable Feedback Design*, Addison Wesley
Smit, S.G. (1990). *MSc Thesis*, Delft University of Technology, under preparation.
Steinbuch, M. (1989). *Dynamic modelling and robust control of a wind energy conversion system, Dissertation.* Delft University of Technology.
Terlouw, J.C. (1990). *MSc thesis,* Delft University of Technology, under preparation.

Selected Topics in Identification,
Modelling and Control; Vol. 1, 1990

Discrete normalized coprime factorization and fractional balanced reduction

Peter M.M. Bongers and Peter S.C. Heuberger

Lab. for Measurement and Control, Dept. Mechanical Engineering and Marine Techn.
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

Abstract. In this paper a reliable algorithm is developed to perform a normalized coprime factorization of proper discrete time finite dimensional linear time invariant systems. Instead of using the bilinear transform the factorization is calculated directly. The system is allowed to have a singular state–space matrix. It is shown that a modified discrete time Riccati equation plays a crucial role to obtain a state–space realization for the factorization. One of the applications of the normalized coprime factorization is in model reduction. In the fractional balanced reduction of a plant a normalized coprime factorization is used. An algorithm is presented to obtain a discrete fractional balanced reduced plant model.

Keywords. Proper discrete time systems; normalized coprime factorizations; fractional balanced reduction; reliable algorithm.

INTRODUCTION

In the theoretical work of Desoer et al. (1980), Vidyasagar et al.(1982), Vidyasagar (1984) the benefits of using coprime representations in stability analysis of controlled systems are shown. In the continuous time case Nett et. al. (1984), Meyer and Franklin (1988) and Vidyasagar (1988) derived state–space representations for the normalized coprime factors. Glover and McFarlane (1988, 1989), McFarlane (1988) showed the importance of normalized coprime factors in the H_∞ controller design. They explicitly solved a continuous time four block H_∞ control problem by using a normalized coprime representation of the plant. In practical situations a continuous time plant is controlled by a discrete controller using sampling and zero order hold circuits. So in order to design accurate discrete time controllers the control design procedure has to be performed in discrete time. The first step in discrete H_∞ control design with normalized coprime factors is to establish whether or not in discrete time normalized coprime factors exist and can be represented in state–space forms. Chu (1988) gave state–space representations for discrete coprime factors with an inner numerator under the condition that the system has no poles in the origin. Poles in the origin are of major importance since in discrete time very often time–delays have to be incorporated in the system model. In this paper we will show the existence of a normalized coprime factorization of a discrete time plant with possibly poles in the origin.

PRELIMINARIES

Stable multivariable linear systems can be studied by considering them as transfer function matrices having all entries belonging to a ring $\mathcal{H}$. Moreover, in many cases (e.g. convolution operators) the ring $\mathcal{H}$ is commutative and is an integral domain (i.e. $\mathcal{H}$ has no divisors of zero). The class of possibly unstable systems are elements of the quotient field $\mathcal{F}$ of $\mathcal{H}$. Throughout this paper we let (Vidyasagar et.al. 1982, Desoer et.al. 1980):

- $\mathcal{F} := \{ a/b \mid a \in \mathcal{H},\ b \in \mathcal{H}\backslash 0 \}$, a quotient field of $\mathcal{H}$
- $\mathcal{G} :=$ a (not necessarily commutative) ring with identity.
- $\mathcal{H} :=$ a subring of $\mathcal{G}$ which includes identity
- $\mathcal{I} := \{ h \in \mathcal{H} \mid h^{-1} \in \mathcal{G} \}$, the set of multiplicative units of $\mathcal{G}$
- $\mathcal{J} := \{ h \in \mathcal{H} \mid h^{-1} \in \mathcal{H} \}$, the set of multiplicative units of $\mathcal{H}$

Note that: $\mathcal{J} \subset \mathcal{I} \subset \mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ (1)

In the sequel of this paper we will study real rational finite dimensional discrete time invariant systems. The ring $\mathcal{G}$ is identified with $\mathbb{RL}_\infty$ the space of proper real–rational functions with no poles on the unit circle with norm $\|.\|_\infty$:

$$\|f(z)\|_\infty = \sup_{0\leq\theta\leq 2\pi} \bar{\sigma}[f(e^{j\theta})] \tag{2}$$

The subring $\mathcal{H}$ is identified with $\mathbb{RH}_\infty$ the subspace of $\mathbb{RL}_\infty$ with no poles outside the open unit disk, and analogously $\mathcal{X}^{\perp}$ is identified with $\mathbb{R}H_{\infty}^{\perp}$. The following notation is used. We will denote transfer functions as G(z) or if there is no confusion G. With a slight abuse of notation a transfer function is given by:

$$G(z) := D + C(zI-A)^{-1}B := \left[\begin{array}{c|c} zI-A & B \\ \hline -C & D \end{array}\right] \qquad (3)$$

A^t denotes the transpose of A and $G^*(z)$ denotes $G^t(z^{-1})$. For minimal plants $G(z) \in \mathcal{X}$, the controllability and observability Grammians P respectively Q are positive definite symmetric solutions of the following Lyapunov equations:

$$APA^t + BB^t = P \qquad (4)$$

$$A^tQA + C^tC = Q \qquad (5)$$

Definition 2.1 (Vidyasagar, 1984; Huang and Liu, 1987)
A plant $G \in \mathcal{F}$ has a right (left) fractional representation if there exist N,M $(\tilde{N},\tilde{M}) \in \mathcal{X}$ such that:

$$G = NM^{-1} \,(= \tilde{M}^{-1}\tilde{N}) \qquad (6)$$

Furthermore we say that the pair M,N $(\tilde{M},\tilde{N})$ is right (left) coprime (rcf or lcf) if there exists U,V $(\tilde{U},\tilde{V}) \in \mathcal{X}$ such that:

$$UN + VM = I \;\; (\tilde{N}\tilde{U} + \tilde{M}\tilde{V} = I) \qquad (7)$$

The pair M,N $(\tilde{M},\tilde{N})$ is normalized right (left) coprime (nrcf or nlcf) if in addition to (6):

$$M^*M + N^*N = I \;\;\; (\tilde{M}\tilde{M}^* + \tilde{N}\tilde{N}^* = I) \qquad (8)$$

□

Meyer and Franklin (1988) gave an explicit method to calculate the normalized right coprime factorization in continuous time. For the discrete time domain, the following proposition gives conditions for a state-space realization of inner transfer functions.

Proposition 2.1 (Heuberger 1990)
A plant $G(z) := D + C(zI-A)^{-1}B \in \mathcal{F}$ is called inner: $G^t(z^{-1})G(z) = I$, if and only if there exist a Q such that:

a) $A^tQA + C^tC = Q, \; Q = Q^t$ (9a)

b) $D^tD + B^tQB = I$ (9b)

c) $C^tD + A^tQB = 0$ (9c)

□

We will show how normalized coprime factorizations can be applied to model reduction. For this purpose we define in the following proposition the graph of a transfer function and show how this concept is related to coprime factorizations.

Proposition 2.2 (Vidyasagar 1985)
All those input-output pairs that are of finite energy define the graph of a plant G(z):

$$\mathcal{G}\{G(z)\} = \{(u,y) \in L_2 \times L_2 \mid y = Gu\}$$

The graph of G(z) can also be expressed in terms of its rcf. Let $(N,M) \in \mathcal{X}$ be a rcf of $G \in \mathcal{F}$, then the Graph of G equals:

$$\mathcal{G}\{G(z)\} = \{\begin{bmatrix} M \\ N \end{bmatrix} w \mid w \in L_2\}$$

□

NORMALIZED COPRIME FACTORIZATION

The following theorem gives sufficient conditions for the existence of a state-space representation of a normalized right coprime factorization of a discrete time plant. In the proof we will frequently use system equivalent operations, described by Rosenbrock (1970).

Theorem 1
Given a minimal realization:

$$G(z) := \left[\begin{array}{c|c} zI-A & B \\ \hline -C & D \end{array}\right] \in \mathcal{F} \qquad (10)$$

and define:

$$\begin{bmatrix} M \\ N \end{bmatrix} = \left[\begin{array}{c|c} zI-A+BF & BH \\ \hline F & H \\ -C+DF & DH \end{array}\right] \qquad (11)$$

then $\begin{bmatrix} M \\ N \end{bmatrix}$ is a normalized right coprime factorization of G(z) if and only if there exist an F, H, Q such that:

a) $F^t = (A^tQB+C^tD)(I+D^tD+B^tQB)^{-1}$ (12a)

b) $HH^t = (I+D^tD+B^tQB)^{-1}$ (12b)

c) $Q-A^tQA-C^tC+(A^tQB+C^tD)\cdot(B^tQB+D^tD+I)^{-1}(B^tQA+D^tC) = 0$ (12c)

d) $Q = Q^t > 0$ (12d)

Proof
Bongers and Heuberger (1990).

□

The procedure to obtain a normalized right coprime factorization for the plant G is to solve the Riccati equation (12c,d) to obtain Q, calculate F and choose an H. The equivalent for the normalized left coprime factorization is a direct result from theorem 1 and is given in the following corollary.

