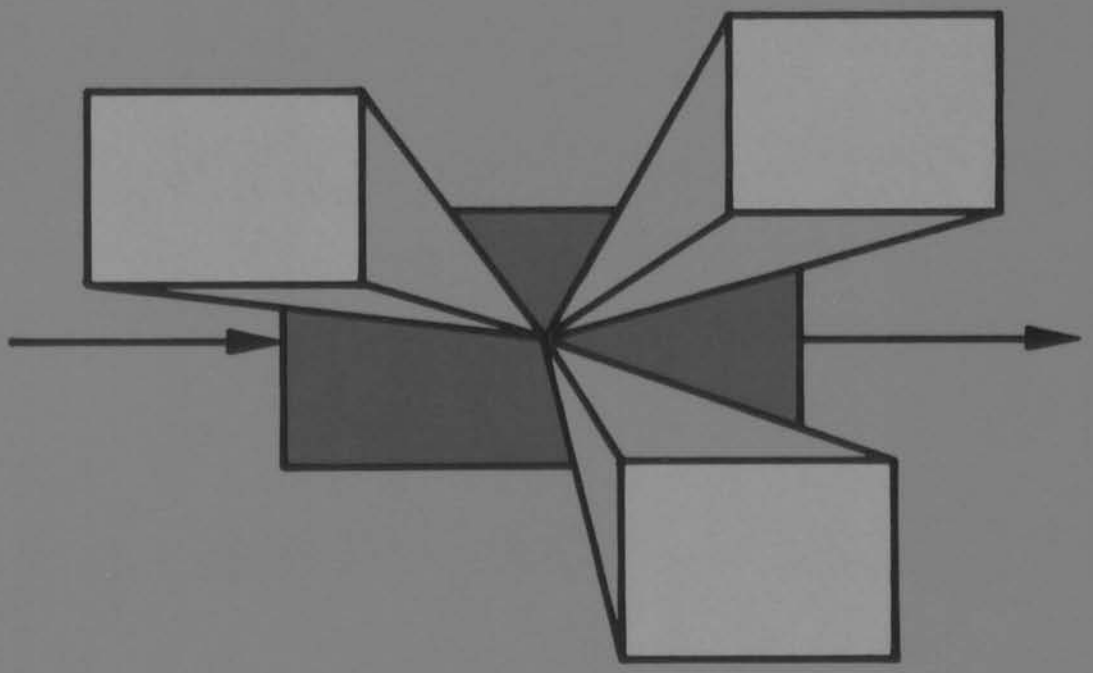
SELECTED
TOPICS
IN

Identification Modelling and Control

Volume 3, 1991

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 3, June 1991

Mechanical Engineering Systems and Control Group
Delft University of Technology

Delft University Press/1991

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

Mechanical Engineering Systems and Control Group
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 identification, modelling and control:
progress report on research activities in the mechanical engineering
systems and control group. – Delft: Mechanical Engineering Systems and Control Group,
Delft University of Technology, Vol. 3 – ed. by O.H. Bosgra and
P.M.J. Van den Hof. – ill. Met lit.opg.
ISBN 90-6275-702-2
SISO 656 UDC 531.7+681.5 NUGI 841

Cover design by Ruud Schrama

Contents

Volume 3, June 1991

Editorial

This is the third issue of Selected Topics in Identification, Modelling and Control. The publication has shown to serve its purpose both as a progress report on research in our group and as an informal and fast means of publishing research results most of which will eventually appear in the open literature in revised or expanded form after the unevitable publication delays.

The present issue contains a balanced mixture of theoretical and application-oriented papers. The application areas include aircraft flight control, the dynamics of mechanical structures, the process dynamics of a power station boiler and the dynamic behaviour of an industrial crystallization plant. Samir Bennani, Bob Mulder (both from the Department of Aerospace Engineering at Delft University) and Ton van der Weiden discuss the design of a robust aircraft flight control system using H_∞ techniques. Pepijn Wortelboer (now with Philips Research Labs., Eindhoven, the Netherlands) presents results on infinity norm computation for very large dynamic models of mechanical systems. Hans Heintze and Henk Aling (now with Integrated Systems Inc., Santa Clara, CA) report on their extensive work involving system identification of a power station boiler under closed-loop experimental conditions. Sjoerd de Wolf and Paul van den Hof report on the system identification of an industrial crystallization plant based on experiments on a large pilot plant located at the Laboratory for Process Equipment at Delft University.

The theoretical part of this issue is covered by papers on robust control and identification. Ruud Schrama and Peter Bongers investigate the possibilities to assess the stability robustness of closed loop systems on the basis of experimental data. Peter Bongers and Okko Bosgra present an algorithm for the computation of normalized coprime factorizations for possibly nonproper systems. Paul Lambrechts and Okko Bosgra provide a parametrization for controllers that achieve output regulation and tracking in the face of persistent disturbances and reference trajectories. A solution for controller reduction with guaranteed stability margins is derived in a paper by Peter Bongers and Okko Bosgra. Finally, Richard Hakvoort presents a theory for the assessment of an upper bound of the model uncertainty in system identification experiments.

Okko Bosgra
Paul Van den Hof
Editors

Selected Topics in Identification, Modelling and Control
Vol. 3, June 1991

Experimental robustness analysis based on coprime factorizations

Ruud J.P. Schrama and Peter M.M. Bongers

Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. The performance of an existing feedback system may be increased by a redesign of the compensator. Before a new compensator is applied to the plant we like to ascertain that the new feedback system will at least not be unstable. In fact we wish to make sure whether the actual performance will be increased. Our approach to this problem is based on the use of a recently developed accurate robustness margin conceived in terms of a coprime factorization of the plant. We apply this robustness margin in conjunction with the frequency response of a coprime factorization of the unknown plant. We indicate how to estimate such a frequency response from input and output measurements of the plant, while the latter still operates under the "old" feedback. We use this frequency domain data to determine the distance between the plant and its model in the sense of the robustness margin. This results in a spectral estimate of an upper bound on the actual distance between plant and model. Stability of the plant under feedback by the new compensator can be guaranteed if this upper bound is smaller one.

Keywords. feedback control; stability criteria; frequency response estimation coprime factorization.

1 Introduction

If an existing feedback system does not meet the performance requirements, then we can try to design a better compensator. Before such a new compensator is applied to the actual plant, we want to be sure that the plant will be stabilized by this new compensator. A common approach to ascertaining stability of the plant under the new feedback starts from describing the plant by a so-called nominal model and a perturbation. The *nominal model* is a simple description of the characteristic dynamics of the plant. The difference between this model and the plant dynamics is regarded as an uncertainty or a perturbation of the nominal model. We denote this difference as the *deficiency* of the nominal model. Further this deficiency is usually characterized by an upper bound. Now we can use the nominal model and the deficiency to make sure whether the new compensator will stabilize the plant. For stability is guaranteed if the feedback system composed of the nominal model and the new compensator is robustly stable in the face of the deficiency.

By conducting robust control design the set of allowable perturbations, i.e. perturbations that still yield an acceptable feedback system, is maximized in terms of a robustness margin. If a reliable upper bound on the deficiency at hand is smaller than this margin, then stability is guaranteed. It is well known that large robustness margins and a good performance are conflicting requirements (Doyle and Stein, 1981; Maciejowski, 1989). Thus in case the upper bound on the deficiency is conservative, it is not likely that stability can be guaranteed under high performance feedback. Since high performance is desired, we need more accurate information on the actual deficiency and we might have to remodel the plant. However first we like to

subject the designed control system to a less conservative robustness analysis. Moreover in case the design paradigm is *not* robust control, i.e. the design does not explicitly provide a robustness margin, then such a robustness analysis is definitely required.

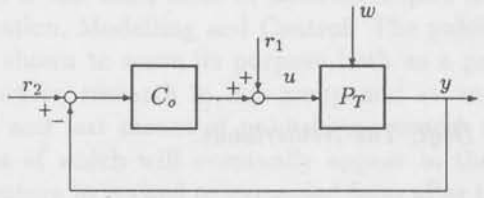


Fig. 1: Feedback system $H(P_T, C_o)$.

We consider the feedback system $H(P_T, C_o)$ of Fig. 1 in which the plant P_T of interest operates under feedback[1] by a known compensator C_o. The subscript T indicates the presence of Two vector inputs to the plant: there is an inner-loop plant input u and an exogeneous signal w. The latter is incorporated just to tolerate possible disturbances. Of main interest is the relation between u and y. This relation is called the inner-loop plant, which is signified by P. We assume that a nominal model P_o of P has been used to design a new and hopefully better compensator C. Prior to the implementation of C we have to assess the robust stability of the feedback system $H(P_o, C)$ in the face of the deficiency of P_o.

A well established means for the ascertainment of stability is the small gain theorem. One of its most widespread applications in linear systems theory lies with multiplicative uncertainties. In Doyle and Stein (1981) a robustness margin has been derived in terms of an upper bound on the multiplicative uncertainty. Notice that there exist plants that are stabilized by C, while the corresponding deficiency is not contained within the largest ball of allowed uncertainties. In this sense robustness *margins* are conservative in view of robustness *analysis*. A typical example hereof is presented in Section 6.

A drawback of the robustness margin based on multiplicative uncertainties is that the nominal model P_o must have as many unstable poles as the inner-loop plant P. However unless $C_o = 0$ we cannot be sure about the number of unstable poles of P. Thus we wish to obviate the above restriction. This is accomplished by expressing the feedback system in terms of coprime factorizations and then applying the small gain theorem. In this paper we will utilize the robustness margin of Bongers (1991a), which is less conservative than a similar margin in gap-metric sense. Furthermore for reasons that are made clear in Section 3, this robustness margin is called the compensator-gap.

The ascertainment of stability by means of the compensator-gap requires the knowledge of a coprime factorization of the unknown plant. Such knowledge can be obtained from experimental data by means of a recently developed framework for the identification of a plant under known feedback Schrama (1991b). This framework enables the estimation of the frequency response of a particular right coprime factorization of the inner-loop plant from measurements of input u and output y, while the plant operates under the initial feedback (see Fig. 1). No information on the signals r_1, r_2 and w is needed. Thus a frequency response is obtained under the initial feedback and then used to conduct robustness analysis in regard of the new compensator. In this paper we do not pursue guaranteed upper bounds or confidence regions for the estimated frequency responses as in Helmicki *et al.* (1991) and Webb *et al.* (1989) respectively Yuan and Ljung (1984), Loh *et al.* (1985) and Zhu (1990). In Section 5 we will make clear that such specific identification techniques can be readily applied within our framework.

The paper is organized as follows. The next section introduces notation and summarizes several concepts of the algebraic theory. In Section 3 we discuss robustness margins in terms of coprime factors and we incorporate the compensator-gap introduced in Bongers (1991a). Subsequently in Section 4 we recall the framework for identification of a plant under feedback from Schrama (1991b). Further we show how this framework can be used to estimate the frequency response of a coprime factorization of the plant. Then Section 5 considers the determination of an upper bound on the compensator-gap from the estimated frequency responses. Section 6 illustrates the utility of our experimental robustness analysis by means of an example. Concluding remarks are listed in Section 7.

2 Notation and preliminaries

The basic results in the algebraic theory of linear time-invariant finite dimensional systems have been derived for the so-called single-variate control system $H(P, C)$ comprising an inner-loop plant P and a compensator C (Desoer *et al.*, 1980). This feedback system equals $H(P_T, C)$ of Fig. 1 except that the plant lacks the second vector input w. Note

[1] Operation in open-loop is a special case: $C_o = 0$.

that both control systems are signified by $H(P_{..}, C)$ and the index of the plant-term obviates any ambiguity. Further the signification P is used to indicate the inner-loop plant as well as its transfer function $P(s)$.

In the algebraic theory of coprime factorizations a plant P is expressed as the ratio of two stable proper mappings: $P = ND^{-1}$, with N and D proper stable rational functions. Particular classes of factors N and D are specified below.

Definition 2.1 (Vidyasagar, 1985)

i. *Let N, D be stable, then the pair (N, D) is called* right coprime *if there exist stable X, Y such that $XN + YD = I$. Analogously $(\tilde{D}, \tilde{N})$ is* left coprime *if there exist stable $\tilde{X}, \tilde{Y}$ such that $\tilde{N}\tilde{X} + \tilde{D}\tilde{Y} = I$.*

ii. *(N, D) is a* right coprime factorization (rcf) *of P if $\det D \neq 0$, $P = ND^{-1}$ and (N, D) is right coprime. Analogously $(\tilde{D}, \tilde{N})$ is a* left coprime factorization (lcf) *of the P if $\det \tilde{D} \neq 0$, $P = \tilde{D}^{-1}\tilde{N}$ and $(\tilde{D}, \tilde{N})$ is left coprime.*

iii. *(N_n, D_n) is called a* normalized *right coprime factorization (nrcf) of P if it is a rcf of P and additionally $N_n^* N_n + D_n^* D_n = I$ with $N_n^*(s)$ the transpose of $N_n(-s)$. Analogously $(\tilde{D}_n, \tilde{N}_n)$ is a normalized lcf of P if it is a lcf of P and additionally $\tilde{N}_n \tilde{N}_n^* + \tilde{D}_n \tilde{D}_n^* = I$.*

Before we recall the main stability theorem we introduce "$\Lambda \in \mathcal{J}$" as a shorthand notation for the paraphrase "Λ is stable and has a stable inverse" (see Vidyasagar (1985) for the ring theoretic meaning of $\mathcal{J}$).

Theorem 2.2 (Vidyasagar *et al.*, 1982) *Let plant P and compensator C have a rcf (N, D) respectively a lcf $(\tilde{D}_c, \tilde{N}_c)$, then the feedback system $H(P, C)$ is stable if and only if $\Lambda \in \mathcal{J}$ with Λ defined as*

$$\Lambda = \tilde{N}_c N + \tilde{D}_c D. \tag{1}$$

With ND^{-1} and $\tilde{D}_c^{-1}\tilde{N}_c$ substituted for P respectively C a block diagram of $H(P, C)$ can be drawn as in Fig. 2. In here $\xi = \tilde{D}_c r_1 + \tilde{N}_c r_2$ and $x = \Lambda^{-1}\xi$. Hence stability of Λ^{-1} and boundedness of ξ guarantee x and all other signals to be bounded.

3 Sufficient conditions for robust stability

This section discusses conditions that guarantee robust stability of the newly designed feedback system $H(P_o, C)$ in the face of the deficiency of the

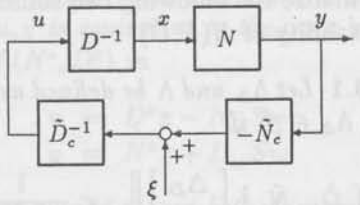


Fig. 2: Fractional representation of $H(P, C)$.

nominal model P_o. In order to examine robust stability in terms of coprime factorizations we adopt the next expression from Vidyasagar and Kimura (1986). The feedback transfer function $T(P, C)$ is defined as

$$T(P,C) = \begin{bmatrix} -P \\ I \end{bmatrix} (I + CP)^{-1} \begin{bmatrix} C & I \end{bmatrix}, \tag{2}$$

which maps r_2 and r_1 into y and u. The feedback system $H(P, C)$ is stable if and only if its transfer function $T(P, C)$ is stable (Vidyasagar and Kimura, 1986). We express $T(P, C)$ in terms of the rcf (N, D) of P and the lcf $(\tilde{D}_c, \tilde{N}_c)$ of C.

$$T(P,C) = \begin{bmatrix} -N \\ D \end{bmatrix} (\tilde{D}_c D + \tilde{N}_c N)^{-1} \begin{bmatrix} \tilde{N}_c & \tilde{D}_c \end{bmatrix}.$$

As proposed in Bongers (1991) we replace the rcf (N, D) by the factorization $P = (NQ)(DQ)^{-1}$ with some stable Q:

$$T(P,C) = \begin{bmatrix} -NQ \\ DQ \end{bmatrix} (\tilde{D}_c DQ + \tilde{N}_c NQ)^{-1} \begin{bmatrix} \tilde{N}_c & \tilde{D}_c \end{bmatrix}. \tag{3}$$

Since Q is stable, the two block-matrices at the right hand side of (3) are stable. Consequently if $(\tilde{D}_c DQ + \tilde{N}_c NQ)^{-1}$ is stable, then $T(P, C)$ is guaranteed stable. Since the inverse of Q is not necessarily stable, this condition is sufficient but possibly not necessary for stability of $H(P, C)$.

Next we implicitly express (NQ, DQ) as a stable perturbation of the rcf (N_o, D_o) of the nominal model P_o:

$$\begin{bmatrix} \Delta_D \\ \Delta_N \end{bmatrix} = \begin{bmatrix} D \\ N \end{bmatrix} Q - \begin{bmatrix} D_o \\ N_o \end{bmatrix} \tag{4}$$

which will be signified as $\mathrm{col}(\Delta_D, \Delta_N)$. Now stability of $H(P, C)$ is guaranteed if the designed $H(P_o, C)$ is robustly stable in the face of the stable factor perturbations Δ_D, Δ_N of (4). In fact these perturbations Δ_D, Δ_N embody the deficiency of the nominal model. By defining Λ_Δ as

$$\begin{aligned} \Lambda_\Delta &= \tilde{D}_c DQ + \tilde{N}_c NQ \\ &= (\tilde{D}_c D_o + \tilde{N}_c N_o) + \begin{bmatrix} \tilde{D}_c & \tilde{N}_c \end{bmatrix} \begin{bmatrix} \Delta_D \\ \Delta_N \end{bmatrix} \end{aligned} \tag{5}$$

we can formalize the following two sufficient conditions for stability of $H(P,C)$.

Lemma 3.1 *Let Λ_Δ and Λ be defined as in* (5) *and* (1). *Then $\Lambda_\Delta \in \mathcal{J}$ if*

$$\left\| \begin{bmatrix} \tilde{D}_c & \tilde{N}_c \end{bmatrix} \begin{bmatrix} \Delta_D \\ \Delta_N \end{bmatrix} \right\|_\infty < \frac{1}{\|\Lambda^{-1}\|_\infty} \qquad (6)$$

provided that the designed compensator C stabilizes the nominal model P_o.

Proof: By Theorem 2.2 stability of $H(P_o,C)$ implies $\Lambda \in \mathcal{J}$. With $\Lambda \in \mathcal{J}$ the inversion lemma 2.2.19 of says that $\Lambda_\Delta \in \mathcal{J}$ if $1/\|\Lambda_\Delta - \Lambda\|_\infty < \|\Lambda^{-1}\|_\infty$. Substitution of (5) and (1) in the left hand side of this inequality yields (6). □

Theorem 3.2 *Let $H(P_o,C)$ be stable and let (N_o,D_o), (N_c,D_c) be rcf's of P_o respectively C such that $\Lambda = I$. Further let plant P have rcf (N,D) and Δ_D, Δ_N be as in* (4). *Then the feedback system $H(P,C)$ is stable if*

$$\|\begin{bmatrix} \tilde{D}_c & \tilde{N}_c \end{bmatrix}\|_\infty \left\| \begin{bmatrix} \Delta_D \\ \Delta_N \end{bmatrix} \right\|_\infty < 1. \qquad (7)$$

Proof: From (3) we know that $H(P,C)$ is stable if $\Lambda_\Delta \in \mathcal{J}$ with Λ_Δ as in (5). Since Lemma 3.1 holds for any (N_o,D_o), $(\tilde{D}_c,\tilde{N}_c)$ we may choose these factorizations such that $\Lambda = I$ without loss of generality. And since all elements in the left hand side of (6) are stable, we may apply the triangle inequality to accomplish (7). □

In Bongers (1991b) it is shown that the robustness margin in (6) is the least conservative if $\Lambda = \alpha I$, where α may be any real scalar. This supports the choice of $\Lambda = I$ in Theorem 3.2. By this theorem we can ascertain robust stability of $H(P_o,C)$ in the face of the deficiency of P_o as follows. We minimize Δ_D, Δ_N of (4) by selecting an appropriate stable Q. Notice that not only the magnitude but also the phase of Q is of importance here. Stability of $H(P,C)$ is guaranteed as soon as a stable Q has been found, such that the inequality (7) is satisfied. Bongers (1991a) showed that the robustness margin of Theorem 3.2 equals the directed gap in case the nominal *model* factors are normalized, and that the robustness margin of (7) is less conservative if the *compensator* factors are normalized. The result is a new robustness margin called the compensator-gap.

Corollary 3.3 (Bongers, 1991a) *Let compensator C stabilize the nominal model P_o. Let $(\tilde{D}_{nc},\tilde{N}_{nc})$ be a nlcf of C and let (N_o,D_o) be a rcf of P_o such that $\tilde{D}_{nc}D_o + \tilde{N}_{nc}N_o = I$. Then C stabilizes P if*

$$\left\| \begin{bmatrix} \Delta_D \\ \Delta_N \end{bmatrix} \right\|_\infty < 1 \qquad (8)$$

with Δ_D, Δ_N as in (4) *for any stable Q. The infimum of the H_∞-norm bound in* (8) *over all stable Q's is called the compensator-gap.*

This new metric is similar to the gap-metric, except that it depends on the compensator at hand. We will use this compensator-gap in our robustness analysis based on experimental data.

4 Frequency response estimation

In this section we show how measurements of u and y can be used to estimate the frequency response of a particular rcf of the inner-loop plant P. These measurements are taken while the plant operates under feedback by the known compensator C_o. In order to provide for cases where the plant of interest is affected by noise we consider the plant P_T of Fig. 1. Since P_T is stabilized by C_o, it is an element of the set of *all* plants, that are stabilized by C_o. In Schrama (1991b) this set has been parameterized in terms of coprime factorizations.

Theorem 4.1 (Schrama, 1991b) *Let the nominal model P_o of the inner-loop plant P be stabilized by compensator C_o. Let (N_o,D_o) and (N_{co},D_{co}) be a rcf of P_o respectively C_o. Define (N_p,D_p) as*

$$\begin{aligned} N_p &\doteq \begin{bmatrix} D_{co}S & N_o + D_{co}R \end{bmatrix} \\ D_p &\doteq \begin{bmatrix} I & 0 \\ -N_{co}S & D_o - N_{co}R \end{bmatrix} \end{aligned} \qquad (9)$$

with stable R,S such that $\det(D_o - N_{co}R) \neq 0$. Then the set of all plants P_T, that make a stable feedback system $H(P_T,C)$ like in Fig. 1, can be parameterized in the form

$$\mathcal{P}(C) = \left\{ P_T = N_p D_p^{-1} \mid (N_p,D_p) \text{ as in } (9) \right\}.$$

This parameterization of the set $\mathcal{P}(C)$ is called the *(R,S)-parameterization* and it has been depicted in Fig. 3. We define the *associated* coprime factorization (N^o,D^o) as

$$\begin{aligned} N^o &= N_o + D_{co}R \\ D^o &= D_o - N_{co}R \end{aligned} \qquad (10)$$

and by expanding $P_T = N_p D_p^{-1}$ we get $P = N^o(D^o)^{-1}$. In Schrama (1991b) it is shown that every pair of C_o and (N_o,D_o) induce an unique rcf

(N^o, D^o) of the inner-loop plant P. In the sequel we consider the identification of this particular rcf of P. This identification is based on the variable x, that appears in Fig. 3 in between D_o^{-1} and N_o. Two important properties of this variable x, called the *intermediate*, are listed below.

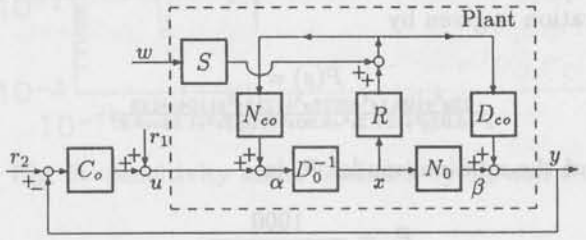


Fig. 3: (R,S)-parameterization of P_T in a stable feedback system $H(P_T, C_o)$.

Lemma 4.2 (Schrama, 1991b) *Let input u and output y in $H(P_T, C_o)$ of Fig. 1 be measured. Let C_o and (N_o, D_o) of Fig. 3 be known. Then the intermediate x can be reconstructed from the measurements of u and y via*

$$x = (D_o + C_o N_o)^{-1}(u + C_o y) \qquad (11)$$

without any knowledge of the plant P_T, except the fact that $H(P_T, C_o)$ is stable.

The latter paraphrase supports the practical utility of our approach. Besides in Schrama (1991b) an alternative expression for (11) has been derived, which enables to reconstruct x by stable filters even if C_o is unstable. Inspection of the contribution of w to y and u reveals the next corollary.

Corollary 4.3 (Schrama, 1991b) *The intermediate x of Lemma 4.2 and outer-loop plant input w of $H(P_T, C_o)$ in Fig. 1 are uncorrelated, provided that each of the signals r_1 and r_2 is uncorrelated with w.*

In summary the artificial intermediate x can be reconstructed from measured u, y and it does not depend on the noise disturbance w. For a full explanation of this phenomenon we refer to Schrama (1991b). The next theorem will enable the frequency response estimation of the unknown plant rcf (N^o, D^o).

Theorem 4.4 (Schrama, 1991) *Let the feedback system $H(P_T, C_o)$ of Fig. 1 be stable and let compensator C_o with lcf $(\tilde{D}_{co}, \tilde{N}_{co})$ be known. Then with x reconstructed via Lemma 4.2 the* closed-loop *identification of the inner-loop plant P from measurements u, y is equivalent to the* open-loop *identification of (N^o, D^o) in*

$$\begin{aligned} u &= D^o x - N_{co} S w \\ y &= N^o x + D_{co} S w \end{aligned} \qquad (12)$$

provided that r_1 and r_2 are both uncorrelated with w.

Since the identification of N^o and D^o is conducted in open-loop we may apply non-parametric identification methods directly to the pairs x, y and x, u.

Corollary 4.5 *Let $H(P_T, C_o)$ with unknown P_T and known C_o be stable. Then the frequency response of the rcf (N^o, D^o) of P as in* (10) *can be estimated from the inner-loop signals u and y and the intermediate x.*

We end up by mentioning that the intermediate x can be specified a priori if the signals r_1 and r_2 are at our proposal (Schrama, 1991a).

5 Estimation of the compensator-gap

In Section 3 we discussed robustness analysis in regard of the new compensator. This ascertainment of stability requires information on a right coprime factorization of the unknown plant. By the framework of Section 4 we can obtain such information, while the plants operates under the initial feedback C_o. We will use the frequency responses of N^o and D^o over the frequency range of interest. In order not to obscure the key objectives we will provisionally assume that exact frequency responses are available. We return to this subject at the end of this section. Besides in this and subsequent section we consider only single-input single-output systems.

By Corollary 3.3 robust stability of $H(P_o, C)$ in the face of the deficiency $\mathrm{col}(\Delta_D, \Delta_N)$ is guaranteed if the H_∞-norm upper bound on the deficiency is smaller than 1. The H_∞-bound of a stable system is the maximum over all frequencies of the largest singular value of its frequency response. Hence we may as well consider the frequency response of the deficiency $\mathrm{col}(\Delta_D, \Delta_N)$, provided that the deficiency is stable. The latter condition can be satisfied as long as the *initial* feedback system is stable. Since stability of $H(P_T, C_o)$ implies the existence of the stable associated rcf (N^o, D^o) of (10), and hence by Corollary 4.5 we can take the (N^o, D^o) for (N, D) in (4). From there on any stable Q yields a stable deficiency $\mathrm{col}(\Delta_D, \Delta_N)$.

By (4) the frequency response of col(Δ_D, Δ_N) depends on the stable term Q, which is at our discretion. Motivated by Corollary 3.3 we seek for a stable Q such that the maximum singular value plot of col(Δ_D, Δ_N) is smaller than 1 for all frequencies. We start our search for an approximate stable Q with the determination of a lower bound of the singular value plot of the deficiency col(Δ_D, Δ_N): we let $\check{N}^o, \check{D}^o$ and $\check{N}_o, \check{D}_o$ be the frequency responses of (N^o, D^o) and (N_o, D_o) and for each frequency point i we minimize

$$|\check{D}_i^o q_i - \check{D}_{oi}|^2 + |\check{N}_i^o q_i - \check{N}_{oi}|^2 \qquad (13)$$

over the scalar $q_i \in \mathbb{C}$. This produces a vector $\check{Q}_q$ of scalars at length of the number of frequency points of interest.

Substitution of $\check{Q}_q$ for $\check{Q}$ in the frequency response analog of (4) yields the smallest maximum value of (13) over all frequencies. With any stable Q the maximum of the singular value plot of the deficiency will be larger or equal. Hence if (13) is larger than one for any frequency point i, then stability of $H(P, C)$ cannot be guaranteed. But if this maximum is smaller than 1, then we model $\check{Q}_q$ as good as possible by a stable Q_1. We substitute Q_1 for Q in (4) and check whether the inequality (8) is satisfied. If not then we model $\check{Q}_q/\check{Q}_1$ by a stable Q_2 and substitute $Q_1 Q_2$ for Q in (4), and repeat this procedure until (8) is satisfied. At every step we obtain an upper bound on the compensator-gap. Since $\check{Q}_q$ provides only a lower bound of this compensator-gap, the latter may happen to be larger than 1 precluding a robust stability assessment.

Finally we comment on the use of estimated frequency response data. As alluded to in the introduction certain identification techniques provide such an estimate together with a bounded region or confidence region, which contains the 'true' frequency response. We point out that these techniques are all designed for open-loop identification. However they can be applied to the feedback controlled plant since the framework of Section 4 enables an open-loop identification of the rcf (N^o, D^o). Such an estimate could be subjected to the above procedure with the regions plugged in the robustness analysis. A prerequisite is that the region is not too conservative, especially in case of outliers in the data (De Vries, 1991). In the example of the next section we will not utilize these techniques. Again we will assume that exact frequency response data is available, or at least that the accuracy of the estimates is such that estimation errors are neglectable. The latter can be achieved e.g. in cases where sine-wave experiments at various frequencies are allowed (Schrama, 1991b).

6 Example

This section describes the application of our approach to robustness analysis based on frequency response data. The inner-loop plant under consideration is given by

$$P(s) = \frac{10s^5+106.1s^4+2016s^3+1212s^2+1104s+129}{s^6+6.02s^5+101.2s^4+8.90s^3+110.0s^2+1.09s+9.87}$$

and the nominal model P_o is

$$P_o = \frac{1000}{s^2+6s+100}.$$

The magnitude Bode-plots of both P and P_o are depicted in Fig. 4. The compensator C_o currently

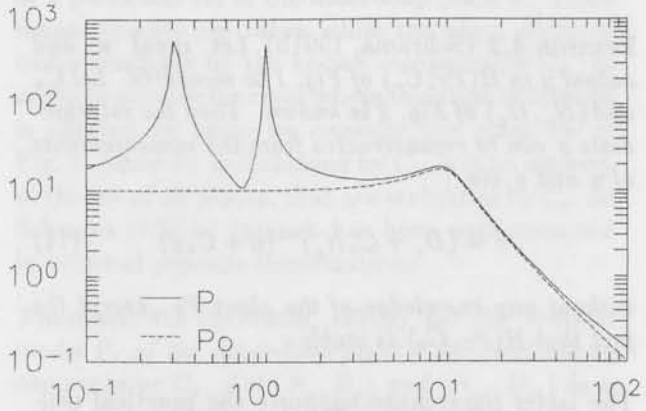


Fig. 4: Magnitude Bode-plots of P and P_o.

stabilizing the plant P is

$$C_o = 0.01\frac{4.70s+0.112}{s+0.0046}$$

and the new compensator is

$$C = \frac{1.971s+27.84}{s+71.81}.$$

Both compensators stabilize P as well as P_o.

The sensitivity function, i.e. the bottom right element of $T(P, C)$ in (2), has been drawn in Fig. 5 for P and P_o under feedback by C_o as well as C. Note that $(I+CP)^{-1}$ is not yet guaranteed to be stable but the "frequency response" can be calculated from $\check{N}^o, \check{D}^o$. We observe that the performance of the controlled plant has been improved (in the sense of a lower sensitivity at low frequencies) provided that $H(P, C)$ is stable. First we verify the robustness by means of a multiplicative uncertainty. From

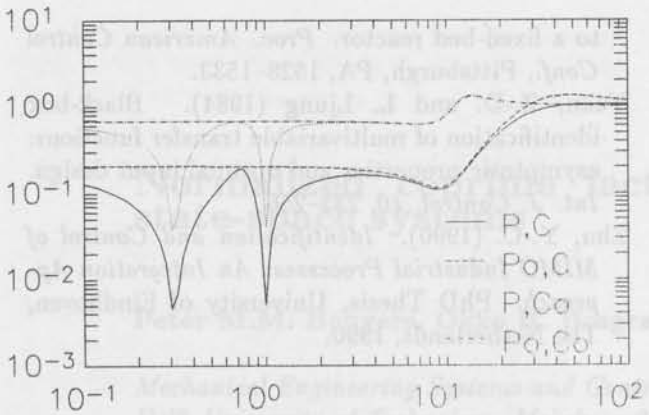


Fig. 5: Sensitivity under feedback by C_o and C.

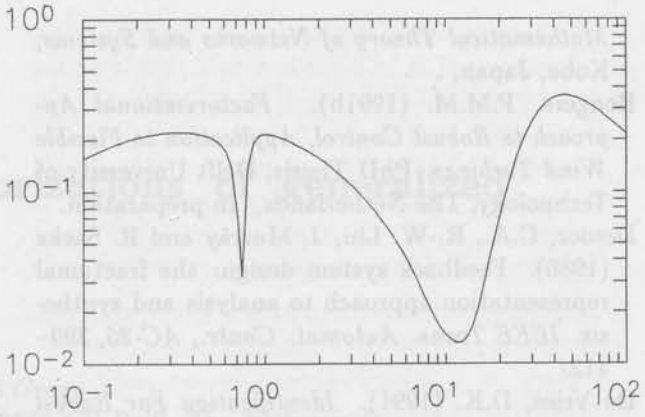


Fig. 7: Minimized coprime factors deficiency.

we know that $H(P,C)$ is stable if the multiplicative uncertainty is smaller than the inverse of the magnitude of the complementary sensitivity of $H(P_o,C)$ (i.e. the negative top left element of $T(P_o,C)$ of (2)). The multiplicative uncertainty has been calculated as $(\breve{P}/\breve{P}_o)-1$ and its magnitude is drawn in Fig. 6 together with the inverse complementary

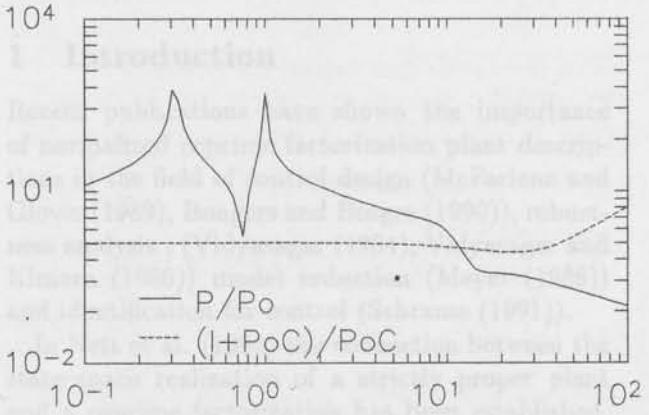


Fig. 6: Multiplicative uncertainty and robustness margin.

sensitivity. Clearly the deficiency of P_o expressed as a multiplicative uncertainty does not satisfy the condition for robust stability.

Now we turn to the robustness analysis in terms of coprime factorizations. By the procedure described in the previous section we estimate a stable Q of order 6 and subsequently another Q of order 4. Since we apply a general identification routine the estimated Q sometimes is unstable. Hence we have to eliminate the unstable part in some ad hoc manner. The corresponding maximum singular value of $|\Delta_D|^2+|\Delta_N|^2$ has been plotted in Fig. 7 (solid line). This curve embodies an upper bound on the compensator-gap. Clearly it is smaller than one over all frequencies and thus stability of $H(P,C)$ is guaranteed.

7 Concluding remarks

In this paper we proposed a solution to the problem of robustness analysis based on experimental data. The solution has been conceived in terms of coprime factorizations. First we have used coprime factors to establish robustness margins and consecutively a sufficient condition for robust stability. By this condition we can ascertain robust stability of a feedback system in the face of stable coprime factor perturbations. Secondly we have provided a means to estimate the frequency response of a particular coprime factorization of an unknown plant. This requires measurements of only the input and output of the plant, while it operates under known feedback. Then these frequency responses have been used to verify the condition for robust stability for this particular plant. That is, the frequency response data of the coprime factors were used to ascertain stability of the unknown plant under feedback by the new compensator. This application of the small gain theorem to the specific frequency response data is legitimate, since the underlying coprime factors of the unknown plant are stable. An example shows the utility of our approach.

Issues for future investigation are the development of a frequency domain identification method that yields only stable models as well as the robustness analysis with the application of identification techniques, that provide a bounded or confidence region for the frequency response estimates.

References

Bongers, P.M.M. (1991a). On a new robust stability margin. To appear at *9th Int. Symp. on*

Mathematical Theory of Networks and Systems, Kobe, Japan, .

Bongers, P.M.M. (1991b). *Factorizational Approach to Robust Control, Application to Flexible Wind Turbines.* PhD Thesis, Delft University of Technology, The Netherlands, In preparation.

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.

De Vries, D.K. (1991). *Identification For Robust Control: What is The Best Approach?.* Report Syst. and Control Group, Dept. Mech. Eng., Delft University of Technology, Delft, The Netherlands 1991.

Doyle, J.C. and G. Stein (1981). Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis. *IEEE Trans. Automat. Contr., AC-26,* 4–16.

Helmicki, A.J., C.A. Jacobson and C.N. Nett (1991). Fundamentals of control-oriented system identification and their application for identification in H_∞. To appear in *IEEE Trans. Automat. Contr.*

Loh, A.P., G.O. Corrêa and I. Postlethwaite (1985). Estimation of uncertainty bounds for robustness. *Control'85, Proc. IEE Int. Conf.,* Cambridge, England, 138–144.

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

Schrama, R.J.P. (1991a). An open-loop solution to the approximate closed-loop identification problem. To appear at *9th IFAC/IFORS Symp. on Identification and System Parameter Estimation,* Budapest, Hungary.

Schrama, R.J.P. (1991b). *Approximate Closed-Loop System Identification for Control Design with Application to Mechanical Systems.* PhD Thesis, Delft University of Technology, The Netherlands, In preparation.

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

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

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

Webb, C., H. Budman and M. Morari (1989). Identifying frequency domain uncertainty bounds for robust controller design - theory with application to a fixed-bed reactor. *Proc. American Control Conf.,* Pittsburgh, PA, 1528–1533.

Yuan, Z.-D. and L. Ljung (1984). Black-box identification of multivariable transfer functions: asymptotic properties and optimal input design. *Int. J. Control, 40,* 233–256.

Zhu, Y.-C. (1990). *Identification and Control of MIMO Industrial Processes: An Integration Approach.* PhD Thesis, University of Eindhoven, The Netherlands, 1990.

Selected Topics in Identification, Modelling and Control
Vol. 3, June 1991

Normalized coprime factorizations of generalized state-space systems

Peter M.M. Bongers, Okko H. Bosgra

Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. This note presents a state–space algorithm for the calculation of a normalized coprime factorization of continuous–time generalized dynamical systems. It will be shown that two Ricatti equations have to be solved to obtain this normalized coprime factorization.

