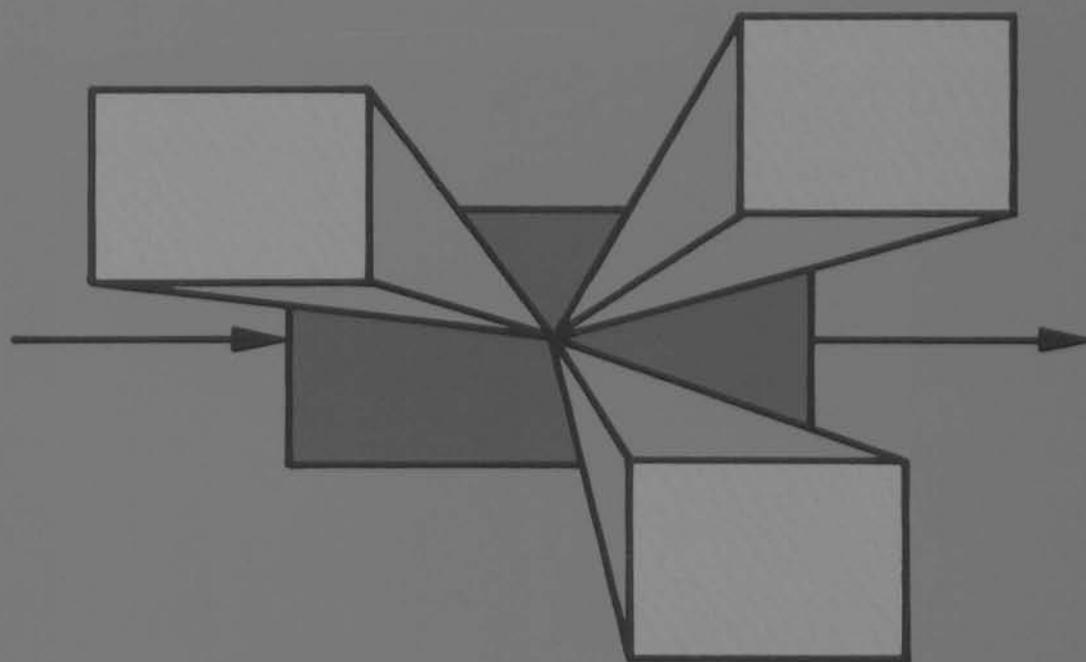


SELECTED
TOPICS
IN

Identification Modelling and Control

Volume 9, 1996

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



Delft University Press

675387
3059167
9R 2995661

IDENTIFICATION, MODELLING AND CONTROL
MODELLING AND CONTROL

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

Edited by O.H. Bosgra, P.M.J. Van den Hof and C.W. Scherer

Volume 6, December 1986



Mechanical Engineering Systems and Control Group
Delft University of Technology

Delft University Press 1986

Contents

Volume 9, December 1996

Maximum likelihood identification: Some practical considerations in real-time identification
P.M.J. Van den Hof and R.A. de Callafon

**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, P.M.J. Van den Hof and C.W. Scherer

Volume 9, December 1996

Mechanical Engineering Systems and Control Group
Delft University of Technology

Delft University of Technology

Mechanical Engineering Systems and Control Group

Delft University of Technology

Mechanical Engineering Systems and Control Group

Delft University of Technology

Mechanical Engineering Systems and Control Group

Delft University of Technology

Mechanical Engineering Systems and Control Group

Delft University of Technology

Mechanical Engineering Systems and Control Group

Delft University of Technology



Published and Distributed by

Delft University Press
Mekelweg 4
2628 CD Delft
The Netherlands
Tel.: (0)15 - 2783254
Telefax: (0)15 - 2781661

By order of

Mechanical Engineering Systems and Control Group
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
Tel.: +31-15-2786400; Telefax: +31-15-2784717
email: e.m.p.arkesteijn@wbmt.tudelft.nl
www: <http://www-mr.wbmt.tudelft.nl/ts>



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. 9-ed. by O.H. Bosgra and
P.M.J. Van den Hof and C.W. Scherer.-ill. Met lit.opg.
ISBN 90-407-1409-6
SISO 656 UDC 531.7 + 681.5 NUGI 841

Cover design by Ruud Schrama

© 1996 Copyright Delft University Press. All rights reserved. No part of this journal may be reproduced,
in any form or by any means, without written permission from the publisher.

Contents

Volume 9, December 1996

Multivariable closed-loop identification: from indirect identification to dual-Youla parametrization <i>P.M.J. Van den Hof and R.A. de Callafon</i>	1
Asymptotic variance expressions for closed-loop identification and their relevance in identification for control <i>M. Gevers, L. Ljung and P.M.J. Van den Hof</i>	9
Analysis of closed-loop identification with a tailor-made parametrization <i>E.T. van Donkelaar, P.M.J. Van den Hof</i>	17
Identification in view of control design of a CD player <i>H.G.M. Dötsch, P.M.J. Van den Hof, O.H. Bosgra and M. Steinbuch</i>	25
Multivariable least squares frequency domain identification using polynomial matrix fraction descriptions <i>R.A. de Callafon, D. de Roover and P.M.J. Van den Hof</i>	31
Model set determination from a batch of plant models <i>T. Zhou and P.M.J. Van den Hof</i>	39
Closed-loop balanced reduction with application to a compact disc mechanism <i>P.M.R. Wortelboer, M. Steinbuch and O.H. Bosgra</i>	47
Alternative parametrization in modelling and analysis of a Stewart platform <i>S. Koekebakker, P.C. Teerhuis and A.J.J. van der Weiden</i>	59
High-performance motion control of a flexible mechanical servomechanism <i>D. de Roover, O.H. Bosgra, F.B. Sperlring and M. Steinbuch</i>	69
Dualization of the internal model principle in compensator and observer theory <i>D. de Roover and O.H. Bosgra</i>	79
Synthesis of a robust iterative learning controller using an H_∞ approach <i>D. de Roover</i>	89
Robust generalized H_2 control for uncertain and linear parametrically varying systems with full block scalings <i>C.W. Scherer</i>	97
Robust linear parametrically varying flight control system design with bounded rates <i>D.M.C. Willemsen, S. Bennani, and C.W. Scherer</i>	105
Parametrically varying flight control system design with full block scalings <i>R.G.E. Njio, C.W. Scherer and S. Bennani</i>	113
CLOSID - A closed-loop system identification toolbox for Matlab <i>P.M.J. Van den Hof, R.A. de Callafon and E.T. van Donkelaar</i>	121
FREQID - Frequency domain identification toolbox for use with Matlab <i>R.A. de Callafon and P.M.J. Van den Hof</i>	129
ORTTOOL - A Matlab toolbox for system identification with generalized orthonormal basis functions <i>P.S.C. Heuberger and P.M.J. Van den Hof</i>	135

Editorial

It is our pleasure to present to you the ninth volume of *Selected Topics in Identification, Modelling, and Control*, giving a report of the ongoing research in our Mechanical Engineering Systems and Control Group.

The current issue again contains a wide variety of subjects, and also shows a number of "new" authors that have contributed to the activities of our research group.

We could like to present Sjirk Koekebakker, who is a Ph.D.-student, working on the modelling and control of a Stewart platform for a motion simulator, in a cooperation project with the SIMONA Research Institute of Delft University of Technology. Within the scope of this project a flight simulator motion system is developed and constructed.

There are two projects in cooperation with the Aerospace Department, in which Dehlia Willemsen and Edwin Njio have worked on their M.Sc.-Theses, dealing with the design of robust control systems for parametrically varying (flight control) systems.

Tong Zhou is a postdoc researcher from Beijing University of Aeronautics and Astronautics, who has joined our group for one year, thanks to a research grant from the Dutch Institute of Systems and Control (DISC). The contribution of Tong reflects his work on uncertainty modelling for batches of dynamical plant models.

Last but not least, we welcome Michel Gevers from the University of Louvain-la-Neuve in Belgium, and Lennart Ljung from Linköping University in Sweden for their contribution to the joint work on closed-loop identification.

This issue is the first one to also contain papers that describe software tools. Three MATLAB toolboxes are presented, all within the area of system identification. As referred to in the respective papers, the corresponding software is available through anonymous ftp from our ftp-site. Interested readers are invited to work with our software, and tell us about their experiences with the tools.

Additional information on the activities of our group, as well as postscript versions of the papers in this and previous volume(s) of our progress report, can be found on our WWW-site: <http://www-mr.wbmt.tudelft.nl/ts>.

For any reactions and discussions on the topics presented, you are welcome to contact one of us.

Finally we would like to wish all our colleagues and friends a happy and prosperous 1997.

Okko Bosgra o.h.bosgra@wbmt.tudelft.nl
Paul Van den Hof p.m.j.vandenhof@wbmt.tudelft.nl
Carsten Scherer c.w.scherer@wbmt.tudelft.nl
Editors

Multivariable closed-loop identification: from indirect identification to dual-Youla parametrization[†]

Paul M.J. Van den Hof and Raymond A. de Callafon[‡]

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: p.m.j.vandenhof@wbmt.tudelft.nl*

Abstract. Classical indirect methods of closed-loop identification can be applied on the basis of different closed-loop transfer functions. Here the multivariable situation is considered and conditions are formulated under which identified approximative plant models are guaranteed to be stabilized by the present controller. Additionally it is shown in which sense the classical indirect methods are generalized by the recently introduced identification method based on the dual-Youla parametrization. For stable controllers the two methods are shown to be basically equivalent to each other.

Keywords. System identification; closed-loop identification; prediction error methods; stability; Youla parametrization; multivariable systems.

1 Introduction

The classical method of indirect identification for handling a closed-loop identification problem is based on the idea of first identifying a closed-loop transfer function, and then calculating the related plant model by using knowledge of the present controller in the loop (see Gustavsson, 1977; Söderström and Stoica, 1989). Attractive properties of this identification scheme are that the method does not suffer from bias effects due to a noise correlation with the input signal, as the input signal for identification is taken to be an external reference signal. The critical part of the indirect identification is the construction of the (open-loop) plant model in the second step, based on the estimated closed-loop transfer. However, if the resulting plant model is not limited in model order, this construction can be done exactly provided that the controller is known and the appropriate closed-loop transfer function has been identified. In this sense the ques-

tion which transfer is "appropriate" is determined - among other things - by the input/output dimensions of the plant, and the location of the external excitation signal.

In recent years several new ideas concerning closed-loop identification of approximate models have been presented, most of them directed towards the ability to identify approximate models of the open-loop plant on the basis of closed-loop data, while the asymptotic bias distribution is not dependent on the noise and thus explicitly tunable by the designer, see e.g. Hansen and Franklin (1988), Lee *et al.* (1992), Van den Hof and Schrama (1993) and Van den Hof *et al.* (1995) as summarized in the survey paper Van den Hof and Schrama (1995). Most of these schemes have been developed in view of the ability to tune the asymptotic bias distribution in order for the identified models to particularly reflect those dynamic aspects of the plant that are most relevant for consecutive model-based control design. One of the newly handled methods is based on a dual-Youla parametrization of the open-loop plant (Hansen and Franklin, 1988; Schrama, 1991; Lee *et al.*, 1992), and this method is suggested to be particularly attractive because of its guarantee that identified (approximate) plant models are guaranteed to be stabilized

[†]This paper is presented at the 35th IEEE Conference on Decision and Control, 11-13 December 1996, Kobe, Japan. Copyright of this paper remain with IEEE.

[‡]The work of Raymond de Callafon is financially supported by the Dutch Systems and Control Theory Network.

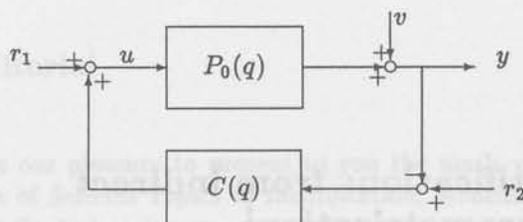


Fig. 1: Closed-loop configuration.

by the present controller.

In this paper we start by summarizing some aspects and results related to the classical indirect identification scheme, particularly addressing the question under which conditions multivariable plant models can be identified. Next it will be shown under which conditions identified plant models are guaranteed to be stabilized by the present controller, and in which sense this classical scheme can be considered as a special -simple- case of the recently used identification in the dual-Youla parametrization.

2 System configuration

The system configuration that will be considered in this paper is sketched in figure 1. P_0 and C are linear time-invariant finite-dimensional but not necessarily stable multivariable transfer functions. The input and output dimensions are determined by $u(t), r_1(t) \in \mathbb{R}^m$, $y(t), r_2(t) \in \mathbb{R}^p$. v is a noise disturbance signal, while r_1, r_2 are external signals that can be either reference (tracking) signals or external disturbances, being uncorrelated to each other and to v .

A particular combination of external signals will be denoted by

$$r(t) := r_1(t) + C(q)r_2(t). \quad (1)$$

The relevant closed-loop transfer functions in the system configuration are reflected by

$$\mathbb{T}(P_0, C) = \begin{bmatrix} P_0 \\ I \end{bmatrix} [I + CP_0]^{-1} \begin{bmatrix} C & I \end{bmatrix} \quad (2)$$

being the mapping from the signals $\begin{bmatrix} r_2 \\ r_1 \end{bmatrix} \rightarrow \begin{bmatrix} y \\ u \end{bmatrix}$.

For notational purposes the following notation for the elements of $\mathbb{T}(P_0, C)$ will be employed:

$$\mathbb{T}(P_0, C) = \begin{bmatrix} T_0 & G_0 \\ Q_0 & S_0 \end{bmatrix}. \quad (3)$$

with

$$T_0 = P_0[I + CP_0]^{-1}C$$

$$\begin{aligned} G_0 &= P_0[I + CP_0]^{-1} \\ Q_0 &= [I + CP_0]^{-1}C \\ S_0 &= [I + CP_0]^{-1}. \end{aligned}$$

It is a standard result from stability theory that the considered closed-loop system is internally stable if and only if $\mathbb{T}(P_0, C) \in \text{IRH}_\infty$, with IRH_∞ the space of real rational transfer functions that are analytic in $z \geq 1$.

As additional notation, I_m will refer to the $m \times m$ identity matrix, and $\det_{\mathbb{R}(z)}(\cdot)$ is the determinant over the field of rational functions in z .

3 Indirect Identification

3.1 Standard approach - scalar situation

The classical method of indirect identification is composed of two steps. For this moment we will just sketch a particular situation in the scalar case.

- (1) Identify the transfer function G_0 from r_1 to y ; this can e.g. be done by applying any of the standard prediction error methods (Ljung, 1987). Note that this identification problem is principally an 'open-loop' type of problem provided that the external signal r_1 is uncorrelated to the noise disturbance term v . The identified model of G_0 is denoted as \hat{G}
- (2) Reconstruct an open loop plant model from the estimated closed-loop transfer function \hat{G} , using knowledge of the controller C .

The second step of this procedure involves the construction of \hat{P} from an available estimate \hat{G} , by solving the equation:

$$\hat{G} = \frac{\hat{P}}{1 + C\hat{P}}. \quad (4)$$

An exact solution for \hat{P} follows by taking

$$\hat{P} = \frac{\hat{G}}{1 - C\hat{G}} \quad (5)$$

which can be calculated when the controller C is known.

When the model \hat{G} is identified using a least-squares output error criterion, i.e.

$$\epsilon(t, \theta) := y(t) - G(q, \theta)r_1(t)$$

and $\hat{G} = G(q, \hat{\theta})$ with $\hat{\theta} := \arg \min_{\theta} \bar{E}\epsilon(t, \theta)^2$, the asymptotic bias-distribution (Ljung, 1987) in the plant model estimate is characterized by:

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{P_0}{1 + CP_0} - \frac{P(\theta)}{1 + CP(\theta)} \right|^2 \Phi_{r_1} d\omega \quad (6)$$

provided that the exact relation (5) is used to construct \hat{P} on the basis of \hat{G} .

One of the problems that is known to occur in an indirect identification approach, is that the order of the identified plant model is not under control. This means that when calculating (5), the order of \hat{P} will be determined by the order n_G of \hat{G} and the order n_C of C , and will generically equal $n_G + n_C$. Limiting the model order to a prespecified value, requires either an additional model reduction step, or the construction of an approximate solution to the equation (4) where the model order of \hat{P} is fixed. However in this latter situation it is not clear how to "solve" this equation properly.

3.2 Indirect identification from closed-loop transfer functions - multivariable case

Actually all four different transfer functions that are present in $\mathbb{T}(P_0, C)$ can be used for identification in the first step of an indirect identification scheme. Dependent on the particular experimental situation, an identifier may have preferences of identifying a particular transfer. This can e.g. be essentially influenced by the possibility of adding external excitation signals at particular locations in the loop (either on the setpoint or on the output of the controller). We will now summarize the possibilities of using any of the four transfer functions, while considering the multivariable situation.

Proposition 3.1 Consider any one of the four transfer functions T_0, G_0, Q_0 , or S_0 to be identified in the first stage of an indirect identification scheme, providing identified models $\hat{T}, \hat{G}, \hat{Q}$, or \hat{S} . Then

(a) $\hat{T} = \hat{P}[I_m + C\hat{P}]^{-1}C$ implies

$$\hat{P} = \hat{T}(I_p - \hat{T})^{-1}C^\dagger \quad (7)$$

under the condition that $p \geq m$ and C has a right inverse C^\dagger .

(b) $\hat{G} = \hat{P}[I_m + C\hat{P}]^{-1}$ implies

$$\hat{P} = \hat{G}[I_m - C\hat{G}]^{-1}. \quad (8)$$

(c) $\hat{Q} = [I_m + C\hat{P}]^{-1}C$ implies that

$$\hat{P} = C^{-1}[(\hat{Q}C^{-1})^{-1} - I_m] \quad (9)$$

under the condition that $p = m$ and $\det_{\mathbb{R}(z)} C \neq 0$.

(d) $\hat{S} = [I_m + C\hat{P}]^{-1}$ implies that

$$\hat{P} = C^\dagger[\hat{S}^{-1} - I_m] \quad (10)$$

under the condition that $p \leq m$ and C has a left inverse C^\dagger .

In the above expressions it is presumed that $\mathbb{T}(\hat{P}, C)$ is well defined.

Proof: Follows by straightforward manipulations of the expressions. \square

It has to be noted that there is only one transfer function (\hat{G}) that provides a unique solution for the related open loop plant model without any conditions on input/output dimensions and controller. For the other transfer functions restrictions apply. Note also that in the scalar case $m = p = 1$, all four transfers can be used without any restrictions.

When taking a look at the relation with available external excitation signals the following can be stated:

- When r_1 is available from measurements (additional to u and y) then one can use $\hat{S}(r_1 \rightarrow u)$ or $\hat{G}(r_1 \rightarrow y)$ and by choosing \hat{G} no restrictions apply.
- When r_2 is available from measurements, then one can use $\hat{T}(r_2 \rightarrow y)$ or $\hat{Q}(r_2 \rightarrow u)$ and one has to face the restrictions $p \geq m$ or $p = m$.

In the second situation considered it can be an alternative to first construct the signal $r(t) = C(q)r_2(t)$ and then using $r(t)$ as if it were added to the loop at the location of r_1 . In this way, one can avoid the dimensional restrictions as mentioned above.

The fact that a unique plant model \hat{P} can be constructed from either of the equations (7)-(10) does not imply that this plant model will be guaranteed to be proper. This will depend on the properties of the estimated closed-loop transfer and of the controller. Properness of \hat{P} is e.g. guaranteed for (8) whenever \hat{G} is proper and $\lim_{|z| \rightarrow \infty} C\hat{G} = 0$, being the commonly considered situation in indirect identification.

4 Stability of controlled models

In this section the question will be addressed under which conditions a plant model \hat{P} that is identified by an indirect identification as described before, will be - a priori - guaranteed to be stabilized by the controller C . To this end the following standard results from stability theory will be exploited.

Proposition 4.1 Consider any linear, time-invariant, finite-dimensional plant P and controller C .

- Let $C \in \text{IRH}_\infty$. Then $\mathbb{T}(P, C) \in \text{IRH}_\infty$ if and only if $P(I + CP)^{-1} \in \text{IRH}_\infty$.
- Let $m = p$ and let C be invertible and satisfy $C^{-1} \in \text{IRH}_\infty$. Then $\mathbb{T}(P, C) \in \text{IRH}_\infty$ if and only if $(I + CP)^{-1}C \in \text{IRH}_\infty$.

- (c) Let $\text{rank}_{\mathbb{R}(z)}(P) = \min(m, p)$, and the Moore-Penrose inverse $P^\dagger \in \mathbb{RH}_\infty$. Then $\mathbb{T}(P, C) \in \mathbb{RH}_\infty$ if and only if $P(I + CP)^{-1} \in \mathbb{RH}_\infty$.

Proof: Part (a) is proven in e.g. Zhou *et al.* (1996). For parts (b) and (c) necessity is obvious. To prove sufficiency for (b), consider $(I + CP)^{-1}C \in \mathbb{RH}_\infty$, so $(I + CP)^{-1}CC^{-1} = (I + CP)^{-1} \in \mathbb{RH}_\infty$. As $P(I + CP)^{-1}C + (I + CP)^{-1} = I$ it follows that $P(I + CP)^{-1}C \in \mathbb{RH}_\infty$ and by postmultiplication of C^{-1} also that $P(I + CP)^{-1} \in \mathbb{RH}_\infty$. Sufficiency for (c) can be shown along similar lines, distinguishing between the situations $p \geq m$, where P^\dagger is a left inverse, and $p \leq m$ when P^\dagger is a right inverse. \square

When applying these results to identified models obtained from indirect identification the following results are direct.

Corollary 4.2 Consider identified models \hat{G} and \hat{Q} of the related closed-loop transfer functions G_0 and Q_0 .

- (a) If C is stable then the plant model estimate (8) is stabilized by C if and only if \hat{G} is stable.
- (b) If $m = p$ and C^{-1} is stable then the plant model estimate (9) is stabilized by C if and only if \hat{Q} is stable.

Particularly, a plant model obtained by indirect identification from estimating the closed-loop transfer function G_0 , will be guaranteed to be stabilized by C in the case that C is stable. The only restriction that the estimate \hat{G} has to satisfy for this result to hold, is that \hat{G} should be stable. Since the closed-loop system is stable, this condition will be naturally satisfied by any sensible identification method.

It would be tempting to formulate a result similar to (a) without any condition on the stability of C or on input/output dimensions. However this will lead to more complex restrictions on \hat{G} as shown next.

Corollary 4.3 Consider a model \hat{G} of the related closed-loop transfer function G_0 , with

$$\text{rank}_{\mathbb{R}(z)}(\hat{G}) = \min(m, p),$$

and satisfying

$$[I_m - C\hat{G}]\hat{G}^\dagger \text{ is stable} \quad (11)$$

where \hat{G}^\dagger is the Moore-Penrose inverse. Then the plant model estimate (8) is stabilized by C if and only if \hat{G} is stable.

Proof: The result follows by manipulation of the expressions in Proposition 4.1(c). \square

When the controller is not stable an additional restriction (11) has to be considered. This constraint on \hat{G} can not simply be incorporated in a parametrization of the closed-loop transfer G_0 to be used during identification. A solution to this problem does exist, as shown in the forthcoming sections.

The stability results shown above, suggest that there is a relationship between these indirect identification methods, and the approach of using a dual-Youla parametrization of all plants that are stabilized by the given controller. This relation is pursued in the next sections.

5 Identification in the dual-Youla form

The Youla-parametrization parametrizes for a given plant $P_0 \in \mathbb{RH}_\infty$ the set of all controllers $C \in \mathbb{RH}_\infty$ that stabilize P_0 . In the dual-Youla parametrization, a similar mechanism is used, but now the set of all plants is considered that is stabilized by a given controller.

In order to formulate this parametrization, the concept of coprime factorizations over \mathbb{RH}_∞ is required. A pair of stable transfer functions $N, D \in \mathbb{RH}_\infty$ is a right coprime factorization (rcf) of P_0 if $P_0 = ND^{-1}$ and there exist stable transfer functions $X, Y \in \mathbb{RH}_\infty$ such that $XN + YD = I$. This implies that two factors are coprime if there are no unstable canceling zeros in the factorization.

Proposition 5.1 (Desoer *et al.* (1980)) Let P_x with rcf (N_x, D_x) be any auxiliary model that is stabilized by the controller C with rcf (N_c, D_c) . Then a plant P_0 is stabilized by C if and only if there exists an $R \in \mathbb{RH}_\infty$ such that

$$P_0 = [N_x + D_c R][D_x - N_c R]^{-1}. \quad (12)$$

For a given plant P_0 , the related dual-Youla parameter $R = R_0$ is given by

$$R_0 = D_c^{-1}[I + P_0 C]^{-1}(P_0 - P_x)D_x. \quad (13)$$

With this parametrization the original system configuration can be resketched into the alternative form as presented in figure 2. In this dual-Youla form the signals $x(t)$ and $z(t)$ are determined by

$$z(t) = (D_c + P_x N_c)^{-1}[y(t) - P_x(q)u(t)] \quad (14)$$

$$x(t) = (D_x + C N_x)^{-1}[r_1(t) + C(q)r_2(t)] \quad (15)$$

while K_0 is given by

$$K_0 = D_c^{-1}(I + P_0 C)^{-1} \quad (16)$$

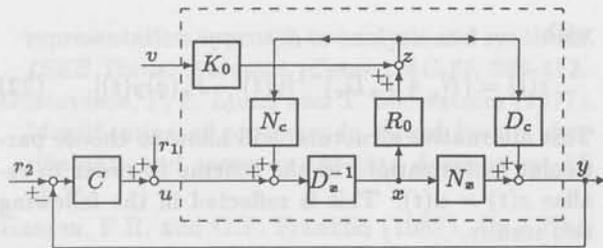


Fig. 2: Dual Youla-representation of the data generating system.

see e.g. Van den Hof and Schrama (1995). In view of the identification problem, one is dealing with the relation

$$z(t) = R_0(q)x(t) + K_0(q)v(t) \quad (17)$$

where the important mechanism is that both signals z and x can be reconstructed from available data y, u, r and by using knowledge of the controller C and of just *any* auxiliary model P_x that is stabilized by C . Moreover as it appears from (15) the signal x is uncorrelated with the noise v , and so relation (17) points to an "open-loop" identification problem of identifying R_0 on the basis of measurement data z, x .

One of the properties of this identification approach is that any identified stable model \hat{R} of R_0 will yield an open-loop plant model

$$\hat{P} = [N_x + D_c \hat{R}][D_x - N_c \hat{R}]^{-1} \quad (18)$$

that is guaranteed to be stabilized by C , because of the dual-Youla parametrization.

A property of this dual-Youla identification method is - similar to the situation of the indirect approach - that the model order of the identified open-loop plant model is not under control. Because of the relation (18), an identified transfer \hat{R} with a specific model order, will lead to an open-loop plant model that has an increased model order, that incorporates the order of the controller and the order of the auxiliary model P_x .

6 Indirect identification as a special case of the dual-Youla method

The question occurs whether the identification of R_0 in the dual-Youla situation is equivalent to the identification of a closed-loop transfer function as present in the first step of an indirect identification scheme. A number of special cases will be pointed out.