Corollary 1
Given a minimal realization (10):

$$G(z) := D+C(zI-A)^{-1}B := \left[\begin{array}{c|c} zI-A & B \\ \hline -C & D \end{array}\right] \in \mathcal{F}$$

and define:

$$[\tilde{M}\;\tilde{N}] = \left[\begin{array}{c|cc} zI-A+KC & K & -B+KD \\ \hline RC & R & RD \end{array}\right] \qquad (13)$$

then $[\tilde{M}\ \tilde{N}]$ is a normalized left coprime factorization of G(z) if and only if there exist a K,R,P such that:

a) $K = (APC^t+BD^t)(I+CPC^t+DD^t)^{-1}$

b) $R^tR = (I+CPC^t+DD^t)^{-1}$

c) $P-APA^t- BB^t+(APC^t+BD^t)\cdot (I+CPC^t+DD^t)^{-1}(CPA^t+DB^t) = 0$

d) $P = P^t > 0$

PROOF

Let $G = \tilde{M}^{-1}\tilde{N}$, with $(\tilde{M},\tilde{N}) \in \mathcal{X}$ a nlcf of G, then $G^t = \tilde{N}^t\tilde{M}^{-t}$ with $(\tilde{N}^t,\tilde{M}^t)$ a nrcf of G^t, so the realization of $[\tilde{M}\ \tilde{N}]$ follows from theorem 1.

□

REMARK

Note that we don't need the assumption that the state matrix A is invertible, as is the case in Chu (1988). Using proposition 2.1 it is straight forward to show that this assumption is indeed superfluous. This is of major importance since in discrete time control design problems very often time-delays are incorporated in the augmented system.

In order to solve the normalized coprime factorization for the plant in discrete time by means of standard techniques, the equation (12c):

$$Q-A^tQA-C^tC+(A^tQB+C^tD)\cdot$$
$$(B^tQB+D^tD+I)^{-1}(B^tQA+D^tC) = 0$$

can be written as a standard Riccati equation. Define:

$$A_1 = \begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix},\ B_1 = \begin{bmatrix} B \\ D \end{bmatrix},\ C_1 = [0\ I],\ R_1 = I,$$

$$Q_1 = \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix}$$

The standard Riccati equation with A_1,B_1,C_1,R_1,Q_1 is :

$$0 = Q_1-A_1{}^tQ_1A_1-C_1{}^tC_1+$$
$$A_1{}^tQ_1B_1(B_1{}^tQ_1B_1+R_1)^{-1}B_1{}^tQ_1A_1$$

Substituting the definitions of A_1 etc. gives:

$$\begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} - \begin{bmatrix} A^t & C^t \\ 0 & 0 \end{bmatrix}\begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix} - \begin{bmatrix} 0 \\ I \end{bmatrix}[0\ I] +$$
$$\begin{bmatrix} A^t & C^t \\ 0 & 0 \end{bmatrix}\begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} B \\ D \end{bmatrix}\left([B^t\ D^t]\begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} B \\ D \end{bmatrix} + I\right)^{-1}\cdot$$
$$[B^t\ D^t]\begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

evaluating this equation gives (12c). This shows that the discrete time normalized coprime factorization problem can be solved by means of standard techniques. Note that the sufficient conditions for the existence of a positive solution of the Riccati equation are still valid.

FRACTIONAL BALANCED REDUCTION

In this section we extend the continuous time fractional balanced model reduction (FBR) method (Liu and Anderson 1986; Meyer 1988) to the discrete time domain. An essential part of the discrete fractional balanced reduction method is the existence of the discrete normalized right coprime factorization. A major advantage of the DFBR method is that plants with and without unstable poles are treated in the same way.

Instead of approximating the full order plant by a reduced order model in a straightforward way:

$$G_n(z) - G_r(z)$$

the graph of the plant is approximated:

$$\mathcal{G}\{G_n\} - \mathcal{G}(G_r\}$$

or:

$$\begin{bmatrix} M_n \\ N_n \end{bmatrix} - \begin{bmatrix} M_r \\ N_r \end{bmatrix}$$

with:

$$G_n = N_nM_n{}^{-1},\ (N_n,M_n) \text{ nrcf}$$

and we define:

$$G_r = N_rM_r{}^{-1}$$

The procedure to obtain a reduced order model in the graph sense is given in the next algorithm.

<u>Algorithm.</u> For a given plant $G_n(z) \in \mathcal{F}$ the construction of a low order approximation $G_r(z) \in \mathcal{F}$ is as follows:

1 Construct a nrcf (N_n,M_n) for the full order plant using eqn. 12.
2 Balance and order the state-space realization of (N_n,M_n):

$$\begin{bmatrix} M_n \\ N_n \end{bmatrix} = \left[\begin{array}{c|c} zI-A_n & B_n \\ \hline -C_n & D_n \end{array}\right]$$

such that:

$$P = Q = \text{diag}(\sigma_1 .. \sigma_r .. \sigma_n)$$

with:

$$\sigma_1 \geq \sigma_r \geq \sigma_n > 0,$$

and partition $\{A_n, B_n, C_n, D_n\}$ as follows:

$$A_n = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},\ B_n = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix},$$

$$C_n = [\ C_1\ C_2],\ D_n = D_n$$

with A_{11} of size rxr, A_{ij}, B_i,C_i of appropriate dimension.

3 The approximation of the coprime factors (N_n,M_n) by (N_r,M_r) is:

$$\begin{bmatrix} M_r \\ N_r \end{bmatrix} = \left[\begin{array}{c|c} zI-A_f & B_f \\ \hline -C_f & D_f \end{array}\right]$$

with:

$$A_f = A_{11} - A_{12}(I+A_{22})^{-1}A_{21}$$
$$B_f = B_1 - A_{12}(I+A_{22})^{-1}B_2$$
$$C_f = C_1 - C_2(I+A_{22})^{-1}A_{21}$$
$$D_f = D_n - C_2(I+A_{22})^{-1}B_2$$

4 Given the construction of a rcf of G and back substituting:

$$D_f = \begin{bmatrix} H_r \\ D_r H_r \end{bmatrix},\ C_f = \begin{bmatrix} -F_r \\ C_r - D_r F_r \end{bmatrix},$$

$$B_f = B_r H_r,\ A_f = A_r - B_r F_r$$

we obtain a state-space realization of the reduced order plant $G_r(z) \in \mathcal{F}$:

$$G_r = \left[\begin{array}{c|c} zI-A_r & B_r \\ \hline -C_r & D_r \end{array}\right]$$

□

Note that although the plant G may have unstable poles its normalized coprime factors (N,M) are stable. By the application of balanced reduction on the coprime factors we are able to reduce plants with or without unstable poles with the same method.

In the standard balance and truncate method an upper bound on the H_∞ approximation error between the full order model and the reduced order model is given by (Heuberger 1990):

$$\left\| \begin{bmatrix} M_n \\ N_n \end{bmatrix} - \begin{bmatrix} M_r \\ N_r \end{bmatrix} \right\|_\infty \leq 2 \sum_{i=r}^{n} \sigma_i$$

CONCLUSIONS

Theorem 1 and Corollary 1 show that with standard mathematical tools the normalized coprime factorization can be calculated, which is necessary to design discrete time controllers, that satisfy H_∞ robustness bounds. Since in practical applications one will in general be dealing with a discrete time problem, this is an important step towards the solution of the H_∞ control problem in discrete time.

Another application of the discrete normalized coprime factorization can be found in a fractional balanced model reduction scheme. An algorithm to calculate the fractional balanced reduced models is given. Using this method plants with or without unstable can be reduced in the same way.

LITERATURE

Bongers, P.M.M. and P.S.C. Heuberger (1990). Discrete normalized coprime factorization, *to appear in Proc. 9th Int. Conf. Analysis ans Optimization of Systems*, Antibes, France.

Anderson, B.D.O. and J.B. Moore (1979). Optimal *Filtering*, Prentice Hall, Inc., Englewood Cliffs, N.J.

Chu, C.C. (1988). On discrete inner-outer and spectral factorizations,*Proc. Amer. Control. Conf.*, Atlanta, Georgia

Desoer, C.A., R-W. Liu, J. Murray and R. Seaks (1980). Feedback System Design: The Fractional Representation Approach to Analysis and Synthesis, *IEEE Trans. Automat. Contr.*, *AC-25*, 399-412

Glover, K. and D. McFarlane (1988). Robust stabilization of normalized coprime factors: an explicit H∞ solution, *Proc. Amer. Control. Conf.*, Atlanta, Georgia

Glover, K. and D. McFarlane (1989). Robust stabilization of normalized coprime factor plant descriptions with H∞-bounded uncertainty, *IEEE Trans. Automat. Contr.*, *AC-33*, 821-830.

Heuberger, P.S.C (1990). *On Model Reduction, Complete Orthonormal Sets and System Identification*, PhD Thesis (to appear), Delft University of Technology, The Netherlands.

Huang, Q. and R. Liu (1987). A necessary and sufficient condition for stability of a perturbed system, *IEEE Trans. Automat. Contr.*, *AC-32*, 337-340.

Liu, Y. and B.D.O. Anderson (1986). Controller reduction via stable factorization and balancing, *Int. J. Control vol 44*, 507-531

McFarlane, D. (1988). *Robust controller design using normalized coprime factor plant descriptions*, PhD. thesis, University of Cambridge.

Meyer, D.G. (1988). A fractional approach to model reduction, *Proc. 27th IEEE Conf. Decision and Control*, 1041-1047.

Meyer, D.G. and G.F. Franklin (1988). A connection between normalized coprime factorization and linear quadratic optimal regulator theory, *IEEE Trans. Automat. Contr.*, *AC-32*, 1041-1047

Nett, C.N., C.A. Jacobson and M.J. Balas (1984). A connection between state-space and doubly coprime fractional representations, *IEEE Trans. Automat. Contr.*, *AC-29*, 831-832.

Rosenbrock, H.H. (1970). *State-space and Multivariable Theory*, Thomas Nelson and sons Ltd. London, Great Brittain.

Vidyasagar, M. (1984). The graph metric for unstable plants and robustness estimates for feedback stability, *IEEE Trans. Automat. Contr.*, *AC-29*, 403-418.