Keywords. normalized coprime factorization, generalized systems, algorithm

1 Introduction

Recent publications have shown the importance of normalized coprime factorization plant descriptions in the field of control design (McFarlane and Glover (1989), Bongers and Bosgra (1990)), robustness analysis , (Vidyasagar (1984), Vidyasagar and Kimura (1986)) model reduction (Meyer (1988)) and identification for control (Schrama (1991)).

In Nett et al. (1984) the connection between the state–space realization of a strictly proper plant and a coprime factorization has been established. The coprime factorization of a generalized dynamical system was presented in Wang and Balas (1989). In Meyer and Franklin (1987) it has been shown that in order to calculate a normalized coprime factorization of a continuous–time strictly proper plant one Ricatti equation has to be solved, Vidyasagar (1988) extended these results to proper plants. For discrete–time proper systems the contruction of a normalized coprime factorization has been set up in Bongers and Heuberger (1990).

In this note we extend the results of (Meyer and Franklin (1987), Vidyasagar (1988)) to generalized dynamical systems. It will be shown that in the calculation of a normalized coprime factorization for systems in a generalized state–space form two Ricatti equations has to be solved instead of one Ricatti equation as in the case for systems in a common state–space form. The procedure to achieve a normalized coprime factorization for systems in generalized state–space form will be given as an explicit algorithm.

2 Preliminaries

In this note we adopt the ring theoretic setting of Desoer et al. (1980), Vidyasagar et al. (1982) to study stable multivariable linear systems. That is we consider a stable system as a transfer function matrix with all its entries belonging to the ring $\mathcal{H}$. We consider the class of possibly non–proper and/or unstable multivariable systems as transfer function matrices whose entries are elements of the quotient field $\mathcal{F}$ of $\mathcal{H}$ ($\mathcal{F} := \{a/b \mid a \in \mathcal{H},\ b \in \mathcal{H}\backslash 0\}$). For the application of our state-space algorithm we will identify the ring $\mathcal{H}$ with $\mathbb{RH}_\infty$, the space of stable real rational finite dimensional linear time–invariant continuous–time systems. The set of multiplicative units of $\mathcal{H}$ is defined as: $\mathcal{J} := \{h \in \mathcal{H} \mid h^{-1} \in \mathcal{H}\}$. In the sequel systems $P \in \mathcal{F}^{m\times n}$ are denoted as $P \in \mathcal{F}$.

Factorizations

Definition 2.1 (Vidyasagar et al. (1982))
A plant $P \in \mathcal{F}$ has a right (left) fractional representation if there exist $N, M(\tilde{N}, \tilde{M}) \in \mathcal{H}$ such that $P = NM^{-1}$ $(= \tilde{M}^{-1}\tilde{N})$.

*The pair $M, N(\tilde{M}, \tilde{N})$ is a right (left) coprime factorization (*rcf *or* lcf *) if it is a right (left) fraction and there exists $U, V(\tilde{U}, \tilde{V}) \in \mathcal{H}$ such that: $UN + VM = I$ $(\tilde{N}\tilde{U} + \tilde{M}\tilde{V} = I)$*

*The pair $M, N(\tilde{M}, \tilde{N})$ is called a normalized right (left) coprime factorization (*nrcf *or* nlcf *) if it is coprime and: $M^*M + N^*N = I$ $(\tilde{M}\tilde{M}^* + \tilde{N}\tilde{N}^* = I)$ with $M^* = M^T(-s)$.*

Proposition 2.2 *Let $P(s)$ be a real rational possibly non-proper transfer function having McMillan degree r. Then $P(s)$ can be represented by $P(s) = C(sE - A)^{-1}B$, where:*

$$E = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \; A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \; C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

with $A_{12}A_{22}^T = 0$, $A_{22}^T A_{21} = 0$ and both $B_2 B_2^T$, $C_2^T C_2$ non-singular. The matrix partitions are assumed to be compatible with the partitioning of E.

Proof: Let $P(s) = P_{sp}(s) + P_p(s)$ with $P_{sp} = \hat{C}(sI - \hat{A})^{-1}\hat{B}$ strictly proper and $P_p = \bar{C}(I - s\bar{J})^{-1}\bar{B}$ the polynomial part with $(\hat{A}, \hat{B}, \hat{C})$ and $(\bar{J}, \bar{B}, \bar{C})$ controllable and observable matrix triples and $\bar{J}$ in Jordan form (Rosenbrock (1974)). Then

$$\left[\begin{array}{c|c} I - s\bar{J} & \bar{B} \\ \hline -\bar{C} & 0 \end{array}\right] \xrightarrow{r.s.e.} \left[\begin{array}{cc|c} sI - \tilde{J}_{11} & -\tilde{J}_{12} & \tilde{B}_1 \\ -\tilde{J}_{21} & -\tilde{J}_{22} & B_2 \\ \hline -\tilde{C}_1 & -C_2 & 0 \end{array}\right]$$

where r.s.e. denotes an operation of restricted system equivalence (Rosenbrock (1974)). In this case the operations only involve interchanging rows and columns containing an s, and sign changes. Controllability/observability of systems in Jordan form implies non-singularity of $B_2 B_2^T$ and $C_2^T C_2$ (Chen and Desoer (1968)) . The Jordan form implies $\tilde{J}_{12}\tilde{J}_{22}^T = 0$ and $\tilde{J}_{22}^T \tilde{J}_{21} = 0$. Defining the partitioned system matrices as:

$$\left[\begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline C_1 & C_2 & 0 \end{array}\right] := \left[\begin{array}{ccc|c} \hat{A} & 0 & 0 & \hat{B} \\ 0 & \tilde{J}_{11} & \tilde{J}_{12} & \tilde{B}_1 \\ 0 & \tilde{J}_{21} & \tilde{J}_{22} & B_2 \\ \hline \hat{C} & \tilde{C}_1 & C_2 & 0 \end{array}\right]$$

proves the proposition □

3 Main result

The main result consists of two parts. First we will show that a nrcf of P is a full rank spectral factor of

$$\begin{bmatrix} I \\ P \end{bmatrix} (I + P^*P)^{-1} \begin{bmatrix} I & P^* \end{bmatrix} \qquad (1)$$

Secondly we will use this result to obtain a state–space realization of a nrcf of P. This will be presented in the form of an algorithm.

Theorem 3.1 *Let $P \in \mathcal{F}$ be given. Then the following statements are equivalent*

a) *(N, M) is a* nrcf *of P*

b) $\begin{bmatrix} M \\ N \end{bmatrix} \in \mathcal{H}$ *is a full rank spectral factor of* $\begin{bmatrix} I \\ P \end{bmatrix} (I + P^*P)^{-1} \begin{bmatrix} I & P^* \end{bmatrix}$

Proof: (a) → (b). Given an (N, M) as a nrcf of P. Then $\begin{bmatrix} M \\ N \end{bmatrix} \in \mathcal{H}$ is full rank and (1) can be written as:

$$\begin{bmatrix} I \\ NM^{-1} \end{bmatrix} \left(I + M^{*^{-1}} N^* N M^{-1}\right)^{-1} \begin{bmatrix} I & M^{*^{-1}} N^* \end{bmatrix}$$
$$= \begin{bmatrix} M \\ N \end{bmatrix} (M^*M + N^*N)^{-1} \begin{bmatrix} M^* & N^* \end{bmatrix}$$
$$= \begin{bmatrix} M \\ N \end{bmatrix} \begin{bmatrix} M^* & N^* \end{bmatrix}$$

which proves the first part of the theorem.

(b) → (a). Conversely, let $(N_i, M_i) \in \mathcal{H}$ be a rcf of P and let $\Lambda \in \mathcal{J}$ be a spectral factor of $(M_i^* M_i + N_i^* N_i)$ and define $\begin{bmatrix} M \\ N \end{bmatrix} = \begin{bmatrix} M_i \\ N_i \end{bmatrix} \Lambda^{-1}$. Then $\begin{bmatrix} M \\ N \end{bmatrix} \in \mathcal{H}$ and is a spectral factor of (1) i.e. (b) holds. Moreover $\begin{bmatrix} M^* & N^* \end{bmatrix} \begin{bmatrix} M \\ N \end{bmatrix} = \Lambda^{*^{-1}} \begin{bmatrix} M_i^* & N_i^* \end{bmatrix} \begin{bmatrix} M_i \\ N_i \end{bmatrix} \Lambda^{-1} = \Lambda^{*^{-1}} \Lambda^* \Lambda \Lambda^{-1} = I$. Hence $\begin{bmatrix} M \\ N \end{bmatrix}$ is a nrcf of P and this proves the second part of the theorem. □

Based on Theorem 3.1 an algorithm is constructed which will lead to a state–space representation of a nrcf of a generalized dynamical system. The proof is given in Appendix 5.

begin Algorithm

step 1 By proposition Proposition 2.2, we realize the system P along the lines of Rosenbrock (1974) in terms of a generalized state–space systems with:

$$E = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \; A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad (2)$$
$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \; C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

step 2 Calculate W_2 as the stabilizing solution of the Riccati equation:

$$C_2^T C_2 + W_2 A_{22} + A_{22}^T W_2 - W_2 B_2 B_2^T W_2 = 0$$

step 3 Define $Y, Z, \bar{C}, \bar{B}, \bar{A}$ to be:

$$\begin{aligned} Y &:= -(W_2 A_{22} + C_2^T C_2)^{-1}(A_{12}^T - W_2 B_2 B_1^T) \\ Z &:= -(W_2 A_{22} + C_2^T C_2)^{-1}(C_2^T C_1 + W_2 A_{12}) \\ \bar{C} &:= C_1 - C_2 Z \\ \bar{A} &:= A_{11} + (A_{12} + Y^T C_2^T C_2) Z \\ \bar{B} &:= B_1 - (A_{12} - B_1 B_2^T W_2)(A_{22} - B_2 B_2^T W_2)^{-1} B_2 \end{aligned}$$

step 4 Calculate W_1 as the stabilizing solution of the Riccati equation:

$$\bar{C}^T \bar{C} + \bar{A}^T W_1^T + W_1 \bar{A} - W_1 \bar{B} \bar{B}^T W_1^T = 0$$

step 5 A state-space realization of $\begin{bmatrix} M(s) \\ N(s) \end{bmatrix}$ is given by:

$$\left[\begin{array}{c|c} sI_r - \bar{A} + \bar{B}\bar{B}^T W_1 & B_1(I - B_2^{\#} B_2) + A_{12} W_2^{-1} B_2^{\#^T} \\ \hline (I - B_2^{\#} B_2) B_1^T W_1 + B_2^{\#} A_{21} & I - B_2^{\#} B_2 \\ -(I - C_2 B_2^{\#}) C_1 + C_2^{\#^T} A_{12}^T W_1 & C_2^{\#^T} W_2 B_2 \end{array}\right] \tag{3}$$

with $B_2^{\#} = B_2^T (B_2 B_2^T)^{-1}$ and $C_2^{\#} = (C_2^T C_2)^{-1} C_2$.

end Algorithm

The connection between a nrcf and a nlcf of a plant is given in the following corollary.

Corollary 3.2 *If (M, N) is a* nrcf *of the system P^T, then (M^T, N^T) is a* nlcf *of P*

4 Example

Assume that our non-proper system is a differentiator: $P(s) = 5s$. A generalized state-space form can be written as:

$$\begin{bmatrix} 0 & 5 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} u$$

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Using Proposition 2.2 we can write the above generalized state-space form into the standard form used in the algoritm:

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{bmatrix} 0 & -.2 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} u$$

$$y = \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Then following the steps outlined in the proposed algorithm a state-space realization of $\begin{bmatrix} M \\ N \end{bmatrix}$ can be written as: $\left[\begin{array}{c|c} s+.2 & .2 \\ \hline -1 & 0 \\ 1 & 1 \end{array}\right]$. Therefore $M(s) = \frac{.2}{s+.2}$, $N(s) = \frac{s}{s+.2}$ and then $M(s), N(s) \in \mathcal{H}$, $N(s)M(s)^{-1} = P(s)$ and $M^*(s)M(s) + N^*(s)N(s) = I$.

Remark 4.1 *For systems $P(s)$ having an proper inverse and using the theory described in Meyer and Franklin (1987), Vidyasagar (1988) we can calculate a* nlcf *of $P^{-1} = \tilde{M}^{-1}\tilde{N}$. Then a* nrcf *of P is given by $M = \tilde{N}, N = \tilde{M}$*

5 Conclusions

In this note a state–space algorithm for the calculation of a normalized coprime factorization of continuous–time generalized dynamical systems is given. It has been shown that two Ricatti equations have to be solved in the calculation of this normalized coprime factorization.

References

Bongers, P.M.M. and O.H., Bosgra. (1990). Low order H_∞ controller synthesis. In *Proc. 29th Conf. Desicion and Control, p.194-199, Hawaii, USA.*

Bongers, P.M.M. and P.S.C., Heuberger. (1990). Discrete normalized coprime factorization. In *Proc. 9th INRIA Conf. on Analysis and Optimization of Systems 1990, p.307-313, Bensoussan, A., J.L. Lions (eds), Lecture notes in Control and Information Sciences, vol.144, Springer Verlag.*

Chen, C.T. and C.A., Desoer. (1968), A proof of controllability of Jordan form state equations. *IEEE Trans. Automat. Contr. AC-13, 195-196.*

Desoer, C.A. and R.W., Liu and J., Murray and R., Saeks. (1980), Feedback systems design: the fractional representation approach to analysis and synthesis. *IEEE Trans. Automat. Contr. AC-25, no.3, pp.399-412.*

Meyer, D.G. (1988). A fractional approach to model reduction. In *Proc. Amer. Contr. Conf. pp. 1041-1047.*

Meyer, D.G. and G.F., Franklin. (1987), A connection between normalized coprime factorization and linear quadratic regulator theory. *IEEE Trans. Automat. Contr. AC-32, no.3, pp.227-228.*

McFarlane, D.C. and K., Glover. (1989), *Robust controller design using normalized coprime factor plant descriptions.* Lecture notes in control and information sciences, vol.138, Springer Verlag, Berlin, Germany.

Nett, C.N. and C.A., Jacobson and M.J., Balas. (1984), A connection between state-space and doubly coprime fractional representation. *IEEE Trans. Automat. Contr. AC-29, no.9, pp 831-832.*

Rosenbrock, H.H. (1974), Structural properties of linear dynamical systems. *Int. J. Control, vol.20, no.2, p.191-202.*

Schrama, R.J.P. (1991). A framework for control-oriented approximate closed-loop identification. *To appear at 9th Int. Symp. on Mathematical Theory of Networks and Systems, Kobe, Japan.*

Vidyasagar, M. and H., Kimura. (1986), Robust controllers for uncertain linear multivariable systems. *Automatica, vol.22, no.1, pp 85-94.*

Vidyasagar, M. and H., Schneider and B.A., Francis. (1982), algebraiç and topological aspects of feedback stabilization. *IEEE Trans. Automat. Contr. AC, vol.27, no.4, pp 880-893.*

Vidyasagar, M. (1984), The graph metric for unstable plants and robustness estimates for feedback stability. *IEEE Trans. Automat. Contr. AC, vol.29, no.5, pp 403-418.*

Vidyasagar, M. (1988), Normalized coprime factorizations for non-strictly proper systems. *IEEE Trans. Automat. Contr. AC-33, p.300-301.*

Wang, F.Y. and M.J., Balas. (1989), Doubly coprime fractional representations of generalized dynamical systems. *IEEE Trans. Automat. Contr. AC-34, pp 733-734.*

Appendix

In this appendix we proof the construction of a nrcf (M, N) of $P \in \mathcal{F}$ as proposed in the algorithm.

Let the generalized state–space realization of the system be partitioned according to Proposition 2.2 and apply a similarity transformation to a generalized state–space realization of $\begin{bmatrix} I \\ P \end{bmatrix} (I + P^*P)^{-1} \begin{bmatrix} I & P^* \end{bmatrix}$:

$$\left[\begin{array}{cc|c} I & -W & 0 \\ 0 & I & 0 \\ \hline 0 & 0 & I \end{array}\right] \times \left[\begin{array}{cc|cc} C^TC & -sE^T - A^T & 0 & C^T \\ sE - A & -BB^T & B & 0 \\ \hline 0 & -B^T & I & 0 \\ -C & 0 & 0 & 0 \end{array}\right] \times \left[\begin{array}{cc|c} I & 0 & 0 \\ -W^T & I & 0 \\ \hline 0 & 0 & I \end{array}\right]$$

$$= \left[\begin{array}{cc|cc} Q & -sE^T - A^T + WBB^T & -WB & C^T \\ sE - A + BB^TW^T & -BB^T & B & 0 \\ \hline B^TW^T & -B^T & I & 0 \\ -C & 0 & 0 & 0 \end{array}\right] \tag{4}$$

with

$$Q = s(E^TW^T - WE) + C^TC + A^TW^T + WA - WBB^TW^T \tag{5}$$

Equation (4) defines a generalized state-space realization of a spectral factor of $\begin{bmatrix} I \\ P \end{bmatrix} (I + P^*P)^{-1} \begin{bmatrix} I & P^* \end{bmatrix}$ provided that Q in

(5) can be made zero.

Define $W = \begin{bmatrix} W_1 & W_{12} \\ 0 & W_2 \end{bmatrix}$ with $W_1 = W_1^T$, $W_2 = W_2^T$, W partitioned according to E. Then the first part of (5): $s(E^TW^T - WE)$ equals zero.
Define $\hat{A}_{11} = A_{11} + A_{12}X$, $\hat{A}_{21} = A_{21} + A_{22}X$, $\hat{C}_1 = C_1 + C_2X$, $X = YW_1 + Z$ with:

$$\begin{aligned} Y &= -(W_2A_{22} + C_2^TC_2)^{-1}(A_{12}^T - W_2B_2B_1^T) \\ Z &= -(W_2A_{22} + C_2^TC_2)^{-1}(C_2^TC_1 + W_2A_{12}) \end{aligned}$$

where W_1, W_2 are the stabilizing solutions to the Ricatti equations (existence can be shown using Proposition 2.2):

$$\begin{aligned} 0 &= C_2^TC_2 + W_2A_{22} + A_{22}^TW_2 - W_2B_2B_2^TW_2 \\ 0 &= \bar{C}^T\bar{C} + \bar{A}^TW_1^T + W_1\bar{A} - W_1\bar{B}\bar{B}^TW_1^T \end{aligned}$$

with:

$$\begin{aligned} \bar{C} &:= C_1 - C_2Z \\ \bar{A} &:= A_{11} + (A_{12} + Y^TC_2^TC_2)Z \\ \bar{B} &:= B_1 - (A_{12} - B_1B_2^TW_2)(A_{22} - B_2B_2^TW_2)^{-1}B_2 \end{aligned}$$

Using $F = \begin{bmatrix} B_1^TW_1 - B_2^TW_2X & B_2^TW_2 \end{bmatrix}$ (4) can be written as:

$$\left[\begin{array}{cc|cc} 0 & -sE^T - A^T + F^TB^T & -F^T & C^T \\ sE - A + BF & -BB^T & B & 0 \\ \hline F & -B^T & I & 0 \\ -C & 0 & 0 & 0 \end{array}\right]$$

which equals a generalized state-space realization of the transfer function $\begin{bmatrix} M \\ N \end{bmatrix} \begin{bmatrix} M^* & N^* \end{bmatrix}$ with $\begin{bmatrix} M \\ N \end{bmatrix} = \left[\begin{array}{c|c} sE - A + BF & B \\ \hline F & I \\ -C & 0 \end{array}\right]$. Now it can be easily checked that $P(s) = N(s)M^{-1}(s)$. Using operations under restricted system equivalence (Rosenbrock (1974)) the generalized state-space realization of $\begin{bmatrix} M \\ N \end{bmatrix}$ is reduced to the state–space form:

with $B_2^{\#} = B_2^T(B_2B_2^T)^{-1}$ and $C_2^{\#} = (C_2^TC_2)^{-1}C_2$.

$$\left[\begin{array}{c} M(s) \\ N(s) \end{array}\right] = \left[\begin{array}{c|c} sI_r - \bar{A} + \bar{B}\bar{B}^TW_1 & B_1(I - B_2^{\#}B_2) + A_{12}W_2^{-1}B_2^{\#^T} \\ \hline (I - B_2^{\#}B_2)B_1^TW_1 + B_2^{\#}A_{21} & I - B_2^{\#}B_2 \\ -(I - C_2B_2^{\#})C_1 + C_2^{\#^T}A_{12}^TW_1 & C_2^{\#^T}W_2B_2 \end{array}\right]$$

Hence $\begin{bmatrix} M(s) \\ N(s) \end{bmatrix}$ is proper and asymptoticly stable.

Which proves that the presented algorithm will lead to a state–space representation of a nrcf of a system in a generalized state–space form.

□

Selected Topics in Identification, Modelling and Control
Vol. 3, June 1991

Robust control design application for a flight control system

Samir Bennani [1] , Jacob A. Mulder [1] , Antonius J.J.vd Weiden [2]

[1] Sec. Stability and Control, Fac. Aerospace Eng.
Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands.
[2] Lab. Measurement and Control, Dept. Mechanical Eng. and Marine Technology
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. This paper studies in a tutorial way the application of Structured Singular value analysis and synthesis concepts, known as $\mu-$synthesis, on the De Haviland Beaver flight control system (FCS). Our goal is to present the general ideas behind the mechanisms involved with $\mu-$synthesis. The main idea is to show how μ as a robust performance index arises from the General Nyquist stability criterion. The power of the method will be demonstrated by bringing this concept into practice on a design example.

Keywords. Robust stability, nominal performance, robust performance.

1 Introduction

No mathematical model can exactly describe a physical system. For this reason we must be aware of how modelling errors might have an adverse effect upon the performance of a control system. In general, a property of a control system such as its stability or performance, is said to be robust if it is preserved under plant perturbations. In the last decade much research has been done on the design of robust control systems. This has led to the H_∞ control design method (Doyle, Glover, Khargonekar and Francis (1988)), which is suited for robust stability and nominal performance problems. For Single-Input-Single-Output (SISO) systems H_∞ control can also be used for the robust performance problems. For multivariable systems however the robust performance design problem is more complicated because of its inherent structure. H_∞ control does not account for this structure and for that reason Doyle (1982) introduced the notion of Structured Singular Values (μ), first for analysis and later for synthesis (Doyle (1984)). Main references on theory and application of the structured singular value concepts can be found in Williams (1990). In order to reveal how μ arises from the general case, we first apply the main ideas on a simple SISO control structure. In figure 1, the basic feedback structure is enough to tackle the SISO robust performance problem that we will review in section 2 of this paper. The generalized robust performance problem cannot be handled within this basic feedback structure. In section 3, we will see how, by introducing an alternative framework proposed by Doyle (1984), consisting of a general interconnection structure as in figure 4 and a matrix norm, a necessary and sufficient condition for the general robust performance problem can be obtained. In section 4 is shown how within this new framework analysis and synthesis can be carried out on a design example of the Beaver DHC-2 aircraft attitude-hold auto-pilot at a speed of $35m/s$ and al-

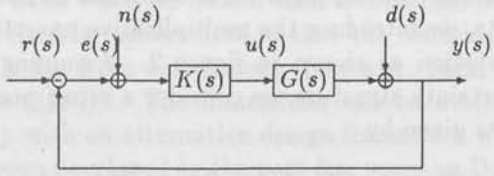


Fig. 1: Basic feedback structure

titude of $6000\,ft$. Two controllers are presented, one achieved by the classical root-locus design method of Evans and the other by $\mu-$synthesis. Rather than only showing superiority of the $\mu-$controller, the design example is meant to show how valuable the information of the characteristics of the classical design can be, for a suitable choice of weighting functions which is not always transparent, these form the starting point of the $\mu-$synthesis.

2 Robust Performance for SISO Systems

2.1 Introduction

Before definitions about robust performance are treated, nominal performance and robust stability are defined. As, if we choose a particular uncertainty model, robust performance is achieved if and only if nominal performance and robust stabilty are achieved. For SISO systems these plant properties are related to the infinity norm,(i.e. maximum over all frequencies of a transfer function), of well known transfer functions, as the sensitivity and complementary sensitvity which satisfy the relation:

$$S(j\omega) + T(j\omega) = 1 \qquad (1)$$

The response $y(s)$ of the closed loop system in figure 1 to a reference $r(s)$ or to a noise signal $n(s)$ is given by the complementary sensitivity function $T(j\omega)$. On the other hand the response of the error signal $e(s)$ to a reference $r(s)$, or the response $y(s)$ to a disturbance signal $d(s)$ as in figure 1, are given by the sensitivity function $S(j\omega)$. Taking into account the fact that generally $|T(j\omega)| \to 0$ as $\omega \to \infty$ and the conservation law of equation (1), a fairly common design rule arises, i.e. keep at low frequencies $|S(j\omega)|$ small and keep $|T(j\omega)|$ small at sensor noise frequencies. In order to trade off among the frequency ranges, weighting functions which form the major design parameters in $\mu-$synthesis are required. For a more thourough treatment on frequency domain properties of feedback systems we refer to Freudenberg and Looze (1988).
In figure 1 the plant $G(s)$ is only a model of reality and to take account for a whole set of possible plants, we introduce the multiplicative uncertainty description as shown in figure 2. Assuming this uncertainty structure we consider a set of possible plants given by :

$$\tilde{G}(j\omega) = G(j\omega)\,(1 + \Delta(j\omega)\,W_2(j\omega))\ . \qquad (2)$$

Here $W_2(j\omega)$ represents the uncertainty profile of the plant and $\Delta(j\omega)$ a scaling factor in magnitude between 0 and 1.

2.2 Nominal Performance

If we want to track a reference signal asymptotically, or reject disturbances at the plant's output, then the sensitivity transfer funtion is involved. This can be expressed in a performance objective as for example:

$$\| S \|_\infty < \epsilon$$

which reflects the desire to keep the maximum amplitude of output errors, which is the ∞-norm of the sensitivity function, i.e. $\max_\omega |S(j\omega)|$ below some level ϵ. Or if we define some weighting function $W_1(s) = \frac{1}{\epsilon}$ the performance specification becomes

$$\| W_1\, S \|_\infty < 1$$

More realistic is the situation when the weighting function is frequency dependent. Assume that $W_1(s)$ is real-rational and stable, we will show how to come to an ∞-norm specification. In several applications, as flight control design, designers have acquired through experience desired shapes of the Bode magnitude plot of S. In particular, suppose good performance is achieved if and only if the plot of $|S(j\omega)|$ lies under some curve. We could rewrite this as

$$|S(j\omega)| < |W_1(j\omega)|^{-1} \quad \forall\, \omega$$

or in other words

Nominal Performance iff

$$\| W_1\, S \|_\infty < 1. \qquad (3)$$

This means that nominal performance is achieved if the worst case response, over frequency, to disturbances (fig. 2) is not amplified.

2.3 Robust Stability

To get a satisfactory characterisation of robust stability we shall use a specific uncertainty model namely, multiplicative perturbations as defined above. In order to obtain a consistent set of models, two conditions have to be satisfied:

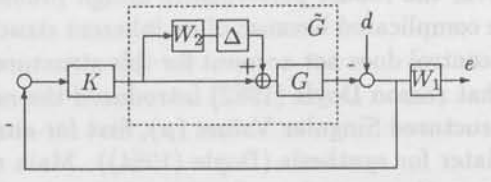


Fig. 2: Multiplicative uncertainty structure

1. $G(s)$ and $\tilde{G}(s)$ have the same number of unstable poles.

2. $\|\Delta(j\omega)\|_\infty \leq 1$.

When both conditions hold the perturbation is called *allowable*. The idea behind the multiplicative uncertainty model is that $\Delta(j\omega)W_2(j\omega)$ represents the relative plant perturbation:

$$\frac{\tilde{G}(j\omega) - G(j\omega)}{G(j\omega)} = \Delta(j\omega)W_2(j\omega)\,. \qquad (4)$$

Hence, if $\|\Delta(j\omega)\|_\infty \leq 1$, then $\forall\,\omega$

$$\left|\frac{G(j\omega)[1+\Delta(j\omega)W_2(j\omega)] - G(j\omega)}{G(j\omega)}\right| \leq |W_2(j\omega)| \qquad (5)$$

in this way the uncertainty profile $|W_2(j\omega)|$ is obtained. The main purpose of $\Delta(j\omega)$ is to account for phase uncertainty and to act as a scaling factor for the perturbation, i.e., $|\Delta(j\omega)|$ varies between 0 and 1.

Assume the nominal feedback system is internally stable. When is stability of a system robust, i.e.,when is internal stability preserved under all allowable perturbations $\Delta(j\omega)$? Using a theorem by Doyle (1984) we get:

Theorem: Robust stability
A system is robustly stable, for all $\Delta(j\omega)$, with $|\Delta(j\omega)| < 1$, if and only if

$$\|W_2(j\omega)\,T(j\omega)\|_\infty < 1\,. \qquad (6)$$

The last result can be obtained by stability considerations of the closed loop system. The question we ask ourselves is, when does the perturbation $\Delta(j\omega)$ destabilize the closed loop system ? Therefore cut the loop at the input and output of $\Delta(j\omega)$ in figure 2 to obtain the transfer function that $\Delta(j\omega)$ "sees". Using the assumptions made on $\Delta(j\omega)$ and the Nyquist stability theorem we obtain the desired result.

2.4 Robust Performance

Once nominal performance and robust stability are achieved, we might ask ourselves if performance can be made robust. Let $\tilde{S}(j\omega)$ be the perturbed sensitivity function, i.e., the sensivity function of the plant under some allowable perturbation $\Delta(j\omega)$. Performance is robust if and only if it preserved under all perturbations. This gives (Doyle (1984)):

Robust Performance A necessary and sufficient condition for robust performance is

$$\sup_{|\Delta|<1} \|\,W_1(j\omega)\,\tilde{S}(j\omega)\,\|_\infty < 1 \qquad (7)$$

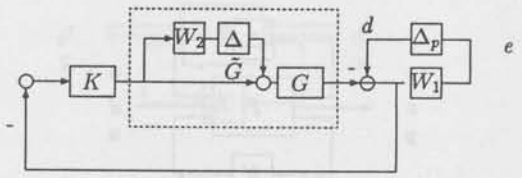


Fig. 3: Robust Performance as a Robust Stability test

The last equation holds iff

$$\|\,|W_1(j\omega)S(j\omega)| + |W_2(j\omega)T(j\omega)|\,\|_\infty < 1 \qquad (8)$$

Which means that robust performance is achieved if and only if nominal performance and robust stability are both satisfied and that their absolute sum is less than unity. Equation (8) follows with some manipulations directly from equation (7). Equation (8) can also be obtained in an other way, namely as a stability test, this forms the crucial idea behind the $\mu-$analysis. Therefore introduce an extra allowable perturbation mostly called performance block Δ_p on the performance as shown in figure 3. To test stability with respect to simultaneous perturbations, again break the loop at the inputs and outputs of both perturbations and determine which transfer function matrix the simultaneous perturbation "sees". Then applying the Nyquist stability criterion we obtain the desired result. This machinery forms the principle of the μ analysis which can handle any complex valued uncertainty and performance as a stability test. A detailled discussion on the material presented in this section can be found in Balas, Packard and Doyle (1990).

3 μ-Synthesis Methodology in a General Framework

3.1 Introduction

In multivariable design performance and stability objectives can not be reflected in terms of the sensitivity and complementary sensitivity alone. More complex transfer functions are involved so that a condition for robust performance is hard to obtain. Even when we obtain such a condition it can be arbitrarly conservative so that the resulting design is far from satisfactory, (see e.g. Stein and Doyle (1990)). The limitations can be overcome partly with an alternative design framework which has been developed in the past few years by Doyle, Wall and Stein (1982), Doyle (1984) and Stein and Doyle (1990). The alternative framework consist of

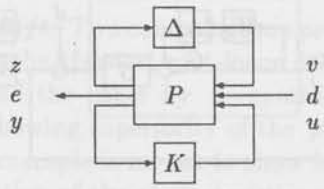


Fig. 4: General interconnection structure

a more general problem description, a more suitable measure of magnitude for matrix transfer functions, and certain key analysis and synthesis results. In this paper we apply the alternative framework on a very simple SISO example in order to reveal the involved mechanisms which can be helpful for the understanding in dealing with more complex multivariable problems. The general problem description is shown in figure 4. It consists of a general system P with three pairs of input/output variables. The first pair consists of the measured outputs y, and control inputs u. The second pair consists of performance variables e, and external input signals d, and the third pair consists of output signals z, and v through which unit-norm perturbations are fed back into the system. Any linear interconnection of inputs, outputs and commands along with the perturbations and a controller can be viewed in this context and can be rearranged to match this diagram so that P can be chosen to reflect many different problem specifications.

3.2 Analysis Review

Beyond its generality, the alternative framework is important because it comes equipped with a non-conservative necessary and sufficient condition for robust performance. In order to describe this condition we first close the compensator feedback-loop in fig 4 to get the loop in fig 5. The system $M(P,K)$ in this figure has a 2×2 block-structured transfer function $M(s)$ whose blocks are defined in terms of the original 3×3 partition of $P(s)$ as follows:

$$M_{ij}(s) = P_{ij}(s) + P_{i3}(s)\left[I - K(s)P_{33}(s)\right]^{-1} K(s)P_{3j}(s) \quad \text{for } i,j = 1,2 \tag{9}$$

Equation (9) is called a *Linear Fractional Transformation* of the system P through K, therefore the notation $M(P,K)$. Suppose that this system is stable, then the following results apply:

Theorem (Doyle (1984)):

1. **Nominal performance is satisfied if and only if**
$$\|M_{22}(j\omega)\|_\infty < 1 \tag{10}$$
2. **Robust stability is satisfied if and only if**
$$\|M_{11}(j\omega)\|_\infty < 1 \tag{11}$$
3. **Robust performance is satisfied if and only if**
$$\mu[M(j\omega)] < 1 \quad \forall\omega \tag{12}$$
(where μ is a function to be defined shortly)

The first result is true by definition since M_{22} represents any nominal performance transfer function block. Already here we see that through this construction we can capture any performance objective not only the sensitivity. The second result follows from a stability consideration with the perturbation loop closed, i.e. $\det(I - \Delta(s)M_{11}(s)) \neq 0$ along the imaginary axis for all allowable perturbations. Again $M_{11}(s)$ can consist of any transfer function not just the complementary sensitivity. The third result is the most significant one. It provides a necessary and sufficient condition for robust performance. It can be established from the definition that performance is robust if and only if the transfer function from d to e with the Δ loop closed remains ∞-norm bounded by unity, that is if and only if

$$\|M_{22} + M_{21}(I - \Delta M_{11})^{-1}\Delta M_{12}\|_\infty < 1 \ \forall\Delta\,. \tag{13}$$

Equation (13) is also a necessary and sufficient condition for the system $M(P,K)$ to remain stable even if we choose to connect a second norm-bounded perturbation Δ_p (performance block) across the e and d terminals. In this view, robust performance is exactly equivalent to robust stability in face of two perturbations Δ and Δ_p connected around the system $M(P,K)$ in the block diagonal arrangement shown in figure 6. The robust stability is assured, if and only if the function $\det(I - diag(\Delta, \Delta_p)M(j\omega))$ remains nonzero along the imaginary axis. This observation brings us to the function μ, called the structured singular value. This function was defined specifically to test the kind of determinant conditions identified above. Its full definition for complex matrices is

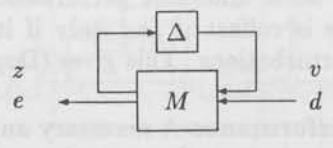


Fig. 5: Analysis part General interconnection structure

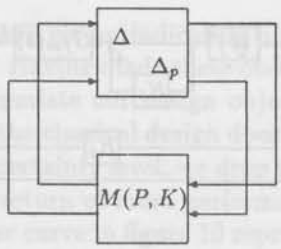


Fig. 6: μ-stability test

the following (Doyle (1984)):

$$\mu[M] \triangleq \left[\min \left\{ \epsilon \;\middle|\; \begin{array}{l} \det[I - \epsilon XM] = 0 \\ \text{for some } X = diag(\Delta_1, \ldots, \Delta_m) \\ \text{with } \|\Delta_i\|_\infty < 1 \text{, for all } i \end{array} \right\} \right]^{-1} \tag{14}$$

In words, μ is the reciprocal of the smallest value of scalar ϵ which makes the matrix $I - \epsilon XM$ singular for some X in a block-diagonal perturbation set. If no such ϵ exists, μ is taken to be zero. It is clear from definition (14) that μ can be applied to the transfer function matrix in figure 3.2 to test wether $\det(I - diag(\Delta, \Delta_p)M)$ remains non-zero along the imaginary axis. In fact the determinant remains non-zero as long as $\mu[M] < 1$. Applying Schur formula for determinants, both conditions, eq(11) and eq(13) can be obtained at once; (in Maciejowski (1989), pp 126-127 is shown how the robust performance condition is obtained). This is a tight condition for robust stability with respect to two perturbation blocks, and equivalently a tight condition for robust performance. Note that the definition is not limited to 2×2 block structures, so that it can be used to test stability with respect to any number of diagonal blocks. This permits to establish robust stability with respect to plant sets characterized by several unstructured perturbations, and simultaneously, to establish robust performance.

For practical use, the function $\mu[M]$ has to be evaluated. This is done across frequency, providing a Bode-like plot to analyse robust stability/performance of any given design.

The algorithms of μ are based on the following inequalities which are proven in Doyle (1982):

$$\sup_{U \in \underline{U}} \rho(MU) \leq \mu(M) \leq \inf_{D \in \underline{D}} \overline{\sigma}(DMD^{-1}) \tag{15}$$

where ρ denotes the spectral radius, $\overline{\sigma}$ the maximum singular value and

$$\underline{U} = \{\text{diag}\,(U_1, U_2, \ldots, U_n) \,|\, U_i^* U_i = I\}$$

$$\underline{D} = \{\text{diag}\,(d_1 I, d_2 I, \ldots, d_n) \,|\, d_i \in R_+\}$$

with block dimensions of U_i and D_i matching those in X.

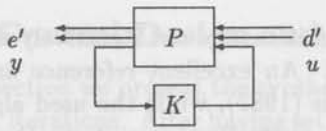


Fig. 7: Synthesis part General interconnection structure

In Doyle (1982) key theorems regarding μ are proven; it is shown that the lower bound is always an equality, while the upper bound is an equality as long as $n \leq 3$. It has to be remarked that the sets $\underline{U}$ and $\underline{D}$ leave perturbations from the set X invariant, in the sense that $\bar{\sigma}(\Delta U) = \bar{\sigma}(U\Delta)$ and $D\Delta D^{-1} = \Delta$, which is equivalent to say for the last expression that $\mu(DMD^{-1}) = \mu(M)$. Noticing the fact that singular values are not invariant under scaling leads to a practical computation scheme for $\mu(M)$ by minimizing $\overline{\sigma}(DMD^{-1})$ *(the upper bound)* over all $D \in \underline{D}$, resulting in the so called optimal $D-$scales. Althoug $\overline{\sigma}(DMD^{-1})$ is convex in $\ln(D)$ the infimum is not necessarily equal to μ, but practice shows that the upper bound is acceptably close to μ.