Proposition 6.1 *If C is stable then there exists a choice for P_x and right coprime factorizations of C*

and P_x such that in the dual-Youla form:

$$\begin{aligned} R_0 &= G_0 \\ z(t) &= y(t) \\ x(t) &= r(t) \end{aligned}$$

and consequently identification of the dual-Youla parameter is identical to identification according to the indirect method (8) on the basis of \hat{G} .

Proof: Since C is stable, one may choose $N_c = C$, $D_c = I$, $N_x = 0$ and $D_x = I$, taking into account that the model $P_x = 0$ is stabilized by a stable controller. The result follows by substitution in the appropriate expressions. \square

It appears that for stable controllers, the dual-Youla identification method is actually equivalent to an indirect identification on the basis of the transfer $r_1 \rightarrow y$ (G_0). A similar result can be formulated for the indirect identification through the transfer $r_2 \rightarrow y$ (T_0).

Proposition 6.2 *If C is stable then there exists a choice for P_x and right coprime factorizations of C and P_x such that in the dual-Youla form:*

$$\begin{aligned} R_0 &= T_0 \\ z(t) &= y(t) \\ x(t) &= r(t) \end{aligned}$$

and consequently identification of the dual-Youla parameter is identical to identification according to the indirect method (7) on the basis of \hat{T} .

Proof: The result follows by choosing $N_c = C$, $D_c = I$, $N_x = 0$ and $D_x = C$, and by substituting this in the appropriate expressions. \square

The closed-loop transfer functions considered in the two propositions above are transfers towards the closed-loop output signal $y(t)$. The question now occurs whether the two other transfer function (Q_0 and S_0) can be considered in a similar way. This appears to be less trivial than expected, most importantly because they are transfers towards the closed-loop input signal $u(t)$. As a consequence, the choices of particular factorizations should be made in such a way that this results in $z(t) = u(t)$. Considering the general expression for $z(t)$ in (14) this seems not possible. A solution for this problem appears to be in considering a dual-Youla parametrization based on the controllers inverse, which is discussed in the next section.

With respect to the asymptotic bias distribution, as indicated in (6) for the indirect method, it is shown

in Lee *et al.* (1992) and Van den Hof and Schrama (1995) that for the dual-Youla method, the corresponding expression is (for the SISO-case):

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{P_0}{1 + CP_0} - \frac{P(\theta)}{1 + CP(\theta)} \right|^2 \frac{\Phi_{r_1}}{|D_c|^2} d\omega$$

which is similar to (6), except for an additional weighting with D_c . In case C is stable, one can always choose $D_c = 1$ leading to equal expressions for both methods. Note that for unstable C the model sets in the two approaches will be slightly different if in the indirect method one does not take account of the parametrization constraint (11).

7 A dual-Youla parametrization on the basis of C^{-1}

In this section attention will be limited to the situation that $m = p$ and controller and plant can be inverted, i.e. they have full rank over $\mathbb{R}(z)$.

Lemma 7.1 Consider the situation $m = p$ and P_0 and C invertible. Then $\mathbb{T}(P_0, C) \in \mathbb{RH}_{\infty}$ if and only if $\mathbb{T}(P_0^{-1}, C^{-1}) \in \mathbb{RH}_{\infty}$.

Proof: By simple manipulations it can be shown that $\mathbb{T}(P_0^{-1}, C^{-1})$ is equal to a permuted version of the original $\mathbb{T}(P_0, C)$. \square

A dual-Youla parametrization can now be formulated on the basis of the inverse controller C^{-1} .

Proposition 7.2 Let P_x with rcf (N_x, D_x) be any auxiliary model that is stabilized by the controller C^{-1} with rcf $D_c N_c^{-1}$. Then a plant P_0 is stabilized by C if and only if there exists an $R \in \mathbb{RH}_{\infty}$ such that

$$P_0 = [D_x - D_c R][N_x + N_c R]^{-1}. \quad (19)$$

Proof: The proof follows by parametrizing P_0^{-1} in a dual-Youla parametrization, and applying lemma 7.1.

Under the conditions of the proposition, it follows that for a given plant P_0 , the related R is given by

$$R = R_0 = D_c^{-1}(I + P_0 C)^{-1}(D_x - P_0 N_x) \quad (20)$$

and the system's equations become:

$$\begin{aligned} y(t) &= (D_x - D_c R_0)x(t) + (I + P_0 C)^{-1}v(t) \\ u(t) &= (N_x + N_c R_0)x(t) - C(I + P_0 C)^{-1}v(t). \end{aligned}$$

Based on these latter equations one can extract R_0 by:

$$z(t) = R_0(q)x(t) + K_0(q)v(t) \quad (21)$$

with

$$z(t) = (N_c + P_x D_c)^{-1}[u(t) - P_x(q)y(t)]. \quad (22)$$

This alternative structure, will allow to choose particular factorizations in the scheme in order to realize $z(t) = u(t)$. This is reflected in the following two results.

Proposition 7.3 Let $p = m$ and let C^{-1} be stable. Then there exist choices for P_x and right coprime factorizations of C and P_x such that in the dual-Youla form of this section:

$$\begin{aligned} \text{either } R_0 &= Q_0 \quad \text{or } R_0 = S_0 \\ z(t) &= u(t) \\ x(t) &= r(t) \end{aligned}$$

and consequently identification of the dual-Youla parameter is identical to identification according to the indirect method (9) on the basis of \hat{Q} or (10) on the basis of \hat{S} .

Proof: The result follows by choosing $N_c = I$, $D_c = C^{-1}$, $N_x = 0$ and either $D_x = I$ (for the case of Q_0) or $D_x = C^{-1}$ (for the case of S_0), and by substituting this in the appropriate expressions. \square

This shows that the two closed-loop transfer functions that are related to the input signal u can also be directly estimated in a dual-Youla framework, provided that we restrict attention to the square situation ($p = m$) and to a stably invertible controller.

8 Conclusions

The classical indirect method for closed-loop identification and the recently discussed approach based on the dual-Youla parametrization appear to be closely related to each other. In the situation of a stable controller, the two methods are algebraically equivalent. In the situation of an unstable controller, the dual-Youla method provides models that are guaranteed to be stabilized by the controller, which goes beyond the capabilities of a simple indirect method. Several relations are given between the two approaches, showing that the dual-Youla method is actually a generalization of the classical indirect approach.

Both approaches share the problem that it is not simply possible to control the model order of the identified plant model.

References

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.

Gustavsson, I., L. Ljung and T. Söderström (1977). Identification of processes in closed loop - identifiability and accuracy aspects. *Automatica*, 13, 59-75.

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

Lee, W.S., B.D.O. Anderson, R.L. Kosut and I.M.Y. Mareels (1992). On adaptive robust control and control-relevant system identification. *Proc. 1992 American Control Conf.*, Chicago, IL, USA, pp. 2834-2841.

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

Schrama, R.J.P. (1991). An open-loop solution to the approximate closed loop identification problem. In: C. Banyasz & L. Keviczky (Eds.), *Identification and System Parameter Estimation 1991*. IFAC Symposia Series 1992, No. 3, pp. 761-766. Sel. Papers 9th IFAC/IFORS Symp., Budapest, July 8-12, 1991.

Söderström, T. and P. Stoica (1989). *System Identification*. Prentice-Hall, Hemel Hempstead, U.K.

Van den Hof, P.M.J. and R.J.P. Schrama (1993). An indirect method for transfer function estimation from closed loop data. *Automatica*, 29, 1523-1527.

Van den Hof, P.M.J., R.J.P. Schrama, R.A. de Callafon and O.H. Bosgra (1995). Identification of normalized coprime plant factors from closed-loop experimental data. *Europ. J. Control*, 1, 62-74.

Van den Hof, P.M.J. and R.J.P. Schrama (1995). Identification and control - closed-loop issues. *Automatica*, 31, 1751-1770.

Zhou, K., J.C. Doyle and K. Glover (1996). *Robust and Optimal Control*. Prentice-Hall, Englewood Cliffs, NJ.

... which plays a central role in many contributions in the area of identification for control. This book has been treated already on the basis of two manuscripts in the form of a "control-relevant" introduction of the book and chapters on (Söderström, 1987; Gustavsson, 1988; Lee et al., 1992; Van den Hof and Schrama, 1993; Söderström, 1994; Van den Hof and Schrama, 1995). Recently it has been argued in (Söderström et al., 1995) that for a particular class of control design methods, also from a practical point of view, closed-loop experiments are preferred over open-loop ones.

... Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden.
... Address to whom correspondence should be addressed.

... reasons for closed-loop relevance in identification

Paul M.J. Van den Hof¹
and
R.J.P. Schrama²

... are the preferred for models that are identified using indirect methods, including the classical direct method, indirect methods, employing multivariable based on the dual Youla/Kucera representation, as well as the open-loop situation, as for subsequent model-based control design. Several experimental situations in identification and control of the resulting model-based controllers, closed-loop identification, asymptotic variance expression and control design.

In this paper we will first present the asymptotic variance expressions for identified models based on several different closed-loop identification methods, including the recently introduced indirect methods using a regular linear model representation (Söderström, 1995; Van den Hof et al., 1995) and the method employing a so-called dual Youla/Kucera representation (Gustavsson and Franklin, 1988; Schrama, 1993; Lee et al., 1992). The results for the classical direct method (Ljung, 1987) are extended to also include variance expressions for the estimated noise model, while they are shown to remain the same for the traditional direct method indirect method.

These variance expressions are compared to related expressions for the open-loop situation, and requirements are shown for the variance of resulting model-based controllers by several types of controller design.

2 Preliminaries

It will consider the closed-loop configuration as depicted in Fig. 1, where G_c and G are linear time-

Asymptotic variance expressions for closed-loop identification and their relevance in identification for control

Michel Gevers[†], Lennart Ljung[§] and Paul M.J. Van den Hof[#]

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: p.m.j.vandenhof@wbmt.tudelft.nl*

Abstract. Asymptotic variance expressions are analysed for models that are identified on the basis of closed-loop data. The considered methods comprise the classical 'direct' and 'indirect' method, as well as the more recently developed indirect methods, employing coprime factorized models and model parametrizations based on the dual Youla/Kucera parametrization. The variance expressions are compared with the open-loop situation, and evaluated in terms of their relevance for subsequent model-based control design. Additionally it is specified what is the optimal experimental situation in identification (open-loop or closed-loop), in view of the variance of the resulting model-based controller.

Keywords. System identification; closed-loop identification; asymptotic variance expressions, prediction error methods; model-based control design.

1 Introduction

When identifying dynamic models for the specific purpose of subsequent model-based control design it is argued that a closed-loop experimental setup during the identification experiments supports the construction of an identified model that is particularly accurate in that frequency region that is relevant for the control design. This mechanism which plays a major role in many contributions in the area of "identification for control", has been motivated mainly on the basis of bias considerations in the form of a "control-relevant" distribution of the bias over frequency (Schrama, 1992; Gevers, 1993; Lee *et al.*, 1993; Van den Hof and Schrama, 1995). Recently it has been shown in Hjalmarsson *et al.* (1996), that for a particular class of control design methods, also from a variance point of view closed-loop experiments are preferred over open-loop ones.

In this paper we will first present the asymptotic variance expressions for identified models based on several different closed-loop identification methods, including the recently introduced indirect methods using a coprime factor model representation (Schrama, 1992; Van den Hof *et al.*, 1995) and the method employing a so-called dual Youla/Kucera parametrization (Hansen and Franklin, 1988; Schrama, 1992; Lee *et al.* 1993). The results for the classical 'direct' method (Ljung, 1993) are extended to also include variance expressions for the estimated noise model, while they are shown to remain the same for the mentioned alternative indirect methods.

These variance expressions are compared to related expressions for the open-loop situation, and consequences are shown for the variance of resulting model-based controllers for several types of controller designs.

2 Preliminaries

We will consider the closed-loop configuration as depicted in Fig. 1, where G_0 and C are linear time-

[†]CESAME, Bâtiment Euler, Louvain University, B-1348 Louvain-la-Neuve, Belgium.

[§]Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden.

[#]Author to whom correspondence should be addressed.

invariant, possibly unstable, finite dimensional systems, while C is a stabilizing controller for G_0 ; e is a white noise process with variance λ_0 , and H_0 a stable and stably invertible monic transfer function. Signals r_1 and r_2 are external reference signals that are possibly available from measurements. For purpose of efficient notation, we will often deal with the signal

$$r(t) := r_1(t) + C(q)r_2(t)$$

being the result of external excitation through either r_1 or r_2 .

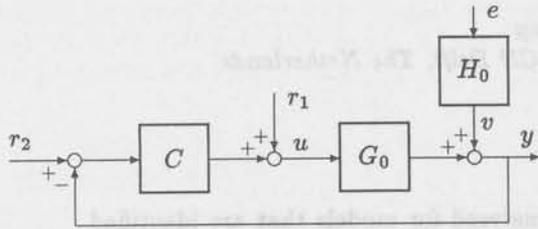


Fig. 1: Closed-loop configuration.

Additionally we will denote:

$$u(t) = u^r(t) + u^e(t) \quad (1)$$

with

$$u^r(t) := S_0(q)r(t), \quad (2)$$

$$u^e(t) := -C(q)S_0(q)H_0(q)e(t), \quad (3)$$

where the sensitivity function S_0 is given by $S_0(q) := \frac{1}{1 + C(q)G_0(q)}$. The signals $u^r(t)$ and $u^e(t)$ refer to those parts of the input signal that originate from, respectively, r and e . For the corresponding spectra it follows that

$$\Phi_u = \Phi_u^r + \Phi_u^e \quad (4)$$

with

$$\Phi_u^r = |S_0|^2 \Phi_r \quad \text{and} \quad (5)$$

$$\Phi_u^e = |CS_0|^2 \Phi_v. \quad (6)$$

In order to simplify notation the arguments q and $e^{i\omega}$ will be omitted when there is no risk of ambiguity. We will consider parametrized models $G(q, \theta)$ for G_0 and $H(q, \theta)$ for H_0 with $\theta \in \Theta$, and in accordance with Ljung (1987) we will use the expressions $S \in \mathcal{M}$ to refer to the situation that there exists a $\theta_o \in \Theta$ such that $G(q, \theta_o) = G_0(q)$ and $H(q, \theta_o) = H_0(q)$; $G_0 \in \mathcal{G}$ will indicate that there exists a $\theta_o \in \Theta$ such that $G(q, \theta_o) = G_0(q)$ only.

The variance expressions that are considered in this paper are asymptotic in both n (model order) and N (number of data), while n/N is supposed to tend to 0, as in the standard framework of Ljung (1987).

3 Direct identification

The direct method of closed-loop identification is characterized by

$$\hat{\theta}_N = \arg \min_{\theta} \frac{1}{N} \sum_{t=0}^{N-1} \varepsilon(t, \theta)^2 \quad (7)$$

with

$$\varepsilon(t, \theta) = H(q, \theta)^{-1} [y(t) - G(q, \theta)u(t)]. \quad (8)$$

For this direct identification method, an expression for the asymptotic variance of the transfer function estimate can be given for the situation that $S \in \mathcal{M}$, and both plant model and noise are estimated. In this case (Ljung, 1987):

$$\text{cov} \begin{pmatrix} \hat{G}(e^{i\omega}) \\ \hat{H}(e^{i\omega}) \end{pmatrix} \sim \frac{n}{N} \Phi_v(\omega) \cdot \begin{bmatrix} \Phi_u(\omega) & \Phi_{ue}(\omega) \\ \Phi_{ue}(\omega) & \lambda_0 \end{bmatrix}^{-1}. \quad (9)$$

With the relation $\Phi_{ue} = -CS_0H_0\lambda_0$ and using the fact that $\Phi_u\lambda_0 - |\Phi_{ue}|^2 = \lambda_0\Phi_u^r$ it follows that

$$\text{cov} \begin{pmatrix} \hat{G} \\ \hat{H} \end{pmatrix} \sim \frac{n}{N} \frac{\Phi_v}{\Phi_u^r} \cdot \begin{bmatrix} 1 & (CS_0H_0)^* \\ CS_0H_0 & \frac{\Phi_u}{\lambda_0} \end{bmatrix}. \quad (10)$$

As a result the variance expressions for \hat{G} and \hat{H} become:

$$\text{cov}(\hat{G}) \sim \frac{n}{N} \frac{\Phi_v}{\Phi_u^r} = \frac{n}{N} \frac{\Phi_v}{\Phi_u} \left[1 + \frac{\Phi_u^e}{\Phi_u^r} \right] \quad (11)$$

$$\text{cov}(\hat{H}) \sim \frac{n}{N} \frac{\Phi_v}{\lambda_0} \frac{\Phi_u}{\Phi_u^r} = \frac{n}{N} \frac{\Phi_v}{\lambda_0} \left[1 + \frac{\Phi_u^e}{\Phi_u^r} \right]. \quad (12)$$

The case of an open-loop experimental situation now appears as a special situation in which $\Phi_u^e = 0$, $\Phi_u^r = \Phi_u$, and $C = 0$, and thus leading to the well known open-loop expressions

$$\text{cov}(\hat{G}) \sim \frac{n}{N} \frac{\Phi_v}{\Phi_u} \quad \text{cov}(\hat{H}) \sim \frac{n}{N} \frac{\Phi_v}{\lambda_0}. \quad (13)$$

As indicated in Ljung (1993), the closed-loop expressions show that only the noise-free part u^r of the input signal contributes to variance reduction of the estimates.

The given expressions are restricted to the situation that $S \in \mathcal{M}$ and that both $G(\theta)$ and $H(\theta)$ are identified; they do not hold true for the situation $G_0 \in \mathcal{G}$, $S \notin \mathcal{M}$.

Remark 3.1 The situation of estimating a plant model in the situation $G_0 \in \mathcal{G}$ and having a fixed and correct noise model $H_* = H_0$ is considered in Ljung (1993). Using the fact that

$$\text{cov} \hat{\theta}_N = \frac{\lambda_0}{N} [E\psi(t)\psi^T(t)]^{-1} \quad (14)$$

where $\psi(t)$ is the negative gradient of the prediction error (8), this leads to

$$\text{cov}(\hat{G}) \sim \frac{n \Phi_v}{N \Phi_u} \quad (15)$$

as it is immaterial whether the input spectrum is a result of open loop or closed loop operation. Note that this expression gives a smaller variance than the situation in which both G and H are estimated, and that in this (unrealistic) case the total input power contributes to a reduction of the estimate variance.

4 Indirect identification

4.1 Introduction

Recently several different indirect approaches to closed-loop identification have been presented, see e.g. Gevers (1993) and Van den Hof and Schrama (1995). These methods have been introduced from considerations related to the bias that occurs in closed-loop identification of approximate models. Here we will briefly illustrate their properties with respect to the variance of the estimates.

4.2 Coprime factor identification

Coprime factor identification is treated in detail in Schrama (1992) and Van den Hof *et al.* (1995). It is a scheme that relates to (and generalizes) the classical joint input/output method of closed-loop identification as e.g. described in Gustavsson *et al.* (1977). It does not require knowledge of the implemented controller C .

The basic principle is that the (two-times-two) transfer function $(r, e)^T \rightarrow (y, u)^T$ is identified, while the plant models (\hat{G}, \hat{H}) are retrieved from these closed-loop estimates.

Consider the system's relations:

$$y(t) = G_0 S_0 r(t) + S_0 H_0 e(t) \quad (16)$$

$$u(t) = S_0 r(t) - C S_0 H_0 e(t). \quad (17)$$

They are rewritten, by using a filtered signal $x(t) := F(q)r(t)$, into the form

$$y(t) = N_{0,F} x(t) + S_0 H_0 e(t) \quad (18)$$

$$u(t) = D_{0,F} x(t) - C S_0 H_0 e(t) \quad (19)$$

with $N_{0,F} := G_0 S_0 F^{-1}$ and $D_{0,F} := S_0 F^{-1}$, constituting a coprime factor representation of G_0 as $G_0 = N_{0,F} D_{0,F}^{-1}$.

The linear and stable filter F can be chosen by the user to serve several purposes, like minimal order properties or normalization of the coprime factorization as discussed in Van den Hof *et al.* (1995); this will not be pursued here any further as it is

immaterial for the variance analysis. The important observation here is that the signals x and e are uncorrelated.

Identification of the 4 transfer functions in (18),(19) from the signals $x(t)$, $y(t)$, $u(t)$ therefore corresponds to a one-input two-output open-loop identification problem. Denote

$$\varepsilon_y(t, \theta) = W_y(q, \theta)^{-1} [y(t) - N(q, \theta)x(t)] \quad (20)$$

$$\varepsilon_u(t, \theta) = W_u(q, \theta)^{-1} [y(t) - D(q, \theta)x(t)]; \quad (21)$$

Least squares minimization of $(\varepsilon_y, \varepsilon_u)^T$ provides estimated models $\hat{N}, \hat{D}, \hat{W}_y, \hat{W}_u$.

Open-loop models \hat{G} and \hat{H} are then retrieved by

$$\hat{G} = \hat{N}(\hat{D})^{-1} \quad (22)$$

$$\hat{H} = \hat{W}_y - \hat{G}\hat{W}_u. \quad (23)$$

In order to guarantee that \hat{H} is a monic transfer function, whenever \hat{W}_y and \hat{W}_u are monic, it will be assumed that \hat{G} is strictly proper.

For obtaining variance expressions of these reconstructed estimates, use can be made of first order approximations: $\hat{G} = G_0 + \Delta G$, $\hat{N} = N_{0,F} + \Delta N$, $\hat{D} = D_{0,F} + \Delta D$ etcetera, leading to

$$\Delta G = \frac{\Delta N}{D_{0,F}} - \frac{N_{0,F} \Delta D}{D_{0,F}^2} \quad (24)$$

$$\Delta H = \Delta W_y - G_0(\Delta W_u) - (\Delta G)W_u. \quad (25)$$

This leads to the result:

$$\text{cov} \begin{pmatrix} \hat{G} \\ \hat{H} \end{pmatrix} \sim \frac{n \Phi_v}{N \Phi_u} \cdot \begin{bmatrix} 1 & (C S_0 H_0)^* \\ C S_0 H_0 & \frac{\Phi_u}{\lambda_0} \end{bmatrix}. \quad (26)$$

A sketch of the derivation of this result is given in the Appendix.

Note that the expression (26) is identical to the expression that was derived for direct identification (10).

4.3 Identification in a dual Youla-Kucera parametrization

Another method that has recently been introduced utilizes a specific parametrization of the plant G_0 . As it is assumed that the controller C stabilizes the plant, G_0 can be parametrized within the class of all plants that are stabilized by C . This parametrization involves the relation

$$G(\theta) = \frac{N_x + D_c R(\theta)}{D_x - N_c R(\theta)} \quad (27)$$

where $N_x/D_x =: G_x$ is any (auxiliary) system that is stabilized by C ; $N_c/D_c = C$, and $R(\theta)$ ranges

over the class of all stable proper transfer functions. The different factors that build up the quotient expressions G_x and C are required to be stable and coprime.

Using an expression like (27) for the plant G_0 with a Youla-Kucera parameter R_0 , and substituting this in the system's relations, shows -after some manipulations- that these can be rewritten as

$$z(t) = R_0 x(t) + K_0 e(t) \quad (28)$$

with

$$R_0 = D_x S_0 (G_0 - G_x) / D_c \quad (29)$$

$$K_0 = H_0 S_0 / D_c \quad (30)$$

$$z = (D_c + G_x N_c)^{-1} (y - G_x u) \quad (31)$$

$$x = (D_x + C N_x)^{-1} r. \quad (32)$$

Since x is not correlated with e , the identification of R_0 and K_0 can again be considered to be an open-loop type of identification problem. Note that the signals z and x can simply be constructed by the user, as they are dependent on known quantities and measured signals. Least-squares identification is performed on the basis of the prediction error

$$\varepsilon_z(t, \theta) = K(q, \theta)^{-1} [z(t) - R(q, \theta)x(t)]$$

and the estimated transfers are denoted by \hat{K} and \hat{R} .

The open-loop model can then be reconstructed from these estimates according to

$$\hat{G} = \frac{N_x + D_c \hat{R}}{D_x - N_c \hat{R}} \quad (33)$$

$$\hat{H} = \hat{K} D_c \hat{S}^{-1} = \hat{K} D_c [1 + C \hat{G}]. \quad (34)$$

In order to guarantee that \hat{H} is monic whenever \hat{K} is monic, it will be assumed that $C \hat{G}$ is strictly proper and D_c is monic.

Variance expressions for the estimates \hat{R} and \hat{K} are available through the standard expressions for (open-loop) identification:

$$\text{cov}(\hat{R}) \sim \frac{n}{N} \frac{|K_0|^2 \lambda_0}{\Phi_x} \quad \text{and} \quad \text{cov}(\hat{K}) \sim \frac{n}{N} |K_0|^2 \quad (35)$$

while $\text{cov}(\hat{R}, \hat{K}) = 0$. In a similar way as has been done for the coprime factor identification method, these results can be utilized to obtain expressions for the variance of (\hat{G}, \hat{H}) , relying on first order approximating expressions. Not surprisingly (see Appendix) the variance expressions for (\hat{G}, \hat{H}) are again given by (26).

Further details on this identification method can be found in Lee *et al.* (1993) and Van den Hof and

Schrama (1995). It can be shown that it is a direct generalization of the classical indirect method of closed-loop identification, see Van den Hof and De Callafon (1996). It has to be stressed that knowledge of the controller C is assumed to be available.

4.4 Two-stage method

A two-stage method for closed-loop identification has been introduced in Van den Hof and Schrama (1993). It operates directly on reference, input and output data, and does not require knowledge of the implemented controller. It can best be explained by considering the system's relations:

$$u(t) = S_0 r(t) - C S_0 H_0 e(t) \quad (36)$$

$$y(t) = G_0 u^r(t) + S_0 H_0 e(t). \quad (37)$$

In the first step, measured signals r and u are used to estimate a model \hat{S} of the sensitivity function S_0 . Next this model is used to construct (by simulation) an estimate \hat{u}^r of u^r according to $\hat{u}^r(t) = \hat{S}(q)r(t)$. In the second stage, the signals \hat{u}^r and y are used as a basis for the identification of a plant model \hat{G} .

The procedure is very much alike the coprime factor identification scheme, albeit that the final plant model is not calculated through division of two identified models; this division is circumvented by constructing the auxiliary simulated signal \hat{u}^r .

If in the first step a consistent estimate of S_0 is obtained, the variance result for \hat{G} will appear to be similar to the previously obtained results

$$\text{cov}(\hat{G}) \sim \frac{n}{N} \frac{\Phi_v}{|S_0|^2 \Phi_r} = \frac{n}{N} \frac{\Phi_v}{\Phi_u^r}. \quad (38)$$

4.5 Summarizing comments

It has been shown that for the considered indirect identification methods, the asymptotic variance expressions for plant and noise model are exactly the same as the known expressions for direct identification. This may not be too surprising, as similar results for the classical indirect and joint i/o methods were already available (Gustavsson *et al.*, 1977). However what has to be stressed here, is that for the indirect type methods the variance expressions for \hat{G} are valid also in the situation that $G_0 \in \mathcal{G}$ but $S \notin \mathcal{M}$, while for the direct identification method the results are only achieved under the stronger condition that $S \in \mathcal{M}$. With indirect identification we can thus e.g. fix the noise model to a predetermined choice, only identifying the plant model \hat{G} , and obtain the same asymptotic variance as would be obtained when indeed estimating a noise model.