Vidyasagar, M. (1985). *Control System Synthesis, a Factorization Approach*, MIT Press, Cambridge, MA. USA.

Vidyasagar, M. (1988). Normalized coprime factorizations for nonstrictly proper systems, *IEEE Trans. Automat. Contr.*, *AC-33*, 300-301.

Vidyasagar, M., H.Schneider and B.A. Francis (1982). Algebraic and topological aspects of feedback stabilization, *IEEE Trans. Automat. Contr.*, *AC-27*, 880-894.

Selected Topics in Identification,
Modelling and Control; Vol. 1, 1990

An analysis of the full information L_2– and H_2–optimal control problem.

Paul F. Lambrechts

Lab. Measurement and Control, Dept. Mechanical Engineering and Marine Techn., Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. Based on a real–rational description of signals and systems in the frequency domain and a state–space description in the time domain, an analysis of the L_2– and H_2–optimal control problem with full (state) information will be given. First both problems will be formulated in the frequency domain, after which the L_2–optimal controller will be derived resulting in the feedback connection of the system to be controlled with its dual. Finally this result will also be considered in the time domain and extended to the H_2–optimal controller. This paper is intended to give some insight in the structure of L_2– and H_2–optimal controllers, especially in an input–output sense. The aim is to give a self–contained derivation that clarifies the advantages of using both frequency domain and time domain arguments. The importance of the solution of the algebraic Riccati equation connected with H_2–optimal control will follow quite naturally from the derivation: it can be considered as an operator from the state–space into the costate–space, leading to an auxiliary input signal that converts the L_2–optimal configuration into an H_2–optimal configuration.

Keywords. system theory; operator theory; multivariable systems; dual systems; L_2– and H_2–optimal control; state–space methods; full information problem.

1 INTRODUCTION

Kalman (1960) introduced the linear quadratic control, or LQ, problem as the dual of a stochastic filtering problem. Since then the LQ problem has been widely studied, especially in the time domain in a state–space setting. From these studies the great importance of the algebraic Riccati equation became apparent; a particular solution to this equation immediately results in the solution of the LQ problem. A thorough investigation of this is given by for instance Brockett (1970). Also Willems (1971) should be mentioned; he considers solvability of LQ problems and shows its dependence on some inequalities that are closely related to the algebraic Riccati equation.
The introduction of H_∞–control theory (Zames, 1981) and its further development (see for instance (Francis, 1987) and the references therein) resulted in a renewed interest in the frequency domain properties of the LQ problem as well as a more operator theoretic approach of some earlier results. In this sense the H_2 problem was introduced as a frequency domain version of the LQ problem, augmented with the possibilities of Wiener–Hopf design (Francis, 1982). Again the algebraic Riccati equation appeared to be of key importance; not only (as to be expected) for the H_2 problem but also for the H_∞ problem (Doyle and others, 1989).

Although the role of the algebraic Riccati equation has been widely recognized, the reason for its importance is not so often considered. MacFarlane (1963, 1969a, 1969b) showed by using a variational approach and Pontryagin's maximum principle (Athans and Falb, 1966) that the LQ optimal trajectory is governed by the behaviour of a pair of interconnected dual dynamical systems with two–point boundary conditions. From this result it is possible to derive the algebraic Riccati equation when a constant stabilizing state–feedback controller is to be found.

This paper will give a self contained derivation of the solution of the full information LQ or H_2 problem, both in the frequency domain and in the time domain. By restricting our attention to linear time–invariant systems, we will derive the L_2– and H_2–optimal controller without explicitly using the aforementioned results. It is intended to show that the use of both frequency domain and time domain arguments can simplify and clarify some well known results and proofs. The L_2–optimal controller is introduced as a convenient intermediate step towards the H_2–optimal controller.

The description of signals and systems is given in an operator theoretic sense and will be considered in section 2. Next the problem formulation for the H_2-optimal control problem will be stated in section 3. Section 4 will then give the solution of the L_2-optimal control problem, followed by the solution of the H_2-optimal control problem in section 5. Finally section 6 will give a discussion of the results.

2 PRELIMINARIES AND NOTATION

A frequency domain description of signals and systems

We will consider signals in $\mathbb{RL}_2$, the Hilbert space of real rational functions of a complex variable $s=\lambda+j\omega$ for which the inner product

$$\langle u_1(s),u_2(s)\rangle := (2\pi)^{-1}\cdot\int_{-\infty}^{\infty} u_1^*(j\omega)u_2(j\omega)d\omega \tag{2.1}$$

is finite, with u^* denoting the complex conjugate transpose of u. This inner product thus defines a norm denoted as:

$$\|u(s)\|_2 := \sqrt{\langle u(s),u(s)\rangle} \tag{2.2}$$

So $u(s)\in\mathbb{RL}_2$ if and only if $u(s)$ is real rational and $\|u(s)\|_2 < \infty$. This implies that $u(s)$ is strictly proper and has no poles on the imaginary axis.
From this we can define two complementary subspaces in $\mathbb{RL}_2$;

$$\mathbb{RH}_2 := \{u(s)\,|\,u(s)\in\mathbb{RL}_2,\text{ no poles in ccrhp}\}$$
$$\mathbb{RH}_2^\perp := \{u(s)\,|\,u(s)\in\mathbb{RL}_2,\text{ no poles in cclhp}\} \tag{2.3}$$

(ccr(l)hp = closed complex right (left) half plane).
So $\mathbb{RH}_2$, $\mathbb{RH}_2^\perp$ consists of signal representations that are real rational, strictly proper and stable, real rational, strictly proper and antistable, respectively.

Next we consider a system as an operator on $\mathbb{RL}_2$ with representation $G(s)\in\mathbb{RL}_\infty$;

$$\mathbb{RL}_\infty:=\{G(s)\,|\,y(s)=G(s)u(s)\in\mathbb{RL}_2,\ \forall u(s)\in\mathbb{RL}_2\} \tag{2.4}$$

Clearly $G(s)$ must be such that:

$$\frac{\|G(s)u(s)\|_2}{\|u(s)\|_2} < \infty \quad \forall\ u(s)\in\mathbb{RL}_2/\{0\} \tag{2.5}$$

so the operator norm or induced norm of $G(s)$ can be defined as:

$$\|G(s)\|_\infty := \sup_{u(s)\in\mathbb{RL}_2/\{0\}} \frac{\|G(s)u(s)\|_2}{\|u(s)\|_2} \tag{2.6}$$

From the characterization of signals in $\mathbb{RL}_2$ given above it follows that $G(s)\in\mathbb{RL}_\infty$ if and only if $G(s)$ is real rational and proper, and has no poles on the imaginary axis.
Furthermore we can define two complementary subspaces in $\mathbb{RL}_\infty$, based on those defined in $\mathbb{RL}_2$ (eq. 2.3):

$$\mathbb{RH}_\infty := \{G(s)\,|\,y(s)=G(s)u(s)\in\mathbb{RH}_2,\ \forall u(s)\in\mathbb{RH}_2\}$$
$$\mathbb{RH}_\infty^\perp := \{G(s)\,|\,y(s)=G(s)u(s)\in\mathbb{RH}_2^\perp,\ \forall u(s)\in\mathbb{RH}_2^\perp\} \tag{2.7}$$

So $\mathbb{RH}_\infty$, $\mathbb{RH}_\infty^\perp$ consists of system representations that are real rational, proper and stable, real rational proper and antistable, respectively.

Based on the inner product given in eq. 2.1, we can now, given $G(s)$, define the adjoint or dual system representation $G^\sim(s)$ as satisfying:

$$\langle G(s)u_1(s),u_2(s)\rangle = \langle u_1(s),G^\sim(s)u_2(s)\rangle$$
$$\forall\ u_1(s),u_2(s)\in\mathbb{RL}_2 \tag{2.8}$$

which is equivalent to $G^\sim(j\omega)=G(j\omega)^*\ \forall\omega\in\mathbb{R}$.
The partitioning of $\mathbb{RL}_2$ in $\mathbb{RH}_2$ and $\mathbb{RH}_2^\perp$ makes it possible to extend the definition of $G^\sim(s)$ to the entire complex plane; because $\mathbb{RH}_2$ and $\mathbb{RH}_2^\perp$ are complementary we have:

$$\langle u_1(s),u_2(s)\rangle=0 \quad \forall u_1(s)\in\mathbb{RH}_2,\ \forall u_2(s)\in\mathbb{RH}_2^\perp \tag{2.9}$$

Now consider $G(s)\in\mathbb{RH}_\infty$, such that $G(s)u_1(s)\in\mathbb{RH}_2$ and:

$$\langle G(s)u_1(s),u_2(s)\rangle=0 \qquad \forall u_1\in\mathbb{RH}_2,\forall u_2\in\mathbb{RH}_2^\perp$$
$$\langle u_1(s),G^\sim(s)u_2(s)\rangle=0 \qquad \forall u_1\in\mathbb{RH}_2,\forall u_2\in\mathbb{RH}_2^\perp \tag{2.10}$$

This implies $G^\sim(s)\in\mathbb{RH}_\infty^\perp$ and therefore we have:

$$G^\sim(s) = G^T(-s) \tag{2.11}$$

with G^T denoting the transpose of G.