3.3 Synthesis Review - H_∞ Optimization

For the purpose of synthesis, the perturbation can be normalized properly to unity so that the normalizing factor can be absorbed into P. This results in the synthesis problem as shown in figure 7. The synthesis problem involves finding a controller K such that performance requirements are satisfied under prescribed uncertainties. The interconnection structure P can be partitioned so that the input-output map from $d' = \begin{bmatrix} v \\ d \end{bmatrix}$ to $e' = \begin{bmatrix} z \\ e \end{bmatrix}$ can also be expressed as the following lower linear fractional transformation denoted $F_l(P,K)$:

$$e' = F_l(P,K)d'$$

where

$$F_l(P,K) = M(P,K) = P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}.$$

For the H_∞ optimal problem, the objective is to find a stabilizing controller K which minimizes $\|F_l(P,K)\|_\infty$. Thus find a controller K such that

$$\|F_l(P,K)\|_\infty < \gamma \tag{16}$$

where $\frac{1}{\gamma}$ is the minimum norm of the perturbation that destabilizes the closed-loop system. The

minimization is carried out iteratively and is called $\gamma-$iteration. An excellent reference on this matter is Francis (1987), while the used algorithms to obtain H_∞ controllers come from Doyle, Glover, Khargonekar and Francis (1988).

3.4 μ-Synthesis Methodology

The μ-synthesis methodology emerges as a practical approach in designing control systems with robust performance objectives. This technique essentially integrates two powerfull theories for synthesis and analysis into a systematic design technique involving H_∞ optimization methods for synthesis and the structured singular value μ for analysis. Recall that μ may be obtained by scaling and applying an infimum over D. Extending this concept to synthesis, the problem of robust controller design becomes that of finding a stabilizing controller K and a scaling matrix D such that the quantity $\| DF_l(P,K)D^{-1} \|_\infty$ is minimized, for more detail see Doyle (1984).

One approach for solving this problem is that of alternaly minimizing the above expression for either K or D while holding the other constant. For fixed D, it becomes an H_∞ optimal control problem and can be solved using the the state-space method of Doyle, Glover, Khargonekar and Francis (1988). On the other hand, with fixed K, the above quantity can be minimized at each frequency as a convex optimization in $\ln(D)$. The resulting data of D can be fit with an invertible, stable, minimum-phase, real-rational transfer function. This proces is called $D-K-$iteration and is carried out until a satisfactory controller is obtained. For a deaper treatement of this procedure, we refer to Balas, Packard and Doyle (1990).

4 Application of μ-Synthesis to a FCS design

4.1 Introduction

This section deals with an application of the presented theory on a flight control system. To be able to perform practical design all desired objectives are translated into suitable weighting functions and absorbed into the general interconnection structure. Then comparison between classical and the achieved $\mu-$controller is made.

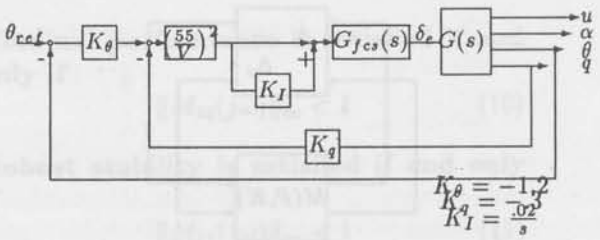


Fig. 8: Class. Feedback Configuration PAH Mode

4.2 Problem Description

We consider the Linear Time Invariant (LTI) model of the DHC2-Beaver aircraft for the longitudinal motion as described in Tjee and Mulder (1988). The plant can be represented in state space form as:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \tag{17}$$

where x represents state vector, u the input vector and y the output vector.

$$x = \begin{bmatrix} u \\ \alpha \\ \theta \\ q \end{bmatrix} \begin{array}{ll} (m/s) & \text{forward speed} \\ (deg) & \text{angle of attack} \\ (deg) & \text{pitch angle} \\ (deg/s) & \text{pitch rate} \end{array}$$

$u = \delta_e \ (deg)$ elevator deflection

$y = \theta$

Also, a third order actuator model denoted $G_{FCS}(s)$ has been included as we can see in figure 8. With the root-locus design method from Evans a two-loop controller has been synthesised. The obtained controller as in figure 8, revealed to have satisfactory characteristics during flight tests. Then performance results of the classical controller are used as requirements for the synthesis of the $\mu-$controller in order to obtain realistic weighting functions. The $\mu-$control structure can be viewed as a disturbance rejection problem as given in figure 2, which means that design can be carried out following the results of section 2. The weighting functions can be obtained from the sensitivity and complementary sensitivity function of the classical design as given in figure 9. This is due to the fact that the complementary sensitivity, in our example, through its inverse, directly can be related to the uncertainty profile. In this view the complementary sensitivity in figure 9 tells us that the design can tolerate an input multiplicative perturbation of 125 percent. On the other hand the sensitivity function tells us how fast and up to which bandwith errors due to disturbances or commands are eliminated. At zero frequency of the sensitivity function

in figure 9, we get an indication of the steady state error level. Having made these observations, we are able to formulate our design objectives. Since we know that the classical design does account for such an high uncertainty level, we drop this down to fifty percent in return of more performance.

The lower curve in figure 10 represents $|W_2(j\omega)|$, i.e., the upper bound on the magintude of the relative plant perturbation with frequency ω. So in our case $|W_2(j\omega)|$ starts at a level of 0.5 crosses the zero dB axis at $100\ rds^{-1}$ and increases at a rate of 20 dB/decade. The transfer function for the input uncertainty is:

$$W_2(s) = \frac{50(s+100)}{(s+10000)} \tag{18}$$

For the performance is chosen a rejection factor 50 to 1 in the operating bandwith, i.e. output disturbances are attenuated over this frequency region. Expressed differently, steady-state tracking error to step references has to be in the order of 0.02 or smaller. This performance requirement gets less and less stringent as frequency increases as shown in figure 10. From equation (1), we know that this performance level is the maximum achievable one in front the chosen uncertainty level. The transfer function associated with the performance goal is:

$$W_1(s) = \frac{.5(s+3)}{(s+.03)} \tag{19}$$

Once, the weighting functions are obtained, the general matrix interconnection structure has to be built up.

In our example we have to break in figure 2 the loop at the compensator and the perturbation Δ to obtain the open loop interconnection structure represented by the transfer function matrix P. Then through an LFT we get the object transfer function $M(P,K)$ for synthesis and analysis purpose:

$$M(P,K) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \begin{bmatrix} -W_2T & -W_2KS \\ W_1SG & W_1S \end{bmatrix}.$$

The following analysis results apply:

$$\begin{aligned} \underline{\text{Nominal performance}} \\ \|M_{22}\|_\infty &= \|W_1S\|_\infty &< 1 \\ \underline{\text{Robust Stability}} \\ \|M_{11}\|_\infty &= \|W_2T\|_\infty &< 1 \\ \underline{\text{Robust performance}} \\ \|M\|_\mu &= \||W_2T| + |W_1S|\|_\infty &< 1 \end{aligned}$$

These results precisely match the earlier statements (see sec. 2.4 and 3.2) and are used after each synthesis step as performance indicators.

4.3 Control Design and Results

In this section we present the synthesis results from the D-K iterations. After having set up the problem into the general form, the first design step consists of synthesising an H_∞ controller. Achieved with the $\gamma-$iteration was $\gamma = 1.43$, this corresponds to an allowable unertainty level of 70 %. To analyze whether robust performance is achieved, μ is calculated and plotted against frequency as a Bode magnitude plot. The resulting μ is over the operating region above one which means that the desired robust performance level yet is not achieved. At this stage optimal dynamic D-scales are introduced and a new controller is synthesised. The new γ value is 1.09, this corresponds to an allowable uncertainty level of 91 %. The robust performance level is not achieved as we see in figure 11. A new D-scale is fitted, leading to $\gamma = .93$ in the third iteration step. At this stage the robust performance level, i.e., $\mu < 1$ is achieved by a 15th order controller

4.4 Comparison to the Classical Design

In order to compare both designs, the classical controller has been also absorbed with the weithing functions into the open loop interconnection structure P.

1. Results Robust Stability.
Let us look at figure 12, where the robust stability level with respect to the specified uncertainty model is represented by the lower curve for the classical design. The curve lies far under unity, which means that the control system is robustly stable. The perturbation which destabilizes the system, can be seen in terms of classical gain margin, which means in our case that the gain could be raised up to 1 over 0.49 which is slightly more than a factor 2 to make the closed loop unstable. For the μ design robust stabiliy is represented in the same figure by the upper curve. The result reflects precisely what we included in the weighting function which intended to allow 50 percent of uncertainty at low frequency, the gain margin in this case is 2 as we expected. Resuming, it can be said that, with respect to the prescribed uncertainty profile, both designs behave similarily.

2. Results Nominal Performance.
Remember that the performance specification for the $\mu-$controller was more stringent than the classical controller. The steady state error for the classical controller is 0.2 while the objective for the

μ−controller was to tolerate only an error of 0.05. So we know in advance that the classical system does not satisfy the nominal performance objective. If we look at the nominal performance plots in figure 13, the above curve goes far above unity at low frequency as expected, while the lower curve represents the achieved nominal performance level of 0.5 as we introduced in the weighting function.

3. Robust Performance.
Since for the classical controller nominal performance is not achieved, neither robust performance could be. This can be seen in figure 14 where the upper figure represents the robust performance level of the classical controller. The lower curve in the plot represents the achieved robust performance level for the μ design, which is everywhere below unity so that it could account for an unity perturbation and still achieve the performance objective.

4. Time Responses.
To give an impression what the frequency bounds reflect in time domain, we made some step responses of the weighted sensitivity and weighted complementary sensitivity functions for both controllers. The weighted complementary sensitivity can be seen as the response of reference disturbances or measurement noise. Let the input signal be a step of level 0.1, then in figure 15 the above curve represents the response of the μ design while the lower curve represents the response of the classical design. These results are in accordance with the chosen weighting function where we intended to obtain a 50 percent robust stability level, reflecting a reduction by a factor 2 of reference disturbances or measurement noise. Figure 16, represents the step responses of the weighted sensitivity function, which is the output error due to external disturbances. In the weighting function we addressed the desire to keep the output error on a level of 50 percent steady-state, which means that an 0.1 step disturbance will be attenuated by a factor 2. We can see on the upper curve classical design nominally already amplifies disturbances in order to attenuate them; we knew this already from frequency domain considerations on performance.

5 Conclusion

This paper reviewed the main mechanisms involved with μ−synthesis. This method is a powerful tool that can handle in a non conservative way the general robust performance problem. We chose this particular simple SISO design configuration in order to demonstrate the mechanism of bringing the basic control structure into the general interconnection structure, from which it is analytically easy to obtain conditions for robust performance. It should be said that for this example the whole μ calculation is not necessary since robust performance can be obtained from equation (8). But if more perturbations are involved and the system is multivariable, everything gets more complex. For example, sensitivity at the systems input is not anymore the same as at the output, also directionality of signals and the systems condition number begin to play an important role. Therefore non conservative analytical expressions for robust performance are hard to obtain. In the view of this context μ turns out to be a non conservative performance measure which means that the methodology reviewed here is very suitable for multivariable design. We showed that selection of weighting functions to represent design ojectives is quite natural when characterics of a classical controller are available. For this SISO system the design specification where captured in terms of S and T. This is certainly not the case for multivariable systems, therefore analysis in the general framework for a suitable choice of the weighting functions is essential. Through a Linear Fractional Transformation on the open loop interconnection structure P, both synthesis and analysis where performed. With respect to robust stability, both designs have the same characteristics. Superiority of μ−synthesis shows up, when also the performance problem is considered. It can be said, that for the same uncertainty level, the μ−design has a four times better performance than the classical design while using only one feedback signal and no gain scheduling. And even more, not only nominal performance but as well robust performance is achieved.

References

Balas, G.J., A. Packard, Doyle, J.C., (1990). Theory and Application of Robust Multivariable Control, MUSYN Short Course.

Balas, G.J., Doyle, J.C., Glover, K., Packard, A.K., Smith, R., (1990). μ−Analysis and Synthesis Toolbox, (Musyn Toolbox). Matlab Functions for the Analysis and Design of Robust Control Systems. Sept, 1990.

Doyle, J.C. and G. Stein (1981). Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis , *IEEE Trans. Auto. Control,* February, 1981.

Doyle, J.C., (1982). Analysis of Feedback Systems with Structured Uncertainty, *Proc. IEE,* November, 1982.

Doyle, J.C., J.E. Wall, and G. Stein, (1982). Performance and Robustness Analysis for Structured Uncertainty, *Proc. 21st IEEE Conf. on Decision and Control,* pp. 629-636, Orlando, FL, Dec.

Doyle, J.C., (1984). *Lecture Notes,* ONR/Honeywell Workshop on Advances in Multivariable Control, Mineapolis.

Doyle, J.C., K. Glover, P. Khargonekar, and B.A. Francis, (1988). State-space Solutions to Standard H_2 and H_∞ Control Problems, *Proc. American Control Conf.*, Atlanta, GA, June, 1988.

Francis, B.A., (1987). *A Course in H_∞ Control Theory.* Lecture Notes Control and information Sciences, vol.88, Springer-Verlag Berlin, Heidelberg, 1987.

Freudenberg, J.S., D.P. Looze, (1988). *Frequency Domain Properties of Scalar and Multivariable Feedback Systems*, Lecture Notes Control and information Sciences, vol.104, Springer-Verlag, NY, 1988.

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

Stein, G., Doyle, J.C., (1990). Beyond Singular Values and Loop Shapes, to appear in *A.I.A.A. Journal of Guidance and Control.*

Tjee R.T.H., Mulder J.A., (1988). Stability and Control Derivatives of the De Havilland DHC-2 'Beaver' aircraft. *Tech. rep. LR-556*, Fac. of Aerospace Eng., Delft University of Technology, Aug 1988.

Williams S.J., (1990). A Review of μ and its applications. IEE Coll. on Successful Industrial Applications of Multivariable Analysis, Feb. 14. 1990, London, *IEE Digest no 1990/052.*

Appendix

We present the system data as in the storage format of the Musyn Toolbox Balas, Doyle, Glover, Packard and Smith (1990). A state space representation (A, B, C, D) is given by a single compact data structure containing all the relevant system information:

$$\begin{bmatrix} A & B & nx \\ C & D & 0 \\ 0 & 0 & -\infty \end{bmatrix}$$

The number of states is given by nx while the $-\infty$ in the last row tells us that we deal with a system state space realization. The system matrix for the symmetric motion of the Beaver rigid body model in the stability reference system at 35m/s, 6000 ft is given by:

$$\begin{bmatrix} -.0756 & .0710 & -.1711 & -.0070 & -.0112 & 4.0000 \\ -.7406 & -.9451 & 0 & .9657 & -.0732 & 0 \\ 0 & 0 & 0 & 1.0000 & 0 & 0 \\ 1.8402 & -4.6441 & 0 & -3.1767 & -6.3391 & 0 \\ 0 & 0 & 1.0000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\infty \end{bmatrix}$$

The model of the hydraulic system is included for design. The system in state space form is given by

$$\begin{bmatrix} -38.8040 & 0 & 0 & 0.0768 & 3.0000 \\ 0 & 0 & 1.0000 & 0 & 0 \\ -656.8851 & -79.9182 & -17.5252 & 1.5199 & 0 \\ 0 & 458.3582 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\infty \end{bmatrix}$$

The states are piston and elevator displacement [m] , elevator translation displacement speed [m/s]. The output is the elevator deflection [deg] and the input is the voltage [Volt].

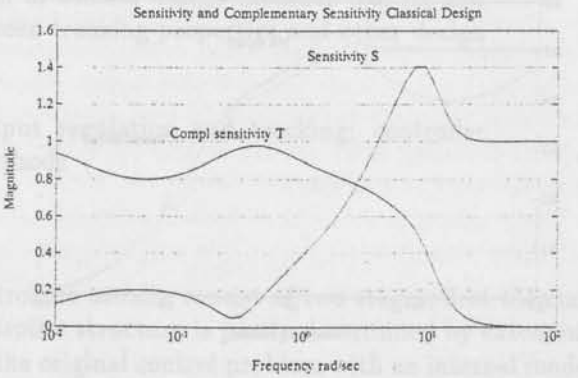


Fig. 9: Sensitivity and Compl. Sensitivity Classical Design

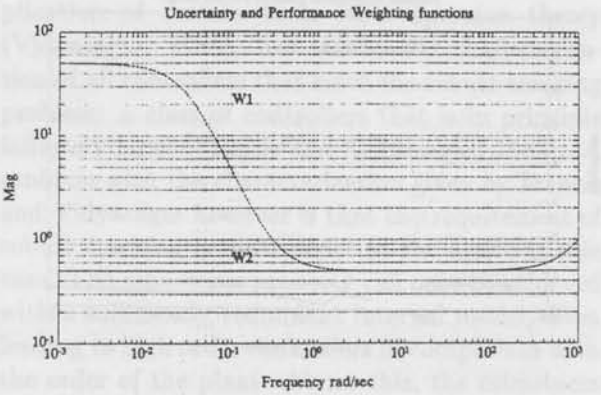


Fig. 10: Weighting functions

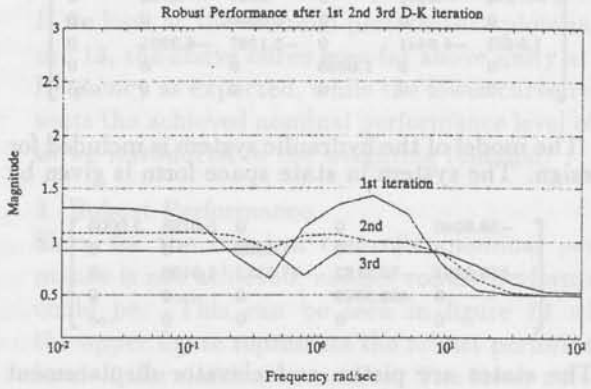


Fig. 11: Robust Performance after D-K iteration

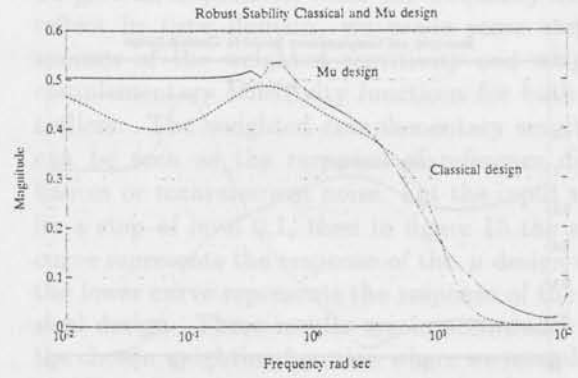


Fig. 12: Robust Stability Classical and μ design

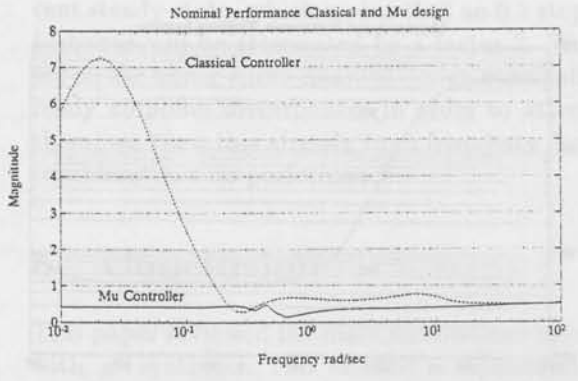


Fig. 13: Nominal performance Classical and μ design

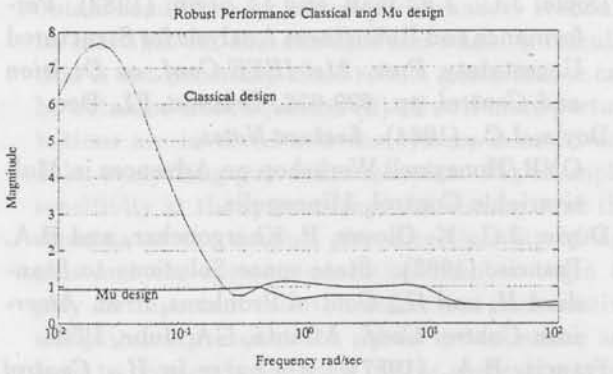


Fig. 14: Robust Performance Classical and μ design

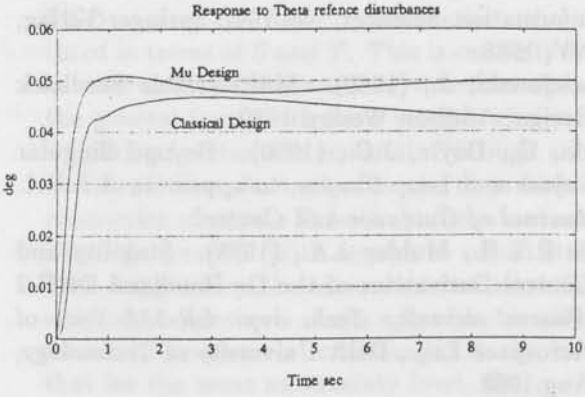


Fig. 15: Step responses weighted compl. sensitivity

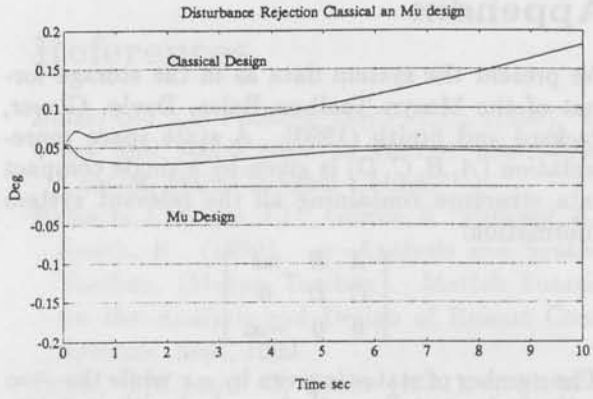


Fig. 16: Step responses weighted sensitivity

Selected Topics in Identification, Modelling and Control
Vol. 3, June 1991

The parametrization of all controllers that achieve output regulation and tracking

Paul F. Lambrechts **Okko H. Bosgra**

Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. This paper considers frequency-domain conditions for feedback compensators that achieve output regulation and tracking in the presence of persistent set point inputs. Using a standard plant approach it is shown that the property of internal stability must be relaxed to what is defined as 'tracking stability'. A characterization of all controllers is formulated in terms of an extended plant description. The formulation provides a framework for compensator design when both regulator and tracking objectives have to be satisfied. It allows application of control design methods like H_2 and H_∞ optimal control to consider trade-offs between tracking properties and other design objectives.

Keywords. linear multivariable systems; output regulation and tracking; controller parametrization; H_2 and H_∞ control design methods

1 Introduction

Recent results in the field of control system design (e.g. H_2 and H_∞ optimal control methods) allow a precise formulation of control performance goals in conjunction with descriptions of the system, the signals and the uncertainties involved. The various control design goals may be conflicting; it may be desirable to achieve stability but also to have a certain amount of bandwidth, disturbance rejection and robustness. Another important objective is output regulation and asymptotically tracking certain classes of reference signals, as has been studied by many authors (Bengtsson 1977, Bhattacharyya and Pearson 1972, Davison and Goldenberg 1975, Davison 1976, Desoer and Wang 1980, Francis 1977). Especially the robust control of the so-called general servomechanism problem has been considered; solvability conditions are available and it is well known that candidate controllers must have certain structural properties according to the *internal model principle* (Francis and Wonham, 1975). Design procedures for this type of controllers usually consist of two stages; first the controller structure is partly determined by extending the original control problem with an internal model of the signals that are to be tracked, next a stabilizing controller is designed to achieve all other control objectives (Desoer and Wang, 1980).

Recently, the (further) development of H_2 and H_∞ optimal control theory, based on a renewed interest in frequency domain methods and the application of certain tools from operator theory (Vidyasagar, 1985), has led to the characterization of all controllers that solve the robust tracking problem: a class of controllers that is in principle infinitely large (Francis and Vidyasagar, 1983). A problem with the characterization given by Francis and Vidyasagar however is that the requirement of robust tracking is quite strict in the multivariable case. The robustness property can only be achieved with a sufficiently redundant internal model, often leading to high order controllers in comparison with the order of the plant. Above this, the robustness property is derived for independent variations of all the plants parameters, which is usually conservative

with respect to the actual perturbations that can occur; this may therefore lead to insufficient performance in other control objectives than tracking and it is even possible that the problem becomes unsolvable (see Grasselli and Longhi, 1991).

This paper will concentrate on deriving a characterization of all acceptable controllers that achieve a prespecified form of tracking and for which the solution of the robust tracking problem can be considered as a special case. We will obtain this characterization based on a 'standard plant' approach, as used in many of the recent studies on regulator problems mentioned above. The main advantage of this approach is the possibility of applying standard H_2 or H_∞ optimization tools to select a controller that not only solves the tracking problem, but also obtains a trade-off between other control design goals. In comparison with another method to incorporate the tracking problem into H_∞ optimization theory, known as $j\omega$-axis shifting (Xu and Mansour 1986, Wu and Mansour 1990), this approach is much less restrictive, and will shift the attention to the problem of choosing appropriate weight functions. As for the servomechanism problem, this then leads to a two-step design procedure; first characterize all possible controllers that achieve tracking, next use H_2 or H_∞ optimization tools to select one of them.

The tracking problem that we will consider here will have some restrictions. We will look at the problem of letting the plants output signal track a prespecified persistent reference signal, although persistent disturbances can be dealt with in the same way. Furthermore we will construct controllers that can only use measured error signals; this implies that we will only find one-degree-of-freedom controllers.

Some preliminaries and notation followed by the exact formulation of the tracking problem can be found in section 2. The actual derivation of a characterization of all controllers that achieve the tracking objective will follow in section 3. Section 4 will then give a high level algorithm and show the possiblity to apply standard optimization methods. Next an example will be given in section 5 to illustrate some of the properties of the developed procedure. Finally, section 6 will give some concluding remarks.

2 Preliminaries and problem formulation

We will adopt the standard plant approach as investigated by for instance Doyle et al. 1984. Therefore

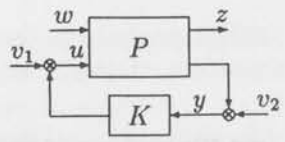


Fig. 1: Standard control configuration

consider the standard control configuration given in fig.2. Here P is the standard plant, K is the controller, z is a vector of control objectives—usually error signals—that are to be minimized, y is a vector of measurement signals, w a vector of external disturbances and u a vector of control inputs. Furthermore the auxiliary signals v_1 and v_2 are added to be able to check internal stability. Many control problems can be brought into this form; the control objective is to find a controller K that minimizes the transfer from w to z in some sense. In many cases this minimization can be sensibly defined by means of operator norms on function spaces in the frequency domain (see Doyle et al. 1984, Francis 1987, Vidyasagar 1985). The function spaces used in this paper may be defined as follows:

- $R[s]$ is the set of polynomials in the indeterminate $s \in C$ with coëfficients in the field R of real numbers.
- $R(s)$ is the field of fractions associated with $R[s]$ and consists of real-rational functions in s.
- RL_∞ is the subspace of $R(s)$ for which $\sup_\omega |F(j\omega)|$ is bounded for all $F \in RL_\infty$. The least upper bound is called the '∞-norm' and makes RL_∞ a Banach space. $F \in RL_\infty$ if it is a proper, real-rational transfer function without poles on the imaginary axis.
- RH_∞ is the subspace of RL_∞ for which F is analytic in the complex closed right half plane C_+ for all $F \in RH_\infty$. $F \in RH_\infty$ if it is a proper, stable, real-rational transfer function.

With abuse of notation we will identify the set of matrices or vectors with elements in a previously defined subspace with the subspace itself. Unless explicitly stated otherwise, transfer function matrices are considered as matrices over the field of real rational functions.

Now consider the standard *tracking* control configuration as given in fig.2. Here we want to minimize the effect of reference signals r and disturbances n_1, n_2 on the weighed error signal z_1 and the weighed control input z_2. We will assume that the weights W_1, W_2, N_1 and N_2 are all in RH_∞. Typically r will consist of a combination of persistent

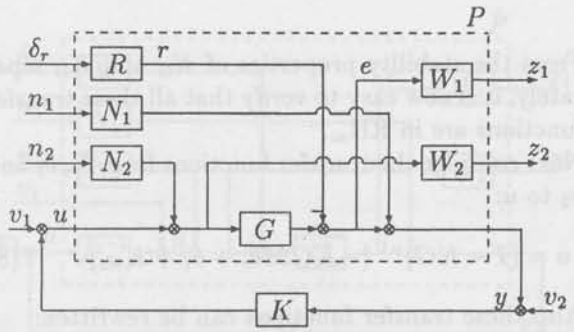


Fig. 2: Standard tracking control configuration

signals like step-functions and sinusoids. These signals can be modelled as the time-domain responses of a linear dynamic system with nonzero initial conditions. In the frequency-domain this results in a transfer function $R \in \mathrm{R}(s)$ that has only imaginary poles and is driven by a q-dimensional constant vector $\delta_r \in \mathrm{R}^q$. We will assume that this vector lies in an arbitrary direction and can be normalized to 1. We will also make a non-restrictive technical assumption; if R is factorized as $\hat{R}\bar{R}$ with $\hat{R}$ strictly proper and $\bar{R}$ polynomial, we assume that $\mathrm{rank}(\hat{R}) = \mathrm{rank}([\hat{R} \mid \bar{R}])$. This assumption ensures that the signal $r = R\delta_r$ is persistent for all $\delta_r \in \mathrm{R}^q$.

It will be the intention to let the outputs y_p asymptotically track any signal r that can be generated this way, or in other words to make sure that the error signals e are not persistent. It is well known that this time-domain demand of tracking a persistent signal can be stated within frequency-domain terminology as the problem of stabilizing the transfer function from δ_r. to $e(s)$. Therefore, if we write down the standard plant of fig.2 without auxiliary signals as:

$$P = \left[\begin{array}{ccc|c} -W_1R & 0 & W_1GN_2 & W_1G \\ 0 & 0 & W_2N_2 & W_2 \\ \hline -R & N_1 & GN_2 & G \end{array}\right] \qquad (1)$$

it can be verified that both the tracking objective and stability of the closed loop system can be achieved with any controller K that internally stabilizes P.

The problem now is that there may not exist any internally stabilizing controller for P, while in practice it is still possible to find acceptable solutions. More specifically, if we want to compensate the effect of the C_+ poles of R in the error signal e we can only do so by creating the same effect in y_p (of course this actually is the objective of the tracking problem). If we assume for the moment that the poles of G are disjoint from those of R, the same effect must also appear in the control input u; this implies that the transfer from δ_r to u must contain C_+ poles which is in contradiction with internal stability. In the next section we will therefore relax the internal stability demand to look for solutions in more general cases.

3 A characterization of all controllers that achieve the tracking objective

From the observations given above it is clear that C_+ poles must be allowed in the transfer function from δ_r to u. To concretise this we will make use of the following two stability concepts.

Definition 3.1 (partial tracking stability)
Consider the standard tracking control configuration in fig.2, and let $G, R \in \mathrm{R}(s)$ *and* $N_1, N_2, W_1, W_2 \in \mathrm{RH}_\infty$ *be given, R having only imaginary poles.*
Then a controller K is said to achieve partial tracking stability *of the closed loop system if:*

- *the transfer functions from δ_r, n_1, n_2, v_1 and v_2 to y are in* RH_∞.
- *the C_+ poles that appear in the transfer functions from δ_r, n_1, n_2, v_1 and v_2 to u are located at the same positions in the complex plane as the poles of R.*

Definition 3.2 (tracking stability)
Consider the standard tracking control configuration in fig.2, and let $G, R \in \mathrm{R}(s)$ *and* $N_1, N_2, W_1, W_2 \in \mathrm{RH}_\infty$ *be given, R having only imaginary poles.*
Then a controller K is said to achieve tracking stability *of the closed loop system if:*

- *it achieves partial tracking stability.*
- *the transfer functions from n_1, n_2, v_1 and v_2 to u are in* RH_∞.

Tracking stability is clearly more restrictive, but also more desirable. We will start however by investigating a necessary and a sufficient condition for existence of controllers that achieve partial tracking stability in lemma 3.3.

Lemma 3.3 *Consider the standard tracking control configuration in fig.2, and let* $G, R \in \mathrm{R}(s)$ *and* $N_1, N_2, W_1, W_2 \in \mathrm{RH}_\infty$ *be given, R having only imaginary poles.*
Then a necessary condition for the existence of a controller K that achieves partial tracking stability is:

$$\exists M \in \mathrm{R}(s) \text{ s.t. } GM = R$$

For sufficiency this can be extended to:
$\exists M \in \mathrm{R}(s)$ *s.t.* $GM = R$ *and*
none of the C_+ *poles of* M *is cancelled in* GM.

Proof:
For necessity first suppose that the condition does not hold. We then must have:

$$\begin{aligned} &GM \neq R,\ \forall M \in \mathrm{R}(s) \quad \Longleftrightarrow \\ &\exists \bar{\delta}_r \in \mathrm{R}^q \text{ s.t. } (GM - R)\bar{\delta}_r \neq 0,\ \forall M \in \mathrm{R}(s) \end{aligned} \tag{2}$$

Given any such $\bar{\delta}_r \in \mathrm{R}^q$, the persistent signal $R\bar{\delta}_r$ can not be tracked by the system G with any control input u, which proves necessity of the first condition.

To prove sufficiency of the extended condition we will construct a suitable controller.
First take any internally stabilizing controller K_G for the system G, such that the transfer functions from v_1 and v_2 to u and y are stable:

$$\begin{matrix} (I - K_G G)^{-1} \\ (I - K_G G)^{-1} K_G \\ (I - G K_G)^{-1} G \\ (I - G K_G)^{-1} \end{matrix} \in \mathrm{RH}_\infty \tag{3}$$

Define $G_K := (I - G K_G)^{-1} G$ as the closed loop transfer function with controller K_G and note that with $GM = R$ we can write the closed loop transfer function from δ_r to e as $G_K M$.
Next take any internally stabilizing controller K_R for $G_K M$; similar to eq.3 we then have:

$$\begin{matrix} (I - K_R G_K M)^{-1} \\ (I - K_R G_K M)^{-1} K_R \\ (I - G_K M K_R)^{-1} G_K M \\ (I - G_K M K_R)^{-1} \end{matrix} \in \mathrm{RH}_\infty \tag{4}$$

From this we will prove that application of the controller $K := K_G + M K_R$ in fig.2 achieves partial tracking stability.
First consider the error signal e as a function of δ_r, v_1 and v_2:

$$e = (I - GK)^{-1} \{ -GM\delta_r + Gv_1 + v_2 \} \tag{5}$$

By using the basic equalities

$$\begin{aligned} (I - GK)^{-1} &= (I - GK)^{-1} GK + I \\ &= G(I - KG)^{-1} K + I \end{aligned} \tag{6}$$

we can rewrite the three transfer functions respectively as:

$$\begin{matrix} -(I - G_K M K_R)^{-1} G_K M, \\ (I - G_K M K_R)^{-1} G_K, \\ (I - G_K M K_R)^{-1} (I - G K_G)^{-1} \end{matrix} \tag{7}$$

From the stability properties of K_G and K_R separately, it is now easy to verify that all these transfer functions are in RH_∞.
Next consider the transfer functions from δ_r, v_1 and v_2 to u:

$$u = (I - KG)^{-1} \{ -KGM\delta_r + v_1 + Kv_2 \} \tag{8}$$

Also these transfer functions can be rewritten:

$$\begin{matrix} M - (I - K_G G)^{-1} M (I - K_R G_K M)^{-1}, \\ (I - K_G G)^{-1} + (I - K_G G)^{-1} \\ \quad \cdot M (I - K_R G_K M)^{-1} K_R G_K, \\ (I - K_G G)^{-1} K_G + (I - K_G G)^{-1} \\ \quad \cdot M (I - K_R G_K M)^{-1} K_R (I - G K_G)^{-1} \} \end{matrix} \tag{9}$$

It can now be verified that all these transfer functions consist of sums and products of elements in which only (some of) the poles of M may appear as unstable poles. Due to the sufficiency condition we may furthermore conclude that all these poles also occur in R. Finally it is straightforward to check that all other transfer functions mentioned in definition 3.1 have the desired properties. We may therefore state that, under the sufficiency condition, the combined controller $K = K_G + M K_R$ achieves partial tracking stability. This concludes the proof. □

With this lemma we now have the possibility to find any controller achieving partial tracking stability in a two-step procedure. However, we will show that not all controllers that can be derived this way will have desirable properties. For this we will bring in a left coprime factorization of M:

$$M = \tilde{D}_M^{-1} \tilde{N}_M \tag{10}$$

The combined controller can then be rewritten as:

$$K = \tilde{D}_M^{-1} \left\{ \tilde{D}_M K_G + \tilde{N}_M K_R \right\} \tag{11}$$

and we can define:

$$\hat{K} := \tilde{D}_M K_G + \tilde{N}_M K_R \tag{12}$$

Now, by incorporating $\tilde{D}_M^{-1}$ in the standard plant and defining a new auxiliary disturbance input signal v_3 and a new control input $\bar{u}$, we arrive at the block diagram given in fig.3 and a modified standard plant $\hat{P}$ given (without auxiliary signals) as:

$$\hat{P} = \left[\begin{array}{ccc|c} -W_1 R & 0 & W_1 G N_2 & W_1 G \tilde{D}_M^{-1} \\ 0 & 0 & W_2 N_2 & W_2 \tilde{D}_M^{-1} \\ \hline -R & N_1 & G N_2 & G \tilde{D}_M^{-1} \end{array}\right] \tag{13}$$

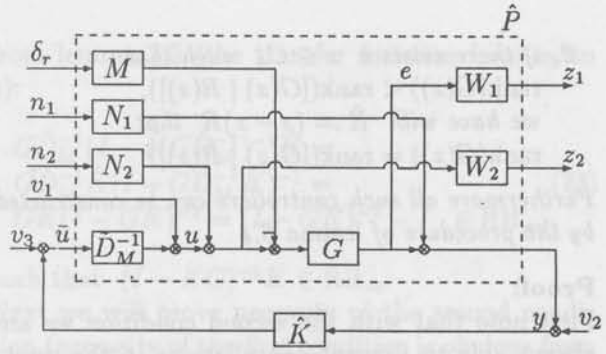


Fig. 3: Modified tracking control configuration

The transfer function from v_3 to y can be found as:

$$\begin{aligned} y &= (I - GK)^{-1} G \tilde{D}_M^{-1} v_3 \\ &= (I - G_K M K_R)^{-1} (I - G K_G)^{-1} G \tilde{D}_M^{-1} v_3 \end{aligned} \tag{14}$$

Left coprimeness of $[\tilde{D}_M \mid \tilde{N}_M]$ implies

$$\left[\tilde{D}_M \vdots \tilde{N}_M\right] \begin{bmatrix} X \\ Y \end{bmatrix} = I \qquad X, Y \in \mathrm{RH}_\infty \tag{15}$$

such that

$$\begin{aligned} y = (I - G_K M K_R)^{-1} (I - G K_G)^{-1} G \\ \cdot \{X + MY\}\, v_3 \end{aligned} \tag{16}$$

and the transfer from v_3 to y is again in RH_∞ (see eq.7). So any set of internally stabilizing controllers, K_G for G and K_R for $G_K M$, will have the property that all transfer functions from $\delta_r, n_1, n_2, v_1, v_2$ and v_3 to y are in RH_∞. The transfer functions from the same inputs to $\bar{u}$ however can be given as:

$$\bar{u} = \hat{K} y + v_3 \tag{17}$$

and may not be stable due to possible unstable poles in $\hat{K}$.