5 Open-loop versus closed-loop experiments

Considering that the variance expressions are identical for all closed-loop identification methods, we can now make a comparison between the variances obtained from open-loop and closed-loop experimental conditions. The appropriate expressions are summarized in table 1.

	Open-loop		Closed-loop
$Var(\hat{G}_N)$	$\frac{n}{N} \frac{\Phi_v}{\Phi_u}$	<	$\frac{n}{N} \frac{\Phi_v}{\Phi_u^r}$
$Var(\hat{H}_N)$	$\frac{n}{N} \frac{\Phi_v}{\lambda_0}$	<	$\frac{n}{N} \frac{\Phi_v}{\lambda_0} \left(1 + \frac{\Phi_u^e}{\Phi_u^r}\right)$

Table 1: Variance expressions under open-loop and closed-loop conditions.

The results show that for both \hat{G} and \hat{H} the variance obtained under closed-loop identification will generally be larger than for open-loop identification. Particularly in a situation where the input power is limited, the difference will become apparent, as in that case only part of the actual input spectrum can be used for variance reduction of \hat{G} and \hat{H} . In case the input power is not restricted, closed-loop identification can achieve the same results as open-loop identification, by choosing a reference signal r such that Φ_u^r is equal to the input spectrum applied in the open-loop situation.

The results suggest that in terms of variance of the model estimates \hat{G}_N and \hat{H}_N , open-loop identification always has to be preferred over closed-loop identification. However, perhaps surprisingly, this is not the case if the objective of the identification is model-based control design, as is explained in the next section.

6 Optimal experiments in view of model-based control

In this section we will consider the situation that the identified transfer functions \hat{G}_N and \hat{H}_N are used as a basis for model-based control design, and we will illustrate the effect of the variance of the identified model on the model application, i.e. the designed controller.

To this end we will first consider the following result from Ljung (1987, Theorem 14.3).

Proposition 6.1 Consider the variance-based identification design criterion

$$J(\mathcal{D}) = \int_{-\pi}^{\pi} tr[P(\omega, \mathcal{D})\Gamma(\omega)]d\omega$$

where

$$P(\omega, \mathcal{D}) = cov \begin{pmatrix} \hat{G}(e^{i\omega}) \\ \hat{H}(e^{i\omega}) \end{pmatrix},$$

\mathcal{D} denotes the design choices with respect to the experimental conditions, represented by $\{\Phi_u, \Phi_{ue}\}$, while $\Gamma(\omega)$ is a 2×2 Hermitian matrix reflecting the intended application of the model.

If $\Gamma_{12}(\omega) \equiv 0$ and the input power is limited, then the experimental condition \mathcal{D} for which $J(\mathcal{D})$ is optimized is given by

$$\Phi_u^{opt} = c \cdot \sqrt{\Gamma_{11}(\omega)\Phi_v(\omega)} \quad (39)$$

$$\Phi_{ue}^{opt} \equiv 0 \quad (40)$$

and c is a constant.

This result shows that open-loop identification is optimal when the intended application is one for which $\Gamma_{12} \equiv 0$. The situation $\Gamma_{12} \equiv 0$, considered in this proposition, reflects the case that a model is evaluated -in view of its intended application- by only considering the variance contributions of \hat{G} and \hat{H} separately, but not penalizing the covariance between the two. This situation applies e.g. to the case where a controller is designed on the basis of \hat{G} only and not considering \hat{H} . This situation is considered in the following corollary.

Corollary 6.2 Consider as model application a control design scheme based on a frequency weighted sensitivity minimization:

$$C_{\hat{G}} = \arg \min_{\hat{G}} \|V(1 + \hat{C}\hat{G})^{-1}\|_2.$$

Then the optimal experiment design in line with the above proposition is given by

- open-loop experiments ($\Phi_{ue}^{opt} \equiv 0$).
- $\Phi_u^{opt} = c \cdot |C_{\hat{G}} V S_0^2| \sqrt{\Phi_v}$

Proof: The application-related error criterion can be written as

$$\|V[(1 + CG_0)^{-1} - (1 + C\hat{G})^{-1}]\|_2$$

which can be shown to be equal to (using first order approximations)

$$\left\| \frac{VC(G_0 - \hat{G})}{(1 + CG_0)^2} \right\|_2.$$

An appropriate choice of Γ_{11} for this model application would thus be

$$\Gamma_{11}(\omega) = \frac{|VC|^2}{|1 + CG_0|^4} \quad (41)$$

leading to the result presented. \square

From the above result one could conclude that -from a variance point of view- an open-loop configuration would be the optimal experimental setup for performing identification for this control design objective in which the controller a function of \hat{G} only, i.e. independent of \hat{H} . It has to be noted, though, that the required input spectrum in this 'open-loop' situation should be proportional to the sensitivity function S_0 of the real plant, being controlled by the yet-to-be-designed controller. Input shaping with S_0 is exactly what is done when closed-loop identification is performed, as in that case $\Phi_u = |S_0|^2 \Phi_r + \Phi_u^e$.

A second related result is present in the recent work of Hjalmarsson *et al.* (1996) on optimal identification for control. In this work the identification criterion is selected to minimize the control performance degradation that results from the random errors on \hat{G}_N and \hat{H}_N . In solving this problem, the authors have quantified the variance error on the designed model-based controller.

Consider a situation where an identified model \hat{G}_N , \hat{H}_N is obtained from a closed-loop experimental situation with a controller C_{id} implemented on the plant. Consider a model-based control design scheme

$$\hat{C}_N = c(\hat{G}_N, \hat{H}_N)$$

and let F_G, F_H reflect the derivatives of c with respect to G, H , i.e. the sensitivity of the controller with respect to changes in G and H . Then the variance of the controller estimate is (see Hjalmarsson *et al.*, 1996)

$$\text{cov}(\hat{C}_N) \sim \frac{n}{N} |H_0|^2 \cdot \left\{ |F_H|^2 + \frac{\lambda_0}{\Phi_r} |F_G + (F_G G_0 + F_H H_0) C_{id}|^2 \right\}$$

leading to the situation that

- If $F_H \neq 0$, then the controller variance is minimized for models identified in closed-loop with an implemented controller C_{id}^{opt} unequal to zero, and the resulting controller variance is

$$\text{cov}(\hat{C}_N) \sim \frac{n}{N} |H_0|^2 |F_H|^2.$$

By comparison, the controller variance obtained with open-loop identification is

$$\text{cov}(\hat{C}_N) \sim \frac{n}{N} |H_0|^2 |F_H|^2 \left(1 + \frac{|F_G|^2}{|F_H|^2} \cdot \frac{\lambda_0}{\Phi_u} \right).$$

We observe that the variance obtained under ideal closed-loop experimental conditions can only be achieved with open-loop identification if the input power is made infinite.

- If $F_H = 0$, then the variance expression for closed-loop identification becomes

$$\begin{aligned} \text{cov}(\hat{C}_N) &\sim \frac{n}{N} |H_0|^2 |F_G|^2 \frac{|1 + C_{id} G_0|^2 \lambda_0}{\Phi_r} \\ &= \frac{n}{N} \frac{\Phi_v}{\Phi_u} |F_G|^2. \end{aligned}$$

The corresponding expression for open-loop identification is

$$\text{cov}(\hat{C}_N) \sim \frac{n}{N} \frac{\Phi_v}{\Phi_u} |F_G|^2.$$

The situation $F_H = 0$ means that the control design depends only on G and not on the noise model. This result is therefore consistent with Proposition 6.1.

We conclude from this analysis that, as far as variance errors are concerned, for model-based control design, closed-loop identification is optimal except when the controller is independent of the noise model.

7 Conclusions

Asymptotic variance expressions have been derived for several closed-loop identification schemes, involving both the (classical) direct method and more recently introduced indirect identification methods. It is shown that the several approaches lead to the same asymptotic variance.

Although asymptotic variance of plant model and noise model generally will increase when performing closed-loop identification, in comparison with open-loop identification, closed-loop identification can still be preferred when the identified model is used as a basis for control design. In the case that a controller is designed on the basis of both plant model and noise model, closed-loop identification is shown to lead to better variance results. When a controller is designed on the basis of a plant model only, the optimal identification experiment is an open-loop experiment with an input signal that has a power distribution that involves the real sensitivity function of the -yet to be designed- closed-loop plant.

Appendix

Proof of (26).

Applying the standard variance expressions to the multivariable situation of (18),(19) it follows that

$$\text{cov} \begin{pmatrix} \hat{N} \\ \hat{D} \end{pmatrix} \sim \frac{n}{N} \frac{|S_0|^2 \Phi_v}{\Phi_x} \begin{bmatrix} 1 & -C^* \\ -C & |C|^2 \end{bmatrix} \quad (\text{A.1})$$

$$\text{cov} \begin{pmatrix} \hat{W}_y \\ \hat{W}_u \end{pmatrix} \sim \frac{n}{N} \frac{|S_0|^2 \Phi_v}{\lambda_0} \begin{bmatrix} 1 & -C^* \\ -C & |C|^2 \end{bmatrix} \quad (\text{A.2})$$

Since (18),(19) reflect an open-loop situation (as x and e are uncorrelated) this implies that the cross-covariance terms between $(\hat{N}, \hat{D})^T$ and (\hat{W}_y, \hat{W}_u) are zero.

Applying the first order approximations in (24) it follows that:

$$\begin{aligned} |\Delta G|^2 &= \left[\frac{\Delta N}{D_{0,F}} - \frac{G_0}{D_{0,F}} \Delta D \right] \left[\frac{\Delta N}{D_{0,F}} - \frac{G_0}{D_{0,F}} \Delta D \right]^* \\ &= \frac{|\Delta N|^2}{|D_{0,F}|^2} + \frac{|G_0|^2}{|D_{0,F}|^2} |\Delta D|^2 + \\ &\quad - \text{Re} \left\{ \frac{G_0(\Delta D)(\Delta N)^*}{|D_{0,F}|^2} \right\}. \end{aligned}$$

Substitution of (A.1) then provides the result for $\text{cov}(\hat{G})$.

For \hat{H} one can similarly write (when neglecting terms that have expectation 0):

$$\begin{aligned} |\Delta H|^2 &= |\Delta W_y|^2 + |G_0|^2 |\Delta W_u|^2 + |W_u|^2 |\Delta G|^2 \\ &\quad - 2 \text{Re} G_0(\Delta W_u) \cdot (\Delta W_y)^* \end{aligned}$$

and the result for $\text{cov}(\hat{H})$ follows after substitution of (A.2).

The expression for $\text{cov}(\hat{G}, \hat{H})$ follows from $\text{cov}(\hat{G}, \hat{H}) = -W_u^* \text{cov}(\hat{G})$.

Variance result for dual Youla-Kucera method

Using (33),(34) the related expressions for the first order approximation errors become

$\Delta G =$

$$\begin{aligned} &\frac{(D_x - N_c R_0) D_c(\Delta R) + (N_x + D_c R_0) N_c(\Delta R)}{(D_x - N_c R_0)^2} \\ \Delta H &= \frac{D_c(\Delta K)}{S_0} + K_0 N_c(\Delta G). \end{aligned} \quad (\text{A.3})$$

For ΔG this leads to

$$\Delta G = \frac{D_c + G_0 N_c}{D_x - N_c R_0} \Delta R = \frac{D_c(\Delta R)}{D_x S_0^2 (1 + C G_x)}$$

and so

$$\text{cov}(\hat{G}) = \left| \frac{D_c}{D_x S_0^2 (1 + C G_x)} \right|^2 \text{cov}(\hat{R}).$$

Substituting the expression for $\text{cov}(\hat{R})$ and using the property that $\Phi_x = |D_x(1 + C G_x)|^2 \Phi_r$ it follows after some manipulation that $\text{cov}(\hat{G}) \sim n/N \cdot \Phi_v / \Phi_u^r$.

For $\text{cov}(\hat{H})$ it follows from (A.3) that

$$\text{cov}(\hat{H}) = \frac{|D_c|^2 \text{cov} \hat{K}}{|S_0|^2} + |N_c K_0|^2 \text{cov} \hat{G}.$$

Substituting the known expressions in the right hand side, will show that $\text{cov}(\hat{H}) \sim n/N |H_0|^2 [1 + \Phi_u^e / \Phi_u^r]$.

For $\text{cov}(\hat{G}, \hat{H})$ it follows from (A.3) that

$$\text{cov}(\hat{G}, \hat{H}) = (K_0 N_c)^* \text{cov}(\hat{G})$$

which leads to the appropriate result.

References

- Gevers, M. and L. Ljung (1986). Optimal experiment design with respect to the intended model application. *Automatica*, 22, 543-554.
- Gevers, M. (1993). Towards a joint design of identification and control? In: H.L. Trentelman and J.C. Willems (Eds.), *Essays on Control: Perspectives in the Theory and its Applications*. Birkhäuser, Boston, pp. 111-151.
- Gustavsson I., L. Ljung and T. Söderström (1977). Identification of processes in closed loop - identifiability and accuracy aspects. *Automatica*, 13, 59-75.
- Hansen, F.R. and G.F. Franklin (1988). On a fractional representation approach to closed-loop experiment design. *Proc. American Control Conf.*, Atlanta, GA, USA, pp. 1319-1320.
- Hjalmarsson, H., M. Gevers and F. De Bruyne (1996). For model-based control design, closed-loop identification gives better performance. To appear in *Automatica*, Vol. 32, no. 12.
- Lee, W.S., B.D.O. Anderson, R.L. Kosut and I.M.Y. Mareels (1993). A new approach to adaptive robust control. *Int. J. Adaptive Control and Signal Proc.*, 7, 183-211.
- Ljung, L. (1985). Asymptotic variance expressions for identified black-box transfer function models. *IEEE Trans. Autom. Control*, AC-30, 834-844.
- Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.
- Ljung, L. (1993). Information contents in identification data from closed-loop operation. *Proc. 32nd IEEE Conf. Decision and Control*, San Antonio, TX, pp. 2248-2252.
- Schrama, R.J.P. (1992). *Approximate Identification and Control Design*. Dr. Dissertation, Delft University of Technology.
- Van den Hof, P.M.J., R.J.P. Schrama, R.A. de Callafon and O.H. Bosgra (1995). Identification of normalised coprime plant factors from closed-loop experimental data. *Europ. J. Control*, 1, 62-74.
- Van den Hof, P.M.J. and R.J.P. Schrama (1995). Identification for control - closed-loop issues. *Automatica*, 31, 1751-1770.
- Van den Hof, P.M.J. and R.A. de Callafon (1996). Multivariable closed-loop identification: from indirect identification to dual-Youla parametrization. To appear in *Proc. 35th IEEE Conf. Decision and Control*, Kobe, Japan.

Analysis of closed-loop identification with a tailor-made parametrization

Edwin T. van Donkelaar[†] and Paul M.J. Van den Hof

Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: e.t.vandonkelaar@wbmt.tudelft.nl

Abstract. An analysis is made of a closed-loop identification scheme in which the parameters of the (open-loop) model are identified on the basis of input and output signals of the closed-loop transfer function. A parametrization of the closed-loop transfer in terms of the parameters of the open-loop plant model is employed, utilizing knowledge of the implemented feedback controller. This is denoted a tailor-made parametrization as it is tailored to the specific feedback structure at hand. To obtain an estimate of the plant model, a dedicated nonlinear optimization algorithm is required as the standard optimization tools for the situation of open-loop models can not be applied. Consistency of the estimate is shown to hold under additional conditions on controller and plant model order. These conditions result from the requirement of a uniformly stable model set. Simulation examples show both the power and the hazard of closed-loop identification with a tailor-made parametrization.

Keywords. Closed-loop identification; tailor-made parametrization; indirect identification; closed-loop stability.

1 Introduction

System identification from closed-loop data has had a lot of attention in literature which has resulted in numerous closed-loop identification schemes. First of all there are the more classical methods like direct identification, indirect identification, instrumental variable methods and joint input/output identification, see e.g. Söderström and Stoïca (1989). More recently particular versions of these closed-loop identification schemes have been developed that are directed towards an explicitly tunable bias expression, which is aiming for an identified model that is particularly suitable for use in control design. Examples of such schemes are the two-stage method (Van den Hof and Schrama, 1993), identification in the dual Youla parametrization ((Lee *al.*, 1993), (Schrama, 1992)) and identification of co-

prime plant factors (Van den Hof *et al.*, 1995). An overview of these closed-loop identification schemes can be found in Gevers (1993) and Van den Hof and Schrama (1995).

In this paper a closed-loop identification method is discussed that has not had a lot of attention in literature: closed-loop identification with a tailor-made parametrization. The basic idea is that the closed-loop transfer function from excitation signal r to output signal y (see Figure 1) is identified using an output predictor

$$\hat{y}(t, \theta) = \frac{G(q, \theta)}{1 + C(q)G(q, \theta)} r(t)$$

using the parameters corresponding to the (open-loop) plant model

$$G(q, \theta) = \frac{b_1 q^{-1} + \dots + b_{n_B} q^{-n_B}}{1 + a_1 q^{-1} + \dots + a_{n_A} q^{-n_A}}$$

with $\theta = [b_1 \dots b_{n_B} \ a_1 \dots a_{n_A}]$.

[†]The work of Edwin van Donkelaar is financially supported by the Dutch Technology Foundation (STW) under contract DWT55.3618

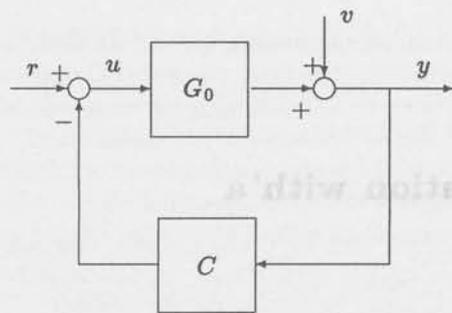


Fig. 1: Closed-loop configuration

Using the open-loop plant parameters, and knowledge of the controller C , a prediction error criterion is used to estimate the plant parameters; this requires a nonlinear optimization procedure.

The parametrization is referred to as a tailor-made parametrization, as it is specifically directed towards (tailored to) the closed-loop configuration at hand, including knowledge of the controller.

This identification approach has been mentioned as an exercise in Ljung (1987). It is also employed in a recursive version in Landau and Boumaïza (1996). In this paper, an analysis will be made of the consistency properties of this method, where in particular we will focus on the connectedness of related parameter sets and the uniform stability of corresponding model sets.

After preliminary notation and the formulation of the problem, in section 4 it will be made clear that the need for uniform stability of the model set, which is adopted in Ljung (1987) to obtain consistency results, imposes additional conditions on the parametrization. Sufficient conditions for consistency are derived which results in a condition on controller and model order. In section 5 compact expressions for the gradient and Hessian of the cost function are given, which are useful if a gradient search method is used for the nonlinear optimization. In section 6 two simulations are given to illustrate both the possible problems and the power of the application of a tailor-made parametrization. Next, in section 7 the relation between this and other closed-loop identification methods is discussed. Finally, section 8 concludes the paper.

2 Preliminaries

Addressed is the problem of obtaining a model of the linear time-invariant discrete-time single input single output plant $G_0(z)$ from measurements of the closed-loop configuration given in figure 1.

The controller is denoted with $C(z)$ and is assumed to be known. The signal $r(t)$ is an external excita-

tion signal, $u(t)$ and $y(t)$ are respectively the plant input and output. It is assumed that measurements of $r(t)$ and $y(t)$ are available. The output noise $v(t)$ is assumed to be generated by filtering of white noise signal $e(t)$ with variance σ^2 using a stable monic filter $H_0(z)$. The output noise is assumed to be uncorrelated with the excitation signal r . Lastly, the loop transfer $C(z)G_0(z)$ is assumed to be strictly proper. The closed-loop transfer function from measured reference to measured output can be written as follows.

$$y(t) = \underbrace{\frac{G_0(q)}{1+C(q)G_0(q)}}_{R_0(q)} r(t) + \underbrace{\frac{1}{1+C(q)G_0(q)} H_0(q)}_{W_0(q)} e(t)$$

where $R_0(q)$ denotes the closed-loop transfer function and $W_0(q)$ the closed-loop noise filter.

The sensitivity function is denoted by $S_0(z) = (1 + C(z)G_0(z))^{-1}$ and the parametrized sensitivity is denoted with $S(z, \theta) = (1 + C(z)G(z, \theta))^{-1}$.

3 Closed-loop identification with a tailor-made parametrization

Consider a parametrized model of the plant $G(q, \theta)$ where the parameter vector lies in a parameter set $\theta \in \Theta$. This parametrized plant model together with knowledge of the controller can be used to parametrize the transfer function between the measured signals $r(t)$ and $y(t)$. This yields the following prediction of the output in case the parametrized closed-loop noise filter is set to $W(q, \theta) = 1$ (output error structure)

$$\hat{y}(t, \theta) = \frac{G(q, \theta)}{\underbrace{1 + C(q)G(q, \theta)}_{R(q, \theta)}} r(t), \quad \theta \in \Theta \quad (1)$$

The corresponding set of closed-loop models is defined as

$$\mathcal{P} := \left\{ R(q, \theta) = \frac{G(q, \theta)}{1 + C(q)G(q, \theta)}, \theta \in \Theta \right\} \quad (2)$$

The parameter estimate is found by least squares minimization of the prediction error by solving $\hat{\theta}_N = \underset{\theta \in \Theta}{\operatorname{argmin}} V_N(\theta)$, in which the criterion function

is given by $V_N(\theta) = \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t, \theta)$ and the prediction error is defined as $\varepsilon(t, \theta) = y(t) - R(q, \theta)r(t)$. The resulting estimation of the plant model will be denoted by $\hat{R}(q) = R(q, \hat{\theta}_N)$.

For this identification method the following consistency result holds (Ljung, 1987).

Proposition 3.1 Get \mathcal{P} be a uniformly stable model set and let the data generating system satisfy the standard conditions in Ljung (1987). Then $\theta_N \rightarrow \theta^*$ w.p. 1 for $N \rightarrow \infty$ with

$$\theta^* = \arg \min_{\theta \in \Theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} |R_0(e^{i\omega}) - R(e^{i\omega}, \theta)|^2 \Phi_r(\omega) d\omega \quad (3)$$

Whenever there exists a θ such that $G(q, \theta) = G_0(q)$ this choice will be a minimizing argument of the integral expression above which is unique provided that $r(t)$ is persistently exciting of sufficient high order.

This proposition states that a consistent estimate is obtained with this parametrization under the condition that the model set \mathcal{P} is uniformly stable. This condition is not trivially satisfied in case the tailor-made parametrization given in (2) is used. Therefore, in the next section the conditions under which the model set (2) is guaranteed to be uniformly stable will be investigated.

4 Uniform stability of the model set

In the previous section it is mentioned that, in case of uniform stability of the model set, a consistent estimate is obtained with closed-loop identification using a tailor-made parametrization. Uniform stability of the model set is defined as follows.

Definition 4.1 (Ljung, 1987) A parametrized model set \mathcal{G} is uniformly stable if

- Θ is a connected open subset of $\mathbb{R}^{(n_A+n_B)}$
- $\mu: \Theta \rightarrow \mathcal{P}$ is a differentiable mapping, and
- the family of transfer functions $\{R(z, \theta), \frac{\partial}{\partial \theta} R(z, \theta)\}$ is uniformly stable.

In this section it will be made clear that in case a tailor-made parametrization is used, the parameter set Θ is possibly not connected due to the specific parametrization of the closed-loop transfer function $R(z, \theta)$. Also a sufficient condition is derived for guaranteed connectedness of the parameter set. Let the strictly proper¹ plant model be parametrized as

$$G(z, \theta) = \frac{B(z, \theta)}{A(z, \theta)} = \frac{b_1 z^{-1} + \dots + b_{n_B} z^{-n_B}}{1 + a_1 z^{-1} + \dots + a_{n_A} z^{-n_A}} \quad (4)$$

¹For simplicity of notation only the case of a strictly proper plant and a proper controller is regarded. However, the case of a strictly proper controller and a proper plant can be described similarly.

where $\theta = [a_1 \dots a_{n_A} \ b_1 \dots b_{n_B}]^T$. The controller of order n_c is given by

$$C = \frac{N_c(z)}{D_c(z)} = \frac{n_0 + n_1 z^{-1} + \dots + n_{n_c} z^{-n_c}}{1 + d_1 z^{-1} + \dots + d_{n_c} z^{-n_c}}$$

where $N_c(z), D_c(z)$ are coprime polynomials. With this notation the parametrization of the output predictor is given by

$$\hat{y}(t, \theta) = \frac{D_c(q)B(q, \theta)}{D_c(q)A(q, \theta) + N_c(q)B(q, \theta)} r(t) \quad (5)$$

All closed-loop models $R(q, \theta)$ are stable if the absolute value of the roots of the denominator $D_c(q)A(q, \theta) + N_c(q)B(q, \theta)$ is strictly less than one. Hence, the parameter set corresponding to closed-loop stable models is given by

$$\Theta = \{\theta \in \mathbb{R}^{n_A+n_B} \mid |\text{sol}\{D_c(z)A(z, \theta) + N_c(z)B(z, \theta) = 0\}| < 1\}. \quad (6)$$

The corresponding set of plant models is denoted by

$$\mathcal{G} := \{G(z, \theta), \theta \in \Theta\}. \quad (7)$$

It can be verified that the parameter set for which the polynomial $A(q, \theta)$ is stable, is pathwise connected². As a result, connectedness of the parameter set when using a (standard) numerator-denominator parametrization of the plant in an open-loop setting, will not be a problem. However, in case the tailor-made parametrization (2) is used, with Θ given by (6), Θ need not be pathwise connected as the following simple example shows.

Example 4.2 Given the 7th order controller defined by the continuous time transfer function $C(s) =$

$$\frac{0.499s^5 + 0.715s^4 + 2.577s^3 + 3.397s^2 + 2.155s + 2.620}{s^7 + 1.717s^6 + 5.100s^5 + 8.410s^4 + 4.198s^3 + 6.631s^2}$$

The plant that is to be identified is parametrized by a simple constant $G = \theta$. The parameter space $\Theta \subset \mathbb{R}$ for which the closed-loop system is stable can be simply derived from a root locus plot and is approximately given by

$$\Theta = \{\theta \mid \theta \in (0, 1.27) \cup (2.64, 4.69) \cup (9.98, \infty)\}$$

This set is a disconnected subset of \mathbb{R} . Therefore the corresponding model set \mathcal{P} is not uniformly stable.

A parameter set that is not connected has not only consequences for the formal proof of consistency as

²A justification of this claim is added in the appendix.

was mentioned before, but also for the nonlinear optimization that has to be performed to obtain an estimate. If, for example, a gradient search method is used and an initial estimate is selected in a region of the parameter set that is disconnected from the region where the optimal parameter vector is located, it will be extremely hard if not impossible to reach the optimum.