A time domain description of signals and systems

It is well known that a system represented in the frequency domain as a transfer function in $\mathbb{RL}_\infty$ can be represented in the time domain by a minimal state-space realization:

$$\dot{x}(t) = Ax(t) + Bu(t) \qquad x(0)=0$$
$$y(t) = Cx(t) + Du(t) \tag{2.12}$$

such that:

$$G(s) = C\cdot(sI-A)^{-1}\cdot B + D \tag{2.13}$$

Here we assume the state-space to be finite dimensional.
Although eq. 2.12 makes it possible to calculate the response of $y(t)$ to any input signal $u(t)$ with $u(t)=0$, $\forall t<0$, we will only consider signals with representations in the frequency domain that are in $\mathbb{RL}_2$. To find a representation of such signals in the time domain we can use the inverse Fourier transformation, which is an isomorphism from the frequency domain into the time domain (Paley and Wiener, 1934).
The procedure is as follows:
Given a signal $u(s)\in\mathbb{RL}_2$, divide it into

$u(s)=u_s(s)+u_a(s)$ with $u_s(s)\in\mathbb{R}H_2$ and $u_a(s)\in\mathbb{R}H_2^\perp$. Perform an inverse Fourier transformation of $u_s(s)$ into the time domain to get a stable realization (A_s,B_s,C_s) such that:

$$\begin{aligned} \dot{x}_s(t) &= A_s x_s(t) + B_s\delta(t) \qquad x_s(0)=0 \\ \begin{bmatrix} u_s(t) \\ u_s(t) \end{bmatrix} & \begin{matrix} = C_s x_s(t), & \forall t\geq 0 \\ = 0, & \forall t<0 \end{matrix} \end{aligned} \tag{2.14}$$

with $\delta(t)$ denoting the unit impulse.
Similarly transform $u_a(s)$, finding an antistable realization (A_a,B_a,C_a) such that:

$$\begin{aligned} \dot{x}_a(t) &= -A_a x_a(t) + B_a\delta(t) \qquad x_a(0)=0 \\ \begin{bmatrix} u_a(-t) \\ u_a(-t) \end{bmatrix} & \begin{matrix} = C_a x_a(t), & \forall t\geq 0 \\ = 0, & \forall t<0 \end{matrix} \end{aligned} \tag{2.15}$$

Finally the time domain representation of $u(s)$ is:

$$u(t) = u_s(t) + u_a(t) \tag{2.16}$$

Note that $u(t)$ can also be found by taking the free responses of the systems in eq. 2.14 and 2.15 with the initial conditions $x_s(0)=B_s$ and $x_a(0)=B_a$ respectively.

By this procedure of splitting $u(s)$ before inverse Fourier transformation we now have a function of time $u(t)$ that is again an element of a Hilbert space defined as follows.
Consider the function space of all real vector–valued functions of time $u(t)$ with $t\in(-\infty,\infty)$ and define the inner product:

$$\langle u_1(t),u_2(t)\rangle := \int_{-\infty}^{\infty} u_1^T(t)u_2(t)\mathrm{d}t \tag{2.17}$$

Then the Hilbert space $\mathscr{L}_2(-\infty,\infty)$ can be defined as:

$$\mathscr{L}_2(-\infty,\infty) := \{u(t)\,|\,\langle u(t),u(t)\rangle < \infty\} \tag{2.18}$$

For all $\forall u(t)\in\mathscr{L}_2(-\infty,\infty)$ we can thus define a norm:

$$\|u(t)\|_2 := \sqrt{\langle u(t),u(t)\rangle} \tag{2.19}$$

Now it follows from eq. 2.14 and 2.15 that $u(t)$ given in eq. 2.16 is bounded on $(-\infty,\infty)$ and approaches 0 towards $+\infty$ and $-\infty$, therefore we have $u(t)\in\mathscr{L}_2(-\infty,\infty)$.
From eq. 2.14 and eq. 2.15 we can see that $u_s(t)$ is nonzero on the interval $[0,\infty)$ and $u_a(t)$ is nonzero on the interval $(-\infty,0]$. It is therefore convenient to define two complementary subspaces in $\mathscr{L}_2(-\infty,\infty)$; $\mathscr{L}_2[0,\infty)$ and $\mathscr{L}_2(-\infty,0]$, such that $u_s(t)\in\mathscr{L}_2[0,\infty)$ and $u_a(t)\in\mathscr{L}_2(-\infty,0]$.
Similar to what was done in the previous section we can now consider a system to be an operator on the time domain space $\mathscr{L}_2(-\infty,\infty)$ having a representation as given in eq. 2.12. In this sense also a dual system representation can be found from the time domain inner product given in eq. 2.17. It is much easier however to substitute eq. 2.13 into eq. 2.11 to get:

$$\begin{aligned} G^{\sim}(s) = G^T(-s) &= \{C\cdot(-sI-A)^{-1}\cdot B + D\}^T = \\ &-B^T(sI+A^T)^{-1}C^T+D^T \end{aligned} \tag{2.20}$$

giving the time domain representation:

$$\begin{aligned} \dot{\xi}(t) &= -A^T\xi(t) - C^Ty(t) \qquad \xi(0)=0 \\ u(t) &= B^T\xi(t) + D^Ty(t) \end{aligned} \tag{2.21}$$

Finally note that it is possible to find the systems response signal $y(t)$ (eq. 2.11) for any input signal $u(t)\in\mathscr{L}_2[0,\infty)$ having a representation as in eq. 2.14 by adding this representation to the one given in eq. 2.12:

$$\begin{aligned} \begin{bmatrix} \dot{x}_s(t) \\ \dot{x}(t) \end{bmatrix} &= \begin{bmatrix} A_s & 0 \\ BC_s & A \end{bmatrix}\begin{bmatrix} x_s(t) \\ x(t) \end{bmatrix} + \begin{bmatrix} B_s \\ 0 \end{bmatrix}\cdot\delta(t) \\ y_s(t) &= \begin{bmatrix} DC_s & C \end{bmatrix}\begin{bmatrix} x_s(t) \\ x(t) \end{bmatrix} \end{aligned} \tag{2.22}$$

3 PROBLEM FORMULATION

The H_2–optimal control problem formulation as considered in this paper is derived from Doyle and others (1989) and starts with the frequency domain system description:

$$\begin{bmatrix} z(s) \\ y(s) \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}\begin{bmatrix} w(s) \\ u(s) \end{bmatrix} \tag{3.1}$$

with $w(s)$ the input vector of disturbance and reference signals, $u(s)$ the input vector of control signals, $z(s)$ the output vector of signals to be controlled and $y(s)$ the output vector of available measurement signals.
In the time domain this system can be represented as:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B_1w(t) + B_2u(t) \qquad x(0)=0 \\ z(t) &= C_1x(t) + D_{11}w(t) + D_{12}u(t) \\ y(t) &= C_2x(t) + D_{21}w(t) + D_{22}u(t) \end{aligned} \tag{3.2}$$

We now want to find a second system that makes use of the measurement signals $y(s)$ and the control inputs $u(s)$ to improve in some sense the behaviour of $z(s)$ given possible disturbances (references) $w(s)$.
The sense in which an H_2–optimal controller aims to improve this behaviour of $z(s)$ can be formulated as follows:

Given a possible nonzero disturbance (reference) vector $w(s)\in\mathbb{R}H_2$, that can be normalized such that $\|w(s)\|_2=1$;

a) make sure that the controlled system is internally stable.
b) make sure that the response of $z(s)$ is an element of $\mathbb{R}H_2$.
c) make sure that $\|z(s)\|_2$ is as small as possible.

To somewhat simplify the problem the following assumptions are made (see th. 2.3 and prop. 3.1 of Wonham (1978) for a definition of stabilizability and detectability):

1. $D_{11}=0$
2. $D_{22}=0$
3. $D_{12}^T \cdot D_{12}=I$
4. $D_{12}^T \cdot C_1=0$
5. (A,C_1) detectable
6. (A,B_2,C_2) stabilizable and detectable
7. $C_2=I$ and $D_{21}=0$
8. $w(t)\in \mathscr{L}_2[0,\infty)$ follows from: (see eq. 2.14)

$$\begin{aligned} \dot{x}_w(t) &= A_w x_w(t) + w_0\delta(t) \quad x_w(0)=0 \\ \begin{bmatrix} w(t) \\ w(t) \end{bmatrix} &\begin{matrix} = C_w x_w(t), & \forall t\geq 0 \\ = 0, & \forall t<0 \end{matrix} \end{aligned} \tag{3.3}$$

This representation of $w(t)$ is assumed to be incorporated in the system description, such that we can take $w_0\delta(t)$ as a new input signal. Note that, although $w_0\delta(t)\notin\mathscr{L}_2(-\infty,\infty)$, we still have a valid minimization because we are considering the transfer from $w(t)$ to $z(t)$.

These assumptions are mainly equal to those made by Doyle and others (1989). They are not very restrictive, with the exception of assumption 4, which is equivalent to not allowing cross-products in the time domain LQ criterion, and assumption 7, which restricts our attention to the full information problem.

Furthermore, based on assumptions 3 and 4, we can assume without further loss of generality that C_1 and D_{12} can be partitioned as:

$$C_1 = \begin{bmatrix} C_1' \\ 0 \end{bmatrix}, \quad D_{12} = \begin{bmatrix} 0 \\ D_{12}' \end{bmatrix} \tag{3.4}$$

Also note that assumption 5 guarantees internal stability of the controlled system if $z(s)\in\mathbb{R}H_2$.

With these assumptions it is now possible to state the H_2–optimal control problem as will be considered in this paper:

Formulation of the H_2–optimal control problem.

Given the system to be controlled:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B_1 w_0\delta(t) + B_2 u(t) \quad x(0)=0 \\ \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} &= \begin{bmatrix} C_1' \\ 0 \end{bmatrix}\cdot x(t) + \begin{bmatrix} 0 \\ D_{12}' \end{bmatrix}\cdot u(t) \\ y(t) &= x(t) \end{aligned} \tag{3.5}$$

with (A,B_2) stabilizable, (A,C_1) detectable and $D_{12}^T D_{12}=I$.