From fig.3 it is however clear that the effect of δ_r can be completely compensated by application of the stable control input $\bar{u} = \tilde{N}_M \delta_r$. Furthermore we have that any C_+ pole appearing in the behaviour of $\bar{u}$ must also appear in u and thus will have no desirable effect in z_2. It seems therefore that such C_+ poles are both unnecessary and undesirable and we will therefore restrict the class of desirable controllers to those that stabilize all transfer functions from the external inputs to both y and $\bar{u}$. The next lemma will show that all internally stabilizing controllers for the product $G\tilde{D}_M^{-1}$ are in this class, such that we can indeed always find a solution without undesirable C_+ poles.

Lemma 3.4 *Consider the standard tracking control configuration in fig.2, and let* $G, R \in \mathrm{R}(s)$ *and* $N_1, N_2, W_1, W_2 \in \mathrm{RH}_\infty$ *be given,* R *having only imaginary poles. Furthermore suppose that* G *and* R *meet the sufficient condition for partial tracking stability given in lemma 3.3, let* $GM = R$ *with* $M \in \mathrm{R}(s)$ *and let a left coprime factorization of* M *be:* $M = \tilde{D}_M^{-1} \tilde{N}_M$. *Finally let the modified standard plant* $\hat{P}$ *be given by fig.3.*

- *Any controller* $\hat{K}$ *for which the closed loop transfer functions from* v_3 *and* v_2 *to* $\bar{u}$ *and* y *are in* RH_∞, *also ensures that the closed loop transfer functions from* n_1, n_2, δ_r *and* v_1 *to* $\bar{u}$ *and* y *are in* RH_∞.
- *For any such controller* $\hat{K}$ *we have that* $K = \tilde{D}_M^{-1} \hat{K}$ *will achieve partial tracking stability according to definition 3.1 for the configuration in fig.2.*

Proof:
To prove the first item, take any controller $\hat{K}$ such that the four transfer functions mentioned are stable:

$$\begin{aligned} &(I - \hat{K} G \tilde{D}_M^{-1})^{-1}, \\ &(I - \hat{K} G \tilde{D}_M^{-1})^{-1} \hat{K}, \\ &(I - G \tilde{D}_M^{-1} \hat{K})^{-1}, \\ &(I - G \tilde{D}_M^{-1} \hat{K})^{-1} G \tilde{D}_M^{-1} \end{aligned} \quad \in \mathrm{RH}_\infty \tag{18}$$

Now define $\bar{v}_3 := v_3 + \tilde{D}_M v_1 + \tilde{D}_M N_2 n_2 + \tilde{N}_M \delta_r$ and $\bar{v}_2 := v_2 + N_1 n_1$ and notice that both signals are in RH_∞. From fig.3 it is then clear that application of all signals seperately is equal to only applying $\bar{v}_3$ instead of v_3 and $\bar{v}_2$ instead of v_2. With this, stability of all transfer functions mentioned is established.

For the second item stability of the transfer functions from all external inputs to y is already proven (eq.7 and 16). Furthermore we have:

$$u = \tilde{D}_M^{-1} \bar{u} + v_1 \tag{19}$$

such that, under the sufficient condition of lemma 3.3, the transfer functions from all external inputs to u can only have the poles of R as unstable poles. With definition 3.1 we may conclude that partial tracking stability is indeed achieved. □

With this lemma we are now able to clarify the specific importance of C_+ poles of M that cancel in the product GM, as mentioned in the the sufficient condition of lemma 3.3. If we consider the equation $GM = R$ with $M = \tilde{D}_M^{-1} \tilde{N}_M$ the cancellation of C_+ poles of M can occur in three different ways:

1. We can find an $\hat{M}$ such that $G\hat{M} = R$, and such that $\hat{M}$ has a lesser McMillan degree than M. Because $G\hat{M} = GM = R$ all poles of M that do not appear in $\hat{M}$ must be cancelled in GM.

2. $\exists z \in C_+$ such that $\text{rank}(G(z)) < \text{rank}([G(z) \mid R(z)])$, but if we define $\hat{R} := (s-z)R$ we have that $\text{rank}(G(z)) = \text{rank}([G(z) \mid \hat{R}(z)])$; z is a C_+ zero of G that does not coincide with a zero of R, nor with a pole of R. M must have C_+ poles that eliminate the effect of such zeros but do not appear in R; they must be cancelled in GM.

3. $\exists z \in C_+$ such that $\text{rank}(G(z)) < \text{rank}([G(z) \mid R(z)])$, and if we define $\hat{R} := (s-z)R$ we still have that $\text{rank}(G(z)) < \text{rank}([G(z) \mid \hat{R}(z)])$; z is a C_+ zero of G that does not coincide with a zero of R, but does coincide with a pole of R. M must have C_+ poles that eliminate the effect of such zeros and must cancel in GM, but M must also have poles at the same location that appear in R.

Cancellations of the first type can be prevented by finding a solution M to the equation $GM = R$ such that M has minimal McMillan degree; this problem is well known as the 'minimal design problem' and solutions are available in literature (see Foster, 1979). Poles of the second type can also be removed. We can redefine R to have C_+ zeros that coincide with those of G, for instance by removing the poles from M that are cancelled in GM and defining $R := GM$. Unfortunately it is not possible to remove poles of the third type. If we attempt to redefine R for these poles, we find that the poles of R at the same location are automatically removed. This then would imply that the persistent signals that were modelled by these specific poles are removed from the problem, thus making it impossible to solve. On the other hand, if we do not redefine R, we can not allow an unstable cancellation occurring in $G\tilde{D}_M^{-1}$; we will therefore consider these cases as unsolvable.

In the following theorem we will now state our main result and show that the necessary condition of lemma 3.3 together with the demand that no pole-zero cancellations of the third kind may occur is necessary and sufficient for existence of controllers that achieve tracking stability.

Theorem 3.5 *Consider the standard tracking control configuration in fig.2, and let* $G, R \in \text{R}(s)$ *and* $N_1, N_2, W_1, W_2 \in \text{RH}_\infty$ *be given, R having only imaginary poles.*
Then necessary and sufficient conditions for the existence of a controller K that achieves tracking stability are:

1. $\exists\, M \in \text{R}(s)$ *such that* $GM = R$
2. *if there exists a* $z \in C_+$ *such that* $\text{rank}(G(z)) < \text{rank}([G(z) \mid R(z)])$, *we have with* $\hat{R} := (s-z)R$ *that* $\text{rank}(G(z)) = \text{rank}([G(z) \mid \hat{R}(z)])$

Furthermore all such controllers can be constructed by the procedure of lemma 3.4

Proof:
First note that with the second condition we are always able to prevent cancellations in the product GM by a correct choice of M and/or an allowable redefinition of R. This then implies that we can meet the sufficient condition for existence of a controller that achieves partial tracking stability as given in lemma 3.3. According to definition 3.2 we therefore only have to prove that with any controller K constructed by the procedure of lemma 3.4, the transfer functions from n_1, n_2, v_1 and v_2 to u are all in RH_∞ if and only if the given conditions are met.
From fig.3 we can find these transfer functions to be $(I-KG)^{-1}KN_1$, $(I-KG)^{-1}KGN_2$, $(I-KG)^{-1}$ and $(I-KG)^{-1}K$ respectively; clearly with $N_1, N_2 \in \text{RH}_\infty$ and $(I-KG)^{-1}KG = (I-KG)^{-1} - I$ we only need to consider the latter two.
First we will look at sufficiency for the transfer function from v_1 to u:

$$(I-KG)^{-1} = \tilde{D}_M^{-1}(I - \hat{K}G\tilde{D}_M^{-1})^{-1}\tilde{D}_M \qquad (20)$$

With $(I - \hat{K}G\tilde{D}_M^{-1})^{-1}$ and $\tilde{D}_M$ in RH_∞ we have from lemma 3.4 that any C_+ poles of this transfer function must be poles of $\tilde{D}_M^{-1}$. Now none of these poles is cancelled by G in the product $G\tilde{D}_M^{-1}$ so any C_+ pole of $\tilde{D}_M^{-1}(I - \hat{K}G\tilde{D}_M^{-1})^{-1}\tilde{D}_M$ is also a pole of $G\tilde{D}_M^{-1}(I - \hat{K}G\tilde{D}_M^{-1})^{-1}\tilde{D}_M$. However, we have:

$$\begin{aligned} &G\tilde{D}_M^{-1}(I - \hat{K}G\tilde{D}_M^{-1})^{-1}\tilde{D}_M = \\ &(I - G\tilde{D}_M^{-1}\hat{K})^{-1}G = (I-GK)^{-1}G \;\in \text{RH}_\infty \end{aligned} \qquad (21)$$

because of lemma 3.3 (the transfer function from v_1 to e). We therefore must have that $(I-KG)^{-1} \in \text{RH}_\infty$.
Next consider sufficiency for the transfer function from v_2 to u:

$$\begin{aligned} (I-KG)^{-1}K &= (I - \tilde{D}_M^{-1}\hat{K}G)^{-1}\tilde{D}_M^{-1}\hat{K} \\ &= \tilde{D}_M^{-1}(I - \hat{K}G\tilde{D}_M^{-1})^{-1}\hat{K} \end{aligned} \qquad (22)$$

Again, but now with $(I - \hat{K}G\tilde{D}_M^{-1})^{-1}\hat{K} \in \text{RH}_\infty$, we have from lemma 3.4 that any C_+ poles of this transfer function must be poles of $\tilde{D}_M^{-1}$. So in this case any C_+ pole of $\tilde{D}_M^{-1}(I - \hat{K}G\tilde{D}_M^{-1})^{-1}\hat{K}$ is also a pole of $G\tilde{D}_M^{-1}(I - \hat{K}G\tilde{D}_M^{-1})^{-1}\hat{K}$. Now we have

from lemma 3.3 (the transfer function from v_2 to e):

$$\begin{aligned} &G\tilde{D}_M^{-1}(I-\hat{K}G\tilde{D}_M^{-1})^{-1}\hat{K} = \\ &G\tilde{D}_M^{-1}\hat{K}(I-G\tilde{D}_M^{-1}\hat{K})^{-1} = \\ &GK(I-GK)^{-1} = (I-GK)^{-1} - I \ \in \mathrm{RH}_\infty \end{aligned} \tag{23}$$

such that $(I-KG)^{-1}K \in \mathrm{RH}_\infty$.

Next we will prove necessity of the second condition (necessity of the first condition is obvious from lemma 3.3). For this, consider the case that an unstable pole-zero cancellation of the type given in condition 2 does occur in the product $G\tilde{D}_M^{-1}$, such that the offending zero has a corresponding pole in R. It can then be verified from the equation:

$$R = GM = G\tilde{D}_M^{-1}\tilde{N}_M \tag{24}$$

that this pole must appear in $\tilde{D}_M^{-1}$ with a multiplicity that is at least one higher than that in R, and that the cancellation does not remove the pole under consideration completely but merely reduces its multiplicity.

Now suppose that in spite of this we can still find a controller K that achieves tracking stability. We will use the fact that—among others—the following transfer function matrices must then be stable:

$$\begin{aligned} \tfrac{u}{v_2} &= K(I-GK)^{-1} \\ &= \tilde{D}_M^{-1}\hat{K}(I-G\tilde{D}_M^{-1}\hat{K})^{-1} \quad \in \mathrm{RH}_\infty \\ \tfrac{e}{\delta_r} &= -(I-GK)^{-1}GM \\ &= -(I-G\tilde{D}_M^{-1}\hat{K})^{-1}G\tilde{D}_M^{-1}\tilde{N}_M \ \in \mathrm{RH}_\infty \end{aligned} \tag{25}$$

First consider the case that none of the C_+ poles of $\tilde{D}_M^{-1}$ is cancelled in the product $\tilde{D}_M^{-1}\hat{K}$. Clearly, if $\tilde{D}_M^{-1}\hat{K}(I-G\tilde{D}_M^{-1}\hat{K})^{-1} \in \mathrm{RH}_\infty$, we then must have for any C_+ pole of $\tilde{D}_M^{-1}$ a corresponding zero of $(I-G\tilde{D}_M^{-1}\hat{K})^{-1}$, that is, a pole of $G\tilde{D}_M^{-1}\hat{K}$. Due to the cancellation in $G\tilde{D}_M^{-1}$ however, we are always at least one zero 'short' to cancel all C_+ poles.

This implies that a pole-zero cancellation in $G\tilde{D}_M^{-1}$ might only be allowable if there is a corresponding pole-zero cancellation in $\tilde{D}_M^{-1}\hat{K}$. Now, because the cancellation in $G\tilde{D}_M^{-1}$ does not remove the pole under consideration completely but merely reduces its multiplicity, this implies that the number of C_+ poles—counting multiplicities—in the product $G\tilde{D}_M^{-1}\hat{K}$ would be at least one less than that in $G\tilde{D}_M^{-1}$. This then would make $-(I-G\tilde{D}_M^{-1}\hat{K})^{-1}G\tilde{D}_M^{-1}\tilde{N}_M \notin \mathrm{RH}_\infty$ according to the same argument that was given before and coprimeness of $[\tilde{D}_M \mid \tilde{N}_M]$.

So any pole-zero cancellations in $G\tilde{D}_M^{-1}$ results in either $-(I-GK)^{-1}GM \notin \mathrm{RH}_\infty$ or $K(I-GK)^{-1} \notin \mathrm{RH}_\infty$ which is in contradiction with tracking stability.

Finally we will prove that *all* controllers that achieve tracking stability can be constructed by the procedure of lemma 3.4. For this, take any controller K that achieves tracking stability and define $\hat{K} := \tilde{D}_M K$. Now note from fig.3 that application of v_3 is equal to application of $v_1 = \tilde{D}_M v_3$, and that $\bar{u} = \tilde{D}_M u$. It is then possible to verify that $\hat{K}$ stabilizes the transfer functions from v_3 and v_2 to $\bar{u}$ and y such that the controller K can always be constructed by the procedure of lemma 3.4. This concludes the proof of the theorem. □

To conclude this section we will consider a special case, resulting in the well-known robust servomechanism problem (see Davison and Goldenberg, 1975). For this, define R in fig.2 to have the following structure:

$$R = \alpha I_q, \quad \alpha \in \mathrm{R}(s), \quad \alpha \text{ having only imaginary poles} \tag{26}$$

The reference signal vector thus consists of q independent signals α, with q being the number of error signals. So any output signal of the plant G must be able to track all possible signals generated by α, and this independent of all other outputs. This leads to the following corollary.

Corollary 3.6 *Consider the standard tracking control configuration in fig.2, and let* $G \in \mathrm{R}(s)^{q\times n}$ *and* $N_1, N_2, W_1, W_2 \in \mathrm{RH}_\infty$ *be given,* $R = \alpha I_q$ *with* $\alpha \in \mathrm{R}(s)$ *having only imaginary poles.*
Then necessary and sufficient conditions for existence of a controller K *that achieves tracking stability are:*

1. $\mathrm{rank}(G) = q$, *q is the number of outputs of G*
2. *G has no transmission zeros that are equal to poles of α*

Furthermore, tracking stability will be maintained for perturbations of the system G as long as the number of C_+ poles in any of the transfer functions mentioned in definition 3.1 remains unchanged.

Proof:
The proof presented here will be using the results of lemmas 3.3 and 3.4 and theorem 3.5; the original proof can be found in Davison and Goldenberg (1975).

First note that condition 1 implies $\mathrm{rank}(G) = \mathrm{rank}([G \mid R])$ which in its turn implies that there exists an $M \in \mathrm{R}(s)$ such that $GM = R$ (see Vidyasagar, 1985). Furthermore G has full row

rank, so it has a right inverse $G^R \in \mathrm{R}(s)$ and we can take $M = G^R R = \alpha G^R$. Now suppose G has no C_+ zeros (otherwise condition 2 ensures that R can be redefined). A left coprime factorization of M then is $M = \tilde{D}_M^{-1}\tilde{N}_M$ with $\tilde{D}_M = \frac{1}{\alpha\beta}I$ and $\tilde{N}_M = \frac{1}{\beta}G^R$ (β is some polynomial with zeros in the open left half plane such that $\tilde{D}_M$ and $\tilde{N}_M$ are proper). Both conditions now ensure that there are no C_+ cancellations in the product $G \cdot \alpha\beta I$ and lemma 3.4 and theorem 3.5 thus prove that all controllers $\hat{K}$ that stabilize $\alpha\beta G$ define a controller $K = \alpha\beta\hat{K}$ that achieves tracking stability.
To prove the second part of the corollary, consider a perturbed system G_p and apply the same controller $K = \alpha\beta\hat{K}$. As long as

1. $\hat{K}$ stabilizes $\alpha\beta G_p$ and
2. G_p has full rank and no transmission zeros equal to poles of α

tracking stability is ensured due to lemma 3.4 and theorem 3.5. It can be verified that violation of 1) destabilizes one of the stable transfer functions mentioned in definition 3.1, and that violation of 2) makes the number of imaginary poles in the transfer from δ_r to u decrease at least by one. The given condition is therefore sufficient to achieve tracking stability for the perturbed system. □

It is now clear that the procedure given thusfar, brings the unstable behaviour defined in R into the control loop by means of $\tilde{D}_M^{-1}$. The resulting incorporation of this term into the final controller K is also referred to as the 'internal model principle' (Francis and Wonham, 1975); from this $\tilde{D}_M^{-1}$ itself can be seen as the 'internal model' of the reference signal's behaviour.

4 Application of optimal control methods

We have shown that all controllers that achieve output regulation and tracking for the control configuration of fig.2 can be found by the procedure given in lemma 3.4. This procedure can be summarized as follows:

1. Construct the standard tracking configuration according to fig.2, with G the plant to be controlled and N_1, N_2, W_1, W_2 some stable weights to account for performance objectives other than tracking.
2. Model all occurring persistent reference signals r as a combination of sinusoids and polynomials and construct a real-rational transfer function matrix R with imaginary poles to generate them.
3. Make sure R has C_+ zeros that coincide with those of G; if it is impossible to do this without removing some of the poles of R the problem is unsolvable (condition 2 of theorem 3.5).
4. Find $M \in \mathrm{R}(s)$ such that $GM = R$ and such that M has minimal McMillan degree; this is a minimal design problem (Foster, 1979).
5. Take a LCF $M = \tilde{D}_M^{-1}\tilde{N}_M$ and construct the modified tracking configuration of fig.3.
6. Find a stabilizing controller $\hat{K}$ for the modified standard plant $\hat{P}$.
7. Construct the controller $K := \tilde{D}_M^{-1}\hat{K}$ that achieves tracking stability.

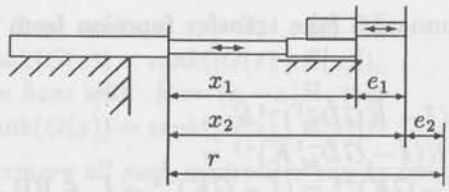


Fig. 4: The two-stage actuator

An explicit parametrization of all controllers that achieve tracking stability can be found if we adjust the weight W_2 such that no persistent signals can occur in z_2. This can be done by defining W_2 as $W_2 := W_2'\tilde{D}_M$ with $W_2' \in \mathrm{RH}_\infty$. The modified standard plant $\hat{P}$ then becomes internally stabilizable and a Youla-parametrization of all internally stabilizing controllers $\hat{K}$ can readily be found (Youla et al. 1976). Next if we take R, N_1, N_2, W_1 and W_2 such that the modified standard plant $\hat{P}$ is proper, we can use standard H_2 or H_∞ optimal control theory to find an optimal $\hat{K}$ (or any other design procedure based on a standard plant formulation). This will then lead to a controller K that under the condition of achieving tracking stability, simultaneously optimizes other performance objectives like disturbance rejection or stability robustness. Although the exact influence on these other performance objectives of the incorporation of C_+ zeros mentioned in step 3 and the extra demand on W_2 is not yet completely clear, the example considered in the next section will show some of the advantages of this unifying approach.

5 Example: the two-stage actuator

This example is derived from a two-stage actuator as given in fig.5. It is intended to use a slow actu-

ator with a large operational range in combination with a fast actuator with a small operational range. The slow actuator is intended to let x_1 asymptotically track step-like reference signals while the fast operator is used for high frequency disturbance reduction. We assume that the system is governed by the following simplified relations and signals:

$$\begin{aligned} x_2 &= x_1 + e_1 \\ x_1 &= \tfrac{1}{s^2+s+1}\,(u_1 + n_{21}) \\ z_{11} = e_1 &= \tfrac{s}{s+10}\,(u_2 + n_{22}) \\ z_{12} = e_2 &= x_2 - r \\ r &= \tfrac{1}{s}\,\delta_r \\ y_1 &= e_1 + .01 n_{11} \\ y_2 &= e_2 + .01 n_{12} \\ z_{21} &= u_1 + n_{21} \\ z_{22} &= u_2 + n_{22} \end{aligned} \tag{27}$$

To bring this system into the standard tracking control configuration of fig.2 we define:

$$G := \begin{bmatrix} 0 & \frac{s}{s+10} \\ \frac{1}{s^2+s+1} & \frac{s}{s+10} \end{bmatrix} \qquad R := \begin{bmatrix} 0 \\ \frac{1}{s} \end{bmatrix}$$

$$N_1 := \begin{bmatrix} .01 & 0 \\ 0 & .01 \end{bmatrix} \quad N_2 := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{28}$$

$$W_1 := \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}$$

With this we can immediately verify the second condition of theorem 3.5, furthermore we can take:

$$M := \begin{bmatrix} \frac{s^2+s+1}{s} \\ 0 \end{bmatrix}$$

$$\tilde{D}_M := \begin{bmatrix} \frac{s}{s^2+s+1} & 0 \\ 0 & 1 \end{bmatrix} \quad \tilde{N}_M := \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{29}$$

to also satisfy the first condition.
As was suggested in the previous section we will define $W_2 := W_2' \tilde{D}_M$ and choose $W_2' = I$ such that the modified standard plant $\hat{P}$ becomes internally stabilizable and optimal control theory can be applied. From eq.13 we can find this modified standard plant as

$$\hat{P} = \left[\begin{array}{ccccc|cc} 0 & 0 & 0 & 0 & \frac{s}{s+10} & 0 & \frac{s}{s+10} \\ -\frac{10}{s} & 0 & 0 & \frac{10}{s^2+s+1} & \frac{10s}{s+10} & \frac{10}{s} & \frac{10s}{s+10} \\ 0 & 0 & 0 & \frac{s}{s^2+s+1} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ \hline 0 & .01 & 0 & 0 & \frac{s}{s+10} & 0 & \frac{s}{s+10} \\ -\frac{1}{s} & 0 & .01 & \frac{1}{s^2+s+1} & \frac{s}{s+10} & \frac{1}{s} & \frac{s}{s+10} \end{array}\right] \tag{30}$$

For this plant a fourth order H_∞ optimal controller $\hat{K}$ was determined, achieving an upper bound of

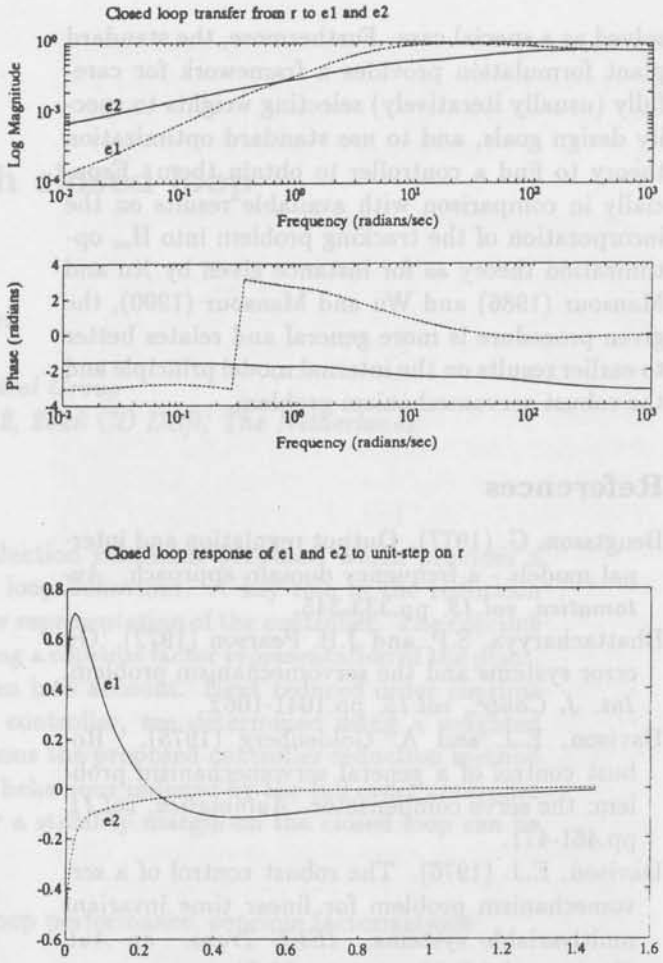


Fig. 5: Closed loop behaviour of two-stage actuator

1.33 for the minimal ∞-norm of the closed loop system transfer from δ_r, n_1 and n_2 to z_1 and z_2. This then gives a fifth order controller $K = \tilde{D}_M^{-1}\hat{K}$ that achieves tracking stability for the original system. Fig.5 shows the tracking property both in the frequency-domain and in the time-domain. Note that a 'robust' solution according to corollary 3.6 does not exist due to the C_+ transmission zero of G at $s = 0$; the structural zero in the upper left corner of G however, still ensures robustness against parameter variations (see Grasselli and Longhi, 1991). Also note that next to the tracking property we can establish a trade-off between properties like disturbance rejection (determined by N_1 and W_1) and stability robustness (determined by N_2 and W_2).

6 Conclusion

We have given necessary and sufficient conditions for the non-robust asymptotic tracking problem and have shown that the robust problem can be

solved as a special case. Furthermore, the standard plant formulation provides a framework for carefully (usually iteratively) selecting weights to specify design goals, and to use standard optimization theory to find a controller to obtain them. Especially in comparison with available results on the incorporation of the tracking problem into H_∞ optimization theory as for instance given by Xu and Mansour (1986) and Wu and Mansour (1990), the given procedure is more general and relates better to earlier results on the internal model principle and the robust servomechanism problem.

References

Bengtsson, G. (1977). Output regulation and internal models - a frequency domain approach. *Automatica, vol.13.* pp.333-345.

Bhattacharyya, S.P. and J.B. Pearson (1972). On error systems and the servomechanism problem. *Int. J. Contr., vol.15.* pp.1041-1062.

Davison, E.J. and A. Goldenberg (1975). Robust control of a general servomechanism problem: the servo compensator. *Automatica, vol.11.* pp.461-471.

Davison, E.J. (1976). The robust control of a servomechanism problem for linear time invariant multivariable systems. *IEEE Trans. on Aut. Contr., vol.AC-21, no.1.* pp.25-34.

Desoer, C.A. and Y.T. Wang (1980). *Linear time-invariant robust servomechanism problem: a self contained exposition.* in: C.T.Leondes ed., Control and Dynamic Systems, Academic Press, New York. pp.81-129.

Doyle, J.C., C.C. Chu, B.A. Francis, P. Khargonekar and G. Stein (1984). *Lecture notes in advances in multivariable control.* ONR/Honeywell Workshop, Minneapolis.

Foster, L.V. (1979). A practical solution to the minimal design problem. *IEEE Trans. on Aut. Contr., vol.AC-24, no.3.* pp.449-454.

Francis, B.A. (1977). The multivariable servomechanism problem from the input output viewpoint. *IEEE Trans. on Aut. Contr., vol.AC-22, no.3.* pp.322-328

Francis, B.A. (1987). *A course in H-infinity control theory.* Lect. Notes in Contr. and Inf. Sc., vol.88, Springer Verlag, Berlin.

Francis, B.A. and M. Vidyasagar (1983). Algebraic and topological aspects of the regulator problem for lumped linear systems. *Automatica, vol.19, no.1.* pp.87-90.

Francis, B.A. and W.M. Wonham (1975). The internal model principle for linear multivariable regulators. *Applied Mathematics and Optimization, vol.2.* pp.170-194.

Grasselli, O.M. and S. Longhi (1991). Robust output regulation under uncertainties of physical parameters. *Sys. and Contr. Lett., vol.16.* pp.33-40.

Lancaster, P. and M. Tismenetsky (1985). *The theory of matrices.* Academic Press, London.

Vidyasagar, M. (1985). *Control system synthesis; a factorization approach.* MIT Press, Cambridge, Massachusetts.

Wu, Q.H. and M. Mansour (1990). H_∞ optimal solutions of robust regulator problem for linear MIMO systems. *Int. J. Contr., vol.52, no.5.* pp.1241-1262.

Xu, J.H. and M. Mansour (1986). Design of H_∞ optimal robust controllers: stability, asymptotic regulation and disturbance rejection. *Proc. IEEE Conf. on Dec. and Contr..* pp.1-6.

Youla, D.C., H.A. Jabr, and J.J. Bongiorno jr. (1976). Modern Wiener-Hopf design of optimal controllers part II: The multivariable case. *IEEE Trans. on Aut. Contr., vol.AC-21, no.3.* pp.319-338.

Selected Topics in Identification, Modelling and Control
Vol. 3, June 1991

Controller reduction with closed loop stability margins

Peter M.M. Bongers, Okko H. Bosgra

Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. In this paper a controller reduction scheme is proposed which provides a guaranteed stability margin of the closed loop behaviour. A key role in the reduction scheme is played by a specific coprime factor representation of the controller. The coprime factors of the controller are constructed using a coprime factor representation of the plant. Thereby the closed loop behaviour is taken into account. Next reduced order coprime factors, parametrizing the reduced order controller, are determined using a weighted Hankel norm approximation. For this reasons the proposed controller reduction method will minimize the difference in closed loop behaviour induced by the full order controller and the reduced order controller, secondly a stability margin on the closed loop can be calculated.

Keywords. Controller reduction, closed loop performance, coprime factorizations

1 Introduction

For various reasons low order controllers are preferred rather than high order controllers:

- The advantage in computational requirements of low order controllers are especially important in consumer electronics applications when a mass production of the controller is necessary.
- Another motivation stems from controller design methods: H_∞ controllers are usually designed on the plant model including weighting functions. These weighting functions are used to shape for example the desired sensitivity function. The order of the controller equals the order of the plant plus the order of the weighting functions. When for example the H_∞ design is based on a normalized coprime factor description of the plant, the order of the controller equals the order of the plant and two times the order of the weighting functions (McFarlane and Glover (1989)). Therefore the order of the H_∞-controller will be in general unnecessary high and reduction seems useful.

The major difficulty in reducing the order of the controller is to ensure that the reduction error does not adversely affect the closed loop objectives. It is assumed that the high order controller is designed such that the closed loop objectives can be specified by the closed loop transfer function $T(C,P)$. For convenience we have denoted the closed loop transfer function of a plant P controlled by the full order controller C by $T(C,P)$, the plant controlled by the reduced order controller C_r will be denoted by $T(C_r,P)$.

An open loop approach like ((Moore (1981), Pernebo and Silverman (1982), Kabamba (1985)) to controller reduction will not take the closed loop objectives into account.

A natural way of relating open loop properties to closed loop properties is by making use of the graph topology (Vidyasagar (1984)). For controller reduction this implies a right coprime factorization of the high order controller is approximated by a right coprime factorization of the low order controller. In the graph-topology the difference between the right coprime factorization of the high order controller and the right coprime factorization of the low or-

der controller is isomorph to the difference between their closed loop transfer functions. Therefore a reduction problem stated in the graph-topology will take the specified closed loop objectives into account.

If the high order controller is represented by a normalized right coprime factorization the reduction error can be stated in the gap-metric and will induce a bound on the order of the low order controller such that the closed loop remains stable. The reason for this is that the low order controller can be seen as a perturbation/deviation of the full order controller, then results on robust stability in the gap-metric (Georgiou and Smith (1990), Bongers and Bosgra (1990)) directly apply to controller reduction. In this respect, if desired, the stability margin derived in Liu et al. (1990) eqn.(21) can be stated in the gap-metric.

It is known that the gap-metric approach results frequently into "very" conservative robustness margins. In this paper we will extend the results of Bongers (1991b) to controller reduction in a closed loop setting. Using these results we obtain a less conservative robustness margin and thereby allow the application of lower order controllers. This margin is based on a coprime factorization of the controller determined by the closed loop transfer function. Hence we take the closed loop objectives into account and are able to guarentee stability of $T(C_r, P)$. A certain amount of performance will also be gained by the graph-topology setting of the whole reduction problem.

The layout of this paper is as follows: after the preliminaries in Section 2 we will formulate the controller reduction problem in Section 3. In Section 4 the main result of this paper is stated followed by an example in Section 6 and Section 7 contains the conclusions.

2 Preliminaries

In this note we adopt the ring theoretic setting of (Desoer et al. (1980), Vidyasagar et al. (1982)) to study stable multivariable linear systems by considering them as transfer function matrices having all entries belonging to a ring $\mathcal{H}$. For the application of state-space algorithms we will identify the ring $\mathcal{H}$ with $\mathbb{RH}_\infty$, the space of stable real rational finite dimensional linear time-invariant continuous-time systems. We consider the class of stable/unstable multivariable systems as transfer function matrices whose entries are elements of the quotient field $\mathcal{F}$ of $\mathcal{H}$ ($\mathcal{F} := \{a/b \mid a \in \mathcal{H},\ b \in \mathcal{H}\backslash 0\}$). The set of multiplicative units of $\mathcal{H}$ is defined as: $\mathcal{J} := \{h \in \mathcal{H} \mid h^{-1} \in \mathcal{H}\}$. In the sequel systems $P \in \mathcal{F}^{m\times n}$ are denoted as $P \in \mathcal{F}$.

Proposition 2.1 *Let $P \in \mathcal{H}$. The Hankel singular values of P are:*

$$\sigma_i^H\{P\} = \lambda_i(W_c W_o)^{\frac{1}{2}} = \sigma_i$$

where W_c, W_o are the symmetric positive definite solutions to the controllable and observable grammians of P. The Hankel norm of P is defined as $\| P \|_H = \sigma_1^H$.

If $P \in \mathcal{F}$ then the Hankel singular values of P are the Hankel singular values of the proper stable part of P i.e. $P|_{\mathcal{H}}$.

The H_∞-norm of P is defined as $\|P\|_\infty = \sup_\omega \max \sigma_i[P(j\omega)]$.

Factorizations

Definition 2.2 (Vidyasagar et al. (1982)) *A plant $P \in \mathcal{F}$ has a right (left) fractional representation if there exist $N, M(\tilde{N}, \tilde{M}) \in \mathcal{H}$ such that $P = NM^{-1}$ $(= \tilde{M}^{-1}\tilde{N})$.*

*The pair $M, N(\tilde{M}, \tilde{N})$ is right (left) coprime (*rcf *or* lcf *) if it is a right (left) fraction and there exists $U, V(\tilde{U}, \tilde{V}) \in \mathcal{H}$ such that: $UN + VM = I$ $(\tilde{N}\tilde{U} + \tilde{M}\tilde{V} = I)$*

*The pair $M, N(\tilde{M}, \tilde{N})$ is called normalized right (left) coprime (*nrcf *or* nlcf *) if it is coprime and: $M^*M + N^*N = I$ $(\tilde{M}\tilde{M}^* + \tilde{N}\tilde{N}^* = I)$ with $M^* = M^T(-s)$.*

3 Closed loop stability

In this section we will study closed loop stability according to Fig. 1, where we assume that a stabilizing controller C has been designed for the plant P. The transfer function $H(C, P)$ mapping the external inputs (e_1, e_2) onto the outputs (u_1, u_2) is

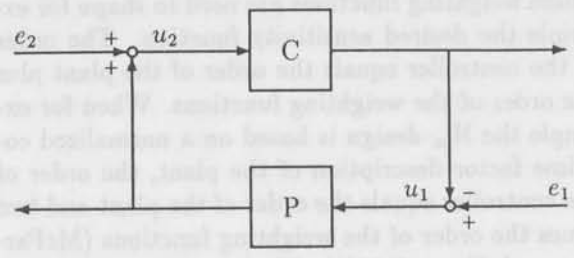


Fig. 1: closed loop configuration

given by:

$$H(C,P) = \begin{bmatrix} I & C \\ -P & I \end{bmatrix}^{-1} = \begin{bmatrix} I - C(I+PC)^{-1}P & -C(I+PC)^{-1} \\ (I+PC)^{-1}P & (I+PC)^{-1} \end{bmatrix}$$

Stability of the closed loop, i.e. the controller C internally stabilizes the plant P, is guaranteed if and only if $H(C,P) \in \mathcal{H}$. Define a closed loop transfer function $T(C,P)$ by

$$T(C,P) = H(C,P) - \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -C \\ I \end{bmatrix} (I+PC)^{-1} \begin{bmatrix} P & I \end{bmatrix}$$

then using the definition of the ring $\mathcal{H}$ we have $H(C,P) \in \mathcal{H} \Leftrightarrow T(C,P) \in \mathcal{H}$. The advantage of studying $T(C,P)$ instead of $H(C,P)$ becomes clear when we use coprime factorizations for the plant and controller. Now let $C = YX^{-1}$ with (X,Y) a rcf of C and let $P = \tilde{M}^{-1}\tilde{N}$ with $(\tilde{M},\tilde{N})$ a lcf of P then:

$$T(C,P) = \begin{bmatrix} -Y \\ X \end{bmatrix} (\tilde{M}X + \tilde{N}Y)^{-1} \begin{bmatrix} \tilde{N} & \tilde{M} \end{bmatrix} \quad (1)$$

Using the coprime representation of plant and controller the closed loop structure of Fig. 1 can be redrawn as in Fig. 2 with $\xi = \tilde{M}e_2 + \tilde{N}e_1$.

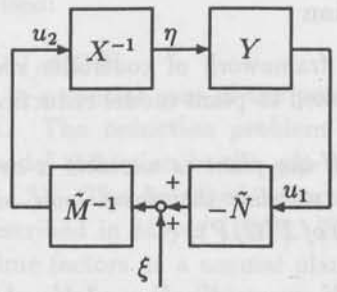


Fig. 2: Closed loop structure with coprime representation

Lemma 3.1 (Vidyasagar (1985)) *Let $C \in \mathcal{F}$ be given as $C = YX^{-1}$ with (X,Y) a rcf of C and let the plant $P \in \mathcal{F}$ be given as $P = \tilde{M}^{-1}\tilde{N}$ with $(\tilde{M},\tilde{N})$ a lcf of P. Then stability of the closed loop $T(C,P)$ is equivalent to:*

$$\Lambda \in \mathcal{J} \quad \text{with} \quad \Lambda = (\begin{bmatrix} \tilde{M} & \tilde{N} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}) \quad (2)$$

Remark 3.2 *Λ^{-1} is the transfer function from ξ to η. Thus stability of the closed loop is equivalent to stability of the transfer function Λ^{-1} from ξ to η, whereas Λ is stable by definition.*

4 Controller reduction framework

Usually a full order controller (with the order of the controller equal to the order of the plant and weightings) is designed based on the plant model P such that the closed loop $T(C,P)$ will be stable and have a certain performance. For controller reduction it is essential that the closed loop transfer function remains stable for reduced order controllers C_r close to C. In a more precise formulation of the controller reduction problem it is usefull to determine the largest class $\mathcal{S}$ such that:

$$C_r \in \mathcal{S} := \{C_r \in \mathcal{F} \mid T(C_r,P) \in \mathcal{H}\}$$

In order to determine a large class $\mathcal{S}$ we will exploid properties of a nlcf of the plant in conjunction with an associated rcf of the full order controller. Furthermore we will parametrize the reduced order controller by the coprime factors of the full order controller and additive transfer functions to describe the difference between the full order controller and the reduced order controller. In a robust control setting these additive transfer functions can be seen as perturbations/deviations on coprime factors of the controller. An upper bound on these coprime factor deviations such that the closed loop remains stable is given in the following theorem.