The denominator of the closed-loop transfer function can be written as a function of the open loop parameter θ as

$$D_c A(z, \theta) + N_c B(z, \theta) = 1 + [z^{-1} \ z^{-2} \ \dots \ z^{-n}] \theta_{cl} \quad (8)$$

where the closed-loop parameter vector is given by $\theta_{cl} := S\theta + \rho$. The order of the closed-loop polynomial of (8) is given by $n = \max(n_A, n_B) + n_c$, $\rho = [p_1 \dots p_{n_c} \ 0 \dots 0]^T \in \mathbb{R}^n$ and $S = [P_D \ P_N] \in \mathbb{R}^{n \times (n_A + n_B)}$ with $P_D \in \mathbb{R}^{n \times n_A}$, $P_N \in \mathbb{R}^{n \times n_B}$ are matrices given by

$$P_D = \begin{bmatrix} 1 & 0 & \dots & 0 \\ d_1 & 1 & & \\ & d_2 & d_1 & \ddots \\ & \vdots & d_2 & \ddots & 1 \\ d_{n_c} & & \ddots & & d_1 \\ & 0 & \ddots & & d_2 \\ & \vdots & \ddots & & \vdots \\ 0 & \dots & 0 & & d_{n_c} \end{bmatrix}, P_N = \begin{bmatrix} n_0 & 0 & \dots & 0 \\ n_1 & n_0 & & \\ & n_2 & n_1 & \ddots \\ & \vdots & n_2 & \ddots & n_0 \\ n_{n_c} & & \ddots & & n_1 \\ & 0 & \ddots & & n_2 \\ & \vdots & \ddots & & \vdots \\ 0 & \dots & 0 & & n_{n_c} \end{bmatrix} \quad (9)$$

The closed loop parameter can vary over a parameter set

$$\Theta_{cl} := \{\theta_{cl} = M\theta + \rho \mid \theta \in \Theta\}$$

where the allowable closed loop parameters are restricted by the affine relation given above. Now, define a parameter set for stable polynomials of order n as follows

$$\Theta_n := \{\theta_n \in \mathbb{R}^n \mid |\text{sol}\{1 + [z^{-1} \ \dots \ z^{-n}] \theta_n = 0\}| < 1\}$$

From connectedness of the parameter set for of stable polynomials (see Appendix) it can be concluded that the parameter set Θ_n is also connected. In the following theorem a sufficient condition for connectedness of the parameter space Θ is given using the connected set Θ_n as a starting point.

Lemma 4.3 Full row rank of the matrix $S = [P_D \ P_N]$ with P_D, P_N given in (9), is a sufficient condition for pathwise connectedness of the parameter set Θ given in (6).

Proof: The closed-loop parameter θ_n can vary over the connected set Θ_n . Now define the set

$$\bar{\Theta}_{cl} = \{\bar{\theta}_{cl} \mid \bar{\theta}_{cl} = \theta_n - \rho, \theta_n \in \Theta_n\}$$

This set is a shifted version of Θ_n and is therefore also pathwise connected. An open loop parameter vector $\theta \in \Theta$ and a parameter vector $\bar{\theta}_{cl} \in \bar{\Theta}_{cl}$ are related via $\bar{\theta}_{cl} = S\theta$, $S \in \mathbb{R}^{n \times (n_A + n_B)}$. If S has full row rank it defines a surjective map, hence $\text{image}(S) = \bar{\Theta}_{cl}$. In the connected set $\bar{\Theta}_{cl}$ a continuous path can be constructed between two parameter vectors. This path can be mapped into a continuous path in Θ using the inverse mapping of S . Therefore Θ is also pathwise connected. \square

This result implies that the parameter set for which the parametrized transfer function (2) is stable, is only a connected set in specific cases. Therefore it is not guaranteed that the model set defined in (2) is uniformly stable following the definition of uniform stability in Definition 4.1. The following lemma gives an easy test for guaranteed uniform stability of the model set with a tailor-made parametrization.

Proposition 4.4 Let a model of order n_s be parametrized as

$$G(q, \theta) = \frac{B(z, \theta)}{A(z, \theta)} = \frac{b_1 z^{-1} + \dots + b_{n_s} z^{-n_s}}{1 + a_1 z^{-1} + \dots + a_{n_s} z^{-n_s}}$$

and let the controller of order n_c be given by

$$C = \frac{N_c(z)}{D_c(z)} = \frac{n_0 + n_1 z^{-1} + \dots + n_{n_c} z^{-n_c}}{1 + d_1 z^{-1} + \dots + d_{n_c} z^{-n_c}}$$

A sufficient condition for connectedness of the parameter set Θ for a tailor-made parametrization given in (2), is given by

$$n_s \geq n_c$$

Proof: From lemma 4.3 it follows that full row rank of S is a sufficient condition for connectedness. By reordering the columns of S a 2×2 upper triangular block matrix can be constructed given by

$$S = \begin{bmatrix} S_1 & S_{12} \\ 0 & S_2 \end{bmatrix} \text{ where}$$

$$S_1 = \begin{bmatrix} 1 & 0 & \dots & 0 & n_0 & 0 & \dots & 0 \\ d_1 & 1 & & & n_1 & n_0 & & \\ & d_2 & d_1 & \ddots & n_2 & n_1 & \ddots & \vdots \\ & \vdots & d_2 & \ddots & \vdots & n_2 & \ddots & n_0 \\ d_{n_c} & & \ddots & d_1 & n_{n_c} & & \ddots & n_1 \\ 0 & d_{n_c} & & d_2 & 0 & n_{n_c} & & n_2 \\ & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & d_{n_c} & 0 & \dots & 0 & n_{n_c} \end{bmatrix}$$

$$S_2 = \begin{bmatrix} d_{n_c} & \dots & d_{2n_c - n_s + 1} & n_{n_c} & \dots & n_{2n_c - n_s + 1} \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & & d_{n_c} & 0 & & n_{n_c} \end{bmatrix}$$

where $S_1 \in \mathbb{R}^{2n_c \times 2n_c}$ and $S_2 \in \mathbb{R}^{(n_s - n_c) \times 2(n_s - n_c)}$. The matrix S has full row rank if S_1 and S_2 have full row rank. The first is a Sylvester matrix which has full row rank if and only if the numerator and denominator of the controller are coprime (Chen, 1984). The second has full row rank if $d_{n_c} \neq 0$ or $n_{n_c} \neq 0$. This is always the case for a controller of order n_c . The number of rows of S is smaller than or equal to the number of columns if $n_a + n_b \geq \max(n_a, n_b) + n_c$. This reduces to $2n_s \geq n_s + n_c$ or equivalently $n_s \geq n_c$. \square

From this it can be concluded that connectedness of the parameter set Θ causes no problem if the order of the controller is smaller than the model order. So for identification of a simple model based on experiments with a complex controller connectedness of the parameter set may be a problem. Note that this is the case in example 4.2.

Apart from connectedness of the parameter set over which is optimized, other issues should be investigated. For example whether local minima and saddlepoints can occur and if so how many can be expected and most of all whether this hampers a good application of this method. To investigate this, functional analysis can be performed on the basis of the expressions for the criterion function, its gradient and the Hessian. In the next section compact expressions for these functions are given.

5 Gradient expressions

To obtain a parameter estimate the optimization problem given in the previous section has to be solved. Due to the used parametrization this is a nonlinear optimization problem. To find a solution to this optimization problem gradient search methods can be used like Newton-Raphson and Gauss-Newton as suggested in Ljung (1987).

However, the character of the function $V_N(\theta)$ that is optimized as well as the parameter set Θ over which is optimized is highly influenced by the controller. Both the function and the set can be extremely non-convex which can make it difficult to apply gradient search methods successfully because the optimization can get stuck in a local minimum or at the boundary of the parameter set. To alleviate these problems it is essential that a good initial estimate is chosen for the iterative search and a good strategy is applied for the choice of the step size.

To apply gradient search methods the gradient of $V_N(\theta)$ needs to be available and for some methods

also the Hessian. In this section these derivatives are derived, where, for convenience of notations, $n_A = n_B = n_s$ is chosen. The more general case, however, can be derived similarly. The derivatives of the cost function can be expressed as

$$\begin{aligned} \frac{\partial V(\theta)}{\partial \theta} &= -\frac{2}{N} \sum_{t=1}^N \varepsilon(t, \theta) \frac{\partial \hat{y}(t, \theta)}{\partial \theta} \in \mathbb{R}^{2n_s} \\ \frac{\partial^2 V(\theta)}{\partial \theta^2} &= 2 \sum_{t=1}^N \frac{\partial \hat{y}(t, \theta)}{\partial \theta} \left(\frac{\partial \hat{y}(t, \theta)}{\partial \theta} \right)^T \\ &\quad - 2 \sum_{t=1}^N \varepsilon(t, \theta) \frac{\partial^2 \hat{y}(t, \theta)}{\partial \theta^2} \in \mathbb{R}^{2n_s \times 2n_s} \end{aligned}$$

Hence, these derivatives can be calculated if the first and second derivative of the output prediction are known. These can be calculated by differentiating (5). Differentiating this expression once yields

$$\begin{aligned} \frac{\partial}{\partial \theta} \{ \hat{y}(t, \theta) + [\hat{y}(t-1, \theta) \dots \hat{y}(t-n, \theta)] ([P_D \ P_N] \theta + \rho) \} = \\ = \frac{\partial}{\partial \theta} \{ [r(t-1) \dots r(t-n)] [0 \ P_D] \theta \} \end{aligned}$$

or equivalently

$$\begin{aligned} \frac{\partial \hat{y}(t, \theta)}{\partial \theta} + \left[\frac{\partial \hat{y}(t-1, \theta)}{\partial \theta} \dots \frac{\partial \hat{y}(t-n, \theta)}{\partial \theta} \right] ([P_D \ P_N] \theta + \rho) + \\ + \begin{bmatrix} P_D^T \\ P_N^T \end{bmatrix} \begin{bmatrix} \hat{y}(t-1, \theta) \\ \vdots \\ \hat{y}(t-n, \theta) \end{bmatrix} = \begin{bmatrix} 0 \\ P_D^T \end{bmatrix} \begin{bmatrix} r(t-1) \\ \vdots \\ r(t-n) \end{bmatrix} \end{aligned}$$

This can be written more concisely as

$$F(q, \theta) \frac{\partial \hat{y}(t, \theta)}{\partial \theta} = M^T \psi(t, \theta) \quad (10)$$

with a filter

$$F(q, \theta) = (1 + [q^{-1} \dots q^{-n}] ([P_D \ P_N] \theta + \rho)),$$

a matrix $M = \begin{bmatrix} P_D & P_N \\ 0 & P_D \end{bmatrix} \in \mathbb{R}^{2n \times 2n_s}$ and a regression vector

$\psi^T(t, \theta) = [-\hat{y}(t-1, \theta) \dots -\hat{y}(t-n, \theta) \ r(t-1) \dots r(t-n)]$. Equation (10) can also be expressed with

$$\frac{\partial \hat{y}(t, \theta)}{\partial \theta} = M^T \psi_F(t, \theta) \quad (11)$$

where $\psi_F(t, \theta) = F^{-1}(q, \theta) \psi(t, \theta)$ is a filtered version of the regression matrix.

The second derivative of the output prediction can be calculated by differentiation of (10), which yields

$$\frac{\partial^2 \hat{y}(t, \theta)}{\partial \theta^2} F(q, \theta) + \frac{\partial F(q, \theta)}{\partial \theta} \left(\frac{\partial \hat{y}(t, \theta)}{\partial \theta} \right)^T =$$

$$= - \left[\frac{\partial \hat{y}(t-1, \theta)}{\partial \theta} \dots \frac{\partial \hat{y}(t-n, \theta)}{\partial \theta} \right] [P_D P_N], \text{ or}$$

$$\frac{\partial^2 \hat{y}(t, \theta)}{\partial \theta^2} =$$

$$-2F^{-1}(q, \theta) \left[\frac{\partial \hat{y}(t-1, \theta)}{\partial \theta} \dots \frac{\partial \hat{y}(t-n, \theta)}{\partial \theta} \right] [P_D P_N]$$

These compact expressions are similar to expressions obtained for nonlinear optimization with a standard input-output parametrization and can be fruitfully used in nonlinear optimization routines for closed-loop identification with a tailor-made parametrization.

6 Simulation examples

In this section two simulation examples are given. One in the case where $G_0 \in \mathcal{G}$ and the parameter set is not connected and the other where $G_0 \notin \mathcal{G}$ with a connected parameter set but with a very bad signal to noise ratio. In the first example the tailor-made parametrization induces an optimization problem which is difficult to solve while in the second example it is demonstrated that closed-loop identification with this parametrization can be very powerful.

Simulation 1

In figure 2 the three separate branches of the cost function $V_N(\theta)$ for the system from Example 4.2 is depicted for a system $G_0 = 3.5$. Here it is assumed that both model and system are a simple constant and an output noise which is driven by a white noise signal $e(t)$ with variance $\sigma = 0.1$ has a noise filter $H_0(q) = 1$. The excitation signal $r(t)$ is white noise with variance 1. The function is clearly discontinuous and has several local minima that are located at the boundary of the stability area which makes it difficult to find the optimum with gradient search methods.

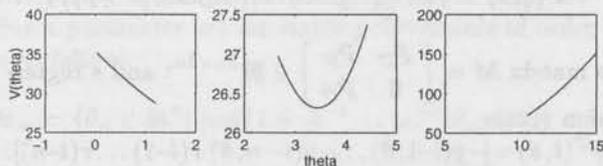


Fig. 2: Criterion function for controller given in example 4.1 and plant $G_0 = 3.5$ for closed-loop identification with a tailor-made parametrization

The global optimum will generally only be found if an initial estimate is selected from the middle of the three branches of the criterion function. In the other ones the iterative search gets stuck in a local optimum which is at the boundary of the stability

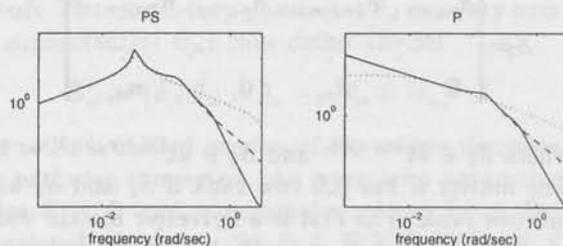


Fig. 3: Closed loop transfer function from $r(t)$ to $y(t)$ (left) and open loop transfer function (right): plant (solid), estimation with tailor-made parametrization (dashed) and direct identification based on an ARX(331)-estimation (dotted).

region.

Note that the parameter regions $(-\infty, 0]$, $[1.27, 2.64]$ and $[4.69, 9.98]$ induce an unstable closed loop system.

Simulation 2

A simulation is made with a fifth order system, which is given by the transfer function

$$G_0(z) = \frac{5.278z^{-1} + 126.7z^{-2} + 299.3z^{-3} + 110.8z^{-4} + 404.2z^{-5}}{1 - 4.391z^{-1} + 7.879z^{-2} - 7.247z^{-3} + 3.430z^{-4} - 4.391z^{-5}}$$

which is a pure integrator with two resonant modes. The controller used in the simulation is a PI-controller which stabilizes the system. The excitation signal $r(t)$ is Gaussian white noise with standard deviation $\sigma_r = 1$ and the output noise $v(t)$ is Gaussian white noise with a standard deviation of $\sigma_v = 0.5$. The data length is $N = 500$. The open loop and closed-loop transfer functions are given in figure 3.

For this system a third order model is estimated with tailor-made parametrization. The non-linear optimization is performed using a Gauss-Newton method where the initial estimate is obtained with use of direct identification. The estimated model is given in figure 3. Also the initial model is given. From this it can be seen that the estimation with a tailor-made parametrization gives a good fit for the integrator and the first resonant mode, despite the bad signal-to-noise ratio and the bad initial estimate for the nonlinear optimization.

7 Relation to other closed-loop identification methods

In this section the relation between closed-loop identification with a tailor-made parametrization and other closed-loop identification methods is dis-

cussed.

An obvious parametrization making further use of knowledge of the closed-loop structure, is given by

$$R(q, \theta) = \frac{G(q, \theta)}{1 + C(q)G(q, \theta)}$$

$$W(q, \theta) = \frac{H(q, \theta)}{1 + C(q)G(q, \theta)}$$

Least squares minimization of the corresponding prediction error yields a criterion function $V_N(\theta) = \frac{1}{N} \sum_{t=1}^N H^{-1}(q, \theta)(y(t) - G(q, \theta)(r(t) - C(q)y(t)))$ which is equal to the cost function for direct identification from $u(t) = (r(t) - C(q)y(t))$ to $y(t)$ which is known to be only consistent in case both the plant G_0 and the noise model H_0 can be modelled exactly within the chosen model set. It is important to note that this inconsistency is due to the dependent parametrization of the closed-loop transfer and the closed-loop noise filter. If $R(q, \theta)$ and $W(q, \theta)$ are parametrized independently, the consistency result given in Proposition 3.1 still holds in case $G_0 \in \mathcal{G}$. The specific approximative properties of closed-loop identification with a tailor-made parametrization can be obtained from (3). This expression can be further specified as

$$\theta^* = \arg \min_{\theta \in \Theta} \frac{1}{2\pi}$$

$$\int_{-\pi}^{\pi} |S_0(e^{i\omega})[G_0(e^{i\omega}) - G(e^{i\omega}, \theta)]S(e^{i\omega}, \theta)|^2 \Phi_r(\omega) d\omega.$$

From this it can be seen that the estimation error is weighted by both the sensitivity function and the estimated sensitivity function. Therefore the crossover region is emphasized in the minimization. This implies that in the case of approximative modelling, $G_0 \notin \mathcal{G}$, the undermodelling error is particularly small in this frequency region which is favourable in case the identified model is used in control design as is pointed out in Van den Hof and Schrama (1995).

Hence, the identification procedure described in the previous sections obtains a control-relevant model because of the implicit weighting. In many control-relevant identification schemes this type of weighting is pursued but can there only be approximated by use of specific filtering strategies, while by using a tailor-made parametrization this weighting is inherent.

The identification method using a tailor-made parametrization resembles the indirect identification method where first the closed-loop transfer function $R(q)$ is identified with a standard numerator-denominator parametrization. Next, a plant model is calculated using knowledge of the controller and

the closed-loop structure with

$\hat{G}(q) = R(q, \hat{\theta})(1 - R(q, \hat{\theta})C(q))^{-1}$. In this calculation the McMillan degree of the model will generally be larger than the McMillan degree of the estimated closed-loop transfer function. Hence, estimation of a model of the open loop transfer function with a prespecified model order is not a trivial task if the indirect method is used. This same mechanism holds true also for identification in the dual Youla parametrization, which is a direct generalization of the classical indirect method (Van den Hof and Callafon, 1996). Using a tailor-made parametrization a plant model can be estimated with prespecified complexity.

8 Conclusions

In this paper identification of a model from closed-loop data with a tailor-made parametrization is discussed. Special attention is given to the possible occurrence of a non-connected parameter set which is induced by the structure enforced on the parametrization.

Sufficient conditions are derived for the model order in terms of the controller complexity such that the parameter set is connected. These conditions indicate that the parameter set may not be a connected set in case a low complexity model is identified from data with a high complexity controller.

From simulations it follows that the approach can yield very accurate models also in case of approximative modelling with a bad signal-to-noise ratio. However, complexity of the optimization problem involved needs to be investigated more thoroughly to assess the possible problem of local minima and saddlepoints and the identification of an accurate initial model.

A compact description of the least squares prediction error criterion function, the gradient and the Hessian thereof is derived using Sylvester matrices. This can be used fruitfully in nonlinear optimization routines which have to be solved to obtain an estimate with a tailor-made parametrization.

Acknowledgements

The authors wish to thank Carsten Scherer for fruitful discussions on the results presented.

Appendix

Lemma 8.1 *The parameter set $\Theta \subset \mathbb{R}^n$ with elements $\theta = [p_1 \dots p_n]^T$, $\{p_i\}_{i=1, \dots, n} \in \mathbb{R}$ for which all polynomials*

$$p(z) = z^n + [z^{n-1} z^{n-2} \dots 1]\theta$$

have stable roots, is a pathwise connected subset of \mathbb{R}^n .

Proof: First the polynomial $p(z)$ is reparametrized as a product of first and second order polynomials

$$p(z) = \begin{cases} \prod_{k=1}^{n/2} (z^2 + a_k z + b_k), \forall k & n \text{ even} \\ (z + c) \prod_{k=1}^{n/2} (z^2 + a_k z + b_k), \forall k & n \text{ odd} \end{cases} \quad (12)$$

Stability of the full polynomial is guaranteed if stability of the second order polynomials and first order polynomial is guaranteed which is guaranteed if and only if $b_k < 1$, $a_k < 1 + b_k$, $-a_k < 1 + b_k$, $\forall k$ and $-1 < c < 1$, see e.g. Åström and Wittenmark (1990). This stability area for the quadratic terms describes a triangular area in the a_k, b_k -plane which is not only pathwise connected but also convex. The stability area for the first order term is also convex. The polynomial coefficients of the original polynomial, $\{p_i\}_{i=1, \dots, n}$, are continuous and continuously differentiable functions in the parameters $\{a_k, b_k\}_{k=1, \dots, n}$. Therefore from pathwise connectedness of the set of admissible coefficients $\{a_k, b_k\}_{k=1, \dots, n}$, pathwise connectedness of the set of admissible parameters $\{p_i\}_{i=1, \dots, n}$ can be concluded. \square

References

- Åström, K. J. and B. Wittenmark (1990). *Computer Controlled Systems - Theory and Design*. Prentice-Hall Inc., Englewood Cliffs, NJ.
- Chen, C. (1984). *Linear System Theory and Design*. Saunders College Publishing, USA.
- Gevers, M. (1993). Towards a joint design of identification and control? *Essays on control: Perspectives in the Theory and its Applications*, Birkhäuser, Boston, pp. 111-151.
- Landau, I.D. and K. Boumaïza (1996). An output error recursive algorithm for unbiased identification in closed loop. *Proc. 13th IFAC World Congress*, San Francisco, USA, Vol. I, pp. 215-220.
- Lee, W.S., B.D.O. Anderson, R.L. Kosut and I.M.Y. Mareels (1993). On robust performance improvement through the windsurfer approach to adaptive robust control. *Proc. 32nd IEEE Conf. Decision and Control*, San Antonio, TX, pp. 2821-2827.
- Ljung, L. (1987). *System identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.
- Schrama, R.J.P. (1992). *Approximate Identification and Control Design*. Ph.D. Thesis, Mech.Eng. Sys. and Contr. Group, Delft Univ. of Techn., The Netherlands.
- Söderström, T. and P. Stoïca (1989). *System Identification* Prentice-Hall, Hemel Hempstead, U.K.
- Van den Hof, P.M.J. and R.J.P. Schrama (1993). An indirect method for transfer function estimation from closed loop data. *Automatica*, 29, 1523-1527.
- Van den Hof, P.M.J. and R.J.P. Schrama (1995). Identification and control - closed loop issues. *Automatica*, 31, 1751-1770.
- Van den Hof, P.M.J. R.J.P. Schrama, O.H. Bosgra and R.A. de Callafon (1995). Identification of normalised coprime plant factors from closed-loop experimental data. *Europ. J. Control*, 1, 62-74.
- Van den Hof, P.M.J. and R.A. de Callafon (1996). Multivariable closed-loop identification: from indirect identification to dual-Youla parametrization. *Proc. 35th IEEE Conf. Dec. and Control*, Kobe, Japan. (See also this issue, pp. 1-7).

Identification in view of control design of a CD player[‡]

Hans G.M. Dötsch[§], Paul M.J. Van den Hof, Okko H. Bosgra and Maarten Steinbuch[¶]

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: H.Dotsch@wbmt.tudelft.nl*

Abstract. Electro mechanical servo systems, as encountered in consumer electronic products, have to keep pace with increasingly high performance demands. As the mechanical construction is a restricting factor regarding the limits of achievable performance, model based control design is proposed to enhance the bandwidth. System identification proves to be an adequate tool to produce nominal models and uncertainty models that are suitable for control design purposes. A method based on performing identification and control design in an iterative manner is proposed in order to systematically enhance the disturbance attenuation properties of a servo system. The proposed method is experimentally verified via application to a compact disc servo mechanism.

Keywords. Control design, system identification, compact disc player.

1 Introduction

A large number of applications of electro mechanical servo systems requires tracking with an increasingly high accuracy at a high speed. Especially in the field of consumer electronic products like audio and video systems the limits of achievable performance are more and more dictated by the mechanical construction of the servo system. In many cases this predominantly results in a desired enhancement of disturbance rejection of the servo system which may be achieved by control design. Design of control systems that establish an improved disturbance attenuation for electro mechanical constructions is however known to be hindered by the presence of resonance modes that are (in most cases) not exactly known. An additional aspect regarding consumer electronic products is the variability of system dynamics due to tolerances in the mass production process. This

motivates development of a tool that on one hand establishes an improved performance of existing constructions having variable dynamical properties and on the other hand explores the physical limitations of electro mechanical constructions in the stage of product development.

In case knowledge of resonance dynamics is sufficiently accurate a high bandwidth controller, designed based on this knowledge, is likely to provide a high bandwidth for the system without causing unstable behaviour. Therefore accurate knowledge of resonance modes is indispensable in the design of a high bandwidth control system. This motivates the use of model based control design as a tool to achieve an enhanced bandwidth for an electro mechanical servo system. Knowledge of resonance dynamics can adequately be described by a mathematical model which serves as a basis for control design. One way to construct such a model is to use relations based on first principles. In general these models are quite elaborate which inevitably leads to a controller of high dynamical order. If measurements can be taken from the system, models can also be obtained from experimental data utilizing system identification techniques. As experimental models are not

[‡]This paper is presented at the Philips Conference on Applications of Control Technology, PACT'96, 29-30 October 1996, Epe, The Netherlands.

[§]The work of Hans Dötsch is financially supported by Philips Research Laboratories, Eindhoven, The Netherlands.

[¶]Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA, Eindhoven, The Netherlands.

based on the physical structure of the system, the order may be kept low in order to describe relations induced by measured data. Therefore experimental modelling is employed in this paper.

As the intended use of the model is control design, the identification problem we are confronted with is to come up with a low order model such that a resulting controller establishes a high performance for the true system. In literature (a.o. Gevers, 1993; Van den Hof and Schrama, 1995) it has been recognized that, in order to establish an enhanced performance for the system through model based control design, identification and control design should be performed in an iterative manner. The topic of this paper is to propose an iterative scheme of identification and control design for systematic enhancement of the closed loop bandwidth and to verify the method experimentally on the servo mechanism of a compact disc (CD) player. A specific feature of the proposed scheme is the utilization of uncertainty models, that may be obtained through recently developed identification techniques (see a.o. de Vries, 1994).