Find a controller that uses measurement signals $y(s)$ and control signals $u(s)$, such that:

1. $z(s)\in\mathbb{R}H_2$
2. $\|z(s)\|_2$ as a result of $w_0\delta(t)$ is minimal. ■

4 THE L_2–OPTIMAL CONTROLLER

Given the problem formulation in the previous section it will appear to be convenient to first consider only the second demand, the minimization of $\|z(s)\|_2$, without considering stability; this will lead to the L_2–optimal control system. The result will then be extended to the H_2–optimal controller in the next section.

In order to somewhat simplify notation we will first define two transfer functions based on eq. 3.5. The transfer from w_0 to $z_1(s)$:

$$H_w(s) := C_1'(sI-A)^{-1}B_1 \tag{4.1}$$

and the transfer from $u(s)$ to $z_1(s)$:

$$H_u(s) := C_1'(sI-A)^{-1}B_2 \tag{4.2}$$

So, as far as $z(s)$ is concerned, we can consider the frequency domain equivalent of eq. 3.5 to be:

$$z(s) = \begin{bmatrix} z_1(s) \\ z_2(s) \end{bmatrix} = \begin{bmatrix} H_u(s)u(s) + H_w(s)w_0 \\ D_{12}'u(s) \end{bmatrix} \tag{4.3}$$

Based on eq. 2.1 and dropping the dependency on s or $j\omega$ we now have:

$$\|z\|_2^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty}(u^{\sim}H_u^{\sim}H_u u + u^{\sim}H_u^{\sim}H_w w_0 + w_0^T H_w^{\sim}H_u u + w_0^T H_w^{\sim}H_w w_0 + u^{\sim}u)\,d\omega \tag{4.4}$$

The following theorem then gives the L_2–optimal control signal $u_{12}(s)$ such that this criterion function is minimized.

Theorem 1:

Given the system to be controlled in eq. 3.5 and given the criterion function to be minimized in eq. 4.4, the following statements hold (dropping the dependency on s when convenient):

1. The L_2–optimal control input $u_{12}(s)\in\mathbb{R}L_2$ that minimizes the criterion function over all $u(s)\in\mathbb{R}L_2$ is:
$$u_{12} = -(H_u^{\sim}H_u+I)^{-1}H_u^{\sim}H_w w_0 \tag{4.5}$$
2. After partitioning $z(s)$ as in eq. 4.3, we have that eq. 4.5 simplifies to:
$$u_{12} = -H_u^{\sim}\cdot z_1 \tag{4.6}$$
3. The minimal value of the criterion function over all $u(s)\in\mathbb{R}L_2$ is:
$$\|z\|_2^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty}(w_0^T H_w^{\sim}(H_u H_u^{\sim}+I)^{-1}H_w w_0)\,d\omega \tag{4.7}$$
□

Proof: see appendix A.1 ■

From this theorem we have that the L_2–optimal control system is given by applying feedback from $z_1(s)$ to $u(s)$ through the dual of the transfer from $u(s)$ to $z_1(s)$. This situation is clarified by the block–diagram given in Fig. 1.

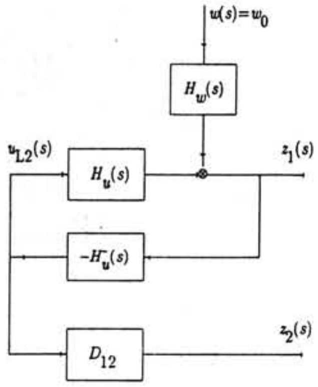


Fig. 1. L_2–optimal control in the frequency domain.

From equations 2.22, 3.5, 4.1, 4.2 and 4.6 we can now find that the time domain description of the L_2–optimal control system is given by:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\xi}(t) \end{bmatrix} = \begin{bmatrix} A & -B_2B_2^{\mathrm{T}} \\ -C_1'^{\mathrm{T}}C_1' & -A^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} x(t) \\ \xi(t) \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} w_0\delta(t)$$

$$\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} C_1' & 0 \\ 0 & -D_{12}'B_2^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} x(t) \\ \xi(t) \end{bmatrix} \tag{4.9}$$

(with $x(0)=\xi(0)=0$), This leads to the block–diagram in Fig. 2.

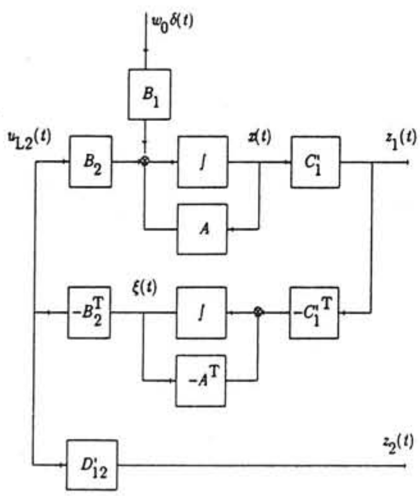


Fig. 2. L_2–optimal control in the time domain.

5 THE H_2–OPTIMAL CONTROLLER

We will now extend the result of the previous section to the H_2–optimal controller, that minimizes $\|z(s)\|_2$ given the extra condition of $z(s)\in\mathbb{R}H_2$. The first step towards the solution of this problem will be to consider more closely the behaviour of the combined state vector $\begin{bmatrix} x(t) \\ \xi(t) \end{bmatrix}$ as given by eq. 4.9. The system matrix in this equation, which from now on will be denoted as H;

$$H := \begin{bmatrix} A & -B_2B_2^{\mathrm{T}} \\ -C_1'^{\mathrm{T}}C_1' & -A^{\mathrm{T}} \end{bmatrix} \tag{5.1}$$

is a Hamiltonian matrix and has the following properties:

__Lemma 1:__
Given the $2n\times 2n$ matrix H as defined in eq. 5.1 with (A,B_2) stabilizable and (A,C_1') detectable, the following statements hold:

1. H has no eigenvalues on the imaginary axis.
2. H has Jordan form $\begin{bmatrix} \Lambda & 0 \\ 0 & -\Lambda^{\mathrm{T}} \end{bmatrix}$ with modal matrix $\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$ and Λ stable. So the stable and antistable modal subspaces of H are $X_-(H)=\mathrm{Im}\begin{bmatrix} M_{11} \\ M_{21} \end{bmatrix}$ and $X_+(H)=\mathrm{Im}\begin{bmatrix} M_{12} \\ M_{22} \end{bmatrix}$ with $\mathbb{R}^{2n}=X_-(H)\oplus X_+(H)$. □

__Proof:__ see appendix A.2 ■

This lemma now immediately leads to:

__Lemma 2:__
Given the $2n\times 2n$ matrix H as defined in eq. 5.1 with (A,B_2) stabilizable and (A,C_1') detectable, and with the Jordan form and modal matrix of H from lemma 1 part 2, the following statements hold:

1. M_{11} is invertible
2. $\mathrm{Im}\begin{bmatrix} M_{11} \\ M_{21} \end{bmatrix} = \mathrm{Im}\begin{bmatrix} I \\ X \end{bmatrix}$ with $X := M_{21}M_{11}^{-1}$
3. X is symmetric
4. X is a solution of the algebraic Riccati equation:
$$A^{\mathrm{T}}X + XA - XB_2B_2^{\mathrm{T}}X + C_1'^{\mathrm{T}}C_1' = 0 \tag{5.2}$$
5. $A-B_2B_2^{\mathrm{T}}X$ is stable □

__Proof:__
See th.7.2.2 and cor.7.2.1 of (Francis, 1987) for a recent and very complete proof; the original proof is given by Potter (1966) and Mårtensson (1971).■

Note that lemmas 1 and 2 do not give all available results on the Hamiltonian matrix and the algebraic Riccati equation; only results necessary for the further exposition in this paper are mentioned. See for instance Richardson and Kwong (1986) and Faibusovich (1987).

We are now able to state the solution of the H_2–optimal control problem as follows:

<u>Theorem 2:</u>
Given the H_2–optimal control problem from section 3, the following statements hold:

1. The H_2–optimal trajectory of $z(t)$ is given by:
$$\begin{bmatrix}\dot{x}(t)\\ \dot{\xi}(t)\end{bmatrix} = \begin{bmatrix}A & -B_2B_2^T\\ -C_1'^TC_1' & -A^T\end{bmatrix}\begin{bmatrix}x(t)\\ \xi(t)\end{bmatrix}+\begin{bmatrix}B_1\\ XB_1\end{bmatrix}w_0\delta(t)$$
$$\begin{bmatrix}z_1(t)\\ z_2(t)\end{bmatrix} = \begin{bmatrix}C_1' & 0\\ 0 & -D_{12}'B_2^T\end{bmatrix}\begin{bmatrix}x(t)\\ \xi(t)\end{bmatrix} \tag{5.3}$$
with X as defined in lemma 2 part 2 and $x(0)=\xi(0)=0$.
2. Equation 5.3 can be simplified to:
$$\dot{x}(t) = (A-B_2B_2^TX)x(t) + B_2w_0\delta(t) \quad x(0)=0$$
$$\begin{bmatrix}z_1(t)\\ z_2(t)\end{bmatrix} = \begin{bmatrix}C_1'\\ -D_{12}'B_2^TX\end{bmatrix}x(t) \tag{5.4}$$
3. The H_2–optimal control input denoted as $u_{h2}(t)$ is given by:
$$u_{h2}(t) = -B_2^TX\cdot x(t) \tag{5.5}$$