Theorem 4.1 *Let the plant $P = \tilde{M}^{-1}\tilde{N}$ be given, with $(\tilde{M},\tilde{N})$ a nlcf of P and let the controller $C \in \mathcal{F}$ (internally stabilizing P) be given as $C = YX^{-1}$ with (X,Y) a rcf such that:*

$$(\begin{bmatrix} \tilde{M} & \tilde{N} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}) = I$$

and let the reduced order controller $C_r \in \mathcal{F}$ be given as:

$$C_r = Y_r X_r^{-1} = (Y - \Delta Y)(X - \Delta X)^{-1} \quad (3)$$

Then a sufficient condition for stability of $T(C_r,P)$ is given by:

$$\| \begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} \|_\infty < 1$$

Proof: $T(C_r,P)$ is stable if Lemma 3.1 holds:

$$\Lambda = \left(\tilde{M}X_r + \tilde{N}Y_r\right) \in \mathcal{J}$$

Using the factorization of the controller given by (3) Λ can be written as:

$$\Lambda = \left(I - \begin{bmatrix} \tilde{M} & \tilde{N} \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} \right)$$

Applying the small gain theorem we have a sufficient condition for stability of $T(C_r, P)$:

$$\left\| \begin{bmatrix} \tilde{M} & \tilde{N} \end{bmatrix} \right\|_\infty \left\| \begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} \right\|_\infty < 1$$

Using the fact that $\left\| \begin{bmatrix} \tilde{M} & \tilde{N} \end{bmatrix} \right\|_\infty = 1$ by definition of normalized left coprimeness, the theorem is proved. □

Next we will use a transfer function $Q \in \mathcal{H}$ to characterize the class of reduced order controller factorizations and determine a $Q \in \mathcal{H}$ such that the coprime factor deviations are minimized. For this reason the reduced order controller can also be factorized as $C_r = (Y_r Q)(X_r Q)^{-1}$ with $Q \in \mathcal{H}$ such that $(Y_r Q, X_r Q)$ is a right fraction. Then the deviation between the full order controller and the reduced order controller are written as:

$$\begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} X_r \\ Y_r \end{bmatrix} Q$$

An uppper bound on all allowable deviations is described by:

$$\left\| \begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} \right\|_\infty = \inf_{Q \in \mathcal{H}} \left\| \begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} X_r \\ Y_r \end{bmatrix} Q \right\|_\infty \quad (4)$$

Theorem 4.2 *Given a full order controller $C = Y_n X_n^{-1}$ with (X_n, Y_n) a* nrcf *stabilizing a plant $P = \tilde{M}^{-1}\tilde{N}$ with $(\tilde{M}, \tilde{N})$ a* nlcf *and define:*

$$\Lambda = \left(\tilde{M} X_n + \tilde{N} Y_n \right) \in \mathcal{J}$$

then the class $\mathcal{S}$ of all stabilizing low order controllers can be written as:

$$\mathcal{S} = \{ C_r \mid T(C, P) \in \mathcal{H} \text{ and } \inf_{\tilde{Q} \in \mathcal{H}} \left\| \left(\begin{bmatrix} X_n \\ Y_n \end{bmatrix} - \begin{bmatrix} X_r \\ Y_r \end{bmatrix} \tilde{Q} \right) \Lambda^{-1} \right\|_\infty < 1 \} \quad (5)$$

Proof: Another coprime factor description (X, Y) of the controller is given by:

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X_n \\ Y_n \end{bmatrix} \Lambda^{-1}$$

Using this factorization sequential in Theorem 4.1 and in (4) proves this theorem. □

Corollary 4.3 *Given Theorem 4.2 and the fact that $\|\Lambda^{-1}\|_\infty = \|T(C,P)\|_\infty$ for normalized coprime factorization of plant and controller a sufficient condition for stability of* (5) *is*

$$\inf_{\tilde{Q} \in \mathcal{H}} \left\| \begin{bmatrix} X_n \\ Y_n \end{bmatrix} - \begin{bmatrix} X_r \\ Y_r \end{bmatrix} \tilde{Q} \right\|_\infty < \frac{1}{\|T(C,P)\|_\infty}$$

The stability condition of Corollary 4.3 is equal to stability in gap-metric sense (Georgiou and Smith (1990), Bongers and Bosgra (1990)).

Remark 4.4 *Only in case of $\Lambda = \alpha I$ the application of the multiplicative H_∞-norm properties on* (5) *implies that the stability margin defined in Theorem 4.2 is as conservative as the margin in gap-metric sense.*

To provide an answer to the question wether $\Lambda = \alpha I$ occurs frequently in control design consider the following remark.

Remark 4.5 *If an $(n-1)^{th}$ order controller has been calculated for an n^{th} order plant using a normalized coprime factor approach (Bongers (1991a)) then $\Lambda = \alpha I$ and there is no difference between the the stability margin in gap-metric sense and the margin derived in Theorem 4.2. Therefore when a lower order controller has been calculated or the controller has been designed with another method Λ will in general not be equal to αI.*

Plant reduction

The presented framework of controller reduction can also be applied to plant model reduction.

Remark 4.6 *If the plant is unstable a controller is necessary to stabilize the closed loop, since we require stability of $T(C,P)$.*

Remark 4.7 *If a controller is available which stabilizes the closed loop then model reduction is equivalent with controller reduction. Then the interchange of plant and controller in Theorem 4.2 yields the closed loop model reduction.*

In the next remark we will show that open loop model reduction is a special case of closed loop model reduction.

Remark 4.8 *We assume a stable plant without a controller, i.e. a controller equal to 0. A normalized left coprime factorization of the zero controller $C = \tilde{X}^{-1}\tilde{Y}$ is $\tilde{Y} = 0, \tilde{X} = I$, then according to*

Theorem 4.2 $\Lambda^{-1} = M^{-1}$. *The coprime factor representation of the plant* $\begin{bmatrix} M \\ N \end{bmatrix} \Lambda^{-1}$ *equals* $\begin{bmatrix} P \\ I \end{bmatrix}$, *which is the open loop plant reduction. Therefore open loop model reduction is a special case of closed loop model reduction.*

5 Application to reduction

We will now apply the framework developed in the previous section in an actual reduction scheme.

The freedom $\tilde{Q}$ in (5) can not be used in the actual reduction procedure but only in the stability analysis phase. For this reasons in the reduction scheme we will use a $\tilde{Q} = 1$. Given a plant and controller by their normalized coprime factorization as in Theorem 4.2 the controller reduction problem is formulated as:

$$\inf_{X_r, Y_r \in \mathcal{H}} \left\| \left(\begin{bmatrix} X_n \\ Y_n \end{bmatrix} - \begin{bmatrix} X_r \\ Y_r \end{bmatrix} \right) \Lambda^{-1} \right\|_\infty \qquad (6)$$

The reduced order controller $C_r = Y_r X_r^{-1}$ will stabilize $T(C_r, P)$ is (6) is less than one. This weighted reduction scheme can be performed by balance and truncate (Enns (1984)), Hankel norm approximation (Latham and Anderson (1985)) or an H_∞-norm based reduction.

We prefer a Hankel norm approximation because it provides the smallest upper bound on the H_∞-norm of the model error which can be calculated easily. In the reduction scheme two cases can be distinghuised:

1. $\Lambda = \alpha I, \alpha \in \mathbb{R}$
   This is a special case of the presented framework. The reduction problem is equivalent to model reduction in the gap-metric (Corollary 4.3). The balanced reduction approach is described in Meyer (1988). If we view the coprime factors as a normal plant description we can apply a Hankel norm approximation (Glover (1984)). In Glover (1984) an H_∞-norm on the reduction error can be given in terms of the neglected Hankel singular values of the plant. This error bound is an upper bound on (6).
2. $\Lambda \neq \alpha I$
   In this general case there is a large benefit compared to the case $\Lambda = \alpha I$ since a lot of conservatism has been removed. The controller reduction can be done by a weighted Hankel norm approximation (Latham and Anderson (1985)) in which Λ^{-1} can be seen as the appropriate closed loop weighting. An H_∞-norm bound of (6) can be calculated (Anderson (1986)) in terms of the neglected Hankel singular values:

Theorem 5.1 *Let P and C be given by their normalized coprime factorizations as in Theorem 4.2 and denote the Hankel singular values* $\sigma_i(\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \Lambda^{*-1})$ *as* $\sigma_1 \geq \ldots \sigma_r \geq \sigma_n > 0$ *then there exists a reduced order controller C_r parametrized by $(Y_r, X_r) \in \mathcal{H}$ such that*

$$\left\| \left(\begin{bmatrix} X_n \\ Y_n \end{bmatrix} - \begin{bmatrix} X_r \\ Y_r \end{bmatrix} \right) \Lambda^{-1} \right\|_\infty < \sum_{i=r+1}^{n} \sigma_i$$

Proof: Denote $G := \begin{bmatrix} X_n \\ Y_n \end{bmatrix}$, $\hat{G} := \begin{bmatrix} X_r \\ Y_r \end{bmatrix}$ then it is straight forward in (Latham and Anderson (1985), Anderson (1986)) to prove this theorem. □

6 Example

In this section we will illustrate the presented framework of controller reduction by an example studied for example in (Enns (1984), Anderson and Liu (1989), Liu et al. (1990)). This four disk example system can be described by an 8^{th} order SISO linear, time-invariant, minimum phase and open loop unstable model. The two unstable poles are located in the origin, therefore a controller is necessary to stabilize the closed loop. In (Anderson and Liu (1989)) different LQG controllers where calculated and reduced by different techniques. In this example we will use one of their LQG controllers (the one with $q_2 = 200$) to be able to compare the results. In the paper by Anderson and Liu (1989) stability of the closed loop is checked for every reduced order controller but no bounds on the stability are determined before the reduced controller is actually applied. However the stability margin derived in (Liu et al. (1990)) eqn.(21) can be stated in the gap-metric if the factorization of the full order controller is chosen to be normalized.

In order to apply the presented framework of controller reduction we need to calculate the Hankel singular values of the weighted controller factorization in order to determine the least allowable order of the controller. The hankel singular values of the weighted controller coprime factors $\sigma_i(\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \Lambda^{*-1})$

are:

(1.1 0.92 0.88 0.75 0.35 0.31 0.16 0.14)

In view of Theorem 5.1 and Theorem 4.2 the sufficient condition for stability of $T(C_r, P)$ is determined by the sum of the neglected singular values. Therefore this sum is plotted in Fig. 3 for different controller orders. It can be seen in Fig. 3 that a

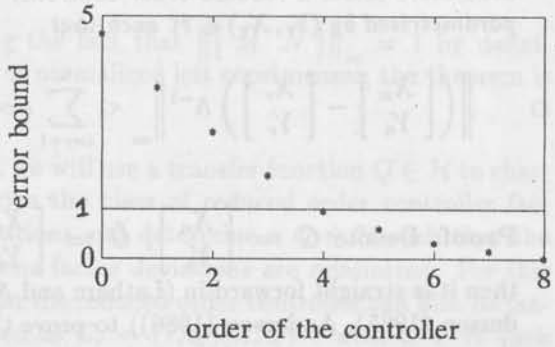


Fig. 3: H_∞-norm bound on the coprime factor controller deviations

4^{th} order controller C_r (which has an error bound smaller than one) will still stabilize the plant P.
In Fig. 4 the (1,2) element of the closed loop transfer function $T(C_i, P)$ being $(I+PC_i)^{-1}P$ is given for the full order controller and the 4^{th} order controller. It can be seen that only a slight performance degradation has occured due to the lower order of the controller. When we calculate stability margins for reduced order controllers using controller reduction in the gap-metric (without the closed loop weighting) it turned out that **no** stability of $T(C_r, P)$ can be guaranteed for **any** reduced order controller. The H_∞-norm on the error between 8^{th} order controller and the 7^{th} order controller (the minimal singular value of $\begin{bmatrix} X_n \\ Y_n \end{bmatrix}$) is $\sigma_{min} = 0.2$, while the allowable deviation is $\|\Lambda^{-1}\|_\infty^{-1} = 0.015$.
Hereby we have illustated the importance of this presented controller reduction framework.

7 Conclusions

In this paper we have shown that using an appropriate closed loop weighting, reduced order controllers can be calculated with a guaranteed stability margin. The closed loop weighting is a function of the

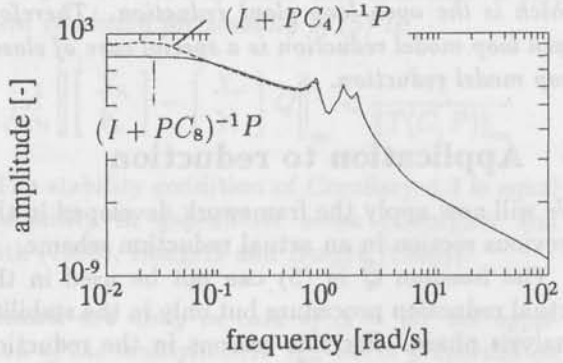


Fig. 4: (1,2) element of the closed loop transfer function$T(C_i, P)$

plant and the full order stabilizing controller. The actual reduction step is performed by a weighted Hankel norm approximation. Compared to a gap metric reduction it has been shown that for significant lower order controllers the closed loop will remain stable.

8 References

Anderson, B.D.O. (1986). Weighted Hankel norm approximation: calculation of bounds. *Syst. Control Lett., vol 7, p.247-255.*

Anderson, B.D.O. and Y. Liu. (1989). Controller reduction: concepts and approaches. *IEEE Trans. Automat. Contr. AC-34, p.802-812.*

Bongers, P.M.M. and O.H. Bosgra. (1990). Low order H_∞ controller synthesis. In *Proc. 29th Conf. Desicion and Control, p.194-199, Hawaii, USA.*

Bongers, P.M.M. (1991a). *Factorizational approach to robust control: application to a flexible wind turbine.* PhD thesis, Delft University of Technology, The Netherlands.

Bongers, P.M.M. (1991b). On a new robust stability margin. In *to appear at the international symp. on the Mathematical Theory of Networks and Systems, june 17-21, Kobe, Japan.*

Desoer, C.A., R.W. Liu, J. Murray, R. Saeks. (1980). Feedback Systems Design: The Fractional Representation Approach to Analysis and Synthesis. *IEEE Trans. Automat. Contr. AC-25, no.3, pp.399-412.*

Enns, D.F. (1984). *Model Reduction for Control System Design.* PhD thesis, Stanford University, Stanford CA, USA.

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

Georgiou, T.T. and M.M. Smith. (1990). Optimal robustness in the Gap metric. *IEEE Trans. Automat. Contr. AC-35, pp.673-686.*

Kabamba, P.T. (1985). Balanced gains and their significance for L^2 model reduction. *IEEE Trans. Automat. Contr. AC-30, p.690-693.*

Latham, G.A. and B.D.O. Anderson. (1985). Frequency-weighted optimal Hankel norm approximation of stable transfer functions. *Syst. Control Lett., vol 5, p.229-236.*

Liu, Y., B.D.O. Anderson, U-L. Ly. (1990). Coprime factorization controller reduction with bezout identity induced frequency weighting. *Automatica, vol.26, p.233-249.*

Meyer, D.G. (1988). A fractional approach to model reduction. In *Proc. Amer. Contr. Conf. pp. 1041-1047.*

McFarlane, D.C. and K. Glover. (1989), *Robust controller design using normalized coprime factor plant descriptions.* Lecture notes in control and information sciences, vol.138, Springer Verlag, Berlin, Germany.

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

Pernebo, L. and L.M. Silverman. (1982). Model reduction via balanced state-space representation. *IEEE Trans. Automat. Contr. AC-27, p.382-387.*

Vidyasagar, M., H. Schneider, B.A. Francis. (1982). algebraic and topological aspects of feedback stabilization. *IEEE Trans. Automat. Contr. AC, vol.27, no.4, pp 880-893.*

Vidyasagar, M. (1984). The graph metric for unstable plants and robustness estimates for feedback stability. *IEEE Trans. Automat. Contr. AC, vol.29, no.5, pp 403-418.*

Vidyasagar, M. (1985), *Control System Synthesis: the Factorization Approach.* MIT Press, Cambridge, MA, USA.

Selected Topics in Identification, Modelling and Control
Vol. 3, June 1991

Infinity norm calculation for large systems

Pepijn Wortelboer

Philips Research Laboratories, P.O.Box 80.000, 5600 JA Eindhoven, The Netherlands
Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. Infinity norm calculation of transfer function matrices is reviewed and modified for usage in large systems analysis. Using the system eigenvalues and an inverse orthogonal iteration technique the recently introduced two-step algorithm is modified to enable the application of infinity norm approaches in high-dimensional systems.

Keywords. large systems; infinity norm; inverse orthogonal iteration.

0 Notation

$\|G\|_\infty$	L_∞ norm of $G(s)$
H	Hamiltonian matrix
$\underline{h}, \bar{h}$	lower- and upper bound of $\|G\|_\infty$
γ	estimate of $\|G\|_\infty$
$\bar{\sigma}$	maximum singular value
$\hat{\omega}$	$\{\omega \in \mathbb{R} \mid \|G(j\omega)\|_\infty = \|G\|_\infty\}$
$\tilde{\omega}$	estimation of $\hat{\omega}$
A^H	Hermitian of A
IOI	inverse orthogonal iteration

1 Introduction

In linear system theory and robust control, the infinity norm of a transfer function matrix is now widely used: both in stability analysis and in performance characterization. The infinity norm (L_∞) of a transfer function matrix $G(s)$ is defined as

$$\|G\|_\infty = \sup_{\omega \in \mathbb{R}} \bar{\sigma}[G(j\omega)] \qquad (1)$$

provided $G(s)$ does not have imaginary poles. For stable systems the H_∞ norm

$$\sup_{Re(s) \geq 0} \bar{\sigma}[G(s)]$$

equals the L_∞ norm. In the sequel the L_∞ norm is subject of discussion and simply denoted by $\|G\|_\infty$.

In most cases the supremum in (1) is reached at one specific frequency, denoted by $\hat{\omega}$.

For G asymptotically stable and proper, $\|G\|_\infty$ can be interpreted as an upper bound on energy transfer from input to output:

$$\sup_u \frac{\|y\|_2}{\|u\|_2} = \|G\|_\infty \qquad (2)$$

with $\|u\|_2 = \sqrt{\int_{-\infty}^{\infty} u^T(t)u(t)dt}$ finite.

The infinity norm obeys standard operator norm properties and besides it satisfies

$$\|G_1 G_2\|_\infty \leq \|G_1\|_\infty \|G_2\|_\infty.$$

These properties are especially useful in robust control applications where a *set* of systems has to be controlled (stabilized in first instance). Representing the system variations in some way by an auxiliary stable perturbation system $\Delta(s)$ which is free within an infinity norm bound, the set of closed-loop systems can be concluded to be stable if $\|T(s)\Delta(s)\|_\infty < 1$, in which $T(s)$ is an auxiliary nominal closed-loop system. The larger $\|\Delta(s)\|_\infty$, the more serious the robust stabilization problem in the infinity-norm setting. In this respect it is important to realize that bounding system variations by means of infinity norm bounds on $\Delta(s)$ is only attractive when little is known about the origin of system variations. When the variations

only depend on a relatively small number of parameters (often called structured uncertainty) the infinity norm bound may be much too crude (conservative).

In the first stages of modelling for controller design purposes, little is known about the variations, and infinity norm bounds on the uncertainty provide a valuable starting point in robustness issues.

In this note the problem of efficient computation of the infinity norm is considered. The purpose is to make infinity norm approaches applicable to large systems which cannot be described accurately by a single low-dimensional model. This means we want to characterize *i)* differences between possibly high-dimensional models in a model set and *ii)* contributions of model reduction errors, both by means of infinity norm bounds.

The difficulty in computing the infinity norm is that no direct methods exist and that the iterative methods are time consuming. Several algorithms have been proposed in the literature to find increasingly tight upper and lower bounds. Basic idea is to first estimate the infinity norm (γ) and then determine whether γ is an upper bound or a lower bound. Robel (1989) used the following relationship: iff $\gamma^2 I - G^T(-s)G(s)$ has no imaginary transmission zeros and $\gamma > \bar{\sigma}[D]$ than $\gamma > \|G\|_\infty$. The transmission zeros are computed by means of the QZ algorithm. Independently Boyd *et al.*(1989) proposed a bisection algorithm in which the lower-bound property was checked by means of the existence of imaginary eigenvalues of a Hamiltonian $H(\gamma)$ in each iteration step. If the Hamiltonian has no imaginary eigenvalues γ is an upper bound. Both methods are strongly related.

An important acceleration was found by Bruinsma & Steinbuch (1990). They showed that it is not only of interest *whether* the Hamiltonian has imaginary eigenvalues, but also at which frequencies these occur. Maximum singular values are computed for a number of intermediate frequencies and these often provide a major increase of the lower bound. The so-called *two step* algorithm of Bruinsma & Steinbuch (1990) is an efficient combination of

- the maximum singular value evaluation over a dense frequency grid (no upper bound) and
- the bisection method that bounds the infinity norm without using any maximum singular values.

Almost at the same time Boyd & Balakrishnan (1990) discovered the improved rate of convergence of the combined method.

This paper proposes some modifications to the two-step algorithm that makes it better suited to high-dimensional systems. The new algorithm makes optimal use of the system poles to obtain a high starting value for the lower bound. This reduces the number of iterations needed. Besides the search for imaginary eigenvalues of the Hamiltonian is performed more efficiently; only eigenvalues close to some value on the positive imaginary axis are calculated.

In section 2 a short summary of the role of the Hamiltonian in L_∞-norm calculation is given, and in section 3 the essentials of the two-step algorithm are pointed out.

Section 4 discusses methods to compute eigenvalues selectively. A Hamiltonian eigenvalue computation scheme is presented that is based on inverse and orthogonal iteration.

In section 5 the new algorithm is discussed, that incorporates the selective eigenvalue computation and the initial search for a high lower bound.

2 The Hamiltonian in infinity norm calculation

Since the infinity norm cannot be computed directly, methods have been designed that compute accurate upper and lower bounds by iteration. Most algorithms make use of a Hamiltonian matrix that is a function of an infinity norm estimate in such a way that the Hamiltonian matrix has no imaginary eigenvalues if the estimate is strictly larger than $\|G\|_\infty$. If the Hamiltonian matrix does have imaginary eigenvalues the estimate is a lower bound. Several types of iteration are suggested to find sufficiently narrow bounds within reasonable time.

First the eigenstructure of the Hamiltonian is introduced. Denote $\gamma > 0$ as infinity norm estimate. Let [A,B,C,D] constitute a minimal stable state-space system. For γ not a singular value of D (note that $\|G\|_\infty \geq \bar{\sigma}[D]$) a Hamiltonian matrix can be defined,

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

where $R(\gamma) = D^H D - \gamma^2 I$ and $S(\gamma) = DD^H - \gamma^2 I$. A crucial property of this matrix is that only if $j\omega$ ($\omega \in \mathbb{R}$) is an eigenvalue of $H(\gamma)$, $G(j\omega)$ has a singular value equal to γ (Boyd *et al.*, 1989 and Bruinsma & Steinbuch, 1990). $G(j\omega)$ may have other singular values larger than γ.

The eigenvalues of H are symmetric with respect to both the real axis and the imaginary axis. The structure in H is revealed by transforming H in the following way:

$$T^{-1}HT = -H^H \text{ with } T = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

Since this transformation does not affect the eigenvalues we have

$$\lambda[H] = \lambda[-H^H] = -\lambda[\text{conj}(H)^T] = -\text{conj}(\lambda[H])$$

from which the double symmetry of the eigenvalues can be concluded. Standard eigensolvers do not take advantage of this given structure; these compute up to four times too many eigenvalues. A few attempts have been made to solve the Hamiltonian eigenproblem while preserving the structure mentioned above (Bunse-Gerstner & Mehrmann, 1986 and Byers, 1990). None of these methods is very effective, and the problem of calculating only eigenvalues on (part of) the positive imaginary axis is not addressed.

3 Infinity norm calculation (two-step algorithm)

The two-step algorithm (Bruinsma & Steinbuch, 1990) is an iterative method consisting of two steps per cycle. Maximum singular value evaluations at specific frequencies are used to obtain lower bounds for the infinity norm. The specific frequencies are obtained from Hamiltonian eigenvalue analysis. The relationship between the two steps is shown in the following scheme:

1. given a lower bound $\underline{h}$, compute the eigenvalues of $H(\gamma)$ with γ an infinity norm estimate

$$\gamma = (1 + 2\epsilon)\underline{h}, \quad 0 < \epsilon \ll 1 \qquad (4)$$

   Store all positive imaginary eigenvalues $j\omega_i$. If no imaginary eigenvalues are found, γ is an *upper* bound: $\bar{h} = (1 + 2\epsilon)\underline{h}$, and the relative difference between upper and lower bound is 2ϵ.

2. evaluate all $\bar{\sigma}[G(j\tilde{\omega})]$ with $\{\tilde{\omega}\}$ a series of intermediate frequencies, $(\omega_i + \omega_{i+1})/2$. The maximum value is taken as new lower bound $\underline{h}$.

Starting value for the lower bound $\underline{h}$ is the maximum of $\bar{\sigma}[D]$ and $\bar{\sigma}[D - CA^{-1}B]$ ($G(j\omega)$ for infinite and zero frequency). If Hamiltonian $H(\gamma)$ has imaginary eigenvalues $j\omega_{min}, \ldots, j\omega_{max}$ we know that $\hat{\omega} \in < \omega_{min}, \omega_{max} >$. The process of increasing γ is accompanied by a decreasing $(\omega_{min}, \omega_{max})$ interval until H has no more imaginary eigenvalues, meaning that the infinity norm estimate γ is an upper bound. Thus this algorithm returns an upper and lower bound on the infinity norm (respectively $\bar{h}$ and $\underline{h}$) and an accurate estimate of $\hat{\omega}$.

The first step in each iteration cycle involves an eigenvalue computation that is quite demanding for large systems. An alternative approach should take advantage of the symmetry structure of the Hamiltonian eigenvalues, and the frequency interval information, that in each cycle becomes more accurate. Selective eigenvalue computation is discussed in the following section.

4 Solution of selected eigenvalues by iteration

The computation of a relatively small number of eigenvalues and eigenvectors can be done efficiently using iteration techniques. The eigenvalue of largest modulus can be computed (together with its eigenvector) by means of the well-known power method. Inverse iteration can be used to calculate the eigenvalue of smallest modulus and the corresponding eigenvector.

Inverse iteration is often used to calculate eigenvectors of a matrix A once some eigenvalues are approximately known. Assume $\tilde{\lambda}$ is an approximation of some eigenvalue of A (not necessarily the smallest or largest in modulus). Then inverse iteration on $(\tilde{\lambda}I - A)$ returns, after convergence, the eigenvalue λ closest to $\tilde{\lambda}$ and the eigenvector v for this λ:

$$(\tilde{\lambda}I - A)v = (\tilde{\lambda} - \lambda)v.$$

For our purpose a generalization of inverse iteration is needed. All imaginary eigenvalues of H inside a certain interval have to be computed. The problem statement is thus: find a sufficient number of eigenvalues of $H - j\tilde{\omega}$ with $\tilde{\omega} > 0$ in the middle of the frequency interval of interest.

Generalizations of the power method are numerous. Golub & van Loan (1989) describe the most important ones. For a general square matrix orthogonal iteration (Golub & van Loan, 1989, p.355) can be used to find the eigenvalues of largest magnitude and the orthonormal vector basis spanning the corresponding invariant subspace. Convergence is linear with rate $|\lambda_{n_1+1}|/|\lambda_{n_1}|$ ($|\lambda|$ in decreasing order and n_1 the number of eigenvalues wanted). Stewart (1976) describes a simultaneous iteration

method in which the convergence is faster for the largest eigenvalues. The rate is

$$|\lambda_{n_1+1}|/|\lambda_i| \text{ with } i = 1, \ldots, n_1.$$

A reduced-order Schur decomposition is performed every now and then to obtain this acceleration. Simultaneous iteration is especially suited to large sparse eigenproblems, since it requires only matrix-vector multiplications, that can be programmed very efficiently for sparse matrices. In these cases standard QR-based methods are less efficient.

To find eigenvalues of H close to $j\tilde{\omega}$ a generalization of *inverse* iteration is needed. Golub & van Loan (1989) give an inverse orthogonal iteration scheme (pp. 359–360) based on orthogonal iteration. Based on Stewart's acceleration of the orthogonal iteration method, an accelerated inverse orthogonal iteration scheme can be derived straightforward.

Suppose the dimension of a general square matrix Z is n. The following inverse orthogonal iteration (IOI) algorithm computes n_1 ($n_1 < n$) eigenvalues with $|\lambda_{1,\ldots,n_1}| < |\lambda_{n_1+1,\ldots,n}|$, and the associated invariant subspace Q_1 (an $(n \times n_1)$ matrix satisfying $Q_1^H Q_1 = I$).

Algorithm $[Q_1,\ R_1] = IOI(Z, n_1)$

choose n_1 orthonormal vectors: $Q_1^{(0)}$ $(n \times n_1)$

for $k = 1, 2, \ldots$

$$ZX^{(k)} = Q_1^{(k-1)} \quad (5)$$

$$\tilde{Q}_1^{(k)}\ R_0^{(k)} = X^{(k)} \quad (6)$$

$$\check{Z}^{(k)} = [\tilde{Q}_1^{(k)}]^H Z \tilde{Q}_1^{(k)} \quad (7)$$

$$\check{Z}^{(k)} V^{(k)} = V^{(k)}\ R_1^{(k)} \quad (8)$$

$$Q_1^{(k)} = \tilde{Q}_1^{(k)}\ V^{(k)} \quad (9)$$

end

The algorithm starts with the solution step (5), showing the inverse (power) character of the algorithm (depending on the sparsity of Z and the expected number of iterations needed, this solution can be obtained directly or after an LU-decomposition). In a QR-factorization step (6) an orthonormal vector basis ($\tilde{Q}_1$) is calculated for the subspace spanned by the vectors in X. Steps (5) and (6) state the original inverse orthogonal iteration, the following steps are added for acceleration. By means of vector basis $\tilde{Q}_1$, the original Z is reduced (7). The reduced-order Schur decomposition (8) yields an upper triangular matrix R_1 with eigenvalue estimates of Z on the diagonal, and an orthogonal transformation matrix V by which an update of the orthonormal vectors in Q_1 can be achieved (9).

The diagonal of R_1 will eventually contain a number of n_1 eigenvalues of Z with smallest magnitude. Besides, Q_1 converges to the orthonormal basis of the associated invariant subspace.

The iteration can be stopped when the maximum relative change of the ordered diagonal elements of R_1 drops below a given bound.

The convergence is best if a set of n_1 eigenvalues exists with significantly smaller modulus than all other $n - n_1$ eigenvalues. Theoretical convergence is linear. In practice this means that convergence is rapid provided $Q_1^{(0)}$ is chosen properly.

IOI can be used to calculate eigenvalues of H in the vicinity of some $j\tilde{\omega}$, but there is no reason to expect that these eigenvalues will be on the imaginary axis. All eigenvalues are calculated that lie within a circle located symmetrically on the positive imaginary axis. Note that the Hamiltonian eigenvalues with negative imaginary part are simply not computed.

5 Algorithm to compute the infinity norm for large systems

As discussed earlier, the two-step approach yields upper and lower bounds on the infinity norm at the expense of a relatively small number of Hamiltonian eigenvalue evaluations (compared to the bisection method that fully relies on Hamiltonian eigenvalue computations).

For large systems it is particularly important to limit the number of Hamiltonian eigenvalue evaluations; IOI may provide some acceleration but even more can be gained by searching a high initial lower bound on the infinity norm in order to reduce the number of iterations. This can be done by evaluating the maximum singular value of $G(j\omega)$ at a large number of frequencies, including the frequencies that are related to the system poles. The maximum of these values is often a good approximation of the infinity norm (lower bound). Besides it gives an idea how each mode in the system contributes to the transfer function matrix. This analysis has only to be done once.

In some system realizations, such as the modal (Jordan) form and Schur form realizations, the system poles are available and their imaginary parts can be used as frequencies in the maximum singular value computations. Besides it can be checked di-

rectly if the infinity norm exists (remember that the L_∞ norm definition excludes systems with imaginary poles).

Another advantage of such realizations is that the maximum singular values can be computed faster, since the inversion of $[sI - A]$ is simplified considerably.

Transformation to some modal or Schur realization may be very effective in infinity norm computation. Fortunately the modal form of a (large) system representation is also used to analyse the (open-loop) physical properties. Modes that are questionable or input-output unimportant can be truncated in this stage. Only if one has confidence in the computed modes the model will be used for controller design.

Nice examples can be found in structural dynamics and particularly in large-space-structure applications. Hundreds of vibration modes may be of importance for controller design and together with a structural dynamicist the control engineer has to decide which modes are reliable and on which modes a controller design can be based. Perturbation analysis and experimental verification are indispensable. As stated earlier it is important to compare different models and the modal form provides a nice basis.

A general (non-modal) model of very high dimension can also be transformed to Schur form; this can be done by unitary transformations thus avoiding most numerical problems in computing the system modes.

It can be concluded that the system poles are directly available in most models of large systems. A first modification of the two-step approach involves the usage of pole-related frequencies in searching a sharp lower bound on the infinity norm by scanning the maximum singular values of $G(s)$ for these frequencies.

The second modification of the two-step algorithm is the incorporation of IOI. In running through the iteration cycles the search for $\hat{\omega}$ can be constrained more and more. As opposed to complete evaluation of the Hamiltonian eigenvalues, IOI takes advantage of the ω interval information built up in previous cycles. The function call

$$[Q_1, R_1] = \mathrm{IOI}(j\tilde{\omega}I - H, n_1) \qquad (10)$$

can be used to determine the eigenvalues of H that deviate minimally from $j\tilde{\omega}$ ($\tilde{\omega}$ chosen in the center of a given interval and n_1 related to the number of eigenvalues expected in the circle with center $j\tilde{\omega}$ and diameter equal to the interval size).

Yet there is no guarantee that IOI will find two imaginary eigenvalues enclosing $j\hat{\omega}$. First there is a risk that a frequency interval is found that bounds a local supremum of $\bar{\sigma}$. Another problem occurs if the IOI algorithm cannot find two imaginary eigenvalues at all. In both cases the remedy is calculating all Hamiltonian eigenvalues. In this way all imaginary eigenvalues are determined, and if there are none, the local supremum is the global supremum.

Since modal form or Schur form realizations do not yield sparse Hamiltonian matrices, IOI is only effective if indeed a very small part of the eigenvalues are calculated.

Algorithm

(modified two-step infinity norm calculation)

1. Search system poles. If non-repeating, compute a modal realization otherwise some Schur form realization.
2. Make $\{\tilde{\omega}\}$, a series of trial values for $\hat{\omega}$ out of the damped frequencies, the zero frequency and all intermediate frequencies.
3. Choose ϵ, the relative accuracy
4. Choose $\tilde{\omega}^{(0)}$ that gives the largest $\bar{\sigma}[G(j\tilde{\omega})]$. $\underline{h}^{(0)} = \bar{\sigma}[G(j\tilde{\omega}^{(0)})]$
5. **if** $(\bar{\sigma}[D] > \underline{h}^{(0)})$, $\gamma^{(0)} = (1+2\epsilon)\bar{\sigma}[D]$, **goto** 8 **end**
6. $\gamma^{(0)} = (1+2\epsilon)\underline{h}^{(0)}$
7. **for** $k = 1, 2, \ldots$

$$Z^{(k)} = j\tilde{\omega}^{(k-1)}I - H(\gamma^{(k-1)})$$

$$\left[Q_1^{(k)},\ R_1^{(k)}\right] = \mathrm{IOI}(Z^{(k)}, 2)$$

$$\begin{bmatrix} \lambda_1^{(k)} & \times \\ 0 & \lambda_2^{(k)} \end{bmatrix} \equiv R_1^{(k)}$$

if $(\ \mathrm{real}(\lambda_1^{(k)}) = 0 \wedge \mathrm{real}(\lambda_2^{(k)}) = 0\ )$

$$\tilde{\omega}^{(k)} = \tilde{\omega}^{(k-1)} - \mathrm{imag}(\lambda_1^{(k)} + \lambda_2^{(k)})/2$$

$$\underline{h}^{(k)} = \bar{\sigma}[G(j\tilde{\omega}^{(k)})]$$

$$\gamma^{(k)} = (1+2\epsilon)\underline{h}^{(k)}$$

else **goto** 8 **end**

end

8. Calculate all $\lambda\,[H(\gamma^{(k)})]$

9. **if** ($\text{real}(\lambda_1), \ldots, \text{real}(\lambda_{2n}) \neq 0$),
$\underline{h} = \gamma$
else
construct a series of $\tilde{\omega}$ from the imaginary eigenvalues and intermediate values, **goto** 3
end

Some additional remarks on the algorithm are presented next.

- To obtain a sharp lower bound $\underline{h}$ the maximum singular value of $G(j\omega)$ is computed for all damped frequencies and all intermediate frequencies. In this way it is almost certain that the two step algorithm starts at the global $\bar{\sigma}$ peak and that subsequent Hamiltonian eigenvalue computation only gives one pair of eigenvalues on the positive imaginary axis (meaning that $\hat{\omega}$ has been located in an interval).

- For simplicity the algorithm given above uses two orthonormal vectors in IOI. This is a minimum, but usually sufficient as a result of the dense search over the frequency in the second step. If IOI does not give two imaginary eigenvalues the algorithm switches to full eigenvalue computation. On the other hand, if the 'for' loop in step 7 continues to yield two imaginary eigenvalues a local supremum will be obtained eventually. In the verification step (8) the character of the supremum is discovered.

- The converged orthonormal vector basis of a previous iteration cycle can be used as starting vectors $Q_1^{(0)}$ in the next IOI calculation.

- The modified two step algorithm has been implemented in MATLAB™. The speed of infinity norm computation for lightly damped high-dimensional systems has increased considerably. A wider range of models may benefit from the modified two step algorithm once the MATLAB restrictions on implementing IOI are removed or circumvented; currently (5) is solved directly in each iteration ($ZX = Q$). One initial LU decomposition ($Z = LU$) and subsequent solution of $LY = Q$ and $UX = Y$ is more efficient provided sophisticated solution schemes are used (not available in MATLAB).

As the case $\|G\|_\infty = \bar{\sigma}[D]$ presents serious problems in computing the Hamiltonian eigenvalues due to near singularity of $D^H D - \gamma^2 I$, reliable upper bounds can only be obtained by substituting a γ which is clearly larger than $\bar{\sigma}[D]$, meaning the relative error in upper and lower bounds is relatively large (large ϵ). In the transmission zero approach of Robel (1989) a QZ algorithm is used that does not require the inversion of $D^H D - \gamma^2 I$.