In section 2 the need for an iterative approach of identification and control design is illustrated in view of achieving an enhanced closed loop bandwidth. The identification of nominal models and uncertainty models is the subject of section 3. The control design method employed is a two-stage procedure that combines a loop shape design with robustness in view of resonance modes. This is the subject of section 4. Results obtained from an experimental set up of a CD player are presented and commented upon in section 5. Conclusions and remarks conclude the paper.

2 Model based performance enhancement

We consider a servo system consisting of an electro mechanical actuator, denoted as P_0 , and a controller C as depicted in the block scheme of figure 1. In many cases the actuator is marginally stable and must therefore operate in closed loop. The signal d represents a reference signal that is not available from measurement but is to be tracked by the actuator output y . As d is presumed to be unknown it is regarded as a disturbance acting on the servo system. The servo error is denoted by e .

The desired performance is achieved in case the tracking error satisfies $|e(t)| < \delta, \forall t$ where the value of δ is determined by physical system properties. Although the translation is not one-to-one, the performance spec is expressed in the frequency domain¹

¹The frequency argument ω is left out for brevity

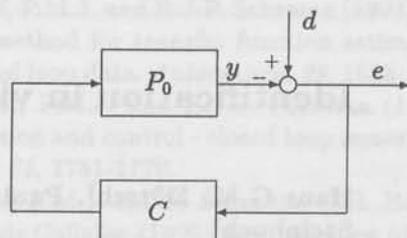


Fig. 1: Block scheme of an electro mechanical servo system

in terms of the following specification regarding the sensitivity function of the closed loop system:

$$\left| \frac{e}{d} \right| = |(1 + CP_0)^{-1}| \leq \beta^{-1}, \quad \forall \omega. \quad (1)$$

where β denotes the minimal disturbance rejection required for the system. We focus on the design of a controller C that establishes the specified disturbance rejection. To that end knowledge of P_0 as well as the disturbance d is indispensable which motivates the need for accurate models of both the actuator and the disturbance. Although it is acknowledged that disturbance modelling should be incorporated in the overall design, here we restrict attention to data based modelling of the actuator dynamics.

In literature the problem of identification of models that are suitable for high performance control design has received a great deal of attention (a.o. Gevers, 1993; Van den Hof and Schrama, 1995). It has been stressed that a model that provides a satisfactory description of the open loop system dynamics might provide a poor basis for control design, in the worst case resulting in controllers that destabilize the closed loop system. The main observation made is that system dynamics that govern the closed loop dynamics in conjunction with a controller often only marginally contribute to the open loop dynamics and vice versa.

This observation has resulted in a widely accepted strategy that identification of models suitable for control design should be performed in a closed loop situation, in the presence of a controller. To do closed loop identification we need a controller that emphasizes the dynamics that are relevant for control design. However, in order to find such a controller a model is required that encompasses control relevant dynamics. Here we are confronted with a circular reasoning that has motivated the proposition of algorithms where identification and control

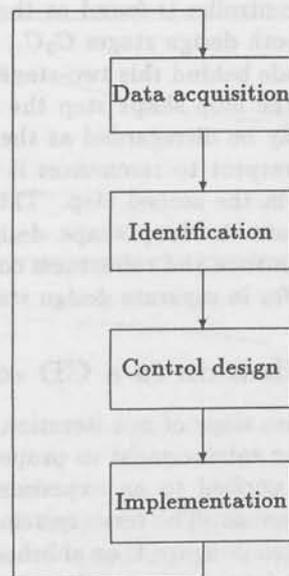


Fig. 2: Iterative approach of model based performance enhancement

design are performed in an iterative manner (Gevers, 1993; Van den Hof and Schrama, 1995; Lee *et al.*, 1995) in order to arrive at an enhanced performance. Basically such a procedure consists of the following steps: data acquisition, identification, control design and controller implementation; this is schematically depicted in figure 2.

In the sequel of this paper the separate steps of identification and control design are addressed. As the procedure is implemented on an experimental set up of a CD servo mechanism, the elaboration from here on is directed towards this application.

3 Identification

In figure 3 a block scheme of the experimental CD player is depicted where time domain signals r, u and e are available from measurement. The signal r is used for excitation, u and e are the input resp. output of the actuator, measured in the presence of a stabilizing controller.

The identification of a parametric model is concerned with estimation of parameters in a predetermined model structure. The data underlying the identification procedure is a frequency domain representation of measured time sequences by means of a discrete Fourier transform in conjunction with periodic excitation. The data are available as (complex valued) data points $\{r(e^{i\omega_j}), u(e^{i\omega_j}), e(e^{i\omega_j})\}$ at a finite number of user specified frequencies $\omega_j, j = 1, \dots, N$. The main motivation for trans-

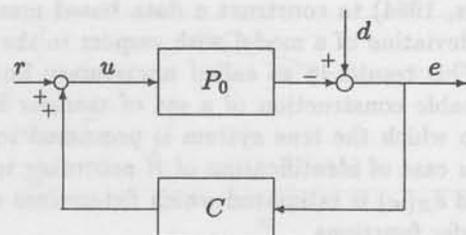


Fig. 3: Experimental configuration of a CD servo system

forming time domain data to the Fourier domain prior to identification is the possibility to establish a considerable compression of the amount of data. Moreover, a frequency domain data representation is compatible with the performance specification (1). Two features are characteristic for the identification problem addressed. Firstly, a model that is employed for control design should be obtained from measurements taken in the presence of a controller, as mentioned in section 2. Secondly, as the identification has to be performed in closed loop, straightforward application of open loop identification methods is hazardous. To that end the identification of a parametric model is performed following a so called indirect approach. A parametric model of the closed loop transfer $P_0(I + CP_0)^{-1}$ is estimated from $\{r(e^{i\omega_j}), e(e^{i\omega_j})\}$ by determining parameters, denoted θ , that minimize the following least squares criterion function:

$$\sum_{j=1}^N |\mathcal{W}(e^{i\omega_j})[e(e^{i\omega_j}) - R(e^{i\omega_j}, \theta)r(e^{i\omega_j})]|^2 \quad (2)$$

where

$$R(e^{i\omega}, \theta) := \frac{b_0 + b_1 e^{-i\omega} + \dots + b_n e^{-in\omega}}{1 + a_1 e^{-i\omega} + \dots + a_n e^{-in\omega}} \quad (3)$$

and \mathcal{W} is a frequency dependent weighting function. A model of the system is constructed from $R(e^{i\omega}, \theta)$, utilizing knowledge of the (stable) controller, as follows:

$$P(e^{i\omega}, \theta) = \frac{R(e^{i\omega}, \theta)}{I - CR(e^{i\omega}, \theta)} \quad (4)$$

It is mentioned that a generalization of this approach is applied, allowing to deal with marginally stable controllers, as is indicated by Van den Hof and de Callafon (1996). The identification of a (low order) model $P(e^{i\omega}, \theta)$ suitable for control design in the SISO case amounts to specifying a suitable weighting \mathcal{W} in (2).

In addition to identification of nominal models, techniques have recently been developed (see a.o.

de Vries, 1994) to construct a data based measure of the deviation of a model with respect to the system. This results in so called uncertainty bounds that enable construction of a set of transfer functions to which the true system is presumed to belong. In case of identification of \hat{R} according to (2) a bound $\delta_R(\omega)$ is estimated which determines a set of transfer functions

$$\mathcal{R} = \{R | R = \hat{R} + \Delta_R, |\Delta_R| < \delta_R(\omega)\} \quad (5)$$

to which the true transfer function R_0 belongs. The motivation for employing uncertainty models in this specific model structure is that they are instrumental in predicting the closed loop dynamics for a set of systems in conjunction with any (stabilizing) controller, as is elaborated by Van den Hof *et al.* (1996). This is a potentially powerful technique to incorporate the aspect of variable system dynamics into the control design.

4 Control design

In this section a nominal control design procedure is presented that is proposed by McFarlane and Glover (1990). The design procedure is solely based on nominal models but has favourable robustness properties and consists of two consecutive stages. The first stage is the determination of a loop shape transfer function C_0 such that the nominal sensitivity function satisfies a minimum prespecified magnitude bound:

$$|(I + C_0 \hat{P})^{-1}| \leq \beta^{-1}, \quad \forall \omega. \quad (6)$$

The determination of C_0 is done by visual inspection of the Bode diagram and Nyquist contour of $C_0 \hat{P}$ where the structure of C_0 is predetermined in terms of a low order lead-lag compensator. Although loop shaping is an appealing technique due to the fact that compensators result from visual inspection, it is not a very robust technique for high bandwidth design especially in case resonance modes are present in the model \hat{P} . Therefore robustness properties are improved in the second stage which consists of a norm based control design, where a controller is determined such that the following criterion function is minimized

$$\|T(C_0 \hat{P}, C)\|_{\infty} \quad (7)$$

where $T(P, C)$ is a 2×2 matrix that comprises the closed loop transfer functions from r to $[e \ u]^T$ in the block scheme of figure 3, defined as²

$$T(P, C) := \begin{bmatrix} P \\ I \end{bmatrix} [I + CP]^{-1} [C \ I]. \quad (8)$$

²Note that the sensitivity is the (2, 2)-element of $T(P, C)$.

The final controller is found as the product of the results of both design stages $C_0 C$.

The rationale behind this two-stage design strategy is that in the loop shape step the presence of resonances may be disregarded as the desired robustness with respect to resonances is supposed to be dealt with in the second step. This may considerably facilitate the loop shape design in the sense that performance and robustness considerations are accounted for in separate design steps.

5 Application to a CD servo system

The separate steps of one iteration of model based performance enhancement as proposed in section 2 have been applied to an experimental CD player servo mechanism. The servo system, as is schematically depicted in figure 1, establishes track following of digital information stored on a rotating optical disc.

Attention is restricted to the radial part of the mechanism (Single Input Single Output case). Controller implementation and data acquisition are carried out utilizing a DSP signal processor (dSPACE GmbH, 1995) at a sample rate of 25 kHz. Measurements are taken of 40 time sequences of $\{r, u, e\}$ each containing 4096 data points where the excitation signal r is chosen as a random phased multisine, exciting the system at 99 logarithmically spaced frequencies between 100 Hz and 10 kHz. A 4th order compensator is present in the loop during measurement.

A nominal parametric model of order 10 is identified according to (2) together with an upper bound of model uncertainty. The frequency response and the nominal model are shown in the Bode diagram of figure 4. The nominal model seems to provide a rather poor description of the data in the low and high frequency region.

Based on this nominal model a 4th order lead-lag compensator C_0 is designed on visual inspection of Bode magnitude diagrams of the nominal sensitivity and the predicted sensitivity, constructed from uncertainty bounds. The compensator is adjusted to a higher bandwidth until the nominal sensitivity function will (inevitably) peak up at frequencies beyond the bandwidth. Figure 5 shows the Bode diagram of the measured sensitivity and the nominally designed sensitivity.

Besides visual inspection of the nominal sensitivity also the actually achieved sensitivity is evaluated in terms of lower and upper magnitude bounds of the sensitivity, constructed from estimated uncertainty bounds of the model; this is depicted in figure 6 together with the nominally designed sensitivity function. The design of the loop shape function is per-

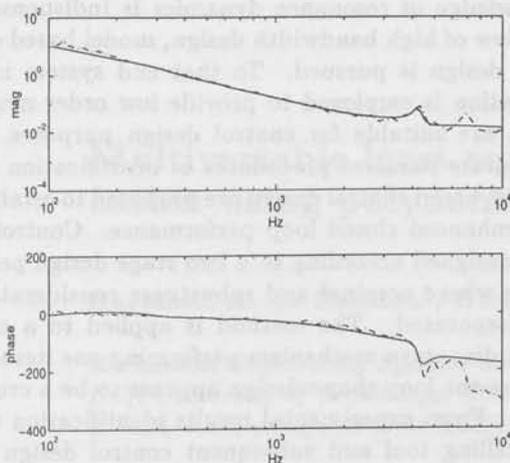


Fig. 4: Measured frequency response (dash-dot) and 10th order nominal model (solid) of the radial CD servo mechanism.

formed such that the nominal sensitivity magnitude remains within the bounds and the bounds are not too large.

To verify the validity of this closed loop system set the sensitivity frequency response measured with the loop shape compensator is added. The measured sensitivity is predicted quite well by the upper and lower magnitude bounds up to 4 kHz, while the designed sensitivity function is not captured by the bounds. This can be attributed to the fact that the (low order) nominal model lacks system dynamics which seem to be relevant in view of the newly designed controller.

The second stage of the control design is performed according to (7). The final controller C_0C is restricted to order 6 due to implementation limitations. This implies that the norm based design step produces controllers of order 2. To analyse the merits of the second control design step, the sensitivity function is measured with the enhanced controller (order 6). The Bode magnitude diagram in figure 8 shows the initial sensitivity function and the enhanced sensitivity. The loop shape compensator and the corresponding final controller are shown in figure 7.

In figure 9 the radial tracking error measured with the low bandwidth compensator and the enhanced compensator is shown. It is evident that increasing the bandwidth is a valid strategy in order to establish a reduction of the tracking error.

An important observation is that the loop shape design is a very crucial stage in the iterative approach. If the nominal design provides a relatively large increase of the bandwidth in comparison to the controller present during measurement (as is illustrated figure 5), then the nominal model may not reliably

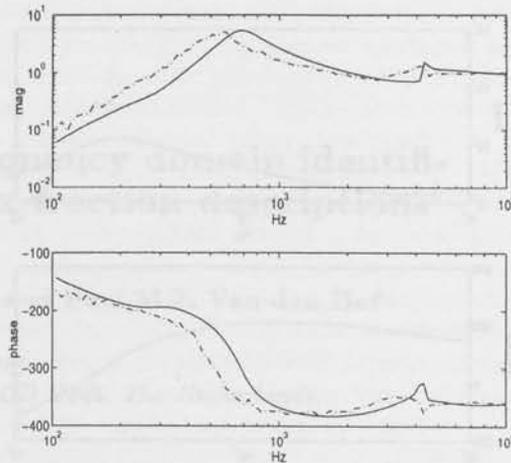


Fig. 5: Sensitivity function: enhanced loop shape design (solid) and measured with low bandwidth compensator in the loop (dash-dot).

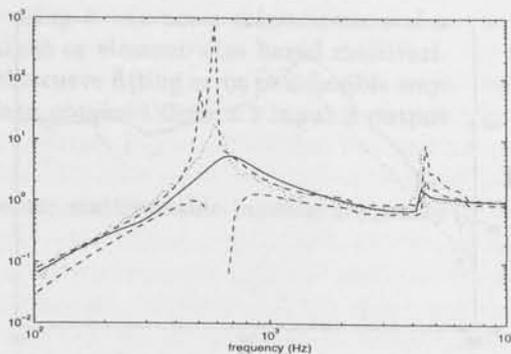


Fig. 6: Magnitude of sensitivity: nominal (solid), lower and upper bounds (dash) and measured (dot).

predict the actual sensitivity. This is in fact illustrated in figure 6 where the nominally designed sensitivity is not completely captured by the magnitude bounds; the nominal design appears not to be very robust. In the line of performing several iterations (here we have only considered one iteration) it is important to take small steps in the nominal loop shape design towards a higher bandwidth in order to maintain a robust design. This has yet to be verified.

6 Conclusions

To comply with increasing higher demands of servo systems as encountered in consumer electronic products, control design is used to obtain a high bandwidth. A crucial issue in designing a high bandwidth control system for electro mechanical servo systems is the presence of (unknown) resonance modes. As

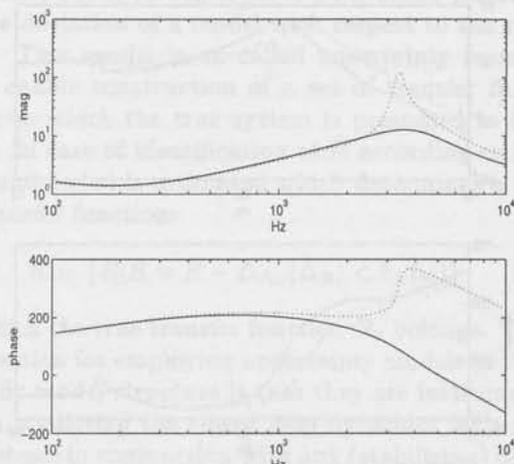


Fig. 7: Loop shape compensator C_0 (solid) and final controller C_0C (dot).

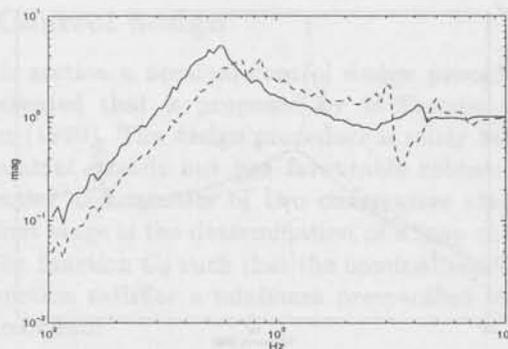


Fig. 8: Magnitude of sensitivity: initial compensator (solid) and enhanced compensator (dash).

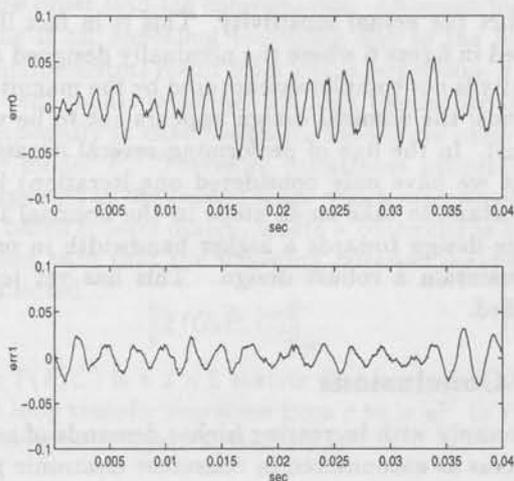


Fig. 9: Radial tracking error: initial compensator (upper) and enhanced compensator (lower).

knowledge of resonance dynamics is indispensable in view of high bandwidth design, model based control design is pursued. To that end system identification is employed to provide low order models that are suitable for control design purposes. In literature iterative procedures of identification and model based control design are proposed to establish an enhanced closed loop performance. Controllers are designed according to a two stage design procedure where nominal and robustness considerations are separated. The method is applied to a compact disc servo mechanism performing one iteration where the loop shape design appears to be a crucial step. From experimental results identification as a modelling tool and subsequent control design appear to be fruitful in order to arrive at an enhanced performance of the servo system.

References

- dSPACE GmbH (1995). *User's Guide Digital Signal Processing and Control Engineering*, Paderborn, Germany.
- de Vries, D.K. (1994). *Identification of Model Uncertainty for Control Design*. Dr. Dissertation, Delft University of Technology, The Netherlands.
- Gevers, M. (1993). Towards a joint design of identification and control? In: H.L. Trentelman and J.C. Willems (Eds.), *Essays on Control: Perspectives in the Theory and its Applications*. Birkhäuser, Boston, pp. 111-151.
- Lee, W.S., B.D.O. Anderson, I.M.Y. Mareels and R.L. Kosut (1995). On some key issues in the Windsurfer Approach to adaptive robust control. *Automatica*, 31, 1619-1636.
- McFarlane, D.C. and K. Glover (1990). *Robust Controller Design Using Normalized Coprime Factor Plant Descriptions*. Lect. Notes in Control and Inform. Sciences, Vol. 138, Springer Verlag.
- Van den Hof, P.M.J. and R.J.P. Schrama (1995). Identification and control - closed loop issues. *Automatica*, 31, 1751-1770.
- Van den Hof, P.M.J., E.T. van Donkelaar, H.G.M. Dötsch and R.A. de Callafon (1996). Control relevant uncertainty modelling directed towards performance robustness. *Proc. 13th IFAC World Congress*, San Fransisco, Vol. I, pp. 103-108.
- Van den Hof, P.M.J. and R.A. de Callafon (1996). Multivariable closed-loop identification: from indirect identification to dual Youla parametrization. To appear in: *Proc. 35th IEEE Conf. on Dec. and Control*, Kobe, Japan.

Multivariable least squares frequency domain identification using polynomial matrix fraction descriptions[‡]

Raymond A. de Callafon[§], Dick de Roover[¶] and Paul M.J. Van den Hof

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: r.a.decalfon@wbmt.tudelft.nl*

Abstract. In this paper an approach is presented to estimate a linear multivariable model on the basis of (noisy) frequency domain data via a curve fitting procedure. The multivariable model is parametrized in either a left or a right polynomial matrix fraction description and the parameters are computed by using a two-norm minimization of a multivariable output error. Additionally, input-output or element-wise based multivariable frequency weightings can be specified to tune the curve fitting error in a flexible way. The procedure is demonstrated on experimental data obtained from a 3 input 3 output Wafer Stepper system.

Keywords. System identification; frequency response; multivariable models; frequency weighting; least squares.

1 Introduction

Formulating a procedure that is able to estimate a model on the basis of frequency domain data has gained considerable attention in the research on system identification. Although the clear distinction between time and frequency domain data is generally overestimated (Ljung, 1993), estimation of models by fitting complex frequency domain data has several advantages compared to time domain approaches. Firstly, representing data in the frequency domain can yield substantial data reduction, see Pintelon *et al.* (1994). Secondly, compressing a huge amount of time domain data into a finite number of frequency points facilitates noise reduction directly. Both aspects are used extensively in commercially available sophisticated test equipment for spectral analysis.

Based on Least Squares (LS) estimation techniques,

as used in Levi (1959) and further refined in Sanathanan and Koerner (1963), multivariable frequency domain curve fitters have been formulated in the literature. One is referred to Lin and Wu (1982), Dailey and Lukich (1987) and the more recently introduced procedure in Bayard (1994). Basically, the procedures differ in the way the multivariable model is parametrized and whether or not the procedure allows for a specification of the model order and a (multivariable) weighting on the curve fit error. As such, in Lin and Wu (1982) a multivariable model is found by the composition of scalar subsystems, while the order of the subsequent transfer functions is determined by testing the residuals. A similar approach can be found in Dailey and Lukich (1987), wherein a Chebyshev polynomial basis is used to improve numerical conditioning of the LS-problem. In Bayard (1994) the model is parametrized directly by means of a matrix numerator polynomial and a scalar common denominator polynomial, whereas only a scalar frequency dependent weighting on the curve fit error is allowed.

Several alternatives to a LS-approach can also be found in the literature. In McKelvey (1995) a subspace based algorithm in the frequency domain is

[‡]This paper is presented at the 35th IEEE Conference on Decision and Control, 11-13 December 1996, Kobe, Japan. Copyright of this paper remains with IEEE.

[§]The work of Raymond de Callafon is financially supported by the Dutch Systems and Control Theory Network.

[¶]The work of Dick de Roover is financially supported by Philips' Research Laboratories, Eindhoven, the Netherlands.

presented that allows the user to specify an additional frequency weighting. In Hakvoort and Van den Hof (1994) a frequency domain curve fitter has been developed in which a maximum amplitude of a (weighted) curve fit error is being considered. Furthermore, so-called \mathcal{H}_∞ -identification procedures, currently applicable to scalar frequency domain data, can guarantee an upper bound on the additive error, see e.g. Gu and Khargonekar (1992) and the references therein. Unfortunately, a maximum amplitude criterion can be highly sensitive to noise, whereas the available \mathcal{H}_∞ -identification procedures might yield high order models for moderately damped processes (Friedman and Khargonekar, 1994).

Based on the LS-approach, this paper presents a multivariable frequency domain curve fitter in which the aim is to minimize the two-norm on a (weighted) curve fit error for a model having a limited McMillan degree. The multivariable model is parametrized by either a left or right polynomial Matrix Fraction Description (MFD). By use of Kronecker calculus it will be shown that both a pre, post or element-wise multivariable frequency weighting on the curve fit error can be handled relatively easily. Furthermore, it will be shown that the iteration described by Sanathanan and Koerner (1963), denoted by SK-iteration, can be generalized to estimate a polynomial MFD. Due to the subsequent convex optimization steps in the SK-iteration, this approach supports the estimation of models with many parameters. Similar to the approach followed by Bayard (1994) and supported by the work of Whitfield (1987), the resulting estimate can be used as an initial value for a Gauss-Newton optimization.

Although cumbersome iterations can be avoided by the use of a realization based algorithm as reported in McKelvey (1995), the possibility to prespecify the McMillan degree of the model and to introduce a flexible element-wise frequency weighting on the multivariable data is quite helpful from a practical point of view. The procedure will be illustrated by fitting a multivariable model on the frequency response obtained from the positioning mechanism present in a wafer stepper.

2 Problem formulation

To formulate the multivariable frequency domain identification problem, consider the following set \mathcal{G} of noisy complex frequency response data observations $G(\omega_j)$, evaluated at N frequency points ω_j .

$$\mathcal{G} := \{G(\omega_j) \mid G(\omega_j) \in \mathbb{C}^{p \times m}, \text{ for } j \in 1, \dots, N\} \quad (1)$$

The aim of the identification problem discussed in this paper is to find a linear time invariant multi-

variable model P of limited complexity, having m inputs and p outputs, that approximates the data \mathcal{G} in (1).

To address the limited complexity, the model $P(\theta)$ is parametrized by either a left or right polynomial MFD that depends on a real valued parameter θ of limited dimension. The specific parametrization of the polynomial MFD of $P(\theta)$ is discussed in the next section. The approximation of the data \mathcal{G} by the model $P(\theta)$ is addressed by considering the following additive error.

$$E_a(\omega_j, \theta) := [G(\omega_j) - P(\xi(\omega_j), \theta)] \text{ for } j \in 1, \dots, N \quad (2)$$

The complex variable $\xi(\cdot)$ in (2) is used to denote the frequency dependency of the model $P(\theta)$. In this way, $\xi(\omega_j) = i\omega_j$ to represent a continuous time model, whereas $\xi(\omega_j) = e^{i\omega_j T}$ (shift operator) or $\xi(\omega_j) = (e^{i\omega_j} - 1)/T$ (δ operator) to represent a discrete time model with sampling time T .

To tune the additive error E_a in (2), both an input-output frequency weighted curve fit error E_w with

$$E_w(\omega_j, \theta) := W_{out}(\omega_j)E_a(\omega_j, \theta)W_{in}(\omega_j) \quad (3)$$

and an element-wise frequency weighted curve fit error E_s with

$$E_s(\omega_j, \theta) := S(\omega_j) .* E_a(\omega_j, \theta) \quad (4)$$

will be considered in this paper. In (4) $.*$ is used to denote the Schur product; an element-by-element multiplication.

Using the notation E to denote the frequency weighted curve fit error E_w in (3) and E_s in (4), the deviation of the data \mathcal{G} is characterized by following the norm function $J(\theta)$.

$$J(\theta) := \sum_{i=1}^N \text{tr}\{E(\omega_j, \theta)E^*(\omega_j, \theta)\} = \|E(\theta)\|_F^2 \quad (5)$$

In (5) $*$ is used to denote the complex conjugate transpose, $\text{tr}\{\cdot\}$ is the trace operator and $\|E(\theta)\|_F$ denotes the Frobenius norm operating on the matrix $E(\theta) = [E(\omega_1, \theta) \dots E(\omega_N, \theta)]$. Consequently, the goal of the procedure described in this paper is to find a real valued parameter $\hat{\theta}$ of limited complexity that can be formulated by the following minimization.

$$\hat{\theta} := \arg \min_{\theta \in \mathbb{R}} J(\theta) \quad (6)$$

3 Parametrization

3.1 Polynomial matrix fraction descriptions

The multivariable model is represented by either a left or right polynomial MFD, respectively given by

$$P(\xi, \theta) = A(\xi^{-1}, \theta)^{-1}B(\xi^{-1}, \theta) \quad (7)$$

$$P(\xi, \theta) = B(\xi^{-1}, \theta)A(\xi^{-1}, \theta)^{-1} \quad (8)$$

where A and B denote parametrized polynomial matrices in the indeterminate ξ^{-1} .