□

<u>Proof:</u>
Consider a control input $u(t)$ given as $u(t)=u_{l2}(t)+v(t)$, with $u_{l2}(t)$ the L_2–optimal control input from th. 1. The L_2–optimal control system given in eq. 4.9 must then be extended to:
$$\begin{bmatrix}\dot{x}(t)\\ \dot{\xi}(t)\end{bmatrix}=\begin{bmatrix}A & -B_2B_2^T\\ -C_1'^TC_1' & -A^T\end{bmatrix}\begin{bmatrix}x(t)\\ \xi(t)\end{bmatrix}+\begin{bmatrix}B_1 & B_2\\ 0 & 0\end{bmatrix}\begin{bmatrix}w_0\delta(t)\\ v(t)\end{bmatrix}$$
$$\begin{bmatrix}z_1(t)\\ z_2(t)\end{bmatrix}=\begin{bmatrix}C_1' & 0\\ 0 & -D_{12}'B_2^T\end{bmatrix}\begin{bmatrix}x(t)\\ \xi(t)\end{bmatrix}+\begin{bmatrix}0 & 0\\ 0 & D_{12}'\end{bmatrix}\begin{bmatrix}w_0\delta(t)\\ v(t)\end{bmatrix} \tag{5.6}$$
(with $x(0)=\xi(0)=0$). Now consider the similarity transformation:
$$\begin{bmatrix}x(t)\\ \xi(t)\end{bmatrix} = \begin{bmatrix}I & 0\\ X & I\end{bmatrix}\begin{bmatrix}x(t)\\ q(t)\end{bmatrix} \tag{5.7}$$
resulting in the transformed system matrix:
$$\begin{bmatrix}I & 0\\ X & I\end{bmatrix}^{-1}\begin{bmatrix}A & -B_2B_2^T\\ -C_1'^TC_1' & -A^T\end{bmatrix}\begin{bmatrix}I & 0\\ X & I\end{bmatrix} =$$
$$\begin{bmatrix}A-B_2B_2^TX & -B_2B_2^T\\ A^TX+XA-XB_2B_2^TX+C_1'^TC_1' & -A^T+XB_2B_2^T\end{bmatrix} =$$
(with lemma 2 part 4)
$$\begin{bmatrix}A-B_2B_2^TX & B_2B_2^T\\ 0 & -A^T+XB_2B_2^T\end{bmatrix} \tag{5.8}$$
and the transformed input and output matrices:
$$\begin{bmatrix}I & 0\\ X & I\end{bmatrix}^{-1}\begin{bmatrix}B_1 & B_2\\ 0 & 0\end{bmatrix} = \begin{bmatrix}B_1 & B_2\\ -XB_1 & -XB_2\end{bmatrix} \tag{5.9}$$
and
$$\begin{bmatrix}C_1' & 0\\ 0 & -D_{12}'B_2^T\end{bmatrix}\begin{bmatrix}I & 0\\ X & I\end{bmatrix}=\begin{bmatrix}C_1' & 0\\ -D_{12}'B_2^TX & -D_{12}'B_2^T\end{bmatrix} \tag{5.10}$$
The controlled system can thus be represented as:
$$\begin{bmatrix}\dot{x}(t)\\ \dot{q}(t)\end{bmatrix} = \begin{bmatrix}A-B_2B_2^TX & -B_2B_2^T\\ 0 & -A^T+XB_2B_2^T\end{bmatrix}\begin{bmatrix}x(t)\\ q(t)\end{bmatrix}$$
$$+\begin{bmatrix}B_1 & B_2\\ -XB_1 & -XB_2\end{bmatrix}\begin{bmatrix}w_0\delta(t)\\ v(t)\end{bmatrix},\quad \begin{bmatrix}x(0)\\ q(0)\end{bmatrix}=0$$
$$\begin{bmatrix}z_1(t)\\ z_2(t)\end{bmatrix} = \begin{bmatrix}C_1' & 0\\ -D_{12}'B_2^TX & -D_{12}'B_2^T\end{bmatrix}\begin{bmatrix}x(t)\\ q(t)\end{bmatrix}$$
$$+\begin{bmatrix}0 & 0\\ 0 & D_{12}'\end{bmatrix}\begin{bmatrix}w_0\delta(t)\\ v(t)\end{bmatrix} \tag{5.11}$$
or in block–diagram as in Fig. 3:

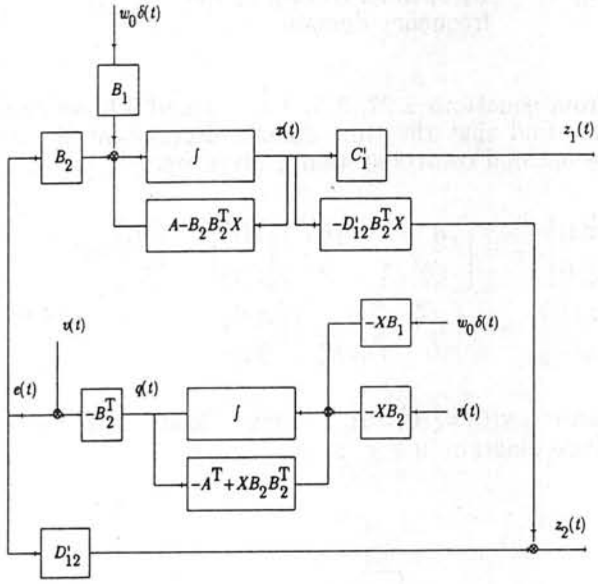


Fig. 3. Block–diagram of eq. 5.11.

In this block diagram the signal $e(t)$ is defined as
$$e(t) := -B_2^Tq(t) + v(t) \tag{5.12}$$
which makes it possible to write down eq. 5.11 as:
$$\dot{x}(t) = (A-B_2B_2^TX)x(t)+[B_1\ B_2]\begin{bmatrix}w_0\delta(t)\\ e(t)\end{bmatrix} \quad x(0)=0$$
$$\begin{bmatrix}z_1(t)\\ z_2(t)\end{bmatrix} = \begin{bmatrix}C_1'\\ -D_{12}'B_2^TX\end{bmatrix}x(t)+\begin{bmatrix}0 & 0\\ 0 & D_{12}'\end{bmatrix}\begin{bmatrix}w_0\delta(t)\\ e(t)\end{bmatrix} \tag{5.13}$$

Now assume that $e(t)$ can be chosen arbitrarily and consider a choice of $e(t)$ in $\mathscr{L}_2(-\infty,0]$ (or $e(s)\in\mathbb{R}H_2^\perp$). In this case $z(t)$ can only be in $\mathscr{L}_2[0,\infty)$ if the transfer from $e(t)$ to $z(t)$ contains right half plane zeros that completely block the influence of $e(t)$. Because the 2–norm of z will then be equal to that in case of $e=0$, we can conclude that only a choice of $e(t)$ in $\mathscr{L}_2[0,\infty)$ can reduce the 2–norm of z without making $z(t)\notin\mathscr{L}_2[0,\infty)$.
Next we have to prove that any nonzero choice of $e(t)$ in $\mathscr{L}_2[0,\infty)$ (or $e(s)\in\mathbb{R}H_2$) will increase the

2–norm of z; this requires a straightforward derivation of the 2–norm of z as a function of e in the frequency domain. This is done in appendix A.3.
We now have that the H_2–optimal situation is given by e=0. It is easy to verify that this situation is given by eq. 5.3 by applying transformation 5.7. Furthermore eq. 5.4, and with that eq. 5.5, follow directly from eq. 5.13 after substitution of $e(t)=0\ \forall t$. ■

Finally, based on the proof of theorem 2, we can find a frequency domain relation between the L_2–optimal and the H_2–optimal control input by the following corollary.

Corollary 1:
Given the H_2–optimal control problem as posed in section 3 and the L_2–optimal control problem as derived from it in section 4.
The H_2–optimal control input $u_{h2}(s)$ is related to the L_2–optimal control input $u_{l2}(s)$ as follows:

$$u_{h2}(s) = u_{l2}(s) - B_2^T(sI+A^T)^{-1}XB_1w_0 \tag{5.14}$$

□

Proof:
The proof of theorem 2 and the block–diagram in Fig. 3 show that $u_{h2}(s)=u_{l2}(s)+v_0(s)$, with $v_0(s)$ such that:

$$e(s) = v_0(s) + B_2^T(sI+A^T-XB_2B_2^T)^{-1}X(B_1w_0+B_2v_0(s))=0 \quad \Leftrightarrow$$

$$v_0(s) = -(I+B_2^T(sI+A^T-XB_2B_2^T)^{-1}XB_2)^{-1}B_2^T \cdot(sI+A^T-XB_2B_2^T)^{-1}XB_1w_0 \quad \Leftrightarrow$$

$$v_0(s) = -B_2^T(I+(sI+A^T-XB_2B_2^T)^{-1}XB_2B_2^T)^{-1} \cdot(sI+A^T-XB_2B_2^T)^{-1}XB_1w_0 \quad \Leftrightarrow$$

$$v_0(s) = -B_2^T\{(sI+A^T-XB_2B_2^T)^{-1}(sI+A^T)\}^{-1} \cdot(sI+A^T-XB_2B_2^T)^{-1}XB_1w_0 \quad \Leftrightarrow$$

$$v_0(s) = -B_2^T(sI+A^T)^{-1}XB_1w_0 \tag{5.15}$$

■

6 DISCUSSION

We have derived the L_2–optimal and H_2–optimal control system using both frequency domain and time domain arguments. It has been shown that the key mechanism behind both solutions consists of a pair of interconnected dual dynamical systems. The L_2–optimal control system appeared to be the basic configuration minimizing the 2–norm of the criterion vector z, without considering stability (in fact it is not hard to prove that the L_2–optimal control system is always unstable). Transformation of this result into the time domain resulted in a state–space model with a system matrix that is Hamiltonian.