In the appendix eigenvalue derivatives of the Hamiltonian are derived and used to classify different types of ω intervals. Based on this information more sophisticated $\tilde{\omega}$'s can be constructed.

In large systems analysis, eigenvalue routines that compute only a specific invariant subspace can reduce the amount of computational work considerably. This note has shown how inverse orthogonal iteration can be used in infinity norm calculation of large systems. In Wortelboer (1990) balanced reduction algorithms for large systems have been proposed based on similar subspace iteration routines.

6 Conclusions

A recently introduced two-step L_∞-norm calculation scheme has been analysed for its applicability in large system analysis. Two modifications have been proposed in this note. These include full usage of the system poles, and the replacement of full eigenvalue computation of the Hamiltonian by selective imaginary eigenvalue computation using inverse orthogonal iteration. Especially for mechanical systems which are often in modal form and have many lightly damped vibration modes this modified L_∞-norm calculation scheme is very economical.

References

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 matrix and a quadratically convergent algorithm for computing its L_∞-norm. *Syst. Control Lett., 15,* 1–7.

Bruinsma, N.A. and M. Steinbuch (1990). A fast algorithm to compute the H_∞–norm of a transfer function matrix. *Syst. Control Lett., 14,* 287–293.

Bunse-Gerstner, A. and V. Mehrmann (1986). A symplectic QR like algorithm for the solution of the real algebraic Riccati equation. *IEEE Trans. Autom. Contr., 31,* pp. 1104–1113.

Byers, R. (1990). A Hamiltonian-Jacobi algorithm. *IEEE Trans. Autom. Contr., 35,* pp. 566–570.

Francis, B.A. (1987). *A Course in H_∞ Control Theory*, Lecture notes in Control and Information Sciences, Springer-Verlag.

Golub, G.H., and C.F. van Loan (1989). *Matrix Computations*, second edition, The John Hopkins University Press.

Robel, G. (1989). On computing the infinity norm. *IEEE Trans. Automat. Contr. 34,* pp. 882–884.

Stewart, G.W. (1976). Simultaneous iteration for computing invariant subspaces of non-Hermitian matrices. *Numer. Math. 25,* 123–136.

Wortelboer, P.M.R. (1990). Subspace balancing. *Selected-Topics in Identification, Modelling and Control, vol.2,* 25–32.

Appendix

The use of Hamiltonian eigenvalue *derivatives* in constructing new $\tilde{\omega}$'s is discussed. This goes beyond the original approach in the two step algorithm that uses only the imaginary eigenvalues of H. By means of a simple example it is shown that using the derivative information, frequency intervals with local suprema can be determined.

In Fig. 1 an example is given. The frequencies $\omega_1, \ldots, \omega_6$ have been derived from the imaginary eigenvalues of H for a γ that is larger than both $\bar{\sigma}[D]$ and $\bar{\sigma}[D - CA^{-1}B]$. Having no derivative information, we can only say that $\hat{\omega} \in < \omega_1, \omega_6 >$. With the derivative information we know exactly how many singular values are larger than γ in a specific frequency interval. This is indicated in Fig. 1 by means of '000', '111' and '222'. Thus it can be concluded that $\hat{\omega} \in < \omega_1, \omega_4 > \cup < \omega_5, \omega_6 >$. Figures 2, and 3 show singular value functions that satisfy both frequency interval and derivative constraints. In Fig. 4 a singular value function with other local suprema is shown that satisfies the frequency interval constraints but violates the derivative constraints.

Only if a 1-type interval is sandwiched between 0-type intervals the existence of a local supremum can be concluded. In that case it makes sense to choose a new $\tilde{\omega}$ in the middle of the interval. However, a sequence of non-zero type intervals should

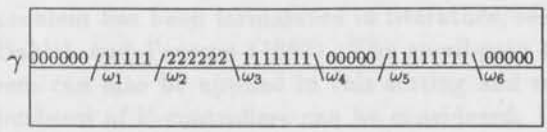


Fig. 1: Frequency intervals and derivatives

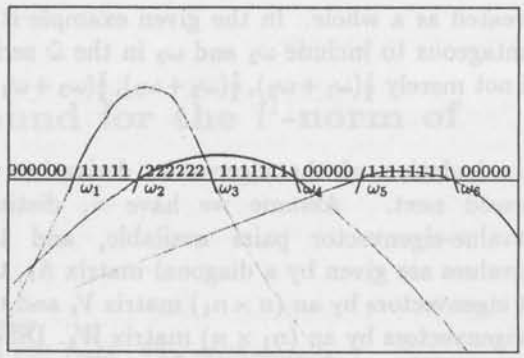


Fig. 2: First singular value function example satisfying derivatives

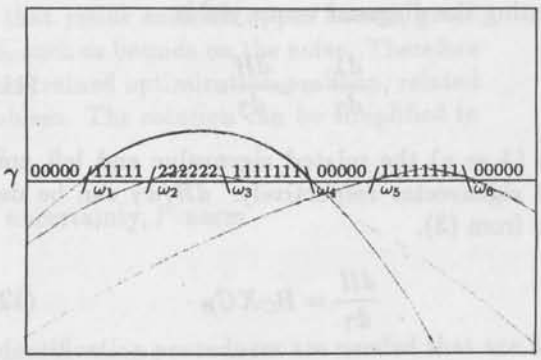


Fig. 3: Second singular value function example satisfying derivatives

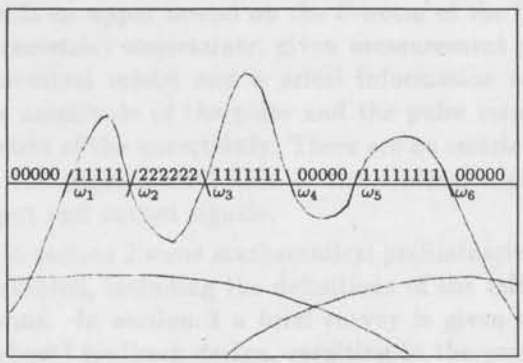


Fig. 4: Singular value function example violating derivatives

be treated as a whole. In the given example it is advantageous to include ω_2 and ω_3 in the $\tilde{\omega}$ series (and not merely $\frac{1}{2}(\omega_1+\omega_2), \frac{1}{2}(\omega_2+\omega_3), \frac{1}{2}(\omega_3+\omega_4)$).

The calculation of the eigenvalue derivatives is discussed next. Assume we have n_1 distinct eigenvalue-eigenvector pairs available, and the eigenvalues are given by a diagonal matrix Λ_1, the right eigenvectors by an $(n \times n_1)$ matrix V_1 and the left eigenvectors by an $(n_1 \times n)$ matrix W_1. Differentiating $HV_1 = V_1\Lambda_1$ with respect to γ and premultiplication with W_1 yields:

$$\begin{aligned}\frac{d\Lambda_1}{d\gamma} &= W_1\frac{H}{d\gamma}V_1 + W_1 H\frac{dV_1}{d\gamma} - W_1\frac{dV_1}{d\gamma}\Lambda_1 \\ &= W_1\frac{H}{d\gamma}V_1 + \Lambda_1 W_1\frac{dV_1}{d\gamma} - W_1\frac{dV_1}{d\gamma}\Lambda_1\end{aligned}$$

Equating the diagonal terms yields

$$\frac{d\lambda}{d\gamma} = w\frac{dH}{d\gamma}v \qquad (11)$$

with (λ, w, v) the related eigenvalue and left and right eigenvector respectively. $dH/d\gamma$ can be derived from (3).

$$\frac{dH}{d\gamma} = B_C X C_B \qquad (12)$$

with

$$B_C = \begin{bmatrix} B & 0 \\ 0 & C^H \end{bmatrix}$$

$$C_B = \begin{bmatrix} C & 0 \\ 0 & B^H \end{bmatrix}$$

and

$$\begin{aligned}X_{11} &= -2\gamma[D^H D - \gamma^2 I]^{-2} D^H \\ X_{12} &= -[D^H D - \gamma^2 I]^{-1} - 2\gamma^2[D^H D - \gamma^2 I]^{-2} \\ X_{21} &= [DD^H - \gamma^2 I]^{-1} + 2\gamma^2[DD^H - \gamma^2 I]^{-2} \\ X_{22} &= 2\gamma D[D^H D - \gamma^2 I]^{-2}\end{aligned}$$

Since the X_{ij} matrices may be of relatively small dimension, the best order in computing (11) is $(wB_C)X(C_B v)$.

The calculation of the eigenvectors for specific imaginary eigenvalues of H is relatively simple. If IOI has been used, a set of orthonormal vectors Q_1 and an upper triangular matrix R_1 for the shifted Hamiltonian are available:

$$[j\tilde{\omega}I - H]Q_1 = Q_1 R_1 \qquad (13)$$

From $Q_1^H H Q_1 = j\tilde{\omega}I - R_1$ it can be concluded that Q_1 also triangularizes H. By calculating the eigenvectors of $j\tilde{\omega}I - R_1$,

$$[j\tilde{\omega}I - R_1]v = v\,\mathrm{diag}(\lambda_1, \ldots, \lambda_{n_1}) = v\Lambda_1 \qquad (14)$$

the n_1 left (W_1) and right (V_1) eigenvectors of H can be obtained straightforward,

$$V_1 = Q_1 v, \qquad W_1 = v^{-1}Q_1^H, \qquad (15)$$

with $W_1 V_1 = I$.

Selected Topics in Identification, Modelling and Control
Vol. 3, June 1991

Identification of an upper bound for the l^1-norm of the model uncertainty

Richard G. Hakvoort

Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. In literature the l^1-control design problem has been formulated. Robustness has been introduced into this setting using the small-gain theorem. For the purpose of robust controller design it is necessary that an upper bound on the l^1-norm of the uncertainty is given as a measure of the difference between the system and the nominal model. In this paper a procedure is developed that yields such an upper bound, given measurement data and some a priori information, such as bounds on the noise. Therefore the problem is reformulated to a (large scale) constrained optimization problem, related to the Generalized Linear Complementarity Problem. The solution can be simplified to linear programming at the price of conservatism.

Keywords. robust identification, unstructured uncertainty, l^1-norm

1 Introduction

In classical (e.g. prediction error) identification schemes a dynamical model of a plant is estimated, using measurement data and usually stochastic assumptions on the noise corrupting the data. However in general a model does not give an exact description of the plant, there always is a model error. In these classical schemes no estimate is given of this model error. Modern control theory however is able to cope with modelling errors, represented by perturbation (or uncertainty) blocks. In H_∞-theory an H_∞-bound on the perturbations is required. So new identification procedures are developed that yield an upper bound on the H_∞-norm of the uncertainty, see e.g. Helmicki *et al.* (1989) and De Vries (1991).

Dual to the H_∞-theory the l^1-controller design problem has been formulated in literature, see e.g. Dahleh and Pearson (1987). The small-gain theorem can also be applied in this setting and so robustness of l^1-controllers can be considered. In l^1-theory an l^1-bound on the perturbation is required, see e.g. Dahleh and Ohta (1988). This implies that identification procedures are needed that are fit for this setting, i.e. yield l^1-optimal models and an estimation of the model error. Some work has already been done in this field by Jacobson and Nett (1991), where single input single output systems are considered and step or pulse responses are assumed to be available.

In the present paper a procedure is presented that yields an upper bound on the l^1-norm of the (non-parametric) uncertainty, given measurement data, a nominal model and a priori information about the amplitude of the noise and the pulse response matrix of the uncertainty. There are no restrictions on the experimental data or the dimension of the input and output signals.

In section 2 some mathematical preliminaries are presented, including the definitions of the relevant norms. In section 3 a brief survey is given of l^1-(robust) feedback design, resulting in the problem statement for this paper. In the next section the required a priori information is established. In section 5 a procedure is developed for the calculation of an upper bound for the l^1-norm of the uncertainty. Three methods to solve the resulting optimization

problem are proposed in the following section. The result is discussed in section 7 and in the final section conclusions are drawn.

2 Mathematical Preliminaries

Operators are denoted by capital letters (e.g. G, Δ), the corresponding pulse response sequences (matrix valued for MIMO systems) by lower case characters (e.g. g, δ) and signal vectors also by lower case letters (e.g. u, y, z). Thus

$$Gu = g * u, \quad G(q^{-1}) = \sum_{k=0}^{\infty} g(k) q^{-k}, \qquad (1)$$

where $*$ denotes the convolution operator and q^{-1} the delay operator, $q^{-1}\, u(t) = u(t-1)$.

Let x be a real-valued $m \times 1$ vector and A a real-valued $m \times n$ matrix. Then x_i is the ith element of x and A_i the ith row of A. A_{ij} is the (i,j) entry of A. Let in addition $|a|$ be the absolute value of the scalar a. Then the following norms are defined (see e.g. Desoer and Vidyasagar, 1975)

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \; 1 \leq p < \infty,$$

the p-norm of the vector x, and

$$\|x\|_\infty = \max_i |x_i|,$$

the ∞-norm of the vector x.

Let h be an infinite sequence of (real) numbers $h(k)$, $k = 0, 1, 2, \ldots$, then the following norms are defined.
If $h \in l^p$

$$\|h\|_p = \left(\sum_{k=0}^{\infty} |h(k)|^p \right)^{\frac{1}{p}}, \; 1 \leq p < \infty,$$

the l^p-norm of the sequence h.
If $h \in l^\infty$

$$\|h\|_\infty = \sup_{k \geq 0} |h(k)|,$$

the l^∞-norm of the sequence h.

Let u be an infinite sequence of $m \times 1$ vectors $u(k)$, and g an infinite sequence of $m \times n$ matrices $g(k)$, $k = 0, 1, 2, \ldots$, then the following norms can be defined.
If $u \in l_m^\infty$

$$\|u\|_\infty = \max_i \sup_{k \geq 0} |u_i(k)| = \max_i \|u_i\|_\infty =$$

$$= \sup_{k \geq 0} \|u(k)\|_\infty ,$$

the l^∞-norm.
And if $g \in l_{mn}^1$

$$\|g\|_1 = \max_i \sum_{j=1}^{n} \sum_{k=0}^{\infty} |g_{ij}(k)| = \max_i \sum_{j=1}^{n} \|g_{ij}\|_1 =$$

$$= \max_i \|g_i\|_1 ,$$

the l^1-norm.

With $\mathcal{A}$ we denote the algebra of BIBO stable, linear, time-invariant, causal operators on l_n^∞. The l^1-norm (or $\mathcal{A}$-norm for operators) is an induced norm, for we have that, if $G \in \mathcal{A}$ (or equivalently $g \in l_{mn}^1$)

$$\|G\|_{\mathcal{A}} = \|g\|_1 = \sup_{\|u\|_\infty = 1} \|Gu\|_\infty . \qquad (2)$$

This means that the maximum peak amplification of the system G is given by the l^1-norm of the system's pulse response matrix.

3 l^1-Optimal Feedback Controllers

The maximum energy amplification of a stable, linear system is equal to the H_∞-norm of the system. So for systems with bounded energy signals the H_∞-norm is the most suitable norm to use. The H_∞-theory is concerned with minimizing the H_∞-norm of transfer functions (e.g. the transfer function of noise signal to output signal) through the choice of a stabilizing controller. Moreover robustness has been introduced into this setting with aid of the small-gain theorem (see e.g. Maciejowski, 1989).

For systems with bounded magnitude signals the more suitable norm is the $\mathcal{A}$-norm or l^1-norm, because of property (2). The l^1-control theory is concerned with minimizing the l^1-norm of a system through the choice of a stabilizing controller. The l^1-problem has been formulated by Vidyasagar (1986) and is totally dual to the H_∞-problem. The problem of designing an l^1-optimal controller has been completely solved for the MIMO discrete time case in Dahleh and Pearson (1987), Mendlovitz (1989) and McDonald and Pearson (1991). In these papers the problem has been reformulated to a linear programming problem. Notice that due to the specific norms used, the H_∞-problem is mainly treated in the frequency domain and the l^1-problem in the time domain.

Robustness has also been introduced into the l^1-problem formulation by Dahleh and Ohta (1988), using a variant of the small-gain theorem. Consider the standard uncertainty configuration of figure 1. In case of an unstructured uncertainty Δ the next

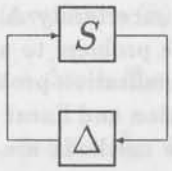


Fig. 1: Standard uncertainty configuration

theorem gives necessary and sufficient conditions for robust stability of this closed loop.

Theorem 1 (Dahleh and Ohta, 1988) *Let $S \in \mathcal{A}$ and suppose Δ is an l^∞-stable, strictly causal (possibly non-linear or time-varying) operator with $\sup_{\|u\|_\infty=1} \|\Delta u\|_\infty \leq 1$. Then the operator $I - S\Delta$ has an l^∞-stable inverse with bounded gain for all Δ if and only if $\|S\|_{\mathcal{A}} = \|s\|_1 < 1$.*

Khammash and Pearson (1991) extended the solution to the general case of a block diagonal structured uncertainty, where each block represents an unstructured uncertainty of the kind of theorem 1. In this setting the robust performance problem can be addressed as well: If a nominal model, a block structured uncertainty and bounds on the amplitude of the input and disturbance signals are given, then a bound on the amplitude of the output signal can be given. This is all analogous to the μ-analysis problem in the H_∞-setting (Maciejowski, 1989). However the problem of l^1-robust controller synthesis is still for the greater part unsolved.

Apparently it is useful to have an identification technique that, starting from time domain data yields a nominal model and a bound on the uncertainty, fit for l^1-robust feedback design. As we are interested in guaranteed robust stability and performance, it is necessary that the uncertainty is represented by an upper bound on its maximum signal peak amplification. On the other hand we do not want an unnecessary conservative bound, because this reduces the achievable performance of a controller. The complete problem of identifying an l^1-optimal model and a non-conservative upper bound on the uncertainty is so far unsolved. In this paper only the problem is addressed of the computation of a (non-conservative) upper bound on the l^1-norm of the uncertainty, given measurement data, a nominal model and a priori knowledge of the system, such as time-domain bounds on the noise corrupting the data. We consider weighted additive and weighted output multiplicative uncertainty. Procedures are derived for the computation of the worst-case uncertainty and the corresponding l^1-norm. The obtained upper bound can then be used for the analysis and design of l^1-robust controllers.

4 A Priori Knowledge

For identification purposes we need of course measurements of the input signal $u(t)$ and the output signal $y(t)$ acting on the system, $t = 0, 1, 2, \ldots, N$. The data may or may not be generated in closed loop.

We only consider discrete time, asymptotically stable, linear, time-invariant systems G with additive bounded output noise. These restrictions are not only meant to make the problem manageable, but these or similar restrictions are also necessary to obtain a sensible problem formulation. If the system is not stable, it does not have a finite l^1-norm. If the system is non-linear, the l^1-norm could also be unbounded, for example if the system contains an on-off switch. If the system is time-varying no upper bound on the l^1-norm can be given, for it is always possible that the dynamics radically change behind the measurement horizon. If finally the noise is unbounded, the data could be explained completely by either the noise or unstructured uncertainty and no conclusions can be drawn from the available data.

The input- output behaviour of the plant is assumed to be given by the equation

$$y = Gu + H_1 e, \quad H_2 e \in [-e_l, e_h], \tag{3}$$

where $u(t)$ is the measured input signal vector, $y(t)$ is the measured output signal vector, $H_1(q^{-1})$ and $H_2(q^{-1})$ are a priori known FIR (Finite Impulse Response) sequences

$$H_1(q^{-1}) = \sum_{k=0}^{h_1} h_1(k) q^{-k}, \; H_2(q^{-1}) = \sum_{k=0}^{h_2} h_2(k) q^{-k}$$

and $e_l(t)$ and $e_h(t)$ are a priori known lower and upper bounds on the noise, possibly functions of time t. The unknown quantities in (3) are the system $G \in \mathcal{A}$, past data $u(t)$ and $y(t)$ (before the measurements started, i.e. for $t < 0$) and the noise $e(t)$ (only bounds on the noise are known a priori). The system G is represented by the (unknown) pulse response matrix g, see equation (1).

Notice that the noise representation is quite general. If we have $H_1 = H_A H_B^{-1}$ and $H_2 = H_C H_D^{-1}$, we can introduce a new H_1 and H_2 as $\tilde{H}_1 = H_A H_D$ and $\tilde{H}_2 = H_C H_B$, which yields the same noise representation under the condition that $H_B H_D = H_D H_B$ (which is always true in the SISO case). For

l^1-feedback design a noise representation is chosen with $H_2 \equiv I$ (Dahleh and Pearson, 1987), in which case the transformation is possible as well.

In this paper we consider the estimation of an upper bound on the l^1-norm of the uncertainty. So we already have a discrete time, asymptotically stable, linear, time-invariant nominal model $\hat{G}$ at our disposal. This nominal model is the result of any existent modelling or identification procedure.

We introduce the $p \times q$ uncertainty matrix Δ and consider the uncertainty structure:

$$G = \hat{G} + \Delta M, \tag{4}$$

where the operators $\hat{G}$, Δ and M are all elements of $\mathcal{A}$ and have pulse response sequences $\hat{g}$, δ and m respectively. The nominal model $\hat{G}$ and weighting function M are known beforehand. Of course the system G and the uncertainty Δ are unknown. In this representation (4) weighted additive uncertainty ($M = W$) and weighted output multiplicative uncertainty ($M = W\hat{G}$) are included, where W is a weighting matrix. In this paper we will treat the case of an unstructured uncertainty Δ. It is however straightforward to extend the result presented to the case that a priori knowledge about the structure of the uncertainty Δ is available.

From a data set of finite length $(N+1)$ we can deduce no knowledge about the system's pulse response sequence $g(k)$ for $k > N$. So we need a priori information about the behaviour of the uncertainty in order to be able to derive an upper bound on the l^1-norm of the uncertainty. Therefore we assume a priori knowledge about the pulse respone sequence of the uncertainty Δ. This means that we have a matrix R and a matrix ρ, $\rho_{ij} > 1$, such that

$$|\delta_{ij}(k)| \leq R_{ij}\rho_{ij}^{-k}, \quad \forall\, k \geq 0, \tag{5}$$

see also Jacobson and Nett (1991). Of course one could use this a priori information (5) directly to derive an upper bound for the l^1-norm of the uncertainty, but in general this will yield a much too conservative upper bound.

Finally, in order to be able to handle initial conditions, we need an upper bound $\bar{u}$ on past (unmeasured) data $u(t)$

$$\bar{u}_i = \sup_{t<0} |u_i(t)|. \tag{6}$$

If the system is at rest at $t = 0$, $\bar{u}_i$ can be chosen to be equal to 0.

5 Problem Reformulation

Given this a priori information we want to compute a non-conservative upper bound on the l^1-norm of the unstructured uncertainty Δ. For that purpose we reformulate the problem to a finite dimensional constrained maximalisation problem with a nonlinear objective function and linear constraints. In the next section three methods are proposed to solve this problem.

In principle Δ has an infinitely long pulse response sequence. We want to find the worst-case uncertainty that is consistent with the data available. This is thus a problem with infinitely many unknowns. In order to reduce the problem to a finite dimensional one we split the uncertainty Δ into two parts:

$$\Delta(q^{-1}) = \tilde{\Delta}(q^{-1}) + \bar{\Delta}(q^{-1}),$$
$$\tilde{\Delta}(q^{-1}) = \sum_{k=0}^{n} \delta(k)q^{-k}, \quad \bar{\Delta}(q^{-1}) = \sum_{k=n+1}^{\infty} \delta(k)q^{-k}, \tag{7}$$

where n is to be chosen by the user. We now try to find the worst-case $\tilde{\Delta}$ that is consistent with the data and estimate the influence of $\bar{\Delta}$ from equation (5). This is a problem with a finite number of unknowns.

We substitute (4) into (3) and get

$$y = \hat{G}u + \Delta M u + H_1 e, \quad H_2 e \in [-e_l, e_l]. \tag{8}$$

We write for the terms appearing in (8):

$$\hat{G}(q^{-1})u(t) = x(t) + a(t), \quad x(t) = \sum_{k=0}^{t} \hat{g}(k)u(t-k),$$
$$a(t) = \sum_{k=t+1}^{\infty} \hat{g}(k)u(t-k), \quad t = 0, 1, \ldots, N. \tag{9}$$

and

$$M(q^{-1})u(t) = w(t) + b(t), \quad w(t) = \sum_{k=0}^{t} m(k)u(t-k),$$
$$b(t) = \sum_{k=t+1}^{\infty} m(k)u(t-k), \quad t = -n, -n+1, \ldots, N, \tag{10}$$

where $w(t) = 0$ for $t < 0$. Using (7) we obtain

$$\Delta M u = (\tilde{\Delta} + \bar{\Delta})(w + b) = \tilde{\Delta}w + c + d,$$
$$c = \tilde{\Delta}b, \quad d = \bar{\Delta}(w + b) \tag{11}$$

In the appendix expressions are given for upper bounds $\bar{a}(t)$, $\bar{b}(t)$, $\bar{c}(t)$ and $\bar{d}(t)$ on the signals $a(t)$, $b(t)$, $c(t)$ and $d(t)$ respectively. In general the functions $\bar{a}(t)$, $\bar{b}(t)$ and $\bar{c}(t)$ vanish for increasing t, but $\bar{d}(t)$ will not vanish for increasing t, though it may remain small if n is sufficiently large.

With these results equation (8) can be written as

$$y = x + \tilde{\Delta} w + H_1 e + r,$$

$$r \in [-\bar{a} - \bar{c} - \bar{d},\ \bar{a} + \bar{c} + \bar{d}],\quad H_2 e \in [-e_l, e_h].$$

By introducing a new noise representation

$$v = \begin{bmatrix} e \\ r \end{bmatrix},\ v_l = \begin{bmatrix} e_l \\ \bar{a} + \bar{c} + \bar{d} \end{bmatrix},\ v_h = \begin{bmatrix} e_h \\ \bar{a} + \bar{c} + \bar{d} \end{bmatrix},$$

$$V_1(q^{-1}) = \begin{bmatrix} H_1(q^{-1}) & I \end{bmatrix}, V_2(q^{-1}) = \begin{bmatrix} H_2(q^{-1}) & 0 \\ 0 & I \end{bmatrix}$$

and introducing $z = x - y$, we finally get the system representation

$$z(t) + \tilde{\Delta}(q^{-1}) w(t) + V_1(q^{-1}) v(t) = 0,\ t = 0, \ldots, N,$$

$$V_2(q^{-1}) v(t) \in [-v_l(t), v_h(t)]. \qquad (12)$$

If $H_1(q^{-1}) \equiv H_2(q^{-1}) \equiv I$ it is simpler just to add $\bar{a} + \bar{c} + \bar{d}$ to $e_l(t)$ and $e_h(t)$ and no new noise vector is needed. In the representation (12) the only unknown quantities are the transfer function $\tilde{\Delta}$ and the noise signal $v(t)$. In order to get to the new system representation (12) extra uncertainty has been added to the noise representation. So the reduction of the number of unknowns in Δ naturally implies an increase of the freedom in the remaining parameters of Δ. This introduces some conservatism which cannot be prevented, but only minimized by using accurate a priori information and choosing n as large as possible.

When we now are able to compute an upper bound on the l^1-norm of each row of Δ, we also have an upper bound on the l^1-norm of Δ. According to the definition of the l^1-norm we namely have that

$$\|\Delta\|_{\mathcal{A}} = \max_i \|\Delta_i\|_{\mathcal{A}}. \qquad (13)$$

Moreover the problem of finding an upper bound on the l^1-norm of the ith row of Δ now boils down to finding an upper bound $\bar{\mu}_i$ on the l^1-norm of the ith row of $\tilde{\Delta}$,

$$\left\|\tilde{\Delta}_i\right\|_{\mathcal{A}} \leq \bar{\mu}_i. \qquad (14)$$

From the definition of the l^1-norm, equation (5) and (7) we namely have that

$$\|\Delta_i\|_{\mathcal{A}} = \left\|\tilde{\Delta}_i\right\|_{\mathcal{A}} + \left\|\bar{\Delta}_i\right\|_{\mathcal{A}} \leq \bar{\mu}_i + \sum_{k=n+1}^{\infty} \sum_{j=1}^{q} R_{ij} \rho_{ij}^{-k} =$$

$$= \bar{\mu}_i + \sum_{j=1}^{q} R_{ij} \rho_{ij}^{-n} (\rho_{ij} - 1)^{-1}. \qquad (15)$$

Finally the value $\bar{\mu}_i$ is obtained by finding the solution to the maximization problem

$$\bar{\mu}_i = \max \sum_{k=0}^{n} \sum_{j=1}^{q} |\delta_{ij}(k)|, \qquad (16)$$

such that (12) and (5) are (still) satisfied. In this way the worst-case situation is calculated and therefore inequality (14) will hold. We can determine the required upper bound directly by solving the nonlinear programming problem (16). This is generally however a difficult job. Therefore in the next section this nonlinear programming problem is reformulated to more tractable programming problems. But first we summarize the procedure we followed till now for the calculation of an upper bound on $\|\Delta\|_{\mathcal{A}}$.

Procedure 1 *Collect the required a priori knowledge (section 4), including measurement data, noise-bounds, uncertainty representation, bounds on the pulse response sequence of the uncertainty and a bound on the amplitude of the input-signal in the past.*

Choose a value of n, the order of $\tilde{\Delta}$, and determine the new system representation (12) by following the steps in section 5 and using the signal-bounds derived in the appendix.

Solve for each row i the nonlinear programming problem

$$\bar{\mu}_i = \max \sum_{k=0}^{n} \sum_{j=1}^{q} |\delta_{ij}(k)|$$

under the (linear) constraints

$$z(t) + \sum_{k=0}^{n} \delta(k) w(t-k) + \sum_{k=0}^{h_1} V_1(k) v(t-k) = 0,\quad \forall\, t$$

$$\sum_{k=0}^{h_2} V_2(k) v(t-k) \in [-v_l(t), v_h(t)],\ \ t = 0, \ldots, N,$$

$$\delta_{ij}(k) \in [-R_{ij}\rho_{ij}^{-k}, R_{ij}\rho_{ij}^{-k}],\ \ k = 0, \ldots, n.$$

This problem has to be solved for the unknown matrices $\delta(k)$, $k = 0, 1, \ldots, n$, and the unknown vectors $v(t)$, $t = 0, 1, \ldots, N$.

Determine an upper bound on $\|\Delta\|_{\mathcal{A}}$ using the equations (13) and (15).

If the a priori information is correct this worst-case l^1-norm is a guaranteed upper bound for the difference between the system and the model. Moreover it is non-conservative in the sense that the worst-case situation can actually occur, consistent with the given a priori information. There is however a source of conservatism not arising from

the presented procedure, but from the definition of the l^1-norm. The result of equation (13) is probably unnecessarily conservative because the maximum is only attained for one value of i. This kind of conservatism can however be banished easily by introducing proper scaling matrices for the input and the output.

Define the input scaling matrix

$$U = diag(\|(Mu)_1\|_\infty, \ldots, \|(Mu)_q\|_\infty), \quad \Delta' = \Delta U$$

and then calculate $\|\Delta_i'\|_{\mathcal{A}}$ for all values of i after proper substitutions. Next define the output scaling matrix

$$T = diag(\|\Delta_1'\|_{\mathcal{A}}, \ldots, \|\Delta_p'\|_{\mathcal{A}}), \quad \Delta'' = T^{-1}\Delta',$$

resulting in the uncertainty representation

$$\Delta = T\Delta''U^{-1}, \quad \|\Delta''\|_{\mathcal{A}} = 1,$$

which is non-conservative in the sense that all entries of input and output are active in the worst-case situation.

6 Methods of Solution

We now reformulate the nonlinear programming problem of procedure 1 to other, more tractable, programming problems. Consider the ith row of $\tilde{\Delta}$ and reparametrise

$$\delta_{ij}(k) = x_j(k) - y_j(k), \quad x_j(k)y_j(k) = 0,$$

$$x_j(k) \geq 0, \; y_j(k) \geq 0, \; k = 0, \ldots, n, \; j = 1, \ldots, q,$$

with the property

$$|\delta_{ij}(k)| = x_j(k) + y_j(k),$$

where the x and y are of course different from those previously used. Of course the matrix $\delta(k)$ is now for each value of k given by

$$\delta(k) = \begin{pmatrix} \delta_1(k) \\ \vdots \\ \delta_{i-1}(k) \\ x^T(k) - y^T(k) \\ \delta_{i+1}(k) \\ \vdots \\ \delta_p(k) \end{pmatrix}$$

If we finally define the parameter μ_i to be the sum of all $x_j(k)$ and $y_j(k)$ we can formulate the first method to solve the nonlinear programming problem of procedure 1.

Method 1 *Determine for each value of i, $\bar{\mu}_i = \max \mu_i$ under the constraints (12) and (5) and the additional constraints*

$$\mu_i - \sum_{k=0}^{n}\sum_{j=1}^{q}(x_j(k) + y_j(k)) = 0,$$

$$\sum_{k=0}^{n}\sum_{j=1}^{q} x_j(k)y_j(k) = 0, \; x_j(k) \geq 0, \; y_j(k) \geq 0.$$

This still is a nonlinear problem but it is closely related to the so-called Generalized Linear Complementarity Problem (GLCP), see De Moor (1988, chapter 3). GLCP is actually concerned with finding a description of the set of unknowns that satisfy constraints of the form appearing in the optimization problem stated here. The optimum is then found by searching for $\bar{\mu}_i$, the maximum value of μ_i appearing in the set of feasible solutions. This largest value of μ_i is then equal to the desired upper bound on $\left\|\tilde{\Delta}_i\right\|_{\mathcal{A}}$.

Another approach is to fix μ_i in the set of constraints and look iteratively for $\bar{\mu}_i$, the largest value of μ_i such that there still is a feasible solution to the set of constraints. This means that we are then interested in existence of solutions and not in an exact description of the whole feasible solution set. We formulate this in the second method for the solution of the nonlinear programming problem of procedure 1.

Method 2 *Determine iteratively for each row i, $\bar{\mu}_i$, the maximum value of μ_i, such that a feasible solution exists for the set of constraints (12) and (5) and the additional constraints*

$$\sum_{k=0}^{n}\sum_{j=1}^{q}(x_j(k) + y_j(k)) \geq \mu_i,$$

$$\sum_{k=0}^{n}\sum_{j=1}^{q} x_j(k)y_j(k) = 0, \; x_j(k) \geq 0, \; y_j(k) \geq 0.$$

This problem can again be solved using GLCP software. Notice that if there is a μ_i for which there is no feasible solution to the set of constraints, then $\bar{\mu}_i < \mu_i$.

We notice that the methods presented require a large computational effort, because the number of constraints can be large (namely proportional to the number of samples N). However it is possible to obtain an approximating solution using linear programming for which efficient software exists, see Luenberger (1984). The procedure is then the same as in the context of parameter-bounding algorithms (Milanese and Belforte, 1982; Milanese, 1989) and

is simply to determine the maximum and minimum possible values of each parameter of the pulse response sequence of $\tilde{\Delta}$ without considering interaction. We formulate the procedure.

Method 3 *Determine for each value of i,j and k the solution to the linear programming problems:*

$$\max \delta_{ij}(k), \ \min \delta_{ij}(k)$$

subject to the (linear) constraints (12) and (5). Call the maximum absolute value of $\delta_{ij}(k)$, $\bar{\delta}_{ij}(k)$. We then obtain the (conservative) upper bound

$$\left\|\tilde{\Delta}_i\right\|_{\mathcal{A}} \leq \sum_{k=0}^{n} \sum_{j=1}^{q} \bar{\delta}_{ij}(k).$$

The advantage of this method is also that the central estimate (the average of the lower and upper bound on $\delta_{ij}(k)$) is optimal in the sense that it minimizes the estimate of the l^1-norm of the uncertainty resulting in method 3. Of course this estimate not necessarily minimizes the upper bound calculated in the methods 1 and 2.

7 Discussion

The procedure presented in section 5 in combination with either of the first two solution methods of the previous section provide a way to determine a guaranteed and non-conservative upper bound on the l^1-norm of the unstructured uncertainty. The non-conservativeness arises from the fact that the actual worst-case situation is calculated and therefore can occur consistent with the a priori information given. The main drawback in the computation is however the amount of unknowns and constraints in the Generalized Linear Complementarity Problem, both of the order $N + n$. This may cause computational problems, especially if the number of samples N is large. Therefore also an approximating method is presented that makes use of linear programming, for which efficient standard software exists. Of course this introduces conservatism in the computation of the upper bound.

The choice of a larger value of n will lead to more unknowns and constraints in the computations, but will also lead to a tighter upper bound on $\|\Delta\|_{\mathcal{A}}$, provided the a priori information is correct. A tighter bound is obtained as well, if the noise is small and the influence of the initial conditions ($\bar{u}$) is small.

If more a priori knowledge is available, it can be taken into account by adding linear constraints to either of the three programming problems. In theorem 1 it has been assumed that Δ is strictly causal, so $\delta_{ij}(0) = 0, \ \forall \ i,j$. This can be taken into account by either removing the $\delta_{ij}(0)$ from the unknowns or adding the constraint that they are equal to zero.

If the structure of Δ is known a priori (e.g. diagonal), this information can be used by again either removing the unknowns or adding linear constraints. In the latter case it is also possible that in the constraints inequalities are used instead of equalities, that means that a certain interval can be specified in which the parameters of the uncertainty have to be. This is of course analogous to equation (5). In the same way it is possible to impose more restrictions on the noise $v(t)$, for example that the mean (absolute) value lies within a certain region, etc.

Finally we notice that the resulting upper bound on the l^1-norm of the uncertainty is an upper bound on the H_∞-norm as well, because the latter one is always smaller than or equal to the former one, see e.g. Boyd and Doyle (1987). This upper bound for the H_∞-norm is of course conservative and tighter bounds should be obtainable, given the same a priori information.

8 Conclusions

In this paper a procedure has been developed that yields a non-conservative upper bound on $\|\Delta\|_{\mathcal{A}}$, starting from measurement data and certain a priori information, such as bounds on the noise. The computation can be performed using GLCP software and is quite complicated due to the large number of constraints, especially if many data points N have to be processed. Also a simplified though approximating method is proposed that requires the solution of a set of linear programming problems. Further research is necessary to investigate the applicability of the proposed methods and to extend the results to identification of l^1-optimal models.