For a model having m inputs and p outputs, the polynomial matrix $B(\xi^{-1}, \theta)$ is parametrized by

$$B(\xi^{-1}, \theta) = \sum_{k=d}^{d+b-1} B_k \xi^{-k} \quad (9)$$

where $B_k \in \mathbb{R}^{p \times m}$, d denotes the number of leading zero matrix coefficients and b the number of non-zero matrix coefficients in $B(\xi^{-1}, \theta)$. For the left MFD in (7), $A(\xi^{-1}, \theta)$ is parametrized by

$$A(\xi^{-1}, \theta) = I_{p \times p} + \xi^{-1} \sum_{k=1}^a A_k \xi^{-k+1} \quad (10)$$

where $A_k \in \mathbb{R}^{p \times p}$ and a denotes the number of non-zero matrix coefficients in the monic polynomial $A(\xi^{-1}, \theta)$. The parameter θ is determined by the corresponding unknown matrix coefficients in the polynomials. Hence,

$$\theta = [B_d \ \cdots \ B_{d+b-1} \ A_1 \ \cdots \ A_a] \quad (11)$$

and $\theta \in \mathbb{R}^{p \times (mb+pa)}$ for the left MFD in (7). Dual results can be formulated for the right MFD in (8). Additionally to the full polynomial parametrization presented here, so-called structural parameters d_{ij} , b_{ij} and a_{ij} with $d := \min\{d_{ij}\}$, $b := \max\{b_{ij}\}$, and $a := \max\{a_{ij}\}$ can be used to specify a none-full polynomial parametrization. In this way, the parameter θ as given in (11) may contain prespecified zero entries at specific locations. This may occur in a discrete time model with $\xi^{-1} = z^{-1}$ where the value of d_{ij} has a direct connection with the number of time delays from the j th input to the i th output.

3.2 Model order

Due to the indeterminate ξ^{-1} , it can be verified that the MFD of (7) or (8) gives rise to a (strictly) proper transfer function matrix $P(\xi, \theta)$, regardless of the value of the integers d_{ij} , b_{ij} or a_{ij} . Hence, there are no restrictions on the size of the structural parameters, other than a limitation on the McMillan degree of the resulting model $P(\xi, \hat{\theta})$. For the connection between the structural parameters and the McMillan degree of $P(\xi, \theta)$, the following result can be given.

Lemma 3.1 Consider a parameter $\hat{\theta}$ such that $A_a \neq 0$ and $B_{d+b-1} \neq 0$. Define

$$\eta := \max\{a, d + b - 1\} \quad (12)$$

and $\bar{A}(\xi, \hat{\theta}) := \xi^\eta A(\xi^{-1}, \hat{\theta})$, $\bar{B}(\xi, \hat{\theta}) := \xi^\eta B(\xi^{-1}, \hat{\theta})$. Let n be used to denote the McMillan degree of

the multivariable transfer function model $P(\xi, \hat{\theta})$ obtained by (7) or (8), then

$$n = \deg \det\{\bar{A}(\xi, \hat{\theta})\}$$

if and only if $\bar{A}(\xi, \hat{\theta})$ and $\bar{B}(\xi, \hat{\theta})$ are left coprime over $\mathbb{R}[\xi]$ in case of (7) and right coprime over $\mathbb{R}[\xi]$ in case of (8).

Proof: The proof is given for (8). With the condition $A_a \neq 0$, $B_{d+b-1} \neq 0$, it follows that $\bar{A}(\xi) := \xi^\eta A(\xi^{-1})$ and $\bar{B}(\xi) := \xi^\eta B(\xi^{-1})$ are polynomial matrices in ξ . In case of (8), $P(\xi) = \bar{B}(\xi)\bar{A}(\xi)^{-1}$ and a state space realization $[A, B, C, D]$ for $P(\xi)$ can be obtained, such that $\dim A = \deg \det\{\bar{A}(\xi)\}$ and $\{A, B\}$ controllable, see e.g. Chen (1984). Furthermore, $\{C, A\}$ is observable if and only if $\bar{A}(\xi)$ and $\bar{B}(\xi)$ are right coprime over $\mathbb{R}[\xi]$, see theorem 6.1 in Chen (1984). Dually, the result can be shown for (7). \square

Under some mild condition on the polynomials $A(\xi^{-1}, \hat{\theta})$ and $B(\xi^{-1}, \hat{\theta})$ being estimated, lemma 3.1 gives a direct relation between the $\deg \det\{\bar{A}(\xi, \hat{\theta})\}$ and the McMillan degree of the resulting estimate $P(\xi, \hat{\theta})$. In case of the left MFD (7), $\deg \det\{\bar{A}(\xi, \hat{\theta})\}$ generally will be equal to ηp . Hence, the structural parameters give rise to (an upper bound) on the McMillan degree of the model being estimated. For a more detailed discussion on the exact relation between the McMillan degree, the row degree of the polynomial matrices $A(\xi^{-1}, \theta)$, $B(\xi^{-1}, \theta)$ and the observability indices of a model computed by a left polynomial MFD, one is referred to Gevers (1986) or Van den Hof (1992).

Compared to a parametrization of the multivariable model $P(\xi, \theta)$ using a scalar common denominator polynomial $d(\xi^{-1}, \theta)$ as presented in Bayard (1994), the parametrization using a (left) MFD is more flexible, as a scalar common denominator restricts $A(\xi^{-1}, \theta)$ to be $I_{p \times p} d(\xi^{-1}, \theta)$. A model with one output that is parametrized by the left MFD of (7), constitutes a scalar common denominator polynomial $A(\xi^{-1}, \theta)$.

4 Computational procedure

4.1 Iterative minimization

In this section, the minimization of (6) will be discussed by means of an iterative procedure of convex optimization steps similar to the SK-iteration of Sanathanan and Koerner (1963). The attention will be restricted to a parametrization of $P(\xi, \theta)$ based on the left MFD (7) as dual results can be obtained for a right MFD. To extend the SK-iteration to the

multivariable case, first consider the (unweighted) additive curve fit error of (2).

For a model $P(\xi, \theta)$ parametrized by left MFD, (2) can be written as

$$E_a(\omega_j, \theta) = A(\xi(\omega_j)^{-1}, \theta)^{-1} \tilde{E}(\omega_j, \theta) \quad (13)$$

where $\tilde{E}(\omega_j, \theta)$ is the equation error defined by

$$\tilde{E}(\omega_j, \theta) := A(\xi(\omega_j)^{-1}, \theta)G(\omega_j) - B(\xi(\omega_j)^{-1}, \theta). \quad (14)$$

Substituting the parametrization (7) for the polynomials A, B , the equation error in (14) can be represented by

$$\tilde{E}(\omega_j, \theta) = G(\omega_j) - \theta \Phi(\omega_j) \quad (15)$$

where θ is given in (11) and

$$\Phi(\omega_j) = \begin{bmatrix} I_{m \times m} \xi(\omega_j)^{-d} \\ \vdots \\ I_{m \times m} \xi(\omega_j)^{-(d+b-1)} \\ G(\omega_j) \xi(\omega_j)^{-1} \\ \vdots \\ G(\omega_j) \xi(\omega_j)^{-a} \end{bmatrix} \quad (16)$$

with $\Phi(\omega_j) \in \mathbb{C}^{(mb+pa) \times m}$.

A matrix $\tilde{E}(\theta)$ can be formed by stacking $\tilde{E}(\omega_j, \theta)$ column-wise for $j \in 1, \dots, N$ and this yields

$$\arg \min_{\theta \in \mathbb{R}} \|\tilde{E}(\theta)\|_F^2 = \arg \min_{\theta \in \mathbb{R}} \|G - \theta P\|_F^2 \quad (17)$$

where G and P are found by stacking the real and imaginary part of respectively $G(\omega_j)$ and $\Phi(\omega_j)$ for $j \in 1, \dots, N$. Due to the linear appearance of the parameter θ , (17) corresponds a standard least squares problem that can be solved by numerical reliable tools as e.g a QR-factorization with (partial) pivoting (Golub and Van Loan, 1989).

Due to the fact that $A(\xi^{-1}, \theta)$ in (13) also depends on the parameter θ , the linear appearance of the parameter θ in (13) is violated. In order to facilitate the convexity in minimizing the two-norm on the equation error in (17), an iterative procedure similar as in Sanathanan and Koerner (1963) can be used. An estimate $\hat{\theta}_t$ in step t is computed by replacing $A(\xi(\omega_j)^{-1}, \theta)$ in (13) by a fixed $A(\xi(\omega_j)^{-1}, \hat{\theta}_{t-1})$ based on an estimate $\hat{\theta}_{t-1}$ obtained from the previous step $t-1$. In this way the Frobenius norm of an output weighted equation error $\tilde{E}_w(\omega_j, \hat{\theta}_{t-1}, \theta) = A(\xi(\omega_j)^{-1}, \hat{\theta}_{t-1})^{-1} \tilde{E}(\omega_j, \theta)$ needs to be minimized repeatedly according to

$$\hat{\theta}_t = \arg \min_{\theta \in \mathbb{R}} \|\tilde{E}_w(\hat{\theta}_{t-1}, \theta)\|_F^2.$$

This generalizes the SK-iteration to multivariable models parametrized by a left polynomial MFD. A dual approach can be formulated for a right polynomial MFD.

The estimate obtained from the SK-iteration is not optimal in the sense of (6) in presence of noise and/or incorrect model order, but it does provide a tool to find an initial estimate for a GN-optimization (Whitfield, 1987). Furthermore, the convex optimization to be solved in each step of the multivariable SK-iteration supports the estimation of models with many parameters. The computational procedure to obtain the parameter $\hat{\theta}$ in case of the (weighted) curve fit errors of (3) and (4) is presented in the subsequent sections.

4.2 Input-output weighting

The input-output weighted curve fit error of (3) can be rewritten into

$$E_w(\omega_j, \theta) = \tilde{W}_{out}(\omega_j, \theta) \tilde{E}(\omega_j, \theta) W_{in}(\omega_j) \quad (18)$$

where $\tilde{W}_{out}(\omega_j, \theta) := W_{out}(\omega_j) A(\xi(\omega_j)^{-1}, \theta)^{-1}$ and $\tilde{E}(\omega_j, \theta)$ is given in (14).

Using a similar approach of iterative minimization steps as used in section 4.1, the parameter θ in $\tilde{W}_{out}(\omega_j, \theta)$ in (18) is fixed to an estimate $\hat{\theta}_{t-1}$ obtained from the previous step $t-1$. Consequently, the weighted equation error \tilde{E}_w defined by

$$\tilde{E}_w(\omega_j, \hat{\theta}_{t-1}, \theta) := \tilde{W}_{out}(\omega_j, \hat{\theta}_{t-1}) \tilde{E}(\omega_j, \theta) W_{in}(\omega_j) \quad (19)$$

again indicates that the parameter θ to be estimated appears linearly in (19).

Although the free parameter θ appears linearly in (19), writing down a matrix representation for the weighted equation error \tilde{E}_w similar to (17) would inevitably lead to additional (large) sparse matrices that need to be stored in order to compute the least squares solution. The sparse matrices arise from the frequency dependent output (and input) weighting that need to be incorporated (Bayard, 1994). Furthermore, the parameter θ might have a structure containing zero entries at prespecified locations if a none-full polynomial parametrization is being used. To avoid the computational and memory storage issues that arise from dealing with (large) sparse matrices and to be able to take into account the specific structure that might be present in the parameter θ , a fairly simple and straightforward computational procedure based on Kronecker calculus is presented here. For this purpose consider the following definition.

Definition 4.1 Consider two matrices $X \in \mathbb{C}^{n_1 \times n_2}$ and $Y \in \mathbb{C}^{m_1 \times m_2}$, then the Kronecker vector

$\text{vec}(X) \in \mathbb{C}^{n_1 n_2 \times 1}$ and the Kronecker product $X \otimes Y \in \mathbb{C}^{n_1 m_1 \times n_2 m_2}$ are respectively defined by $\text{vec}(X) := [x_1 \cdots x_{n_2}]^T$ and

$$X \otimes Y := \begin{bmatrix} x_{1,1}Y & \cdots & x_{1,n_2}Y \\ \vdots & \cdots & \vdots \\ x_{n_1,1}Y & \cdots & x_{n_1,n_2}Y \end{bmatrix}$$

where $x_{i,j}$ and x_j for $i \in 1, \dots, n_1$ and $j \in 1, \dots, n_2$ are used to denote respectively the (i, j) th entry in X and the j th column in X .

The Kronecker product is a well known concept (Bellman, 1970) and by stacking the columns of a matrix to obtain the corresponding Kronecker vector as mentioned in definition 4.1, the following result can be obtained.

Proposition 4.2 Consider (complex) matrices X , Y and Z with appropriate dimensions, such that the matrix product $C := XYZ$ is well defined. Then $\text{vec}(C)$ satisfies

$$\text{vec}(C) = [Z^T \otimes X] \text{vec}(Y).$$

Proof: The proof can be found in Bellman (1970). \square

On the basis of proposition 4.2, the Kronecker vector of the input/output weighted equation error $\tilde{E}_w(\omega_j, \hat{\theta}_{t-1}, \theta)$ in (19) can be written as

$$\text{vec}(\tilde{E}_w) = \text{vec}(\tilde{W}_{out} G W_{in}) - [[\Phi W_{in}]^T \otimes \tilde{W}_{out}] \text{vec}(\theta)$$

wherein the arguments ω_j , $\hat{\theta}_{t-1}$ and θ are left out, to avoid notational issues. As the Frobenius-norm satisfies $\|X\|_F^2 = \|\text{vec}(X)\|_F^2$ for an arbitrary matrix X , the Frobenius-norm on \tilde{E}_w can be characterized by a matrix representation formed by stacking $\text{vec}(\tilde{E}_w(\omega_j, \hat{\theta}_{t-1}, \theta))$ row-wise for $j \in 1, \dots, N$. This yields the following estimate

$$\begin{aligned} \hat{\theta} &= \arg \min_{\theta \in \mathbb{R}} \|\text{vec}(\tilde{E}_w(\hat{\theta}_{t-1}, \theta))\|_F^2 \\ &= \arg \min_{\theta \in \mathbb{R}} \|G_w - P_w \theta\|_F^2 \end{aligned} \quad (20)$$

where $\tilde{\theta} = \text{vec}(\theta) \in \mathbb{R}^{p(mb+pa) \times 1}$ according to (11). Furthermore, $G_w \in \mathbb{R}^{2pmN \times 1}$ and $P_w \in \mathbb{R}^{2pmN \times p(mb+pa)}$ are matrices that can be found by row-wise stacking of the real and imaginary part of respectively $\text{vec}(\tilde{W}_{out}(\omega_j, \hat{\theta}_{t-1})G(\omega_j)W_{in}(\omega_j))$ and $\text{vec}([\Phi(\omega_j)W_{in}(\omega_j)]^T \otimes \tilde{W}_{out}(\omega_j, \hat{\theta}_{t-1}))$ for $j \in 1, \dots, N$.

The regression matrix P_w in (20) does not exhibit any sparse matrix structure as occurs e.g. in the method of Bayard (1994). In fact, $2pmN \times p(mb +$

$pa)$ entries is the smallest dimension of the regression matrix P_w in order to compute a least squares parameter $\hat{\theta}$ that has $p(mb + pa)$ unknown entries (for a left full polynomial parametrization) on the basis of N complex frequency domain points of a $p \times m$ multivariable system. In this way memory storage problems are avoided directly as much as possible.

As the parameter θ is converted into a column parameter $\tilde{\theta} = \text{vec}(\theta)$, any prespecified zero entries in $\tilde{\theta}$ can be incorporated in the estimation of the parameter relatively easy. This can be done by omitting the columns in the regression matrix P_w that correspond to the zero entries in $\tilde{\theta}$ and thereby reducing the size of the parameter to be estimated directly.

4.3 Schur weighting

Consider the Schur or element-wise frequency weighted curve fit error in (4) and rewrite this into

$$E_s(\omega_j, \theta) = S(\omega_j) * [A(\xi(\omega_j)^{-1}, \theta)^{-1} \tilde{E}(\omega_j, \theta)] \quad (21)$$

where the equation error $\tilde{E}(\omega_j, \theta)$ was defined in (14). Using a similar approach of iterative minimization steps as used in section 4.1, the parameter θ in $A(\xi(\omega_j)^{-1}, \theta)^{-1}$ in (21) is fixed to an estimate $\hat{\theta}_{t-1}$ obtained from the previous step $t - 1$. Consequently, the weighted equation error \tilde{E}_s defined by

$$\begin{aligned} \tilde{E}_s(\omega_j, \hat{\theta}_{t-1}, \theta) &:= \\ S(\omega_j) * [A(\xi(\omega_j)^{-1}, \hat{\theta}_{t-1})^{-1} \tilde{E}(\omega_j, \theta)] \end{aligned}$$

again indicates that the parameter θ to be estimated appears linearly. Finally, it can be verified (leaving out the arguments ω_j , $\xi(\omega_j)^{-1}$, $\hat{\theta}_{t-1}$ and θ) that $\text{vec}(\tilde{E}_s)$ can be rewritten into

$$\text{vec}(S * [A^{-1}G]) - \text{diag}(\text{vec}(S))[\Phi^T \otimes A^{-1}] \text{vec}(\theta) \quad (22)$$

by using the result of proposition 4.2. Hence, stacking $\text{vec}(\tilde{E}_s(\omega_j, \hat{\theta}_{t-1}, \theta))$ row wise for each $j \in 1, \dots, N$ will yield a similar expression for the minimizing argument $\hat{\theta}$ as given in (20). However, the matrix G_w in (20) now contains real and imaginary part of $\text{vec}(S(\omega_j) * [A(\xi(\omega_j)^{-1}, \hat{\theta}_{t-1})G(\omega_j)])$, whereas P_w in (20) will consist of the real and imaginary part of $\text{diag}(\text{vec}(S(\omega_j)))[\Phi(\omega_j)^T \otimes A^{-1}(\xi(\omega_j)^{-1}, \hat{\theta}_{t-1})]$ for $j \in 1, \dots, N$. Hence, the same computational procedure can be used to incorporate an element-by-element weighted curve fit error (4) by a slight modification of the matrices in (20).

5 Application to experimental data

5.1 Description of the wafer stepper system

The multivariable curve fit procedure discussed in this paper is illustrated by curve fitting experimental data obtained from a positioning system of a wafer stepper.

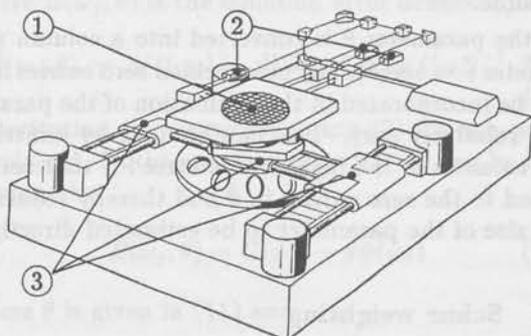


Fig. 1: Schematic view of a wafer stage; 1: wafer chuck, 2: laser interferometers, 3: linear motors.

A wafer stepper is a high accuracy positioning machine, used in chip manufacturing processes and a schematic view is depicted in figure 1. The wafer carries approximately 80 chips and is placed on a moving table, called the wafer chuck, which needs to be positioned accurately. The position of the wafer chuck on the horizontal surface of a granite block is measured by means of three laser interferometry measurements, whereas three linear motors are used to position the wafer chuck. In this way, the positioning system is considered to be a multivariable system, having three currents to the linear motors as inputs and three position measurements as outputs of the process.

5.2 Experimental results

Periodic random noise signals of 1024 points are used to excite the system. Using the resulting averaged time series, a spectral estimate is computed, resulting in a finite number of frequency domain data points that constitutes a suitable starting point for the subsequent curve fit procedure.

As the resulting model has to be used for discrete time control design purposes, the aim is to estimate a possibly *low order* discrete time multivariable model, that describes the dynamical behaviour of the positioning system in the frequency domain till approximately 400 Hz. For frequencies smaller than 100 Hz, the positioning system acts like a double integrator. To illustrate the usage of weighting functions in order to shape the curve fit error, an output weighting is used that emphasizes the frequency range between 200 and 300 Hz and starts to

roll off at 300 Hz. The order of the resulting multivariable model (without the 3 double integrators) is chosen to be 12, represented by a full left polynomial matrix fraction description having 81 parameters.

The SK-iteration is started up by first estimating a high order model to compute an initial value for the modified output weighting \bar{W}_{out} in (19). After this initialization, the SK-iteration is invoked 8 times. The Bode amplitude plot and phase plot of the 18th order estimate (including the 3 double integrators) is depicted respectively in figure 2 and figure 3. It should be noted that the multivariable output weighting applied during the estimation procedure emphasizes the frequency domain area of interest.

6 Conclusions

An approach is presented to estimate a linear multivariable model on the basis of noisy frequency domain data using a two-norm minimization of a weighted curve fit error. The weighting on the curve fit error can be specified by either an input/output or an element-by-element frequency dependent multivariable weighting function. The multivariable model is parametrized in either a left or right polynomial matrix fraction description wherein structural parameters allow the specification of both full polynomial or none-full polynomial descriptions. The computational procedure is able to estimate complex models by using an iterative procedure of solving weighted multivariable least squares problems and exploits the structure of the least squares problem, thereby reducing any computation and memory requirements directly. The curve is demonstrated on experimental multivariable frequency domain data obtained from a Wafer Stepper system having 3 inputs and 3 outputs.

7 Acknowledgements

The authors would like to acknowledge the support of a wafer stepper experimental setup by the Philips' Research Laboratory in Eindhoven, the Netherlands. Furthermore we like to thank Erwin Walgers, for his contribution to the experimental part of this paper.

References

- Bayard, D.S. (1994). High order multivariable transfer function curve fitting: algorithms, sparse matrix methods and experimental results. *Automatica*, 30, 1439-1444.
- Bellman, R. (1970). *Introduction to Matrix Computations*. McGraw-Hill, NY.

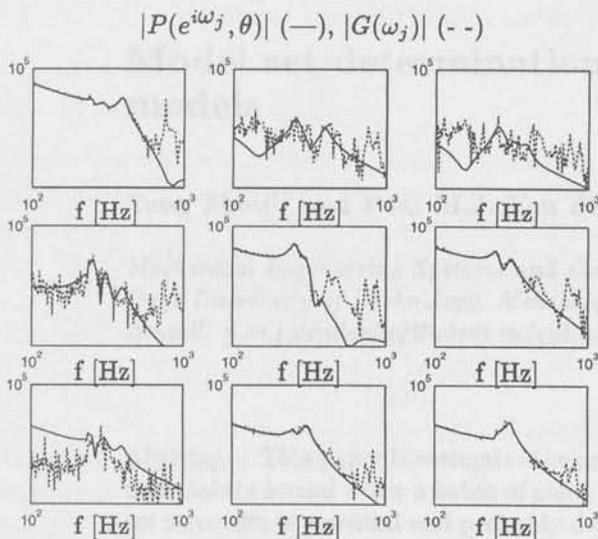


Fig. 2: Amplitude Bode plot of 18th order discrete time model $P(e^{i\omega_j}, \hat{\theta})$ and the data $G(\omega_j)$.

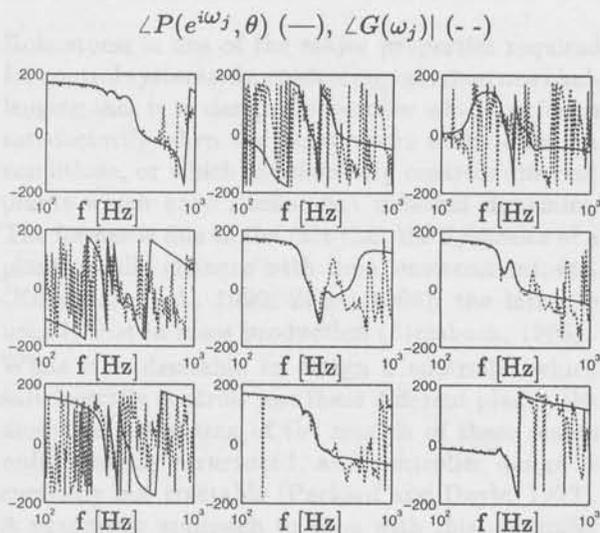


Fig. 3: Phase Bode plot of 18th order discrete time model $P(e^{i\omega_j}, \hat{\theta})$ and the data $G(\omega_j)$.

Chen, C.T. (1984). *Linear System Theory and Design*. CBS College Publishing, NY.

Dailey, R.L. and M.S. Lukich (1987). MIMO transfer function curve fitting using Chebyshev polynomials. In *SIAM 35th Anniversary Meeting*, Denver, USA.

Friedman, J.H. and P.P. Khargonekar (1994). Identification of lightly damped systems using frequency domain techniques. In *Prepr. 10th IFAC Symp. on System Identification*, pp. 201–206, Copenhagen, Denmark.

Gevers, M.R. (1986). ARMA models, their Kronecker indices and their McMillan degree. *Int. J. Control*, 43, 1745–1761.

Golub, G.H. and C.F. Van Loan (1989). *Matrix Computations*. Johns Hopkins Univ. Press, Baltimore.

Gu, G. and P.P. Khargonekar (1992). A class of algorithms for identification in H_∞ . *Automatica*, 28, 299–312.

Hakvoort, R.G. and P.M.J. Van den Hof (1994). Frequency domain curve fitting with maximum amplitude criterion and guaranteed stability. *Int. J. Control*, 60, 809–825.

Levi, E.C. (1959). Complex-curve fitting. *IEEE Trans. Autom. Control*, AC-4, 37–44.

Lin, P.L. and Y.C. Wu (1982). Identification of multi-input multi-output linear systems from frequency response data. *J. Dynam. Syst., Measurement. Control*, 104, 58–64.

Ljung, L. (1993). Some results on identifying linear systems using frequency domain data. In *Proc. 32nd IEEE Conf. Decis. Control*, San Antonio, TX, pp. 3534–3538.

McKelvey, T. (1995). Frequency weighted subspace based system identification in the frequency domain. In *Proc. 34th IEEE Conf. Decis. Control*, New Orleans, USA, pp. 1228–1233.

Pintelon, R., P. Guillaume, Y. Rolain, J. Schoukens, and H. Van hamme (1994). Parametric identification of transfer functions in the frequency domain – a survey. *IEEE Trans. Autom. Control*, AC-39, 2245–2260.