Although the exact proof is rather lengthy, the procedure of extending the L_2–optimal control system to the H_2–optimal control system has been shown to be quite simple: given any disturbance, just map it onto the stable trajectories of the L_2–optimal control system. To do this the disturbance should be represented in the time domain as an initial condition of –or similarly as an impulsive input on– a state–space model. This is possible because the disturbance is assumed to be in $\mathbb{R}H_2$. The resulting initial condition is an element of the $2n\times2n$ state–space of the L_2–optimal control system and can be mapped into the stable modal subspace of the Hamiltonian system matrix. The influence of this mapping can then be considered as the result of an auxiliary input (disturbance) signal, as is most clearly seen in theorem 2 (eq. 5.3) and is given in the frequency domain by cor. 1.

The exact form of this auxiliary signal is determined by the solution of the algebraic Riccati equation connected with the Hamiltonian system matrix. The reason for this is that the initial condition of the plant $x(0)\in\mathbb{R}^n$ (representing the disturbance w) can not be changed by a control input u that is in $\mathbb{R}L_2$. In order to map the combined initial condition into $X_-(H)$, it is therefore only possible to change the initial condition of the dual system $\xi(0)\in\mathbb{R}^n$. This implies that $x(0)$ must be embedded in $X_-(H)$ by choosing $\xi(0)$. Because from lemma 2 part 2 we have that $X_-(H)=\mathrm{Im}\begin{bmatrix}I\\X\end{bmatrix}$ it is easy to see that $\begin{bmatrix}I\\X\end{bmatrix}x(0)\in X_-(H)$ and so $\xi(0)=X\cdot x(0)$ is a correct choice for all $x(0)\in\mathbb{R}^n$. The solvability of the H_2–optimal control problem is thus determined by the following conditions:

1. $\dim\{X_-(H)\} = n$
2. M_{11} invertible

If these conditions are not met, then there exists an initial condition $x(0)\in\mathbb{R}^n$ for which there is no (finite) $\xi(0)$ such that $\begin{bmatrix}x(0)\\\xi(0)\end{bmatrix}\in X_-(H)$. Satisfaction of these conditions is proven in lemma 1 part 2 and lemma 2 part 1.

APPENDICES

A.1 Proof of theorem 1

First part. The criterion function $\|z\|_2$ reaches its minimum if (dropping the dependency on s and $j\omega$ and with δx denoting the variation of x)

$$\delta(\|z\|_2^2) = \delta\Big[\frac{1}{2\pi}\int_{-\infty}^{\infty}(u^\sim H_u^\sim H_u u + u^\sim H_u^\sim H_w w_0 +$$
$$w_0^T H_w^\sim H_u u + w_0^T H_w^\sim H_w w_0 + u^\sim u)\mathrm{d}\omega\Big] = 0 \qquad (A.1)$$

Because the integrand is real and nonnegative for all $\omega\in\mathbb{R}$ we can minimize eq. A.1 by minimizing the integrand at every frequency:

$$\delta(u^\sim H_u^\sim H_u u + u^\sim H_u^\sim H_w w_0 + w_0^T H_w^\sim H_u u +$$
$$w_0^T H_w^\sim H_w w_0 + u^\sim u) = 0 \qquad \forall\omega\in\mathbb{R} \qquad \Leftrightarrow$$

$$\delta u^\sim H_u^\sim H_u u + u^\sim H_u^\sim H_u \delta u + \delta u^\sim H_u^\sim H_w w_0 + w_0^T H_w^\sim H_u \delta u +$$
$$\delta u^\sim u + u^\sim \delta u = 0 \qquad \forall\omega\in\mathbb{R} \qquad \Leftrightarrow$$

$$\delta u^\sim[(H_u^\sim H_u + I)u + H_u^\sim H_w w_0] +$$
$$[u^\sim(H_u^\sim H_u + I) + w_0^T H_w^\sim H_u]\delta u = 0 \quad \forall\omega\in\mathbb{R} \qquad (A.2)$$

Now define

$$g := (H_u^\sim H_u + I)u + H_u^\sim H_w w_0 \qquad (A.3)$$

so that eq. A.2 simplifies to

$$\delta u^\sim g + g^\sim \delta u = 2\cdot\mathrm{Re}(\delta u^\sim g) = 0 \qquad \forall\omega\in\mathbb{R} \qquad (A.4)$$

($\delta u^\sim g$ is a scalar function).
This implies that $\delta u^\sim g$ is imaginary for all $\omega\in\mathbb{R}$ and so g must be imaginary. Furthermore we have that $H_u\in\mathbb{R}L_\infty$ and $H_w w_0\in\mathbb{R}L_2$ (see eq 3.3; H_w is strictly proper), therefore we must have that $g\in\mathbb{R}L_2$. Now if g is a real rational function, it can only be imaginary for all $\omega\in\mathbb{R}$ if all its poles are on the imaginary axis. This clearly contradicts $g\in\mathbb{R}L_2$ unless g=0, thus the only possible minimum of $\|z\|_2$ is found when:

$$(H_u H_u + I)u + H_u^\sim H_w w_0 = 0 \quad \forall\omega\in\mathbb{R}$$

$$u_{12} = -(H_u^\sim H_u + I)^{-1} H_u^\sim H_w w_0 \quad \forall\omega\in\mathbb{R} \qquad (A.5)$$

Furthermore u_{12} is an element of $\mathbb{R}L_2$ because $H_u^\sim H_u \geq 0$ $\forall\omega\in\mathbb{R}$ and therefore $(H_u^\sim H_u + I)^{-1}\in\mathbb{R}L_\infty$.

<u>Second part.</u> Consider z when applying u_{12} and make use of $(I+AB)^{-1}A = A(I+BA)^{-1}$:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} H_u\cdot\{-(H_u^\sim H_u + I)^{-1} H_u^\sim H_w w_0\} + H_w w_0 \\ D_{12}'\cdot\{-(H_u^\sim H_u + I)^{-1} H_u^\sim H_w w_0\} \end{bmatrix} =$$

$$\begin{bmatrix} \{-H_u H_u^\sim + (H_u H_u^\sim + I)\}(H_u H_u^\sim + I)^{-1} H_w w_0 \\ -D_{12}'\cdot(H_u^\sim H_u + I)^{-1} H_u^\sim H_w w_0 \end{bmatrix} =$$

$$\begin{bmatrix} (H_u H_u^\sim + I)^{-1} H_w w_0 \\ -D_{12}'\cdot(H_u^\sim H_u + I)^{-1} H_u^\sim H_w w_0 \end{bmatrix} =$$

$$\begin{bmatrix} I \\ -D_{12}' H_u^\sim \end{bmatrix}(H_u H_u^\sim + I)^{-1} H_w w_0 \qquad (A.6)$$

So $z_1 = (H_u^\sim H_u + I)^{-1} H_w w_0$ and

$$u_{12} = -H_u^\sim (H_u H_u^\sim + I)^{-1} H_w w_0 = -H_u^\sim\cdot z_1 \qquad (4.6')$$

<u>Third part.</u> From eq. A.6 we can write down the criterion function as:

$$\|z\|_2^2 = \frac{1}{2\pi}\cdot$$
$$\int_{-\infty}^{\infty}(w_0^T H_w^\sim (H_u H_u^\sim + I)^{-1}[I + H_u H_u^\sim](H_u H_u^\sim + I)^{-1} H_w w_0)\mathrm{d}\omega$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}(w_0^T H_w^\sim (H_u H_u^\sim + I)^{-1} H_w w_0)\,\mathrm{d}\omega \qquad (A.7)$$

A.2 Proof of lemma 1

<u>First part.</u> It will be shown that all possible imaginary eigenvalues of H must be poles of the L_2–optimal transfer from w to z for at least one admissable disturbance w (given by w_0). This would then imply $z\notin\mathbb{R}L_2$ and therefore contradict theorem 1.
So we have to prove that imaginary eigenvalues of the system $(H, \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \begin{bmatrix} C_1' & 0 \\ 0 & D_{12}' B_2^T \end{bmatrix})$ are controllable and observable. Necessary and sufficient conditions for this are (Rosenbrock, 1970):

$$\mathrm{rank}\begin{bmatrix} j\omega - A & B_2 B_2^T & B_1 \\ C_1'^T C_1' & j\omega + A^T & 0 \end{bmatrix} = 2n \quad \forall\omega\in\mathbb{R} \quad (A.8)$$

and:

$$\mathrm{rank}\begin{bmatrix} j\omega - A & B_2 B_2^T \\ C_1'^T C_1' & j\omega + A^T \\ -C_1' & 0 \\ 0 & D_{12}' B_2^T \end{bmatrix} = 2n \quad \forall\omega\in\mathbb{R} \quad (A.9)$$

We have by assumption that (A, B_2, C_1') is stabilizable and detectable (section 3), so:

$$\mathrm{rank}[j\omega - A\ \ B] = \mathrm{rank}\begin{bmatrix} j\omega - A \\ C_1' \end{bmatrix} = \mathrm{rank}[j\omega + A^T\ \ C_1'^T]$$
$$= \mathrm{rank}\begin{bmatrix} j\omega + A^T \\ B_2^T \end{bmatrix} = n \quad \forall\omega\in\mathbb{R} \qquad (A.10)$$

First consider controllability of $(H, \begin{bmatrix} B_1 \\ 0 \end{bmatrix})$. It is clear from eq. 4.9 and Fig. 2 that any choice of B_1 must give an L_2–optimal trajectory; so we can choose $B_1 = I$ and consider:

$$\mathrm{rank}\begin{bmatrix} j\omega - A & B_2 B_2^T & I \\ C_1'^T C_1' & j\omega + A^T & 0 \end{bmatrix} = 2n \quad \forall\omega\in\mathbb{R} \quad (A.11)$$

which holds if $\mathrm{rank}[j\omega + A^T\ \ C_1'^T] = n$.