Appendix Upper Bounds for a, b, c and d

Using the a priori information of section 4 we can give upper bounds for the signals $a(t)$, $b(t)$, $c(t)$ and $d(t)$ appearing in the equations (9) till (11).
Combining (9) and (6) yields

$$|a_i(t)| \leq \bar{a}_i(t) = \sum_{k=t+1}^{\infty} \sum_{j=1}^{q} |\hat{g}_{ij}(k)| \bar{u}_j, \ t = 0, \ldots, N,$$

so $\bar{a}_i(t)$ is a decreasing function of t.
In the same way we get from (10)

$$|b_i(t)| \leq \bar{b}_i(t) = \sum_{k=t+1}^{\infty} \sum_{j=1}^{q} |m_{ij}(k)| \bar{u}_j, \ t = -n, \ldots, N$$

with the property that for any $t > 1$

$$\bar{b}_i(-t) = \cdots = \bar{b}_i(-1) \geq \bar{b}_i(0) \geq \bar{b}_i(1) \geq \cdots \geq \bar{b}_i(t).$$

Combining (11) and (5) yields

$$|c_i(t)| \leq \sum_{k=0}^{n} \sum_{j=1}^{q} |\delta_{ij}(k)||b_j(t-k)| \leq \bar{c}_i(t) =$$

$$= \sum_{k=0}^{n} \sum_{j=1}^{q} R_{ij} \rho_{ij}^{-k} \bar{b}_j(t-k), \ t = 0, 1, \ldots, N,$$

so $\bar{c}_i(t)$ is also a decreasing function of t.
Finally we find

$$|d_i(t)| \leq \sum_{k=n+1}^{\infty} \sum_{j=1}^{q} |\delta_{ij}(k)|(|w_j(t-k)| + |b_j(t-k)|)$$

$$\leq \sum_{k=n+1}^{\infty} \sum_{j=1}^{q} R_{ij} \rho_{ij}^{-k} (|w_j(t-k)| + \bar{b}_j(t-k)) =$$

$$= \sum_{k=n+1}^{t} \sum_{j=1}^{q} R_{ij} \rho_{ij}^{-k} (|w_j(t-k)| + \bar{b}_j(t-k)) +$$

$$+ \sum_{k=t+1}^{\infty} \sum_{j=1}^{q} R_{ij} \rho_{ij}^{-k} \bar{b}_j(-1) := \bar{d}_i(t) =$$

$$= \sum_{k=n+1}^{t} \sum_{j=1}^{q} R_{ij} \rho_{ij}^{-k} (|w_j(t-k)| + \bar{b}_j(t-k)) +$$

$$+ \sum_{j=1}^{q} R_{ij} \rho_{ij}^{-t} (\rho_{ij} - 1)^{-1} \bar{b}_j(-1), \ t = 0, 1, \ldots, N,$$

that will not vanish for increasing t, especially due to the contribution of $|w_j(t-k)|$.

References

Boyd, S. and J. Doyle (1987). Comparison of peak and RMS gains for discrete-time systems. *Systems & Control Letters,* Vol. 9, pp. 1–6.

Dahleh, M.A. and Y. Ohta (1988). A necessary and sufficient condition for robust BIBO stability. *Systems & Control Letters,* Vol. 11, pp. 271–275.

Dahleh, M.A. and J.B. Pearson (1987). l^1-Optimal feedback controllers for MIMO discrete-time systems. *IEEE Transactions on Automatic Control,* Vol. AC-32, pp. 314–322.

De Moor, B. (1988). *Mathematical Concepts and Techniques for Modelling of Static and Dynamic Systems,* Ph.D. Thesis, Dept. Electrical Eng., Catholic University Leuven, Belgium.

Desoer, C.A. and M. Vidyasagar (1975). *Feedback Systems: Input-Output Properties,* Academic Press, New York.

De Vries, D.K. (1991). *Identification for Robust Control: What is the Best Approach?,* internal report, Lab. Measurement and Control, Dept. Mechanical Eng. and Marine Techn., Delft University of Technology, The Netherlands.

Helmicki, A.J., C.A. Jacobson and C.N. Nett (1989). H_∞ identification of stable LSI systems: a scheme with direct application to controller design. *Proceedings of the American Control Conference,* pp. 1428–1434.

Jacobson, C.A. and C.N. Nett (1991). Worst case system identification in l_1: optimal algorithms and error bounds. (Submitted for publication).

Khammash, M. and J.B. Pearson (1991). Performance robustness of discrete-time systems with structured uncertainty. *IEEE Transactions on Automatic Control,* Vol. AC-36, pp. 398–412.

Luenberger, D.G. (1984). *Linear and Nonlinear Programming,* Addison Wesley, U.K., 2nd edition.

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

McDonald, J.S. and J.B. Pearson (1991). l^1-Optimal control of multivariable systems with output norm constraints. *Automatica,* Vol. 27, pp. 317–329.

Mendlovitz, M.A. (1989). A simple solution to the l^1 optimization problem. *Systems & Control Letters,* Vol. 12, pp. 461–463.

Milanese, M. (1989). Estimation and prediction in the presence of unknown but bounded uncertainty: a survey. In: *Robustness in Identification and Control,* eds. M. Milanese , R. Tempo and A. Vicino, Plenum Press, New York, pp. 3–24.

Milanese, M. and G. Belforte (1982). Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: linear families of models and estimators. *IEEE Transactions on Automatic Control,* Vol. AC-27, pp. 408–414.

Vidyasagar, M. (1986). Optimal rejection of persistent bounded disturbances. *IEEE Transactions on Automatic Control,* Vol. AC-31, pp. 527–534.

Selected Topics in Identification, Modelling and Control
Vol. 3, June 1991

Closed loop identification of a 600 MW Benson Boiler.[‡]

J. Heintze [§] **and H. Aling** [♯]

[§] *Mechanical Engineering Systems and Control Group*
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
[♯] *Integrated Systems Inc.*
2500 Mission College Boulevard, Santa Clara, CA 95054-1215, USA.

Abstract. Experiments have been performed for MIMO system identification of a Benson Boiler. Five input variables were perturbed simultaneously by excitation of the set points with independent binary sequences. After preprocessing of the data, several signal analysis and prediction error identification methods have been applied. Identified models have been cross validated, with satisfactory results. The main issue of this research is to investigate the feasibility of closed loop identification for large scale industrial processes.

Keywords. Closed loop identification; Signal analysis; Application to power plant.

1 Introduction

In spite of the attention which has been given to the development of methods for system identification in the past decades, the number of practical applications on a real plant is relative small. This applies especially to multivariable closed loop system identification.

Most of the applications deal with the open loop case (Katayama *et al.* (1977), Looy (1988), Sutherland *et al.* (1976), Swaanenburg *et al.* (1985), Tyssø (1981)), or with simulated data (Greco *et al.* (1982)). Successful applications have been obtained by Otomo, Nakagawa and Akaike (1972), Nakamura and Akaike (1981) who have applied AR modeling methods to a cement rotary kiln and a power plant. In both cases, successful results were obtained by application of optimal control techniques based on the open loop identified models.

References Zee (1981) and Eklund *et al.* (1973) deal with closed loop identification of real data. The former is based on stochastic realization and prediction error methods applied to data of a pilot plant, the latter deals with prediction error methods applied to a boiler system, but without explicit reference to the closed loop nature of their experiments. Surveys which contain a number of recommendations and some general information about experiments performed on real plants, are given in Isermann (1980), Gustavsson (1972).

We have applied identification to an industrial powerplant, explicitly addressing the multivariable and closed loop situation. Therefore, we have performed a number of closed loop experiments on the boiler system of a 600 MW coal-fired power plant of the EPON in Nijmegen, the Netherlands. In this paper we report the results of the final identification experiment which was performed under partial load operation. We intend to present a number of the problems encountered in our particular application, which we believe are fundamental from either a practical or a theoretical point of view, and which are common to most other large scale industrial applications of system identification. For details on the underlying theory we will refer to the literature.

Because of the closed loop situation under which

[‡]Part of this work has been supported by the Department of Trade and Industry. Experimental facilities: Stork Boilers and EPON.

the experiments had to be performed, we will first discuss the closed loop identification problem in section 2. Some characteristics of the boiler are described in section 3. The experiment design is briefly treated in section 4. Especially in the multivariable closed loop situation, the choice of a correct timestep and the selection of signals for identification is complex. These problems and ways to attack them are discussed comprehensively in sections 5 and 6. Results of the parametric identification are given in section 7, and we will finish with conclusions (section 8).

2 Closed loop system identification

In application of system identification techniques to industrial processes, experiments are almost inevitably performed under closed loop conditions. This is necessary for maintaining a sufficient level of security and productivity. In most cases, three kinds of observed variables can be distinguished. **Set point variables**, which serve as references for the output variables, **controlled input variables** which can be set directly to any desired value and which are used as controller outputs, and the system **output variables**.

The disturbed feedback system, as depicted in figure 1, is symbolically described by the following equations:

$$y = Gu + d \quad , \quad u = Ky + n \quad , \quad n = n_1 + Kn_2 \quad .$$

Here, G and K represent transfer functions for the deterministic part of the forward and backward system. For ease of notation, the frequency-dependence of these transfer functions and of the signals has not been indicated explicitly.

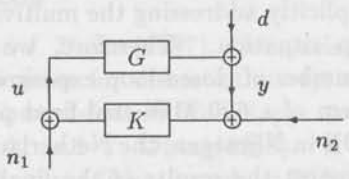


Fig. 1: Feedback system.

The measurable variables are: y (process output variables), u (controlled input variables), n_2 (set-point variables). The unmeasurable variables are: n_1 (input disturbances), d (output disturbances). n_2 can be used to introduce specific disturbances to the system, determined by the experiment design for identification. Therefore, the variable n is considered as a disturbance with a measurable and an unmeasurable part. Sometimes d is described by $d = He$. H represents the transfer function for the stochastic part of the forward system, e is noise. The problem of identifying a model which represents the open loop behaviour of G and sometimes also H, based on measured signals of the process, is called the **closed loop system identification** problem. In Anderson *et al.* (1982), Söderström *et al.* (1976) and Ng *et al.* (1977), the problem is treated in the frequency domain. in Aling (1990), a state space approach towards the problem has been formulated.

We will assume that some basic conditions are satisfied. First, all identifiability results have been proved under ideal conditions: the model structure is assumed to be compatible with the underlying data generating system. Secondly, the unmeasurable noise e is assumed to be independent in time (white), to have a time-independent variance and, with prediction error based identification methods, to be Gaussian. Thirdly, the closed loop system is assumed to be stable. We will use direct methods, which are subject to a number of associated identifiability conditions:

Direct methods. The identification is performed using open loop methods with u and y as measured model in- and output variables. A number of assumptions with respect to the delay structure and correlation between the noise sources on controlled and output variables are required (Söderström *et al.*, 1976):

1) The product of transfer functions GK is strictly proper. In other words, there is a delay of one discrete time step in the control loop.
2) The unmeasurable disturbance d is uncorrelated with n.
3) The external disturbance n is **persistently exciting**. Persistency of excitation means that the spectral density function of the disturbances has a sufficiently broad bandwidth.

Recently condition 1 is relaxed. In Hof *et al.* (1990) is shown that the closed loop transfer function GK may not have any algebraic loops, which is less severe then strict properness.

The noise conditions are hard to verify. Some other conditions can be checked a posteriori; the correlation between input excitation and noise (prediction errors) is an example. We must bear in mind that the results hold asymptotically, i.e. for an infinite amount of data. Thus, small input excitations will not automatically lead to good models for a moderate observation time.

3 Process description

The boiler can be separated in the combustion circuit and the water-steam circuit. The behaviour of the combustion circuit is primarily described by the following signals: fuel flow, combustion gas damper, primary and secondary air flow.

We will concentrate our identification efforts on the water-steam circuit of the boiler, which is schematicly presented in figure 2. Some properties are discussed below.

- The Benson boiler is of the once-feedthrough type, and therefore contains no drum. After the medium pressure turbine the steam is condensed and recirculated via the accumulator by the feed water pump.
- The high pressure part consists consecutively of economizer, evaporator, Benson bottle (see next item) and three superheaters. After the high pressure turbine the steam is reheated before it enters the medium pressure turbine. A valve for pressure controle is placed in front of each turbine.
- The Benson bottle is a water buffer, with the purpose of maintaining a minimal amount of circulation in the evaporator during startup or partial load. This minimal circulation flow is necessary for preservation of the evaporator pipes of which the temperature must remain below a safe limit. Without the Benson circuit the evaporator would not be able to evaporate the feed water with this minimum flow and the feed water level would finally end up in the superheaters. The circulation flow is controlled by a level controller which regulates the water level in the Benson bottle and is effectuated by the circulation pump. During normal full load operation this circulation circuit should not affect the process.
- The boiler operates subcritical. As a consequence, the temperatures in the evaporator are directly related to the pressure.

Although we do not go into details of the combustion circuit, we have to realize the following: The sequence in which the steam passes through different parts of the water-steam circuit does not correspond with the flow of the combustion gases through the combustion circuit. This leads to complicated couplings via the combustion circuit, which are not easily described. Basically, these couplings will introduce additional dynamics in the model of the water-steam circuit, caused by physical feedback or feedforward. We have assumed that the influence of the combustion channel to the water-steam circuit is one-sided.

4 Experiment design

In order to maintain a 40% load operation, only three of the six burners where used. The feedwater flow was increased slightly in order to enforce circulation flow.

Contrary to the standard way of operation, where the pressure is controlled primarily by the fuel flow, a control scheme was chosen by which the pressure was regulated by the turbine valve. This offered the opportunity to excite the turbine valve by injecting pressure set point perturbations to the control system, which would have been impossible otherwise. Unfortunately, the fuel flow had to maintain a constant level by which it was not possible to excite this input variable.

To avoid discontinuities in the data, the process of cleaning the combustion area from smuts at discrete time instances was stopped. The hereby introduced slowly time-dependent effect of pollution is expected to be removed from the data by means of trend removal.

Almost all important input and output variables are directly related to the high pressure part. Compared to this, the influence of the dynamics of the medium pressure part is considerably less. We will summarize the most important signals related to the high pressure part of the boiler and discuss them later.

Input variables:

QFW	feed water flow
QSC1	flow spray cooler 1
QSC2	flow spray cooler 2
VHP	turbine valve aperture high pressure
CGD	combustion gas damper aperture
QC	circulation flow

Output variables:

TSH3	steam temperature after third superheater
Δ_T	temp. difference over the second spray coole
PHP	average pressure in the high pressure part
LC	circulation water level

Intermediate variables:

QIE	water flow inlet evaporator
TCJ	temperature after circulation junction
TSH1	temperature after the primary superheater
TSH2	temperature after the secondary superheater
PFW	pressure feedwater
Δ_P	PFW-PHP
QF	fuel flow
QPA	primary air flow
QSA	secondary air flow

Remark 1: The subdivision of the signals in input, output and intermediate is based on a priori knowledge, and is to some extent arbitrary. The results

of signal analysis and also the identification will finally prescribe which variables can be regarded as being inputs etcetera.

Remark 2: The turbine valve apertures which are used for the model are not the actual apertures, but so-called linearized apertures which relate in a linear way to the pressure difference over the valves.

Chosen was for a 2 Hertz sampling rate. This rate will certainly cover all the dynamics.

We have employed a new method for the generation of independent Pseudo Random Binary Sequences (PRBS, plural: PRBS's), which requires one shift register only and no specific initial state information. This method, which is very practical for multivariable real-time applications, is described in Aling (1990). The bandwith of each of the following PRBS's is according to the expected dynamics in the subsystem it has to perturbe, and is determined by the clock period Δ, given a basic timestep of one second for the shifregister, which has order 16, giving a PRBS length of 65535 (maximum length sequence).

PRBS QIE: excites the feed water flow through the evaporator. $\Delta = 53$.

PRBS QSC1 and PRBS QSC2: affects different temperatures in the process. $\Delta = 141$, $\Delta = 89$ respectively.

PRBS VHP: excites the turbine valve. $\Delta = 17$.

PRBS QC: excitation of the circulation flow. $\Delta = 23$.

The combustion gas damper and the reheater spray cooler were kept at fixed levels. Unfortunately a **step disturbance on the CGD** took place halfway the experiment, which forced us to include the damper as an additional regression variable into some of the models.

5 Data preprocessing

The total experimentation time was circa 18 hours, including load change from full load to partial load and the adjustment of PRBS disurbances. After the data preprocessing time series of 7.4 hours of the process with circulation circuit active and perturbed by the 5 pertubations simultaneously where at our disposal for signal analysis and identification. In the following we will give a comment on trend removal, and discuss extensively the choice of a correct timestep for identification.

A difficulty with **trend removal** is the lack of a definition of a trend. We carried out trend removal by fitting a 15-th order Chebychev polynomial to each time serie and subsequently substract this polynomial from the time serie.

The **choice of sampling interval** for identification was done based on the assumed time constants of different parts of the boiler. Pressure phenomena involve time constants in the order of seconds, whereas thermal effects are much slower. With a time step of 2 seconds, both high and low frequent dynamics can be modeled. Unfortunately, such a time step would probably be bad for the quality of the low frequent part of the estimated models, for which a sampling interval of 20 seconds would be better. With this choice, fast dynamics would have to be represented by a feedthrough term in the model structure. As a compromise two models could be identified, one of which is better for the high frequencies and the other one for the low frequencies.

For the closed loop case however, such a procedure is not possible. As was explained in section 2, we need a delay of one time step in the loop for identifiability. Consequently, in case there is a feedthrough term in the controller one should always incorporate a one time step delay in the model structure for the forward part. A sampling interval of 20 seconds would then be impossible because it would lead to an intolerable bias for the high frequent parts. For our closed loop case there are two possibilities:

- A small sampling interval of (say) 2 seconds with an imposed delay of one time step for the forward model. The estimates are likely to be poor in the low frequent region.
- Partitioning of the model into different parts which contains either high or low frequent dynamics and identification of models with a delay of one time step based on sampling intervals of 2 or 20 seconds, respectively.

The conclusion is that in the closed loop case with different time constants, we would be inclined to perform the identification in parts, while in the open loop case this is not necessary. In practice this implies that, in the closed loop case, MISO identification of different parts of the system is to be preferred over MIMO identification of the total system. Indeed, this is the case as will be shown in section 7.

As a result of the former we performed the post sampling twice to get a data set with sampling intervals of 2 seconds and a data set with sampling intervals of 20 seconds. Before sampling the data

has to be filtered to prevent aliasing. The design of such a anti aliasing filter involves choices as filter order and relative cutoff frequency. In our case, an autoregressive 4-th order Chebychev filter with a relative cutoff frequency of 0.1 and a pass band ripple of 0.05, gave satisfactory results for the post sampling operation from 2 to 20 seconds dataset.

6 Signal analysis

The purpose of signal analysis is to get a rough impression of the relationship between different system variables in the form of coherencies, covariance functions, nonparametric transfer function estimates etcetera prior to performing parametric system identification. Although the estimates may be far from exact, the role of signal analysis is significant. Especially for the choice of variables which should be incorporated into the model and for the choice of the sampling interval signal analysis may be a key factor to success.

Usually, signal analysis is performed on a pair of scalar signals. In multivariable applications however, contributions from an input variable to an output variable will be considered as noise for any of the other relevant input/output relationships with the same output variable, when scalar methods are applied. Thus, low coherencies are measured although the relationship may be perfectly linear in a multivariable sense. Rather than using scalar methods, we would like to use a method which decomposes the power of an output variable into contributions from all input variables and a noise component, on a relative scale from 0 to 1 over all frequencies.

In this section we will describe such a parametric method for multivariable signal analysis, which was successfully applied by Otomo *et al.* (1972) for an open loop problem, and we give an interpretation of this method for the closed loop case. Then an application of this method to our data set, in order to find a signal selection for the total system, is described. In the last part of this section the process partitioning, which is motivated by the reasoning at the end of section 5, is discussed.

The method is based on a simple autoregressive (AR) model structure for the observed data y_t ($y_t \in R^p$, $t \in Z$). Let us assume that the data can be described by an autoregressive model (ϵ_t represents the innovation):

$$A(q)y_t = \epsilon_t$$

This relationship is expressed in the frequency domain by

$$y_t = H(z)\epsilon_t = A^{-1}(z^{-1})\epsilon_t$$

Let $H_{ij}(z)$ denote the (i,j)-th element of $H(z)$, and let $y_t^{(i)}$ and $\epsilon_t^{(i)}$ be the i-th element of the output and innovation vector, respectively. Also, assume that $\epsilon_t^{(i)}$ and $\epsilon_s^{(j)}$ be independent for any values $i \neq j$, $s \neq t$ and that $\epsilon_s^{(j)}$ has variance σ_j^2. The spectral density function of $y_t^{(i)}$ is given by

$$S_i(z) = \sum_{j=1}^{p} |H_{ij}(z)|^2 \sigma_j^2$$

and the relative power contribution of $\epsilon_t^{(j)}$ to $y_t^{(i)}$ by

$$s_{ij}(z) = \frac{|H_{ij}(z)|^2 \sigma_j^2}{S_i(z)}$$

For fixed i these relative contributions sum up to 1 and are the frequency-dependent relative power contribution from all innovation elements, including innovation element i. In other words, the functions s_{ij} represent a parametric multiple coherence estimate of output element i with the innovation vector. The quantities s_{ij} are scaling-independent, which has been shown in Aling (1990). If y_t is a joint input/output vector of an open loop system, these s_{ij} are an measure for the influence of the inputs (e.g. the excitation due to the testsignals) on the outputs (e.g. the system variables). After having identified the AR model, the independence of the components of the innovation vector, which was assumed, can be checked. With our open loop application of this method, this asumption was always satisfied.

Let us consider what the interpretation of $s_{ij}(z)$ is for multivariable closed loop situation. Again, the analysis is done in the frequency domain. For conciseness, we will drop the dependence on z and use a new notation for a partitioned output vector: Let the output y_t be partitioned as

$$y_t = \begin{pmatrix} y_t^{(1)} \\ y_t^{(0)} \end{pmatrix}$$

Here, $y_t^{(1)}$ is the first output component and $y_t^{(0)}$ is the vector consisting of components $2, \ldots, p$ of y_t. Let the closed loop system be described by

$$y_t = \begin{pmatrix} y_t^{(1)} \\ y_t^{(0)} \end{pmatrix} = \begin{pmatrix} G_1 \\ G_0 \end{pmatrix} u_t + \begin{pmatrix} H_1 \\ H_0 \end{pmatrix} e_t^{(y)}$$

$$u_t = \begin{pmatrix} K_1 & K_0 \end{pmatrix} \begin{pmatrix} y_t^{(1)} \\ y_t^{(0)} \end{pmatrix} + Le_t^{(u)}$$

The vectors $e^{(y)}$ and $e^{(u)}$ are white noise disturbances on the output and input vectors, respectively. The system is shown in figure 3a.

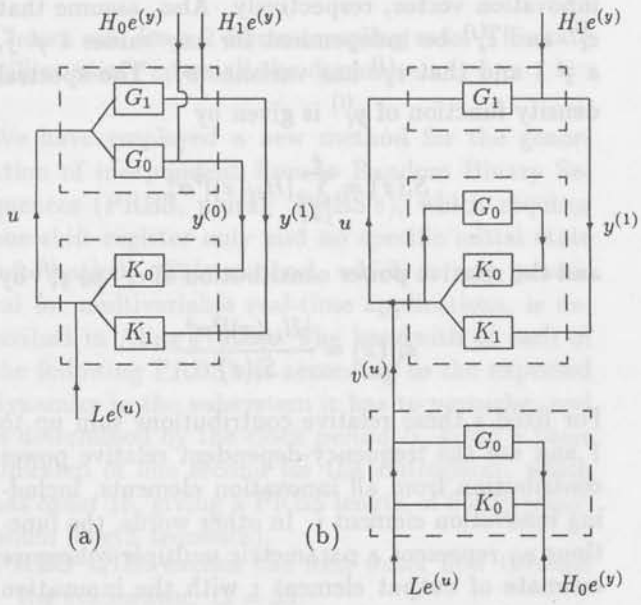


Fig. 3: Feedback and alternative feedback representation of output element $y^{(1)}$.

After some calculations the following is derived: $(I - G_1(I - K_0G_0)^{-1}K_1)\,y^{(1)} = G_1(I - K_0G_0)^{-1}\left(Le^{(u)} + K_0H_0e^{(y)}\right) + H_1e^{(y)}$. In accordance with this formula, the system can also be represented as in figure 3b, which represents the same system as figure 3a.

Let us define
$v^{(u)} = (I - K_0G_0)^{-1}\left(Le^{(u)} + K_0H_0e^{(y)}\right)$ Then, $v^{(u)}$ represents disturbances entering the loop at the input signals. What matters for the identification of MISO models with $y^{(1)}$ as output is the relative power of the signals $G_1v^{(u)}$ and $H_1e^{(y)}$. The first signal represents the effect of independent input disturbances at the output signal, the second represents the amount of additional noise. If $G_1v^{(u)}$ is sufficiently large, we will have a good signal to noise ratio on the output and we may expect good identification results.

The relevant issue here is that, if the parametric signal analysis procedure explained above is applied to the joint input/output vector $y_{I/O}$ where $y_{I/O}^T = \begin{pmatrix} y^T & u^T \end{pmatrix}^T$, the power of $y^{(1)}$ effectuated by these two disturbances, is expressed exactly by $s_{1j}(z)$. For this, a number of assumptions must be made:

- $H_1e^{(y)}$ and $H_0e^{(y)}$ are independent. Therefore, if the analysis is to be performed for all output elements, this condition can only be satisfied if H is diagonal and all components of $e^{(y)}$ are independent. This is a rather restrictive assumption which needs to be checked.
- $H_1e^{(y)}$ and $Le^{(u)}$ are independent. Again, if the analysis is to be performed for all output elements, this implies that $e^{(y)}$ and $e^{(u)}$ must be independent.

Note that the first assumption vanish when the joint i/o vector consists of only one output and (more) inputs (dimension H_0 is zero). This is of importance for closed loop MISO analysis. If the first condition mentioned does not hold, then a lower bound for the power of $G_1v^{(u)}$ is obtained by considering the total power due to the term $G_1(I - K_0G_0)^{-1}Le^{(u)}$ only.

Under these assumptions, noting that $(I - G_1(I - K_0G_0)^{-1}K_1)$ is a scalar quantity which eliminates in the computation of $s_{1j}(z)$, it is easily seen that the quantities $s_{1j}(z)$ represent the relative influence of the disturbances $G_1v^{(u)}$ and $H_1e^{(y)}$ decomposed in their individual innovation components. If the $s_{ij}(z)$ are computed for all $i = 1, \ldots, p$ then an impression is obtained of the excitation by other in- and output variables, which is relevant for the corresponding estimates.

We have applied this parametric signal analysis to our data set with 2 seconds sampling interval (highest frequency is 0.25 Hz), using as the joint I/O vector the variables: QFW, QSC1, QSC2, QC, VHP (inputs) and TSH3, TSH2, TSH1, TCJ, PHP, LC (outputs). An 10-th order autoregressive model is estimated. The innovation components have been checked on their mutual independence. For example, such a check in a application previous to the one discussed here, showed us clearly that the steam flow and the steam pressure can not be seen as two independent system variables, which is logical from a physical point of view. Consequently, the steam flow is not considered as a system variable.

Some of the results are shown in figure 4. The power contributions, which sum up to 1, are represented in the figure by the height of the band between adjacent pairs of plotted lines. The horizontal axis is the logarithm of the frequency. As expected, the power of signal PHP is obviously for a great part due to VHP, and for a smaller amount due to QSC1 and QSC2. The remainder represents the power due to the PHP innovation. In the second figure, the power distribution of QFW is shown. The largest influence in the low frequencies is from

QC. Indeed, there is a strong interaction between QFW and QC caused by the control system. As QC is strongly related with LC because of the level control, also the effect of LC is present. The contributions from QSC1 and QSC2 are due to the fact that the spray cooler flows are tapped directly from the feed water pipe. Note that in the high frequency part, the power contribution to QFW is merely due to its own innovation, whereas the power contribution to PHP is not only due to its own innovation, but also significantly due to VHP. In general the pressure phenomena contain relatively high frequent dynamics, whereas most other system parts have bandwidths which are about ten times smaller. In the present closed loop situation, this observation supports the idea of model partitioning as proposed at the end of section 5, and consequently, to perform MISO identification instead of the estimation of one MIMO model.

With the knowledge over the correct system inputs and outputs, gained by the foregoing parametric signalanalysis, we will divide the process into parts. At this moment we need the intermediate variables, as they are declared in section 4. The partitioning rests on: knowledge of the physical phenomena in the process, results of (parametric) signal analysis, also performed for the joint i/o vectors of each submodel, and results of the identification itself. The latter means that if for example the process variable TSH1 is predicted much better by a model taking also CGD as an input, then CGD is nominated as being an input. Besides the foregoing the following preconditions must be valid: the signal choice for a certain model must not contradict with the signal analysis, checkable conditions as independence of the innovations of the inputs must be fullfilled and the connection of all the submodels, in which outputs of some submodels are inputs for other submodels, must result in a total model. This total model has as inputs the possible controlled input variables and as outputs the defined output variables. The names and some aspects of the ten submodels are:

I: Circulation junction, **II:** Water flow generation, **III:** Economiser,

IV: Evaporator, Benson bottle and first superheater. This is one of the most complicated and crucial parts of the boiler system. The evaporator stretches over the full height of the combustion chamber, and therefore interacts with almost every part of the water-steam and combustion circuit. One of the problems for identification is that many important variables cannot be measured. Temperatures in the combustion chamber are not available at all, and temperatures in the evaporator are equivalent with pressure measurements because in the evaporating process temperature and pressure are related by the phase transient curve of the pressure/temperature diagram for water. For these reasons we were more or less forced to consider the temperature after the primary superheater as the relevant output variable. Although there is no exact water level in the evaporator, we think that there is an effective one which is related with Δ_P. Indeed, inclusion of Δ_P as an input variable resulted in a significant improvement of the model quality. Also CGD was needed as an additional regression variable because of the step disturbance which was mentioned earlier, and turned out to be of considerable influence. Also inclusion of QF, QPA and QSA resulted in improved validation results. The independency in the latter three process inputs is in this case probably due to the nature of the fuel; the quality of coal expressed in combustion energie is irregular. Afterwards identification of this submodel, the innovations can be checked on their mutual independency.

V: Δ_P - subsystem. Because Δ_P is included as a model input for subsystems IV and VII, we have to describe it as the output of this auxiliary subsystem. It turns out that Δ_P can be described fairly well by a model with QIE, QC, QSC1, QSC2, VHP and PHP as input variables.

VI: Spray cooler 1, **VII:** Secondary superheater, **VIII:** Pressure subsystem, **IX:** Spray cooler 2, and **X:** Third superheater.

The data set with 20 seconds sampling interval is used to identify the submodels I, III, IV, VII and X.

Altough the results of the parametric signal analysis give the impression that there should be information related to the varaible LC in the data set, we were not able to find a model to describe the behaviour of LC, by means of closed loop identifcation.

7 Identification results

In this section results of the parametric identification are given. First we discuss the used model structures, identification- and validation-techniques. Then we show time responses of 3 identified models. Two models are the result of MISO closed loop identification, the last one concerns MIMO identification in the open loop.

We use so-called "Talmon and van den Boom" ABFCD model structure (see e.g. Aling (1990)), which is of the form:
$A(q)y_t = F^{-1}(q)B(q)u_t + D^{-1}(q)C(q)\epsilon_t$
A, B, F, C and D are polynomial matrices in the backward shift operator q. The identification was based on the direct method, using Least Squares (LS) and Recursive Prediction Error (RPE) techniques. Equation error AB-models were identified by means of LS, RPE was mostly used to estimate BFCD models. However, several kinds of MIMO ABFCD model structures have been identified. In the closed loop situation, always a delay of one time step is imposed on the forward part of the deterministic part of the model.

The models were cross-validated on an independent part of the data set. For this, the data set was split into two parts of 3.7 hours, 666 or 6660 samples depending on the sampling interval. The first part, where the step disturbance of CGD took place, was used for identification and the second part for validation based on the so-called output error RMS values (Relative Mean Square, the sample variance of the output errors divided by that of the corresponding output). The output error is defined as follows. Consider a model structure of the form $y_t = Gu_t + H\epsilon_t$ where G and H represent the deterministic and the stochastic part of the model respectively. Then $\epsilon_t^o = y_t - Gu_t$ is the output error. Although the identified models could have an equation error structure, output errors are always used for validation because we think that this gives a better measure for the quality of the deterministic part of the model. In many practical cases the user is interested in the deterministic transfer function G. Surprisingly, the models identified with an equation error structure often performed better in output error validation sense then the identified models with an output error structure, for example BF-models. We will not give an interpretation of this phenomenon.

Once the correct inputs and outputs for a model are selected, the determenistic parts of the LS and RPE models (AB and BFCD structures respectively) performed equal, as well with the output error validation as with the shape of the step responses. Often even the static endvalues of the step responses of the different models are close to each other. (See figures 6 and 8).

Figure 5, subsystem IV: Evaporator, Benson bottle and first superheater. A 13-th order BFCD model and a 15-th order LS model have been identified. Only a record of the output error validation is shown, over a time span of 8000 seconds. RMS = 0.26 for both models. The solid line is the measured output, the dashed line is the predicted one by the model.

Figures 6, and 7, subsystem V: Δ_P. Step responses are displayed over 100 seconds of a 6-th order LS model (solid line) and a 4-th order BFCD model (dashed line). An increase of QIE indeed leads to an increase of Δ_P or, in other words, an increase of the water mass in the evaporator. Also, after opening the turbine valve (VHP), the pressure (PHP) decreases by which again Δ_P increases. We have no interpretation for the remaining responses. The RMS value of the LS model is 0.14, 800 seconds of the output error time serie is shown.

Remark: The representation af a submodel in the form of a multivariable step response should not lead to the conclusion that the auxiliary variables used as input variable for the submodel can be manipulated directly and independently. Prudence is called for by physical interpretations of the auxiliary variables as controlled input variables.

Figures 8, and 9, MIMO open loop model. A model has been estimated with the five set point disturbances as inputs and LC, TSH3 and PHP as output variables. This concernes an open loop identification, and the model does not only contain the transfer function of the forward system G, but also an unknown part of the backward system K (which may be undesired). The resulting model step responses are shown in figure 8 over 1000 seconds for a 7-th order BFCD model (solid line) and a 7-th order LS model (dashed line). The RMS values for the BFCD model are 0.40, 0.14 and 0.09 for LC, TSH3 and PHP respectively (validation: figure 9, over 10000 seconds). Note that, contrary to the closed loop situation, no problems arised in estimating a MIMO model and also the behaviour of the variable LC is modelled.

Altogether, it seems that closed loop identification of carefully selected subsystems as well as open loop MIMO identification using RPE methods yields satisfactory results, even for complicated industrial processes.

8 Conclusions

An application of closed loop system identification has been performed on a coal-fired boiler system. Conditions have been formulated under which identification of (partial) models lead to meaningful results. A signal analysis method proposed by Akaike

has been reinterpreted and applied to the closed loop case. Based on signal analysis and physical insight, we made specific choices for the input and output variables of different parts of the boiler. The closed loop identifiability conditions, in combination with the fact that the time constants of the system were quite different, required partitioning of the boiler system and identification of partial models (MISO). For these models, as well as for MIMO open loop models, satisfactory validation results have been obtained. Identification of closed loop systems is considerably more complicated than identification of open loop systems. It requires a lot of expertise, not merely about the process itself, but especially about the potentials and limits of closed loop system identification.

References

Aling, H. (1990). *Identification of Closed Loop Systems.* ISBN 90-370-0029-0, Ph.D. thesis, Delft University Technology, the Netherlands.

Anderson, B.D.O. and M. Gevers (1982). Identifiability of linear stochastic systems operating under feedback. *Automatica vol. 18 no. 2*, 195-213.

Eklund, K. and I. Gustavsson (1973). Identification of drum boiler dynamics. *Proc. 3rd IFAC Symposium, the Hague/Delft*, 1, 87-108.

Greco, C. and V. Marchis (1982). A steam power plant identification. *IASTED Davos*, 113-117.

Gustavsson, I. (1972). Comparison of different methods for identification of industrial processes. *Automatica, 8*, 127-142.

Hof, P.M.J. van den, D.K. de Vries and P. Schoen (1990). On relaxed delay structure conditions in closed loop identifiability problems. *Selected topics in Identification, Modelling and Control, 2*, 53-60, Delft University Press, the Netherlands.

Isermann, R. (1980). Practical aspects of process identification. *Automatica, 16*, 575-587.

Katayama, T., K. Akimoto and Y. Sawaragi (1977). Application of maximum likelihood method to boiler system identification. *L. Memoirs Fac. of Eng. Kyoto University, vol. 39, pt. 1*, 148-167.

Looy, J.M.P. v.d. (1988) *Dynamic Modeling and Control of Coal Fired Fluidized Bed Boilers.* ISBN 90-370-0017-7, Ph.D. thesis, Delft University Technology, the Netherlands.

Nakamura, H. and H. Akaike (1981). Statistical identification for optimal control of supercritical thermal power plants. *Automatica, vol. 17, no. 1*, 143-155.

Ng, T.S., G.C. Goodwin and B.D.O. Anderson (1977). Identifiability of MIMO linear dynamic systems operating in closed loop. *Automatica, 13*, 477-485.

Otomo, T., T, Nakagawa and H. Akaike (1972). Statistical approach to computer control of cement rotary kilns. *Automatica, 8*, 35-48.

Söderström, T., L. Ljung and I. Gustavsson (1976) Identifiability conditions for linear multivariable systems operating under feedback. *IEEE-AC, vol. 21, no. 6*, 837-840.

Sutherland, P. and R.W. Walker (1976). Correlation techniques applied to the identification of boiler dynamics. *Proc. 6th IFAC World Congress, Boston/Cambridge, Ma.*, 1-8, Instrument Society of America, Pittsburgh, PA.

Swaanenburg, H., W. Schinkel, G.A. van Zee and O.H. Bosgra (1985). Practical aspects of industrial multivariable process identification. *Preprints 7th IFAC/IFORS Symposium, York*, 201-206.

Tyssø, A. (1981). Modelling and parameter estimation of a ship boiler. *Automatica, vol. 17, no. 1*, 157-166.

Zee, G.A. van (1981). *System Identification for Multivariable Control.* Ph.D. thesis, Delft University of Technology, the Netherlands.

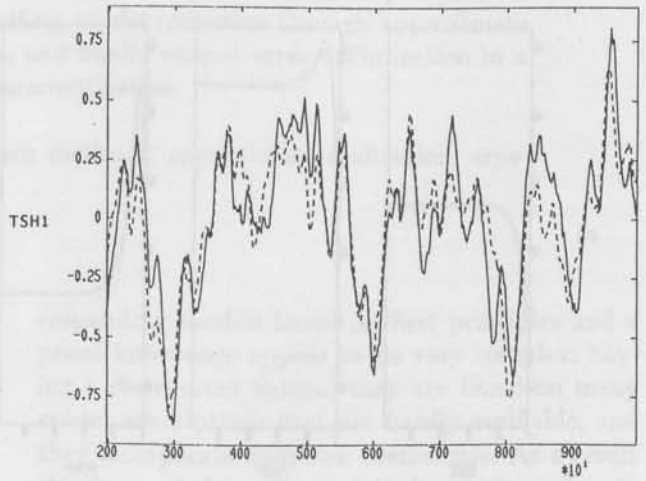


Fig. 5: Validation (submodel IV).

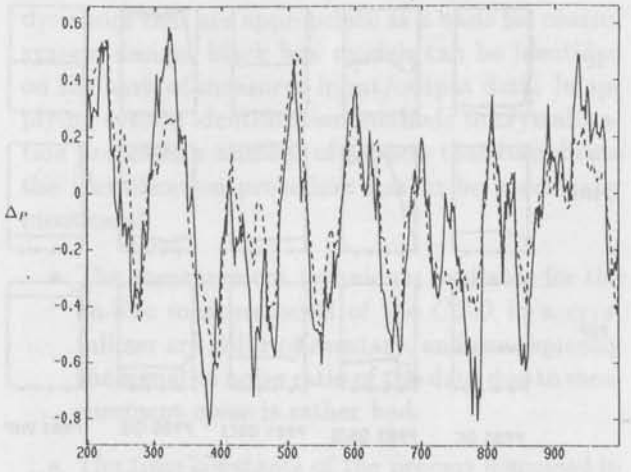


Fig. 7: Validation (submodel V).