Sanathanan, C.K. and J. Koerner (1963). Transfer function synthesis as a ratio of two complex polynomials. *IEEE Trans. Autom. Control*, AC-8, 56–58.

Van den Hof, P.M.J. (1992). System order and structure indices of linear systems in polynomial form. *Int. J. Control*, 55, 1471–1490.

Whitfield, A.H. (1987). Asymptotic behaviour of transfer function synthesis methods. *Int. J. Control*, 45, 1083–1092.

Model set determination from a batch of plant models

Tong Zhou^{†§} and Paul M.J. Van den Hof

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: p.m.j.vandenhof@wbmt.tudelft.nl*

Abstract. This paper investigates the problem of obtaining a nominal model and its error uncertainty bound when a batch of plant models is provided. While the selection of model set structure is essential and probably depends on the intended model application, in this paper, it is assumed that a nominal model is perturbed by its error through a Homographic transformation. A necessary and sufficient condition is obtained for the existence of a suboptimal nominal model. Moreover, an algorithm is proposed to obtain a nominal model which is suboptimal and has a low complexity. Furthermore, the extraction of structured nominal model error is also discussed. The efficiency of the proposed algorithms is confirmed by a simulation example.

Keywords. Error bound, Homographic transformation, nominal model, robust control.

1 Introduction

Robustness is one of the major properties required for control systems. In control engineering, one challenging task is to design a controller which performs satisfactorily when the plant works under different conditions, or which satisfactorily controls different plants which have similar but different dynamics. The former is due to the fact that the dynamics of a plant usually changes with time, environment, etc. (Kuraoka, et al., 1990; Zhou, 1996), the latter is usually met in mass production (Steinbuch, 1996). While it is desirable to design a controller which satisfactorily controls *just* these different plants, the model set consisting of the models of these plants *only* is highly structured, and controller design is currently not tractable (Packard and Doyle, 1993). A pragmatic approach to cope with this controller design problem is to find a model set which contains all of these plant models and can be handled

by the available robust control theories (Abrishamchian and Barmish, 1996; Zhou, 1996).

In the last decade, H_∞ control theory has been well developed, in which nominal model errors are regarded as unstructured, and Riccati equation or linear matrix inequality based solutions have been established (Packard and Doyle, 1993). In this unstructured uncertainty setting, additive error, multiplicative error, relative error, Homographic transformation error, coprime factorization error, linear fractional transformation error, etc., have been applied (Packard and Doyle, 1993). The investigation on the suitability of model set structure is essential in both identification and robust control.

In this research, we deal with the problem of model set determination when a batch of plant models is provided. Former results are extended to the case in which plant nominal model is perturbed by its error through a Homographic transformation. This model set structure is one of the most general model set descriptions utilized in robust control theory, and additive, multiplicative, relative, etc., model set structures can be regarded as a special form. It is proved that this problem can be converted into a model matching problem. Moreover, an algorithm

[†]On leave from the Seventh Research Division, Beijing University of Aeronautics and Astronautics, Beijing, P.R.China.

[§]The work of Tong Zhou is financially supported by the Dutch Institute of Systems and Control (DISC).

based on Hankel norm model reduction is proposed for the determination of a nominal model which is suboptimal and has a low complexity.

To reduce the conservatism in controller design, sometimes, structured error uncertainty bound is preferable (Packard and Doyle, 1993; Ariaans, et al., 1996; Zhou and Kimura, 1994). A necessary condition is obtained for the existence of structured nominal model errors, and an algorithm is proposed for their extraction.

The proposed algorithms are illustrated by a simulation example, and their efficiency is confirmed through a comparison with the intuitively determined plant nominal models.

2 Problem formulation

When a nominal model is perturbed by unstructured errors through a Homographic transformation, the model set determination problem can be formulated as follows.

Problem. Assume that plant models $G_1(s), \dots, G_n(s)$, weighting functions $w_1(s), \dots, w_n(s)$, transfer function matrices $N_1(s), D_1(s)$ and a positive number γ are given. Moreover, assume that $G_i(s)$, $i = 1, \dots, n$, $N_1(s), D_1(s)$ are stable, while $w_i(s)$, $i = 1, \dots, n$, are both stable and invertibly stable. Find stable transfer function matrices $N_0(s), D_0(s)$, such that

- (1) $N_0(s), D_0(s)$ are right coprime;
- (2) $D_0^{-1}(s)D_0(s) = I$;
- (3) there exists at least one stable $\Delta_i(s)$ satisfying $G_i(s) = [N_0(s) + N_1(s)\Delta_i(s)][D_0(s) + D_1(s)\Delta_i(s)]^{-1}$, $i = 1, \dots, n$;
- (4) $J = \|[w_1(s)\Delta_1^T(s) \dots w_n(s)\Delta_n^T(s)]^T\|_\infty < \gamma$

A nominal model $N_0(s)D_0^{-1}(s)$ satisfying these four conditions is called suboptimal, and it achieves the optimal one with the diminution of γ .

Several remarks on the control engineering significance of the above problem are now in order.

Remark 1. When $N_1(s) - G_i(s)D_1(s)$, $i = 1, \dots, n$ is invertibly stable, it will become clear in the subsequent discussion that there exists a minimum phase transfer function $w(s)$, such that all the plant models $G_i(s)$, $i = 1, \dots, n$, are included in transfer function matrix set \mathcal{G} defined as

$$\mathcal{G} = \{G(s) \mid G(s) = [N_0(s) + w(s)N_1(s)\Delta(s)] \times [D_0(s) + w(s)D_1(s)\Delta(s)]^{-1}, \|\Delta(s)\|_\infty \leq 1\} \quad (1)$$

Remark 2. When $N_1(s) = I, D_1(s) = 0$, the above transfer function matrix set \mathcal{G} can be expressed as

$$\mathcal{G} = \{G(s) \mid G(s) = [N_0(s) + w(s)\Delta(s)]D_0^{-1}(s), \|\Delta(s)\|_\infty \leq 1\} \quad (2)$$

That is, in this case, the nominal model is perturbed by an additive error.

Remark 3. When $N_1(s) = 0, D_1(s) = I$, transfer function matrix \mathcal{G} has an expression

$$\mathcal{G} = \{G(s) \mid G(s) = N_0(s)[I + w(s)D_0^{-1}(s)\Delta(s)]^{-1} \times D_0^{-1}(s), \|\Delta(s)\|_\infty \leq 1\} \quad (3)$$

which means that the nominal model has a relative error.

Remark 4. It is a direct result of the small gain theorem that the transfer function matrix set \mathcal{G} is robustly stabilized by a controller $C(s)$ iff

$$\|w(s)D_0^{-1}(s)\{D_1(s) + C(s)[I + N_0(s)D_0^{-1}(s)C(s)]^{-1} \times [N_1(s) - N_0(s)D_0^{-1}(s)D_1(s)]\}\|_\infty < 1 \quad (4)$$

Based on this condition, a controller can be designed on the basis of \mathcal{H}_∞ control theory (for robust stability and nominal performance) or structured singular value control theory (for robust stability and robust performance).

Remark 5. An essential problem in this model set determination problem is the selection of transfer function matrices $N_1(s)$ and $D_1(s)$. It is $N_1(s)$ and $D_1(s)$ that determine the structure of the model set. It will become clear in the following discussion that to make the resulted model set compatible with the available robust control theories, it is desirable that $N_1(s), D_1(s), (N_1(s) - G_i(s)D_1(s))^{-1}$, $i = 1, \dots, n$, are stable. However, the determination of $N_1(s)$ and $D_1(s)$, needs further investigation.

It is worthwhile to mention that from the viewpoint of controller design, it is more suitable to minimize the cost function $\max_i \|w_i(s)\Delta_i(s)\|_\infty$ in the above model set determination problem. The minimization of this cost function, however, is currently not tractable.

3 Main results

To solve the model set determination problem, some properties of the cost function J are investigated first.

Theorem 3.1 Assume that $N_1(s) - G_i(s)D_1(s)$ has no purely imaginary zeros. Then, there exists a stable and invertibly stable $X_i(s)$, such that $(N_1(s) - G_i(s)D_1(s))(N_1(s) - G_i(s)D_1(s))^{-1} = X_i(s)X_i^{-1}(s)$. Moreover,

$$J = \|T_1(s) - T_2(s)N_0(s)D_0^{-1}(s)\|_\infty$$

in which $T_1(s)$ and $T_2(s)$ respectively represent

$$\begin{bmatrix} w_1(s)X_1^{-1}(s)G_1(s) \\ \vdots \\ w_n(s)X_n^{-1}(s)G_n(s) \end{bmatrix}, \begin{bmatrix} w_1(s)X_1^{-1}(s) \\ \vdots \\ w_n(s)X_n^{-1}(s) \end{bmatrix}$$

Proof: The existence of $X_i(s)$ is a direct result of spectral factorization theory (Francis, 1987). From

$$G_i(s) = [N_0(s) + N_1(s)\Delta_i(s)][D_0(s) + D_1(s)\Delta_i(s)]^{-1} \quad (5)$$

we have

$$[N_1(s) - G_i(s)D_1(s)]\Delta_i(s) = G_i(s)D_0(s) - N_0(s) \quad (6)$$

On the other hand, from the definition of $X_i(s)$, the next relation can be established.

$$\Delta_i^{\sim}(s)\Delta_i(s) = [X_i^{-1}(s)(N_1(s) - G_i(s)D_1(s))\Delta_i(s)]^{\sim} \times [X_i^{-1}(s)(N_1(s) - G_i(s)D_1(s))\Delta_i(s)] \quad (7)$$

Hence

$$\begin{aligned} & \sum_{i=1}^n w_i^{\sim}(s)w_i(s)\Delta_i^{\sim}(s)\Delta_i(s) \\ &= \sum_{i=1}^n w_i^{\sim}(s)w_i(s)[X_i^{-1}(s)(N_1(s) - G_i(s)D_1(s)) \times \\ & \Delta_i(s)]^{\sim}[X_i^{-1}(s)(N_1(s) - G_i(s)D_1(s))\Delta_i(s)] \\ &= \begin{bmatrix} w_1(s)X_1^{-1}(s)[G_1(s)D_0(s) - N_0(s)] \\ \vdots \\ w_n(s)X_n^{-1}(s)[G_n(s)D_0(s) - N_0(s)] \end{bmatrix}^{\sim} \times \\ & \begin{bmatrix} w_1(s)X_1^{-1}(s)[G_1(s)D_0(s) - N_0(s)] \\ \vdots \\ w_n(s)X_n^{-1}(s)[G_n(s)D_0(s) - N_0(s)] \end{bmatrix} \quad (8) \end{aligned}$$

Therefore,

$$\begin{aligned} J &= \left\| \begin{bmatrix} w_1(s)X_1^{-1}(s)[G_1(s)D_0(s) - N_0(s)] \\ \vdots \\ w_n(s)X_n^{-1}(s)[G_n(s)D_0(s) - N_0(s)] \end{bmatrix} \right\|_{\infty} \\ &= \|T_1(s)D_0(s) - T_2(s)N_0(s)\|_{\infty} \\ &= \|T_1(s) - T_2(s)N_0(s)D_0^{-1}(s)\|_{\infty} \quad (9) \end{aligned}$$

This completes the proof. \square

For brevity, define $G_0(s) = N_0(s)D_0^{-1}(s)$. Then

$$J = \|T_1(s) - T_2(s)G_0(s)\|_{\infty} \quad (10)$$

On the other hand, let $T_{20}(s)$ be the square stable transfer function matrix which satisfies

$$T_{20}^{\sim}(s)T_{20}(s) = T_2^{\sim}(s)T_2(s), \quad T_{20}^{-1}(s) \in \mathcal{H}_{\infty} \quad (11)$$

and define transfer function matrix $T_{2I}(s)$ as

$$T_{2I}(s) = T_2(s)T_{20}^{-1}(s) \quad (12)$$

Then, it is obvious that

$$T_{2I}^{\sim}(s)T_{2I}(s) = I, \quad T_2(s) = T_{2I}(s)T_{20}(s) \quad (13)$$

Moreover, there exists a transfer function matrix $T_{2\perp}(s)$, which belongs to \mathcal{H}_{∞} and satisfies (Francis, 1987)

$$\begin{aligned} & \begin{bmatrix} T_{2I}(s) & T_{2\perp}(s) \end{bmatrix}^{\sim} \begin{bmatrix} T_{2I}(s) & T_{2\perp}(s) \end{bmatrix} = \\ & \begin{bmatrix} T_{2I}(s) & T_{2\perp}(s) \end{bmatrix} \begin{bmatrix} T_{2I}(s) & T_{2\perp}(s) \end{bmatrix}^{\sim} = I \quad (14) \end{aligned}$$

Hence

$$\begin{aligned} J &= \left\| \begin{bmatrix} T_{2I}(s) & T_{2\perp}(s) \end{bmatrix}^{\sim} \left\{ T_1(s) - \begin{bmatrix} T_{20}(s) \\ 0 \end{bmatrix} G_0(s) \right\} \right\|_{\infty} \\ &= \left\| \begin{bmatrix} T_{20}(s)[T_{20}^{-1}(s)T_{2I}^{\sim}(s)T_1(s) - G_0(s)] \\ T_{2\perp}^{\sim}(s)T_1(s) \end{bmatrix} \right\|_{\infty} \quad (15) \end{aligned}$$

According to Equation (15) and the results in Francis (1987), the following theorem is established, which is the main result of this paper.

Theorem 3.2 *The cost function J is smaller than γ , if and only if*

$$\begin{aligned} & \|T_{2\perp}^{\sim}(s)T_1(s)\|_{\infty} < \gamma \\ & \|T_{20}(s)[T_{20}^{-1}(s)T_{2I}^{\sim}(s)T_1(s) - G_0(s)]R^{-1}(s)\|_{\infty} < 1 \end{aligned}$$

Here, the transfer function matrix $R(s)$ is both stable and invertibly stable and satisfies

$$R^{\sim}(s)R(s) = \gamma^2 I - T_1^{\sim}(s)T_{2\perp}(s)T_{2\perp}^{\sim}(s)T_1(s)$$

4 Model set determination algorithm

From Theorem 3.2, it is obvious that $G_0(s) = T_{20}^{-1}(s)T_{2I}^{\sim}(s)T_1(s)$ is one of the suboptimal plant nominal models. From the viewpoint of control engineering, however, a simple nominal model is preferable. Hence, it is more suitable to select $G_0(s)$ through frequency weighted L^{∞} norm model reduction of $T_{20}^{-1}(s)T_{2I}^{\sim}(s)T_1(s)$. On the other hand, it is well known that model reduction based on the criterion of frequency weighted L^{∞} norm is currently not tractable, while Hankel norm model reduction has been well developed (Glover, et al., 1992). Based on these arguments, a pragmatic algorithm is proposed for model set determination.

- (1) Compute the L^{∞} norm of $T_{2\perp}^{\sim}(s)T_1(s)$. If it is smaller than γ , go to the next step; otherwise, γ is not achievable and it must be increased.
- (2) Fix the Smith-McMillan degree of $G_0(s)$ to be k . Perform frequency weighted Hankel norm model reduction $\|T_{20}(s)[T_{20}^{-1}(s)T_{2I}^{\sim}(s)T_1(s) - G_0(s)]R^{-1}(s)\|_H$, and obtain a stable transfer function matrix, say, $\tilde{G}_0(s)$, with Smith-McMillan degree k (To reduce nominal model error bound, convex optimization can be applied to the determination of the numerator coefficient matrices of $\tilde{G}_0(s)$).

(3) Compute the L^∞ norm of $T_{20}(s)[T_{20}^{-1}(s)T_{2I}^{\sim}(s)T_1(s) - \bar{G}_0(s)]R^{-1}(s)$. If it is smaller than 1, let $G_0(s) = \bar{G}_0(s)$; Otherwise, let $k+1 \rightarrow k$, repeat step (2).

(4) Let $N_0(s) = G_0(s)$, $D_0(s) = I$.

(5) Define $\Delta_i(s) = [N_1(s) - G_i(s)D_1(s)]^{-1} \times [G_i(s)D_0(s) - N_0(s)]$, $i = 1, \dots, n$. Find a minimum phase transfer function $w(s)$, such that $|w(j\omega)| \geq \max_{1 \leq i \leq n} \bar{\sigma}(\Delta_i(j\omega))$, $\omega \in [0, +\infty)$.

In the above model set determination algorithm, it is also possible to permit $\bar{G}_0(s)$ to be unstable. In this case, transfer function matrices $N_0(s)$ and $D_0(s)$ will be obtained as follows.

Assume that $C(sI - A)^{-1}B + D$ is the minimal realization of transfer function matrix $G_0(s)$. Then

$$N_0(s) = D + (C - DK)(sI - A + BK)^{-1}B \quad (16)$$

$$D_0(s) = I - K(sI - A + BK)^{-1}B \quad (17)$$

are one of the right coprime factorization of $G_0(s)$, provided that K is a stabilizing matrix (Nett, et al., 1984). Let P be a positive definite matrix which satisfies the following Lyapunov equation

$$AP + PA^T - BB^T = 0 \quad (18)$$

Moreover, define $K = B^T P^{-1}$. Then, $A - BK$ is stable and $D_0^{\sim}(s)D_0(s) = I$.

When Equation (18) has no positive definite solutions, assume that matrix A has no purely imaginary eigenvalues and K is one of its stabilizing matrices. Then, $\bar{N}_0(s) = N_0(s)M^{-1}(s)$ and $\bar{D}_0(s) = D_0(s)M^{-1}(s)$ are the desirable transfer function matrices, in which $M(s)$, $M^{-1}(s) \in \mathcal{H}_\infty$ and $M^{\sim}(s)M(s) = D_0^{\sim}(s)D_0(s)$.

5 Structured error extraction

In the previous sections, we discussed the problem of determining a nominal model and its error bound from a batch of plant models. A model set has been obtained which includes all the plant models. In this transfer function matrix set, nominal model error is regarded as unstructured. Sometimes, however, it is preferable to represent nominal model error as structured one, in order to reduce the conservatism in controller design (Packard and Doyle, 1993; Arias, et al., 1996; Zhou and Kimura, 1994). Ideally, it is desirable to simultaneously obtain the bounds of structured and unstructured nominal model errors from the provided plant models. Unfortunately, this problem is not tractable at the moment. To improve the performance of control systems, a two step approach is suggested. Firstly, a nominal model is

obtained, regarding its error as unstructured. Secondly, the structured information of the nominal model error is extracted.

Another reason for structured error extraction is as following.

Generally, the nominal models that respectively minimize cost functions $\max_i \|w_i(s)\Delta_i(s)\|_\infty$ and $\|[w_1(s)\Delta_1^T(s) \dots w_n(s)\Delta_n^T(s)]^T\|_\infty$ are different. While the former criterion is more natural for controller design, the latter one is easier to cope with. In our model set determination problem, the latter criterion is applied for its mathematical tractability. In consequence, structured error is generally introduced into the obtained model set due to the selection criterion. To make the unstructured error bound of the model set as small as possible, one approach is to suitably adjust the weighting functions $w_1(s), \dots, w_n(s)$, another approach is to extract the structured error from the nominal model errors. While the former heavily depends on the provided plant models, the latter is investigated in this section.

To simplify discussion, we assume, without loss of generality, that the $m \times p$ nominal model errors $\Delta_i(s)$, $i = 1, \dots, n$, satisfy $m \geq p$. If $m < p$, the problem can be solved by just transposing $\Delta_i(s)$, $i = 1, \dots, n$.

At first, we have the following results.

Theorem 5.1 Let $\Delta(s) = [\Delta_1^T(s) \dots \Delta_n^T(s)]^T$. Assume that $\omega_0 = \arg \max_\omega \bar{\sigma}(w(j\omega)\Delta(j\omega))$. Moreover, assume that

$$w(j\omega_0)\Delta(j\omega_0) = [u_1^* \dots u_{nm}^*] \begin{bmatrix} \text{diag}\{\sigma_i |_{i=1}^p\} \\ 0 \end{bmatrix} [v_1 \dots v_p]$$

in which, $\sigma_1 \geq \dots \geq \sigma_p \geq 0$, and $[u_1^* \dots u_{nm}^*]$, $[v_1 \dots v_p]$ are unitary matrices. Then, there exists a $\Delta_0(s) \in \mathcal{H}_\infty$ and $\delta_i \in \mathcal{R}$, $i = 1, \dots, n$, such that

$$\Delta_i(s) = \delta_i \Delta_0(s) + \hat{\Delta}_i(s), \quad \sum_{i=1}^n \delta_i^2 = 1$$

$$\|w(s)[\hat{\Delta}_1^T(s) \dots \hat{\Delta}_n^T(s)]^T\|_\infty < \|w(s)\Delta(s)\|_\infty$$

only if vector u_1 can be expressed as

$$[\alpha_1 \ k_1 \alpha_1 \ \dots \ k_{m-1} \alpha_1 \ \alpha_2 \ k_1 \alpha_2 \ \dots \ k_{m-1} \alpha_2 \\ \alpha_3 \ \dots \ k_{m-1} \alpha_n], \quad \alpha_i \in \mathcal{R}$$

A proof of this theorem is given in the appendix. Based on the conclusions of Theorem 5.1, the following algorithm is proposed for structured error extraction.

(1) Verify whether the conditions of Theorem 5.1 are satisfied. If the answer is negative, stop the computation; Otherwise, go to the next step.

(2) Define $\delta_i, i = 1, \dots, n$ and $\Delta_0(s)$ as

$$\delta_i = \frac{\alpha_i}{\sqrt{\sum_{i=1}^n \alpha_i^2}}, \quad \Delta_0(s) = \sum_{i=1}^n \delta_i \Delta_i(s)$$

(3) Perform convex optimization to find a $\bar{\delta}_i, i = 1, \dots, n$, such that $\|w(s)[\Delta_i(s) - \bar{\delta}_i \Delta_0(s)]\|_\infty$ is minimized. Assume the desirable $\bar{\delta}_i$ is $\bar{\delta}_i^0$.

(4) Find a minimum phase transfer function $\hat{w}(s)$ satisfying $|\hat{w}(j\omega)| \geq \max_{1 \leq i \leq n} \bar{\sigma}(\Delta_i(j\omega) - \bar{\delta}_i^0 \Delta_0(j\omega))$ for all $\omega \in [0, +\infty)$.

(5) Define $\hat{\Delta}_0(s) = (\max_{1 \leq i \leq n} |\bar{\delta}_i^0|) \Delta_0(s)$.

From the above algorithm, it is obvious that all the nominal model errors, $\Delta_i(s), i = 1, \dots, n$, are contained in transfer function matrix set Δ defined as

$$\Delta = \left\{ \Delta(s) \mid \Delta(s) = [I_m \quad I_m] \begin{bmatrix} \delta I_m & \hat{\Delta}(s) \\ \hat{\Delta}_0(s) & I_p \end{bmatrix} \times \delta \in \mathcal{R}, |\delta| \leq 1, \|\hat{\Delta}(s)\|_\infty \leq 1 \right\} \quad (19)$$

These conclusions can be extended to the case in which the parametric perturbation δ is permitted to be complex.

While it is possible to extract the structured error in $\hat{\Delta}(s)$ by the proposed algorithm, it is worthwhile to note that with the increment of the number of parametric error blocks, robust controller design will become difficult (Packard and Doyle, 1993).

6 A simulation example

In this section, a simulation example is provided to illustrate the proposed model set determination algorithm and structured error extraction algorithm. With a little abuse of terminology, in this section, the error bound of a nominal model is referred to $\max_i \bar{\sigma}(\Delta_i(j\omega))$ for a specific frequency ω .

Assume that we have two poorly damped plants which are well met in mechanical engineering and their models are

$$G_1(s) = \frac{n_1(s)}{d_1(s)} = \frac{s^2 + 2 \times 0.1 \times 6s + 6^2}{s^2 + 2 \times 0.1 \times 20 + 20^2}$$

$$G_2(s) = \frac{n_2(s)}{d_2(s)} = \frac{s^2 + 2 \times 0.1 \times 10s + 10^2}{s^2 + 2 \times 0.1 \times 40 + 40^2}$$

Moreover, assume that the frequency weighting functions in the model set determination have been provided. They may be determined from the requirements on the performance of the closed loop control systems.

The frequency responses of $G_1(s)$ and $G_2(s)$ are shown in Fig.1.a.

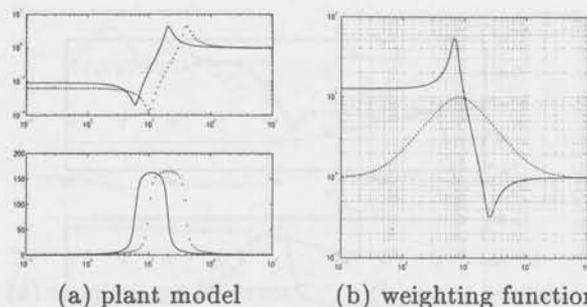


Fig. 1: Frequency response of plant model and weighting function's magnitude. —: $G_1(s)$; \cdots : $G_2(s)$; —: $w_a(s)$; \cdots : $w_r(s)$ and $w_f(s)$.

Intuitively, the following methods can be considered in the determination of a nominal model. One of them is to select a nominal model such that its frequency response is equal to the center of the frequency responses of the provided plant models at every frequency. Another method is to select a nominal model such that its coefficients are the same as the centers of the corresponding coefficients of the provided plant models. Denote these plant nominal models by $G_{01}(s)$ and $G_{02}(s)$. Obviously,

$$G_{01}(s) = \frac{1}{2}[G_1(s) + G_2(s)], \quad G_{02}(s) = \frac{n_1(s) + n_2(s)}{d_1(s) + d_2(s)}$$

When $N_1(s) = 1, D_1(s) = 0, w_1(s) = w_2(s) = w_a(s) = \frac{s^2 + 7.2s + 608}{s^2 + 1.4s + 48}$, according to the algorithm of Section 4, the next four plant nominal models are obtained, which have Smith-McMillan degree as 1, 2, 3, 4, respectively.

$$G_{a1}(s) = \frac{6.8950 \times 10^{-3}s + 3.2437 \times 10^{-2}}{s + 0.39203}$$

$$G_{a2}(s) = \frac{1.4718s^2 + 1.1748s + 69.6732}{s^2 + 19.65s + 530.96}$$

$$G_{a3}(s) = \frac{2.1031s^3 + 5.3731s^2 + 112.20s + 188.12}{s^3 + 34.733s^2 + 631.74s + 5151.2}$$

$$G_{a4}(s) = \frac{s^4 + 7.6s^3 + 1076.8s^2 + 1704s + 48800}{s^4 + 12s^3 + 2032s^2 + 9600s + 640000}$$

The frequency responses of these nominal models and their error bounds are shown in Fig.2.