Next consider observability of $(H, \begin{bmatrix} C_1' & 0 \\ 0 & -D_{12}' B_2^T \end{bmatrix})$.

Suppose eq. A.9 does not hold; then there exists a certain $\omega=\omega_1$ and a vector $x=\begin{bmatrix}x_1\\x_2\end{bmatrix}$ such that:

$$\begin{aligned}(j\omega_1-A)x_1 \;+\; & B_2B_2^Tx_2 &= 0\\ C_1'^TC_1'x_1 \;+\; & (j\omega_1-A^T)x_2 &= 0\\ -C_1'x_1 & &= 0\\ & D_{12}'B_2^Tx_2 &= 0\end{aligned}\tag{A.12}$$

So the last two equations give $C_1'x_1=0$ and $B_2^Tx_2=0$ such that from eq. A.10 we must have $(j\omega_1-A)x_1\neq0$ and $(j\omega_1+A^T)x_2\neq0$. This then clearly contradicts the first two equations of A.12.

Second part. Define:

$$J := \begin{bmatrix}0 & I\\ -I & 0\end{bmatrix}\tag{A.13}$$

such that:

$$\begin{aligned}JHJ^{-1} &= \begin{bmatrix}-A^T & -C_1'^TC_1'\\ -B_2B_2^T & A\end{bmatrix} = -H^T \quad\Leftrightarrow\\ H &= J^{-1}(-H^T)J\end{aligned}\tag{A.14}$$

Now suppose H can be decomposed as $H=M\Lambda M^{-1}$ with Λ in blockdiagonal Jordan form. Substitution of eq. A.14 then gives:

$$\begin{aligned}H &= J^{-1}(-H^T)J = M\Lambda M^{-1} \quad\Leftrightarrow\\ H^T &= JM\cdot(-\Lambda)\cdot M^{-1}J^{-1}\end{aligned}\tag{A.15}$$

Furthermore we also have

$$H^T = (M\Lambda M^{-1})^T = (M^{-1})^T\cdot\Lambda^T\cdot M^T\tag{A.16}$$

such that:

$$\begin{aligned}JM\cdot(-\Lambda)\cdot(JM)^{-1} &= (M^{-1})^T\cdot\Lambda^T\cdot M^T \quad\Leftrightarrow\\ \Lambda &= (M^TJM)^{-1}\cdot(-\Lambda^T)\cdot M^TJM \quad\Leftrightarrow\\ \Lambda &= T^{-1}(-\Lambda^T)T\end{aligned}\tag{A.17}$$

This implies that $-\Lambda^T$ and Λ are similar; so if Λ_i is a Jordan block in Λ with eigenvalue λ_i, then $-\Lambda^T$ must have a diagonal block $-\Lambda_j^T$ such that $\Lambda_i = T_i^{-1}(-\Lambda_j^T)T_i$. Now $-\Lambda_j^T$ came from a Jordan block Λ_j of Λ, so it has $-\lambda_j$ on the diagonal. Therefore we must have $\lambda_j=-\lambda_i$, such that the first part of the lemma implies $\Lambda_j\neq\Lambda_i$. Λ thus contains for each Jordan block Λ_i a second, equally large Jordan block Λ_j such that $\lambda_j=-\lambda_i$.

A.3 Proof of $e=0$ in theorem 2

Given eq. 5.13 it will be proven that there is no nonzero signal $e(t)\in\mathcal{L}_2[0,\infty)$ ($e(s)\in\mathbb{R}H_2$) that results in a smaller 2–norm of z than with $e=0$.

Consider eq. 5.13 in the frequency domain:

$$\begin{aligned}z_1(s) &= C_1'(sI-A+B_2B_2^TX)^{-1}\cdot\{B_1w_0+B_2e(s)\}\\ z_2(s) &= -D_{12}'B_2^TX(sI-A+B_2B_2^TX)^{-1}\cdot\{B_1w_0+B_2e(s)\}\\ &\quad + D_{12}'e(s)\end{aligned}\tag{A.18}$$

and define $S := (sI-A+B_2B_2^TX)^{-1}$.
The influence of $e(s)$ on the 2–norm of $z(s)$ is then determined by $z^*(j\omega)z(j\omega)$ (dropping the dependency on $j\omega$):

$$\begin{aligned}z^*z = {}& w_0^TB_1^TS^*C_1'^TC_1'SB_1w_0 + w_0^TB_1^TS^*C_1'^TC_1'SB_2e +\\ & e^*B_2^TS^*C_1'^TC_1'SB_1w_0 + e^*B_2^TS^*C_1'^TC_1'SB_2e +\\ & w_0^TB_1^TS^*XB_2B_2^TXSB_1w_0 - w_0^TB_1^TS^*XB_2(I-B_2^TXSB_2)e\\ & -e^*(I-B_2^TS^*XB_2)B_2^TXSB_1w_0 +\\ & e^*(I-B_2^TS^*XB_2)(I-B_2^TXSB_2)e\end{aligned}\tag{A.19}$$

So e decreases the 2–norm of z if and only if:

$$\begin{aligned}& w_0^T\{B_1^TS^*[C_1'^TC_1'+XB_2B_2^TX]SB_2 - B_1^TS^*XB_2\}e +\\ & e\{B_2^TS^*[C_1'^TC_1'+XB_2B_2^TX]SB_1 - B_2^TXSB_1\}w_0 +\\ & eB_2^TS^*[C_1'^TC_1'+XB_2B_2^TX]SB_2e + e^*e - e^*B_2^TS^*XB_2e\\ & - e^*B_2^TXSB_2e < 0\end{aligned}\tag{A.20}$$

The algebraic Riccati equation (lemma 2 part 4) now gives:

$$\begin{aligned}& C_1'^TC_1'+XB_2B_2^TX=\\ &\qquad -(j\omega I+A^T-XB_2B_2^T)X + X(j\omega I-A+B_2B_2^TX)\\ &\qquad = S^{*-1}X + XS^{-1}\end{aligned}\tag{A.21}$$

Substitution of eq. A.21 in eq. A.20 then gives:

$$w_0^TB_1^TXSB_2e + eB_2^TS^*XB_1w_0 + e^*e < 0\tag{A.22}$$

Now consider the signal $w_0'(s) := B_2^TS^{\sim}XB_1w_0$. From lemma 2 part 5 we know that $A-B_2B_2^TX$ is stable, thus $S=(sI-A+B_2B_2^TX)^{-1}\in\mathbb{R}H_\infty$ and $S^{\sim}\in\mathbb{R}H_\infty^\perp$. Therefore it is clear that $w_0'(s)\in\mathbb{R}H_2^\perp$ such that:

$$\langle e(s),w_0'(s)\rangle=\langle w_0'(s),e(s)\rangle=0 \quad \forall e(s)\in\mathbb{R}H_2\tag{A.23}$$

This implies that there is no choice of $e(s)\in\mathbb{R}H_2$ such that eq. A.22 holds and thus the 2–norm of z reaches its minimum when $e=0$.

REFERENCES

Athans, M. and P.L.Falb (1966). *Optimal control, an introduction to the theory and its applications.* McGraw–Hill, New York.

Brockett, R.W. (1970). *Finite dimensional linear systems.* Wiley, New York.

Doyle, J.C., K.Glover, P.P.Khargonekar and B.A.Francis, (1989). State–space solutions to standard H_2 and H_∞ control problems, *IEEE Trans. Aut. Contr.*, vol.AC–34, no.8, pp.831–847.

Faibusovich, L.E. (1987). Matrix Riccati inequality: existence of solutions. *Syst. Control Lett.*, vol.9, pp.59–64.

Francis, B.A. (1982). On the Wiener–Hopf approach to optimal feedback design. *Syst. Control Lett.*, vol.2, pp.197–201.

Francis, B.A. (1987). A course in H–infinity control theory. *Lect. Notes Cont. Inf. Sc.*, vol.88, Springer Verlag, Berlin.

Kalman, R.E. (1960). On the general theory of control systems. *Proc. first IFAC Congr., Moscow,* vol.1, Ed.J.F.Coales, Butterworths, London, pp.481–491.

MacFarlane, A.G.J. (1963). An eigenvector solution of the optimal linear regulator problem. *J. Electron. Contr.*, vol.14, pp.496–501.

MacFarlane, A.G.J. (1969a). Dual–system methods in dynamical analysis. Part 1; variational principles and their application to nonlinear–network theory. *IEE Proc.*, vol.116, no.8, pp.1453–1457.

MacFarlane, A.G.J. (1969b). Dual–system methods in dynamical analysis. Part 2; optimal regulators and optimal servomechanisms. *IEE Proc.*, vol.116, no.8, pp.1458–1462.

Mårtensson, K. (1971). On the matrix Riccati equation. *Inf. Sc.*, vol.3, pp.17–49.

Paley, R.E.A.C. and N.Wiener (1934). *Fourier transforms in the complex domain*, Am. Math. Soc., Providence, Road Island.

Potter, J.E. (1966). Matrix quadratic solutions. *SIAM J. Appl. Math.*, vol.14, no.3, pp.496–501.

Richardson, T.J. and R.H.Kwong (1986). On positive definite solutions to the algebraic Riccati equation. *Syst. Control Lett.*, vol.7, pp.99–104.

Rosenbrock, H.H. (1970). *State–space and multivariable theory*, Nelson, London.

Willems, J.C. (1971). Least squares stationary optimal control and the algebraic Riccati equation. *IEEE Trans. Aut. Contr.*, vol.AC–16, no.6, pp.621–634.

Wonham, W.M. (1978). *Linear multivariable control: a geometric approach.* Springer Verlag, Berlin.

Zames, G. (1981). Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Trans. Aut. Contr.*, vol.AC–26, no.2, pp.301–320.

CONTENTS Volume 1, April 1990