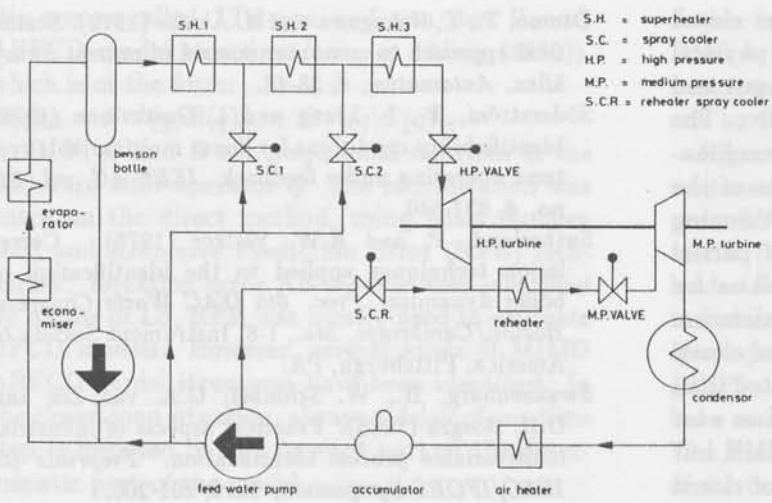


Fig. 2: Water-steam circuit.

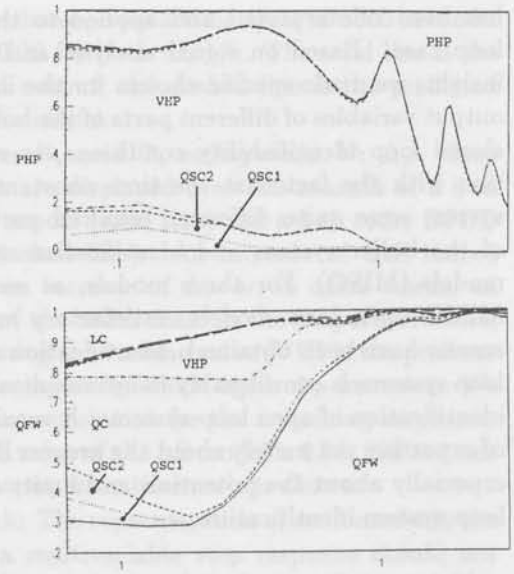


Fig. 4: Results of parametric signal analysis.

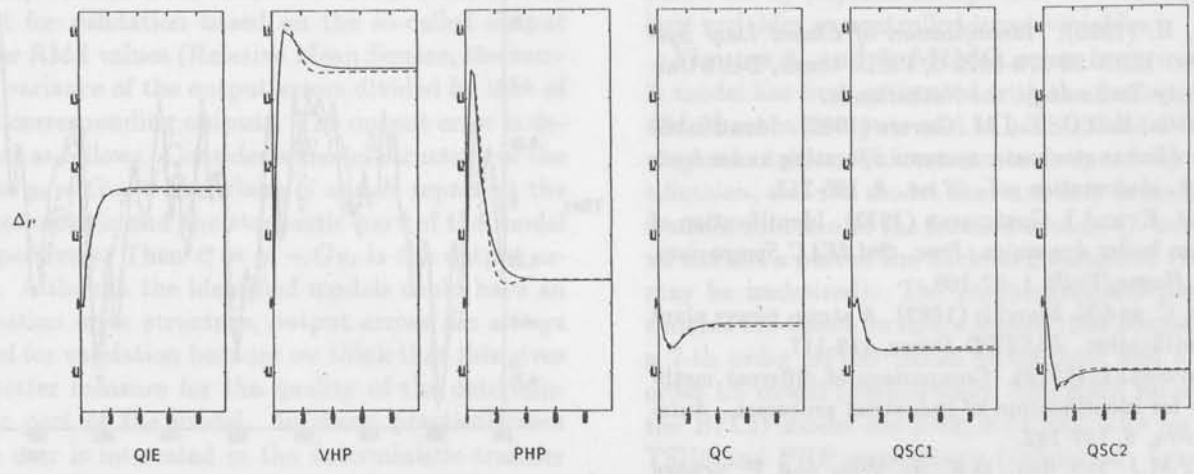


Fig. 6: Step responses (submodel V).

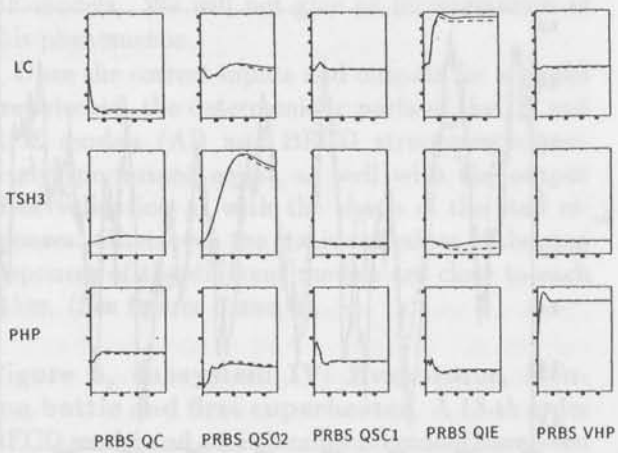


Fig. 8: Step responses (Open loop MIMO model).

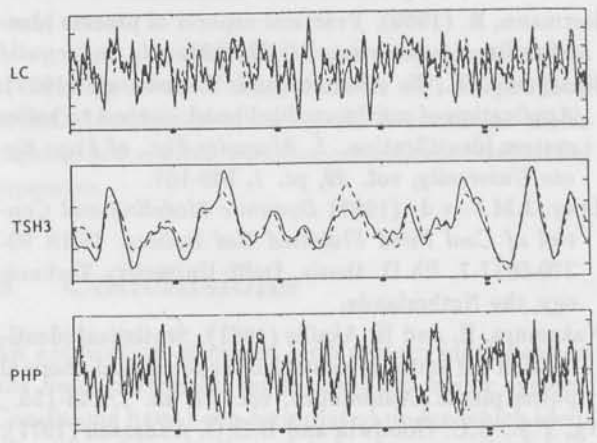


Fig. 9: Validation (Open loop MIMO model).

Selected Topics in Identification, Modelling and Control
Vol. 3, June 1991

Identification of a pilot plant crystallization process with output error methods ‡

Sjoerd de Wolf § **and Paul M.J. Van den Hof**

Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

Abstract. Results are shown of the identification of a three input two output pilot plant evaporative crystallizer (970 litre), in which the *Crystal Size Distribution (CSD)* is the main characteristic of the process to be modelled, and subsequently to be controlled. The experimental situation is characterized by short observation times and a bad signal to noise ratio of the CSD measurement device. A three step identification procedure is applied, consisting of high order ARX modelling, model reduction through approximate realization based on step response matrices, and finally output error optimization in a pseudo-canonical (overlaping) state space parametrization.

Keywords. system identification, state space methods, approximate realization, crystallization, industrial production systems.

1 Introduction

Crystallization from solution is an important separation and purification technique in chemical industry. It is characterized by the formation of a spectrum of differently sized crystals. This spectrum, called the Crystal Size Distribution or CSD, dictates the behaviour of the crystals in succeeding operations, such as filtration, drying, storage and transportation and is also important for the marketability of the crystals produced. In many industrial crystallizers the observed CSD's show transients and oscillations due to (external) disturbances or instability of the process itself (Randolph and Larson, 1988). Therefore it is desirable to control the CSD produced in the crystallizer.
In order to design a controller for the CSD a compact model of the CSD dynamics is required. Corresponding models based on first principles and a priori knowledge appear to be very complex, having a distributed nature; they are based on many model assumptions that are hardly verifiable, and they incorporate unknown coefficients. As a result this type of theoretic models does not produce a sufficiently accurate description of the process dynamics that can be used as a basis for control design on an actual plant (De Wolf, 1990).
In order to arrive at compact models of the CSD dynamics that are appropriate as a basis for control system design, black box models can be identified on the basis of measured input/output data. In applying system identification methods to crystallization processes a number of aspects that complicate the identification procedure has to be specifically mentioned:

- The measurement techniques available for the on-line measurements of the CSD in a crystallizer are still rudimentary, and consequently the signal to noise ratio of the data due to measurement noise is rather bad.
- The time constants of the process discussed in

‡The original version of this paper is presented at the 9^{th} IFAC/IFORS Symposium on Identification and System Parameter Estimation, July 8-12, 1991, Budapest, Hungary. Copyright of this paper remains with IFAC.

§Present address: Koninklijke/Shell-Laboratorium, Amsterdam (Shell Research B.V.), P.O. Box 3003, 1003 AA Amsterdam, The Netherlands

this paper range from 1 to 10 hours, which makes it difficult to measure sufficiently long data sets that are obtained under constant operating conditions. This means that the length of the available data sets will be relatively small.

It has to be stressed that this situation puts fundamental limitations on the achievable accuracy of the identified models.

This paper discusses the identification of a three input two output pilot plant evaporative continuous crystallizer (970 litre), available at the Delft University of Technology. This system was build in order to investigate control strategies for industrial crystallizer systems. A brief sketch of this process is given in section 2. In order to arrive at a compact model suitable for control system design, a specific - three step - identification procedure is presented, resulting in the use of an output error model structure. This procedure, which is discussed in section 3, starts with the estimation of a a high order initial model, which - after model reduction - can be used as an initial estimate for a nonlinear optimization algorithm, minimizing the sum of squared output errors directly in a state space representation. In this identification strategy a similar philosophy is pursued as in Swaanenburg *et al.* (1985), Wahlberg (1987), Backx and Damen (1989), starting the identification procedure with high order modelling. The second step in this strategy, being the model reduction step, will be performed by applying a weighted approximate realization algorithm based on step responses in stead of Markov parameters, showing an improved preformance in the low frequency range (Van Helmont *et al.*, 1990). Special attention will be given to the output error identification algorithm in a state space form. Finally some of the identification results will be presented and will be commented upon.

2 The Crystallization Process

Figure 1 shows a simplified process scheme of the 970 litre pilot plant crystallizer. The crystallizer is operated with the ammonium sulfate water system. The production rate is approximately 3500 kg of crystals per 24 hours. Crystal growth and nucleation (the birth of new crystals) are driven by supersaturation which is created by the evaporation of the solute (water). The crystallizer is fed by an undersaturated feed. Product is isokinetically removed by a discharge tube in order to prevent that the CSD in this flow differs from the one in the crystallizer. The large cilindric zone around the crystallizer (the annular zone) is used to remove small crystals present in the crystallizer. These small particles removed from the crystallizer by the annular zone are heated, send to the dissolving tank and subsequently returned to the crystallizer. This preferential removal of small crystals and their dissolution is an important input to affect the dynamics of the CSD (Randolph and Larson, 1988). Besides this flow, which is referred to as the fines flow Q_f , two other inputs are available for control of the CSD. These are the product flow rate Q_p , which affects the residence time of the crystals in the crystallizer, and the nett heat input to the crystallizer P_{in}. The CSD in the product flow rate is measured by a combined system of a dilution unit and a Malvern 2600C particle sizer. The principle of operation of the particle sizer is based on diffraction. The diffraction pattern of light emitted by a monochromatic laser beam will depend on the size distribution of the particle sample present in this beam. The measured diffraction pattern is used to reconstruct the CSD of the sample. The dilution unit is required to dilute the slurry concentration of a slurry sample below the maximum allowable concentration that can be handled by the Malvern, which is 1 volume percent of particles. The combined system of the dilution unit and the Malvern allows for a sample rate of one CSD measurement per two minutes. For purpose of modelling and control we distinguish two output variables that act as measures for the performance of the CSD: m_3, being a measure for the slurry concentration in the crystallizer, and vol_{25}, the volume fraction of crystals in the size range $(87.5 - 100\mu m)$. More details on the crystallizer and the on-line CSD measurement system can be found in De Wolf (1990) and Jager (1990).

3 Identification Strategy

The identification strategy that will be pursued in this paper is a three step procedure, where in the ultimate step an output error type of model structure is applied. In view of the general class of prediction error model structures, see e.g. Ljung (1987), we will consider a linear, time-invariant finite dimensional system S:

$$S: \; y(t) = G(q)u(t) + H(q)e(t) \qquad (1)$$

with $y(t) \in \mathbb{R}^p$ the p-dimensional output signal, $u(t) \in \mathbb{R}^m$ the m-dimensional input and $e(t) \in \mathbb{R}^p$, with $\{e(t)\}$ a sequence of independent random vectors with zero mean and covariance $Ee(t)e^T(t) = \Lambda$.

$G(q)$ and $H(q)$ represent the proper rational transfer function matrices of the transfers from input to output, and from noise contribution to output. In (1) q denotes the forward shift operator: $qu(t) = u(t+1)$.
A general type of prediction error model set is characterized by:

$$\mathcal{M}: \; y(t) = G(q,\theta)u(t) + H(q,\theta)\varepsilon(t) \quad \theta \in \Theta \subset \mathbb{R}^d \tag{2}$$

with $G(q,\theta)$, $H(q,\theta)$ proper transfer functions of appropriate size, depending on a real-valued parameter vector θ that is varying over a set Θ of admissible values, and $\varepsilon(t) \in \mathbb{R}^p$ the one step ahead prediction error (Ljung, 1987). Throughout this paper we will consider a quadratic type of identification criterion:

$$\Theta_N = arg \min_{\theta \in \Theta} \frac{1}{N} \sum_{t=1}^{N} \varepsilon^T(t,\theta)\varepsilon(t,\theta) \tag{3}$$

The output error model structure is represented by the fixed choice: $H(q,\theta) \equiv I$. The choice for applying an output error model structure is motivated by the fact that - in contrast with the very popular linear regression type of model structures - an output error model approximates a system to be identified much more "balanced" over the whole frequency range. It has been shown by various authors that when considering identification in terms of approximate modelling, linear regression type of models may lead to very bad approximations, or in other terms they often require very high order models in order to provide good approximations, see e.g. Wahlberg and Ljung (1985), Damen *et al.* (1985), Van den Hof and Janssen (1987). However the use of output error type of model structures shows one main disadvantage: the numerical procedure to generate optimal output error models generally relies on nonlinear optimization routines, and consequently problems like occurrence of nonglobal minima, and large computation times (especially in the multivariable situation) have to be dealt with. In this respect it is important to be able to start the output error identification procedure with an accurate initial estimate. The finale step in our strategy will be directed towards the output error optimization; the first two steps will be preformed in order to generate such an accurate initial estimate.
The strategy that we will follow in this paper has close relations with strategies as suggested by Swaanenburg *et al.* (1985), Wahlberg (1987) and Backx (1987), in the sense that an initial model is constructed having a very high order, using simple identification methods based on linear regression models. Subsequently a model reduction procedure will be applied in order to reduce the high dimension of the model to a tractable one. In fact this means that the "approximating part" of the identification procedure has been deferred to a model reduction algorithm. Next we will briefly describe the three different steps in our strategy, which are implemented in the software packages of Aling (1989) and Van den Hof (1991).

a. First Step - High Order *ARX*-Modelling

In the first step an *ARX* model structure is applied:

$$\mathcal{M}_{\mathrm{ARX}} : \mathrm{A(q,\theta)y(t) = B(q,\theta)u(t) + \varepsilon(t)} \tag{4}$$

for $\theta \subset \Theta \subset \mathbb{R}^d$, with $A(q,\theta)$, $B(q,\theta)$ polynomial matrices in the rings $\mathbb{R}^{p\times p}[q^{-1}]$, $\mathbb{R}^{p\times m}[q^{-1}]$, while the entries of the coefficient matrices of the polynomials act as the unknown parameters θ:

$$\begin{aligned} A(q,\theta) &= I + A_1 q^{-1} + A_2 q^{-2} + \cdots + A_{n_a} q^{-n_a} \\ B(q,\theta) &= B_0 + B_1 q^{-1} + B_2 q^{-2} + \cdots + B_{n_b} q^{-n_b} \end{aligned}$$

Unless otherwise stated, the polynomial degrees n_a and n_b will be chosen equal. Applying the *ARX* model structure has the advantage that the identification result (3) can be calculated analytically, and results will be available very fast even for large (multivariable) models having very high polynomial orders. In this first step the polynomial order will be chosen very high in order to guarantuee that the model transfer function $A(z,\theta)^{-1}B(z,\theta)$ is an appropriate representative of the system transfer G(z).

b. Second Step - Model Reduction through Approximate Realization

The model transfer function $A(z,\theta)^{-1}B(z,\theta)$ generated by the high order ARX model obtained in the first step can be realized by an equivalent state space representation:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) & (5) \\ y(k) &= Cx(k) + Du(k) & (6) \end{aligned}$$

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, and $D \in \mathbb{R}^{p\times m}$. Generically this equivalent state space model will have a state space dimension equal to $n = n_a \times p$, see e.g. Gevers and Wertz (1987). In this second step of our identification strategy a model reduction algorithm will be applied to this deterministic model, in order to reduce its dimension to a realistic and tractable value.

Approximate realization based upon a block Hankel matrix built from Markov parameters is a powerful and easy-to-use method for generating reduced order models (Kung, 1979; Damen and Hajdasinski, 1982), being very closely related to the balanced model truncation method of model reduction (Pernebo and Silverman, 1982). A test on the Hankel singular values also provides a simple means for determining the order of the reduced order model to be constructed. However, it is geneally known that this approximate realization scheme shows a preference for the high frequency behaviour of the model, very often leading to a bad approximation in the low frequency region, and specifically to a bad approximation of the static model gain. In order to improve this situation we apply an approximate realization scheme based upon a weighted block Hankel matrix, as equivalently presented in Van Helmont *et al.* (1990). While a standard approximate realization method is based on a block Hankel matrix, $H =$

$$\begin{bmatrix} M(1) & M(2) & M(3) & . & M(N_c) \\ M(2) & M(3) & M(4) & . & M(N_c+1) \\ M(3) & M(4) & . & . & . \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ M(N_r) & M(N_r+1) & . & . & M(N_r+N_c-1) \end{bmatrix}$$

a weighted matrix is constructed according to:

$$H_w = HU \text{ with } U = \begin{bmatrix} I_m & I_m & . & . & I_m \\ 0 & I_m & . & . & . \\ . & 0 & . & . & . \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & . & . & . & I_m \end{bmatrix} \quad (7)$$

leading to $H_w =$

$$\begin{bmatrix} S(1) & S(2) & S(3) & . & S(N_c) \\ S(2) & S(3) & S(4) & . & S(N_c+1) \\ S(3) & S(4) & . & . & . \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S(N_r) & S(N_r+1) & . & . & S(N_r+N_c-1) \end{bmatrix}$$

$$- \begin{bmatrix} S(0) \\ S(1)) \\ S(3) \\ \vdots \\ S(N_r) \end{bmatrix} [I_m I_m .... I_m] \quad (8)$$

with $S(k) := \sum_{i=0}^{k} M(i)$, the step response matrix at time k. Since H_w and H have equal rank, an ordinary approximate realization algorithm based on singular value decomposition can now be applied to the weighted matrix H_w. Because of the specific weighting of the Hankel matrix an improved low frequency behaviour of the aproximate models is obtained.

c. Third Step - Output Error Identification in State Space Form

As mentioned before, the model obtained so far is used as an initial estimate in an output error identification method, dealing with the following model structure:

$$x(k+1) = A(\theta)x(k) + B(\theta)u(k);\ x(0) = x_0(\theta) \quad (9)$$

$$\hat{y}(k,\theta) = C(\theta)x(k) + D(\theta)u(k) \quad (10)$$

$$\varepsilon(k,\theta) = y(k) - \hat{y}(k,\theta) \quad (11)$$

where the elements of the parameter vector θ occur as entries of the coefficient matrices A, B, C, D, and possibly in the initial state vector $x(0)$. Since the output prediction $\hat{y}(k,\theta)$ generally will be nonlinear in the parameter vector θ, solving the identification criterion (3) will require a nonlinear optimization algorithm.

In order to guarantee that the model structure (9) - (11) is uniquely parametrized, the state space form has to be restricted to a uniquely identifiable parametrization, see e.g. Gevers and Wertz (1987). In our present implementation we are able to use three different identifiable parametrizations: the canonical observability form, the pseudo-canonical observability (overlapping) form and the minimal polynomial representation. The first two parametrizations are surveyed in Gevers and Wertz (1987); the latter form is applied and advocated in Backx (1987) and Backx and Damen (1989). In this paper we will show results obtained with the second (overlapping) form. This parametrization has the advantage that, within one single continuous parametrization, it covers almost all models of the specified minimal state space dimension. In order to solve the optimization problem (3), a quasi-Newton method has been implemented as the core of the optimization algorithm. This algorithm employs the following special features:

- The first derivative of the loss function in (3) is analytically determined using the fast algorithm of Van Zee and Bosgra (1982).
- Parametrization of the initial state vector $x_0(\theta)$, which is of importance especially when dealing with short data sets.

- The parametrized state space form contains parameters in $A(\theta)$, $B(\theta)$, $D(\theta)$ and $x_0(\theta)$. Since the prediction error $\varepsilon(k,\theta)$ is a linear function of the parameters in $B(\theta)$, $D(\theta)$ and $x_0(\theta)$, the latter set of parameters can be optimized using linear (one shot) optimization techniques. In the algorithm implemented it is possible, in a very flexible way, to alternatingly optimize nonlinearly with respect to a subset of $A(\theta)$, $B(\theta)$, $D(\theta)$, $x_0(\theta)$, and linearly with respect to a subset of $B(\theta)$, $D(\theta)$, $x_0(\theta)$. This linear optimization of part of the total parameter set has appeared to be of major importance with respect to both speeding up the optimization procedure, and getting out of areas with local minima.

The approach chosen guarantees that we will end the identification procedure with a model having a prespecified minimal state space dimension, being determined during the approximate realization in the second step.

4 Results

The three step identification procedure presented above will be applied to the data of a typical experiment conducted with the 970 litre crystallizer. A three input two output MIMO model (the heat input P_{in}, the fines flow rate Q_f and the product flow rate Q_p being the process inputs and the third moment m_3 and the volume fraction vol_{25} being the outputs) is identified. The experiment lasted 85 hours, of which 68 hours were recorded with input excitation applying PRBS (Pseudo Random Binary Sequence) signals. The sampling rate employed in the experiment was 1 sample per two minutes (the maximum sampling rate of the CSD measurement device). For modelling purposes, a sampling interval of 6 minutes was used, being equal to the clock period of the PRBS chosen. This resulted in 686 samples recorded with PRBS excitation.

The 68 hours of PRBS excitation is relatively short compared to the dynamics of the process. More data could not be obtained because of limitations on the experiment duration. Approximately 85 hours after start-up, incrustation in the crystallizer blocks the slurry circulation and shutdown is required. In order to maximize the amount of data with PRBS excitation, this excitation was started 17 hours after start-up, although at that moment the process was not yet at steady state. Therefore, the initial state vector x(0) is parametrized in the state space model structure that is applied in the third step.

The quality of the estimated model will highly depend on the quality of the data. Therefore, a data preprocessing step is required to reduce the defects in the recorded signals not being a part of the response to the input signals. Such a preprocessing procedure consists of peak shaving, detrending, filtering, decimation, offset correction, scaling and correction for time delays. The raw signals vol_{25} and m_3 as recorded during the experiment are depicted in figure 2. The large peaks are caused by failures of the CSD measurement device and periodic cleaning of the product flow line and the dilution unit. First these peaks were removed using a semi-automatic procedure (De Wolf, 1990) which replaces each peak by a linear interpolation between two samples around the peak. Detrending was not applied as low-frequency disturbances causing trends were not detected. Next, the signals were decimated to the sample interval to be used in the identification. In this decimation step, extra digital filtering was applied in order to prevent aliasing effects and to reduce the measurement noise on the output signals. Decimation was required because the signals were recorded at the maximum frequency allowed by the measurement system. The resulting abundancy of information was used in the repairs of the signals (peak shaving and extra noise reduction by off-line filtering). Finally, the signals were corrected for offset (zero mean) and scaled (variance of 1).

In the first step of the identification procedure, ARX models were estimated for different values of the polynomial orders. An example of the results is depicted in figure 3, which shows the response of the model output vol_{25} on a step applied at input P_{in} for estimated ARX models having polynomial orders ranging from 10 - 50. The responses in this figure show that even for very high orders the static gain of the estimated models does not converge, and remains to be rather uncertain. This problem obviously is caused by the fact that the number of data does not tend to infinity fast enough with increasing polynomial orders. The example shown in figure 3 is characteristic for the different transfers estimated in the model.

Several of the ARX models identified have been used as a basis for the next two steps in the identification strategy. However in view of the results finally obtained, the specific choice for an ARX polynomial order did not appear to be a very critical choice. Results presented in the sequel of this paper are based on ARX models having a polynomial degree of 40.

In the second step approximate realization was ap-

plied to the step response of the estimated ARX model. The singular values of the weighted Hankel matrix (8) in the approximate realization algorithm, gave rise to a choice for a 5th order model. The result of this approximate realization is illustrated in figure 4 showing a comparison of the step response of the original ARX model and of 5th order approximations being obtained by either the standard approximate realization algorithm and by the weighted method. As indicated before, figure 4 shows an improved static gain approximation for the realization algorithm based on the weighted Hankel matrix (indicated by Step).

From earlier MISO identification results obtained for the output signal m_3, it was found that this output can be well described with a first-order state space model. Consequently, this information was used in the choice for appropriate structure indices of the MIMO model to be used as an initial estimate in the third step: optimization in an output error model structure in state space form. The initial state space model, obtained from approximate realization is transformed to a pseudo-canonical form having structure indices 1 and $k-1$, where k is the order of the MIMO state space model resulting from the approximate realization. Figure 5 shows the three step responses related to output vol_{25} of the finally estimated output error model for state space dimensions 4 to 7. In order to show the improved convergence properties of this identification strategy, in relation to the first ARX step, also other model dimensions than the suggetsed value of 5 have been tried out. The figure shows a good convergence of both fast and slow responses in the model if the model dimension is chosen at least equal to 5. It is therefore concluded that a more powerful result is obtained by the output error method as compared to the ARX model. Moreover, equivalent results were found for the other ARX model orders depicted in figure 3, which shows that the addition of the third step in the identification strategy is very essential for arriving at appropriate models. Finally figure 6 shows the simulation result of the identified state space model with $n = 5$ on the available data set. Because of the very restricted length of this data set it was not possible to validate the identification result on an independent data set. For final acceptance of the models obtained for the crystallizer as a basis for controller design, additional experiments will be required in order to provide a means for model validation.

Conclusions

In this paper a three step identification strategy is presented which is applied to data of a pilot plant crystallization process. The avaialable data set of the process is characterized by a low signal to noise ratio, and by a relatively short observation time. In the three step identification strategy, first high order ARX modelling is applied. The resulting model is reduced in order through a modified approximate realization scheme, after which it serves as an initial estimate in an output error identification algorithm, implemented in a state space parametrization. As a model reduction procedure use is made of an approximate ralization algorithm based on a weighted Hankel matrix, showing improved low frequency performance of the approximate model. The results obtained in this paper illustrate that the third step in the procedure (output error optimization) is a necessary step which shows substantial improvements of the identified models compared to the models obtained after the second step.

Acknowledgments

The research of the first author is financially supported by the Netherlands Technology Foundation (STW), AKZO, DOW, DSM, Dupont de Nemours, Rhone Poulenc, Suiker Unie and Unilever. This support is gratefully acknowledged. The authors would like to thank Ronald Spijkerman for assistence in running the identification programs.

References

Aling, H. (1989). *RPEPAC User's Manual.* Report N-304, Lab. Measurement and Control, Departm. Mechanical Engineering and Marine Techn., Delft University of Technology.

Backx, A.C.P.M. (1987). *Identification of an Industrial Process, A Markov Parameter Approach.* Dr. Dissertation, Eindhoven University of Technology, Dept. Electrical Engineering.

Backx, A.C.P.M. and A.A.H. Damen (1989). Identification of industrial MIMO processes for fixed controllers. *Journal A, 30,* Part I: 3-12; Part II: 33-43.

Damen, A.A.H. and A.K. Hajdasinski (1982). Practical tests with different approximate realizations based on the singular value decomposition of the Hankel matrix. In: G.A. Bekey (Ed.), *Identification and System Parameter Estimation,* Proc. 6th IFAC Symp., Arlington, Virginia, pp. 903-908.

Damen, A.A.H., Y. Tomita and P.M.J. Van den Hof (1985). *Equation Error Versus Output Error Methods in System Identification.* Report ER 1985/06, Dept. Electr. Eng., Eindhoven Univ. Techn.

De Wolf, S. (1990). *Modelling, System Identification and Control of an Evaporative Continuous Crystallizer.* Dr. Dissertation, Delft University of Technology, Dept. Mechanical Engineering.

Gevers, M. and V. Wertz (1987). Techniques for the selection of identifible parametrizations for multivariable linear systems. In: C.T. Leondes (Ed.), *Control and Dynamic Systems, 26.* Academic Press, New York, pp. 35-86.

Jager, J. (1990). *Control of Industrial Crystallizers - The Physical Aspects.* Dr. Dissertation, Delft University of Technology.

Kung, S. (1979). A new identification and model reduction algorithm via singular value decomposition. *Proc. 12th Ann. Asilomar Conf. Circuits Syst. Comp.*, pp. 705-714.

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

Pernebo, L. and L.M. Silverman (1982). Model reduction via balanced state space representations. *IEEE Trans. Automat. Contr., AC-27,* 382-387.

Randolph, A.D. and M.A. Larson (1988). *Theory of Particulate Processes.* Academic Press, New York, Second Edition.

Söderström, T. and P. Stoica (1989). *System Identification.* Prentice Hall, Hemel Hempstead, UK.

Swaanenburg, H.A.C., W.M. Schinkel, G.A. van Zee and O.H. Bosgra (1985). Practical aspects of industrial multivariable process identification. In: H.A. Barker and P.C. Young (Eds.), *Identification and System Parameter Estimation.* IFAC Proceedings Series 1985, no. 7, pp. 201-206. Proc. 7th IFAC Symp. Identif. and System Param. Estim., July 1985, York, UK.

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.

Van den Hof, P.M.J. (1991). *DUMSI-Package for Off-Line Multivariable System Identification.* Report Mech. Eng. Systems and Control Group, Delft Univ. Technology. (In preparation).

Van Helmont, J.B., A.J.J. van der Weiden and H. Anneveld (1990). Design of optimal controller for a coal fired Benson boiler based on a modified approximate realisation algorithm. In: R. Whalley (Ed.), *Application of Multivariable System Techniques (AMST90).* Elsevier Publ. Comp, London, pp. 313-320.

Van Zee, G.A. and O.H. Bosgra (1982). Gradient computation in prediction error identification of linear discrete-time systems. *IEEE Trans. Automat. Contr., AC-27,* 738-739.

Wahlberg, B. (1987). *On the Identification and Approximation of Linear Systems.* Ph.D.-Thesis, Dept. Electrical Engineering, Linköping University, Sweden.

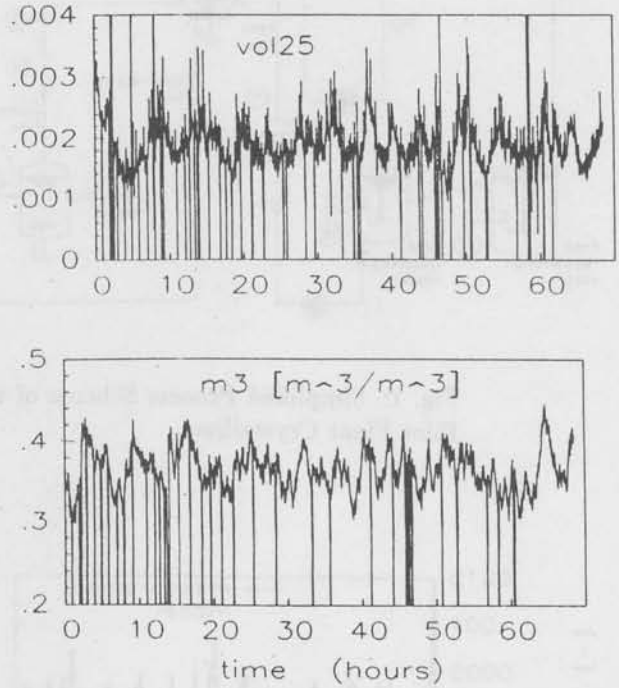


Fig. 2: Measured data for output signals vol_{25} and m_3 before preprocessing.

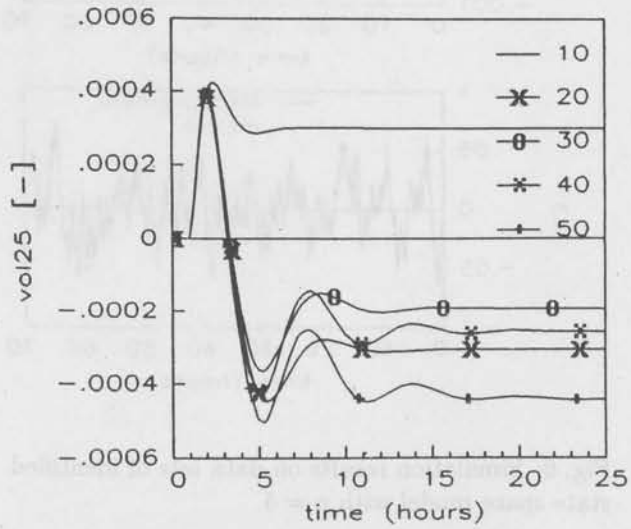


Fig. 3: Step responses of estimated high order *ARX* model of transfer function $P_{in} \rightarrow vol_{25}$ for ARX polynomial orders 10, 20, 30, 40, 50.

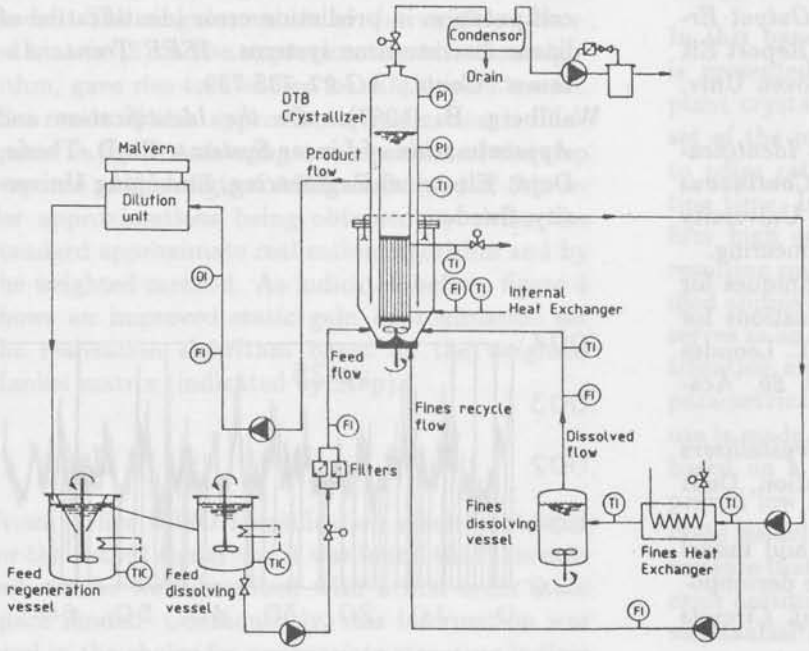


Fig. 1: Simplified Process Scheme of the 970 litre Pilot Plant Crystallizer.

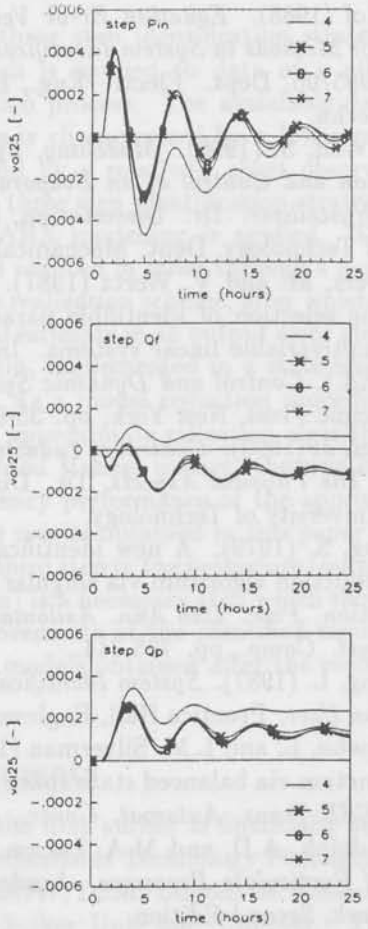


Fig. 5: Step responses of output vol_{25} of identified state space models with output error algorithm for state space dimensions 4 to 7.

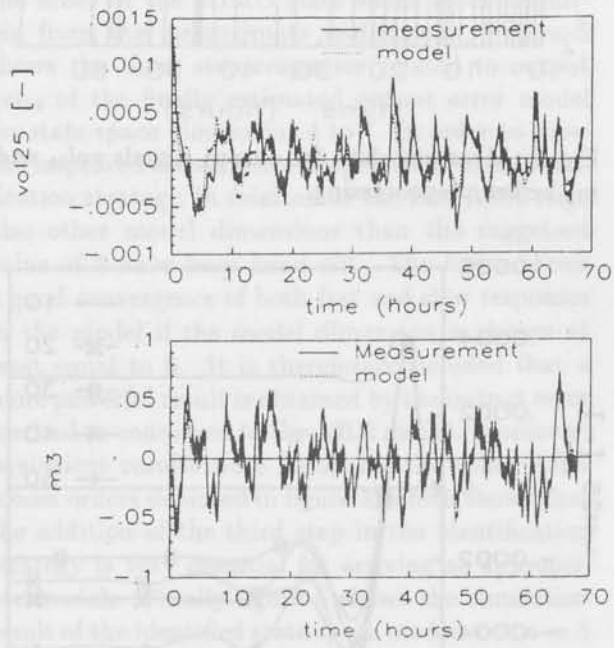


Fig. 6: Simulation results on data sets of identified state space model with $n = 5$.

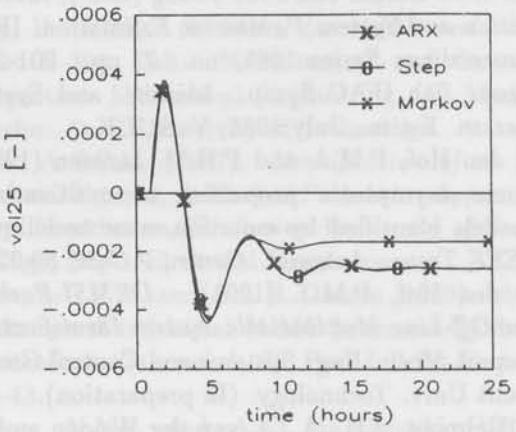


Fig. 4: Step responses of estimated high order *ARX* model of transfer function $P_{in} \rightarrow vol_{25}$ for ARX polynomial order 40, and of approximate realizations with $n = 5$ based on the standard approximate realization algorithm (Markov) and on the weighted approximate realization (Step).

CONTENTS Volume 3, June 1991

TU Delft

Delft University of Technology