When $N_1(s) = 0, D_1(s) = 1, w_1(s) = w_2(s) = w_r(s) = \frac{s^2 + 128.5s + 64}{s^2 + 12.8s + 64}$, the next four plant nominal models are obtained.

$$G_{r1}(s) = \frac{3.5232 \times 10^{-2}s + 5.2712 \times 10^{-1}}{s + 8.0683}$$

$$G_{r2}(s) = \frac{0.5226s^2 + 2.9554s + 59.94}{s^2 + 16.985s + 1187.7}$$

$$G_{r3}(s) = \frac{0.555s^3 + 24.5s^2 + 193s + 2482}{s^3 + 64.2s^2 + 1965s + 53048}$$

$$G_{r4}(s) = \frac{0.88s^4 + 16s^3 + 173s^2 + 1480s + 4669}{s^4 + 36s^3 + 1545s^2 + 20923s + 67190}$$

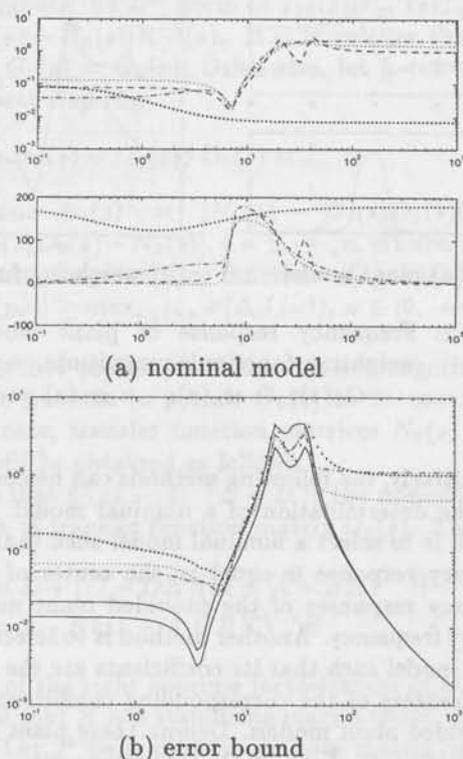


Fig. 2: Frequency response of nominal model and its error bound. $\bullet\bullet$: $G_{a1}(s)$; \cdots : $G_{a2}(s)$; $- \cdot -$: $G_{a3}(s)$; $- -$: $G_{a4}(s)$.

The frequency responses of these nominal models and their error bounds are shown in Fig.3.

When $N_1(s) = \frac{0.8s+64}{s+100}$, $D_1(s) = \frac{4s+1200}{s+1000}$, $\frac{s^2+34s+64}{s^2+4.8s+64}$, and $w_1(s) = w_2(s) = w_f(s) = \frac{s^2+128.5s+64}{s^2+12.8s+64}$, the following four plant nominal models are obtained.

$$G_{f1}(s) = \frac{4.1157 \times 10^{-3}s + 8.4463 \times 10^{-3}}{s + 0.12292}$$

$$G_{f2}(s) = \frac{1.1704s^2 + 6.2641s + 110.1}{s^2 + 51.056s + 1606.7}$$

$$G_{f3}(s) = \frac{0.25s^3 + 4.1s^2 + 28.6s + 118}{s^3 + 12s^2 + 1192s + 1634}$$

$$G_{f4}(s) = \frac{0.93s^4 + 35s^3 + 738s^2 + 3790s + 22964}{s^4 + 165s^3 + 2922s^2 + 160910s + 297930}$$

The frequency responses of these nominal models and their error bounds are shown in Fig.4.

For comparison, the error bounds of the intuitively determined nominal models, $G_{01}(s)$ and $G_{02}(s)$, are also shown in Fig.2, Fig.3 and Fig.4. They are represented by solid lines.

The magnitude frequency responses of the weighting functions $w_a(s)$, $w_r(s)$ and $w_f(s)$ are shown in Fig.1.b.

From the simulation results, it is clear that with the increment of the nominal model complexity, the

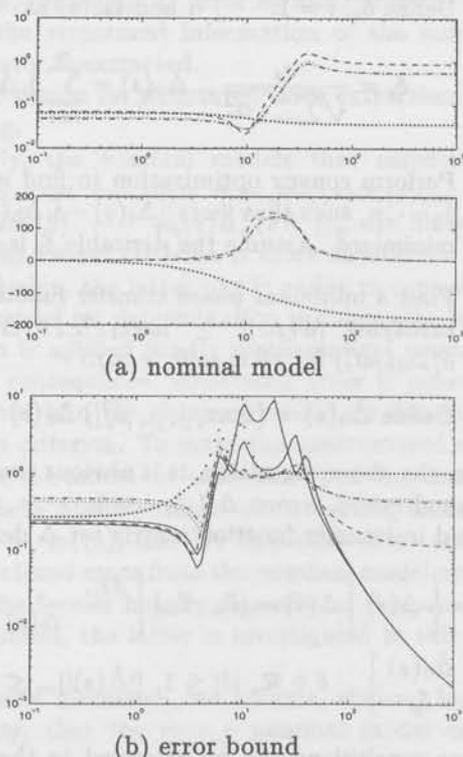


Fig. 3: Frequency response of nominal model and its error bound. $\bullet\bullet$: $G_{r1}(s)$; \cdots : $G_{r2}(s)$; $- \cdot -$: $G_{r3}(s)$; $- -$: $G_{r4}(s)$.

bound of the nominal model error in the interested frequency range is reduced. Hence, the complexity of a nominal model can be determined from the requirements on the performance of the closed loop system.

When nominal models are represented in additive form, the error bounds of all the nominal models obtained by the proposed algorithm are smaller than that of $G_{02}(s)$, in the interested frequency range. When the Smith-McMillan degree of the plant nominal model is increased to 4, its error bound equals that of $G_{01}(s)$ at every frequency. It is obvious that, in this case, $G_{01}(s)$ is the optimal nominal model, in the sense that at every frequency, the error bound of a nominal model can not be reduced less than that of $G_{01}(s)$.

When nominal models are represented in relative form or Homographic transformation form, the simulation results show that in the interested frequency range, almost all the nominal models obtained by our approach have a smaller error bound than that of $G_{01}(s)$ or $G_{02}(s)$.

It is worthy to point out that although in the interested frequency range, the Homographic transformation error bound of a nominal model is smaller than the relative one, while the relative error bound

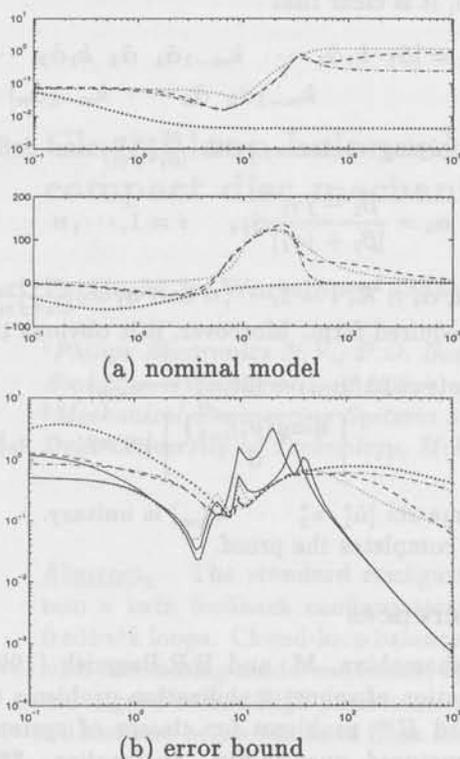


Fig. 4: Frequency response of nominal model and its error bound. $\bullet\bullet$: $G_{f1}(s)$; \cdots : $G_{f2}(s)$; $- \cdot -$: $G_{f3}(s)$; $- -$: $G_{f4}(s)$.

is smaller than the additive one, it *does not* imply that Homographic transformation error representation will result in the best controller design. This is because different model set descriptions lead to different conditions for robust stability and robust performance.

Next, the proposed structured error extraction algorithm is applied to all the obtained nominal models, using the same weighting functions as those in nominal model determination. The result is that there does not exist a structured nominal model error. However, when the weight of the nominal model error at low frequencies is reduced, a structured error appears for nominal model $G_{f1}(s)$.

In Fig.5.a, the frequency responses of the nominal model errors and the extracted structured error are presented, while in Fig.5.b, the frequency responses of the unstructured error bounds are given. In the structured error extraction, the weighting function is selected as $w(s) = \frac{s^2}{(s+2)^2} w_f(s)$.

From Fig.5, it is obvious that the unstructured error bound is significantly reduced at the middle and high frequencies by structured error extraction.

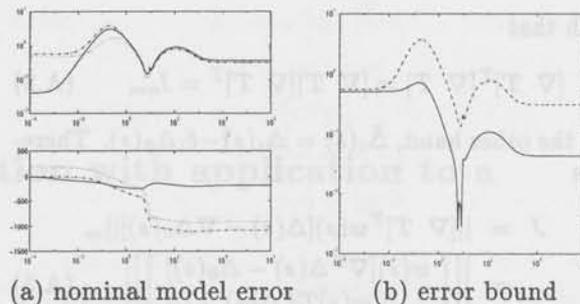


Fig. 5: Frequency response of nominal model error and unstructured error bound. $- \cdot -$: $\Delta_1(s)$; \cdots : $\Delta_2(s)$; $- -$: structured error; $- -$: before extraction; $-$: after extraction.

7 Concluding remarks

In this paper, we discussed the problem of model set determination when a batch of plant models is provided. Previous results have been extended to the case in which a plant model is described by a Homographic transformation. It has been proved that this model set determination problem can be reduced to a frequency weighted L^∞ norm model reduction problem. An algorithm is proposed to obtain a suboptimal nominal model which has a low Smith-McMillan degree, as well as its error bound. In addition, a necessary condition is obtained for the existence of structured nominal model errors, and an algorithm is suggested for their extraction. Simulation results show that the error bound of the nominal model determined by the proposed algorithm is generally smaller than that of the intuitively determined one, and the bound of the unstructured nominal model error can be significantly reduced by structured error extraction.

Recently, the proposed model set determination algorithm has been successfully applied to the simultaneous spiral control of two compact disc players. The results will be reported in some other places. However, some important issues concerned with this model set determination problem, still remain unsolved. One of them is about the selection of $N_1(s)$ and $D_1(s)$. Another one is to extend the established results to the case in which a nominal model is perturbed by its error through a linear fractional transformation.

Appendix A proof of Theorem 5.1

For brevity, let $J = \|w(s)[\hat{\Delta}_1^T(s) \cdots \hat{\Delta}_n^T(s)]^T\|_\infty$. Since $\delta_1^2 + \cdots + \delta_n^2 = 1$, we have

$$\nabla^T \nabla = I_m \quad (A.1)$$

in which, $\nabla = [\delta_1 I_m \quad \delta_2 I_m \quad \cdots \quad \delta_n I_m]^T$. As a consequence, there exists a $nm \times (n-1)m$ matrix,

such that

$$[\nabla T]^T [\nabla T] = [\nabla T][\nabla T]^T = I_{nm} \quad (\text{A.2})$$

On the other hand, $\hat{\Delta}_i(s) = \Delta_i(s) - \delta_i \Delta_0(s)$. Therefore

$$\begin{aligned} J &= \|[\nabla T]^T w(s)[\Delta(s) - \nabla \Delta_0(s)]\|_\infty \\ &= \left\| \begin{bmatrix} w(s)[\nabla^T \Delta(s) - \Delta_0(s)] \\ w(s)T^T \Delta(s) \end{bmatrix} \right\|_\infty \end{aligned} \quad (\text{A.3})$$

Hence, J is smaller than a positive number, say, γ , if and only if (Francis, 1987)

$$\|w(s)T^T \Delta(s)\|_\infty < \gamma \quad (\text{A.4})$$

$$\begin{aligned} &\|w(s)[\nabla^T \Delta(s) - \Delta_0(s)]\|_\infty < \gamma \\ &[(w(s)T^T \Delta(s))^{-1} (w(s)T^T \Delta(s))]^{-\frac{1}{2}} \|_\infty < 1 \end{aligned} \quad (\text{A.5})$$

Now, assume that there exist $\delta_i \in \mathcal{R}$, $i = 1, \dots, n$, and a $\Delta_0(s) \in \mathcal{H}_\infty$, such that $\Delta_i(s) = \delta_i \Delta_0(s) + \hat{\Delta}_i(s)$, $i = 1, \dots, n$; $\delta_1^2 + \dots + \delta_n^2 = 1$, and $J < \gamma < \|w(s)\Delta(s)\|_\infty$. Then, $\|w(s)T^T \Delta(s)\|_\infty < \gamma$, which implies that

$$w(j\omega_0)w^*(j\omega_0)T^T \Delta(j\omega_0)\Delta^*(j\omega_0)T < \gamma^2 I_{(n-1)m} \quad (\text{A.6})$$

From the definition of matrix T , it is obvious that $T^T T = I_{(n-1)m}$. Hence

$$T^T \{ \gamma^2 I_{nm} - [w(j\omega_0)\Delta(j\omega_0)][w(j\omega_0)\Delta(j\omega_0)]^* \} T > 0 \quad (\text{A.7})$$

Define matrix U as $U = [u_1^T \dots u_{nm}^T]^T$. Then,

$$(UT)^* \begin{bmatrix} \text{diag}\{(\gamma^2 - \sigma_i^2)_{i=1}^p\} & 0 \\ 0 & \gamma^2 I_{nm-p} \end{bmatrix} UT \quad (\text{A.8})$$

Since $\gamma < \|w(s)\Delta(s)\|_\infty = \sigma_1$, to guarantee that Inequality (A.7) is satisfied, it is necessary that

$$u_1 T = 0, \quad \text{or} \quad T^T u_1^T = 0 \quad (\text{A.9})$$

From Equation (A.2), it is obvious that

$$T^T \nabla = 0, \quad \text{rank}(T) = (n-1)m, \quad \text{rank}(\nabla) = m \quad (\text{A.10})$$

Hence, there exist real numbers β_i and γ_i , $i = 1, \dots, n$, such that

$$u_1 = ([\beta_1 \dots \beta_m] + j[\gamma_1 \dots \gamma_m])\nabla^T \quad (\text{A.11})$$

and there is at least one i , $1 \leq i \leq n$, such that $\beta_i + j\gamma_i \neq 0$. Without loss of generality, assume that $\beta_1 + j\gamma_1 \neq 0$. Define

$$\hat{\alpha}_i = (\beta_1 + j\gamma_1)\delta_i, \quad i = 1, \dots, n \quad (\text{A.12})$$

$$k_i = \frac{\beta_{i+1} + j\gamma_{i+1}}{\beta_1 + j\gamma_1}, \quad i = 1, \dots, m-1 \quad (\text{A.13})$$

Then, it is clear that

$$u_1 = [\hat{\alpha}_1 \quad k_1 \hat{\alpha}_1 \quad \dots \quad k_{m-1} \hat{\alpha}_1 \quad \hat{\alpha}_2 \quad k_1 \hat{\alpha}_2 \quad \dots \quad k_{m-1} \hat{\alpha}_2 \quad \hat{\alpha}_3 \quad \dots \quad k_{m-1} \hat{\alpha}_n] \quad (\text{A.14})$$

Multiplying vector u_1 with $\frac{\beta_1 - j\gamma_1}{|\beta_1 + j\gamma_1|}$ and define

$$\alpha_i = \frac{\beta_1 - j\gamma_1}{|\beta_1 + j\gamma_1|} \hat{\alpha}_i, \quad i = 1, \dots, n \quad (\text{A.15})$$

Then, $\alpha_i \in \mathcal{R}$, $i = 1, \dots, n$ and $\hat{u}_1 = \frac{\beta_1 - j\gamma_1}{|\beta_1 + j\gamma_1|} u_1$ has the required form. Moreover, it is obvious that

$$w(j\omega_0)\Delta(j\omega_0) = [\hat{u}_1^* \quad u_2^* \quad \dots \quad u_{nm}^*] \times \begin{bmatrix} \text{diag}\{\sigma_i^p\}_{i=1}^p \\ 0 \end{bmatrix} [v_1 \quad v_2 \quad \dots \quad v_p] \quad (\text{A.16})$$

and matrix $[\hat{u}_1^* \quad u_2^* \quad \dots \quad u_{nm}^*]$ is unitary.

This completes the proof. \square

References

- Abrishamchian, M. and B.R.Barmish (1996). Reduction of robust stabilization problems to standard H_∞ problems for classes of systems with structured uncertainty. *Automatica*, 32, 1101-1115.
- Ariaans, L.J.J.M., A.A.H.Damen and S.Weiland (1996). A flexible algorithm for structuring and reducing model error bounds. Submitted to *35th IEEE Conf. Dec. Control*, Kobe, Japan.
- Francis, B.A. (1987). *A Course in H_∞ Control Theory*. Springer-Verlag.
- Glover, K., D.J.N. Limebeer and Y.S.Hung (1992). A structured approximation problem with application to frequency weighted model reduction. *IEEE Trans. Autom. Control*, AC-37, 447-465.
- Kuraoka, H., N.Ohka, M.Ohba, S.Hosoe and F.Zhang (1990). Application of H-infinity design to automotive fuel control. *IEEE Control System Magazine*, 10, 102-106.
- Nett, C.N., C.A. Jacobson and M.J.Balas (1984). A connection between state space and doubly coprime fractional representations. *IEEE Trans. Autom. Control*, AC-29, 831-832.
- Packard, A. and J.C.Doyle (1993). The complex structured singular value. *Automatica*, 29, 71-109.
- Steinbuch, M. (1996). Personal communications.
- Zhou, T. and H. Kimura (1994). Simultaneous identification of nominal model, parametric uncertainty and unstructured uncertainty for robust control. *Automatica*, 30, 391-402.
- Zhou, T. (1996). Nominal model selection for control plant based on Hankel-norm model reduction. *Chinese Journal of Automation, in Chinese*, to appear.

Closed-loop balanced reduction with application to a compact disc mechanism

Pepijn M.R. Wortelboer[†], Maarten Steinbuch[†] and Okko H. Bosgra[§]

[†]Philips Electronics N.V., P.O. Box 218, 5600 MD, Eindhoven, The Netherlands

E-mail: wortel@nlcm01.cft.philips.nl, steinbuc@natlab.research.philips.com

[§]Mechanical Engineering Systems and Control Group

Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

Abstract. The standard configuration for model-based control design is reformulated into a twin feedback configuration with the model and controller written in separate feedback loops. Closed-loop balanced reduction is developed in this framework. A scheme with alternating model reduction, optimal control synthesis and controller reduction steps is designed to find high-performance low-order controllers. CD-player tracking controllers with orders below ten have been found starting from a 120th-order model.

Keywords. Linear dynamic systems; optimal control design; order reduction.

1 Introduction

This paper proposes a tractable iterative procedure to incorporate order reduction of both the model and controller in the control design.

Consider the optimization-based control design problem of Fig. 1, where a specific norm is minimized with respect to K , the controller that closes the lower loop around the so-called *standard plant* N . This general representation will be denoted the Standard Controller Synthesis Configuration (SCSC).

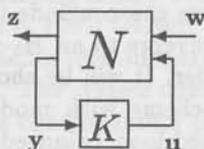


Fig. 1: Lower feedback configuration

The standard plant N comprises the model (G) of the system to be controlled and weights that are used to ensure that minimization of the influence of w on z leads to the desired controlled behaviour. Often, the order of the model is high and a reduction is needed before any controller can be calculated. Here we assume that the original model has

high accuracy and is the best we can get; the only problem is its order. The weights that are involved, however, are to be created and are auxiliary in arriving at a satisfactory closed-loop behaviour. This strongly suggests that the dynamics of the real system that is given by nature and the weights that are to be tuned in the control design phase, should be isolated from one and another. Hence, we introduce the *twin feedback configuration* (Wortelboer, 1994). In the twin feedback configuration the interconnecting system matrix M exclusively contains the weights involved in the controller synthesis problem. We will call it the *master weight*.

The controller synthesis configuration can be thought of as the twin feedback configuration with freedom to choose the controller in the lower feedback loop. The performance is related to the norm of the twin feedback configuration: the lower this closed-loop system norm, the better the performance.

Robustness can be defined in a similar way as (nominal) performance. In this paper, we will not try to model realistic variations and to achieve robust performance, but merely achieve nominal performance and some basic robustness property. This is also motivated by the fact that the robust performance problem has no straightforward solution.

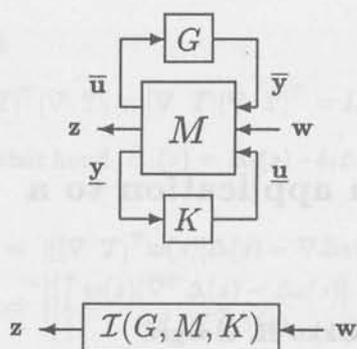


Fig. 2: Twin feedback configuration and its symbolic representation (underlined symbols for lower feedback loop and over-lined symbols for upper feedback loop)

1.1 Order reduction

The reasons for applying order reduction have already been given. The methods available for doing this are reviewed briefly and the order reduction objectives for our control design purpose are explained. We classify order reduction methods as follows (Wortelboer, 1994):

Norm minimizing model reduction A reduced order model is sought that is closest to the original model measured in some system norm. Only for the optimal Hankel-norm, a straightforward solution is available (Glover, 1984). For the more important H_∞ -norm and H_2 -norm approximation there is no such solution.

Reduction based on parameter matching A reduced order model is sought that has some key parameters in common with the full-order model.

Projection of dynamics A state-space realization of the full-order model is sought that can be truncated to the desired reduced-order. Modal reduction and balanced reduction are the main methods. Interpolation (de Ville-magne and Skelton, 1987) is also characterized by a projection. In (Hyland and Bernstein, 1985) it is shown that for the solution of the H_2 -norm minimum distance problem it is necessary to base the reduction on a projection principle.

Shortcomings

None of these methods can be applied safely for the purpose of low-order control design. Of course the norm-minimizing method has much in common with the optimal control objective, but we cannot isolate

an optimal order reduction problem from the optimal control problem. In 1982 it has already been explained that LQG-controllers can destabilize the closed-loop (Balas, 1982). This phenomenon is now known as 'spill-over'. Reduced-order modelling and control design are strongly coupled problems (Skelton, 1989; Liu and Skelton, 1993). The reduction results have to be evaluated in the closed-loop configuration. Another problem is that there is no method to find the optimal order itself.

Since there are no direct methods for this specific closed-loop reduction problem, our interest is in indirect methods exploiting order reduction that work fast and provide insight into the order selection issue. It will turn out that projection-based reduction is very efficient and can be applied step by step to the model or controller. To be more specific, a closed-loop balanced reduction method (Ceton et al., 1993), that basically extends frequency weighted balanced reduction, is adopted as the main order reduction technique.

We refer to (Wortelboer, 1994a; Anderson and Liu, 1989; Zhou et al., 1996; de Ville-magne and Skelton, 1988) and the references therein for more information on existing order reduction methods.

The iteration path and the choice of the reduced orders are the design freedom once the SCSC is fixed. The iteration has to be performed on a computer with strong numerical and graphical properties. We developed a user interface in MATLAB to support this process. Together with the key routines for closed-loop balanced reduction this interface is implemented in the so-called WOR-toolbox (which refers to Weighted Order Reduction (Wortelboer, 1994b)). This toolbox is linked to the μ -toolbox (Balas et al., 1994), and parts of it are used in the QFT-toolbox (Borghesani et al., 1994). The graphical input of additional frequency weighting functions to further direct the reduction process in a promising direction is not discussed in this paper. For that part we refer to (Wortelboer, 1994a, 1994b). In the H_2 control case, we can conclude the iteration with a final step to converge to an H_2 -norm minimizing low-order controller. It will be shown that the proposed iteration scheme with model and controller reduction by closed-loop balanced reduction yields several candidate reduced order controllers (of very specific order) that form a good starting point for H_2 -norm optimal fixed-order control. The basics of this method have been published in (Wortelboer and Bosgra, 1994; Wortelboer, 1994a).

1.2 Organisation

After the introduction of the notation for basic operations (Section 2), the balanced order reduction in

the twin feedback configuration is introduced in Section 3. Section 4 gives a detailed description of the application of the reduction procedure to a Compact Disc Mechanism.

2 Preliminaries

2.1 Truncation and projection of realizations

Continuous-time finite-dimensional time-invariant linear systems can be written in state-space as

$$G = (A, B, C, D) \quad \begin{cases} \dot{x} = Ax + Bu \\ y = Cx + Du \end{cases}$$

For clarity we will write G_n for a realization of n^{th} -order.

Truncation of a system realization underlies both balanced reduction and modal reduction. Let

$$\Gamma_r = \begin{bmatrix} I_r \\ O \end{bmatrix}$$

with $O \in \mathbb{R}^{(n-r) \times r}$ a zero matrix, then the truncation of realization G_n to order r is:¹

$$G_r = \mathcal{R}_{[\Gamma_r, \Gamma_r]}(G_n) \triangleq (\Gamma_r^T A \Gamma_r, \Gamma_r^T B, C \Gamma_r, D) \\ = (A_{(1:r, 1:r)}, B_{(1:r, :)}, C_{(:, 1:r)}, D) \triangleq \mathcal{R}_r(G_n)$$

In the same format a state transformation can be written as

$$\tilde{G}_n = \mathcal{R}_{[T^*, T]}(G_n)$$

and a projection of dynamics that is governed by the projection pair $[L_r, R_r]$ obeying $L_r^* R_r = I_r$ can be written as

$$\tilde{G}_r = \mathcal{R}_{[L_r, R_r]}(G_n)$$

2.2 Balancing

The theory of balancing is now well established: see for instance (Moore, 1981; Glover, 1984; Zhou et al., 1996). The balancing idea hinges explicitly on the state coordinates. The system dynamics is analysed in two parts. The controllability part measures the influence of input u on the state coordinates assuming $x_0 = 0$, and the observability part measures the influence of x_0 on the output y assuming $u = 0$.

The controllability Gramian P_n and observability Gramian Q_n can be solved uniquely for stable systems from the following Lyapunov equations

$$AP_n + P_n A^* + BB^* = O \quad (1.a)$$

$$A^* Q_n + Q_n A + C^* C = O \quad (1.b)$$

A balancing transformation $\check{T} = \check{R}_n$ (with $\check{T}^{-*} = \check{L}_n$) satisfies

$$\check{L}_n^* P_n \check{L}_n = \check{R}_n^* Q_n \check{R}_n = \text{diag}(\sigma_n)$$

with

$$\sigma_n = \sqrt{\lambda(P_n Q_n)}$$

the so-called Hankel Singular Value (HSV) vector. The HSVs are system invariants (realization-independent). This transformation is exclusively based on P_n and Q_n :

$$[\check{L}_n, \check{R}_n] = \check{T}(P_n, Q_n)$$

By definition, $\check{G}_n = \mathcal{R}_{[\check{T}^{-*}, \check{T}]}(G_n)$ is a balanced realization, and

$$\check{G}_r = \mathcal{R}_r(\check{G}_n) = \mathcal{R}_{[\check{L}_r, \check{R}_r]}(\check{G}_n) \triangleq \text{bal} \mathcal{R}_r(G_n)$$

defines balanced reduction, with $\check{R}_r = \check{T}_{(:, 1:r)}$, $\check{L}_r = [\check{T}^{-*}]_{(:, 1:r)}$ satisfying $\check{L}_r^* \check{R}_r = I_r$.

For frequency-weighted balanced reduction (Enns, 1984), we start from frequency weighted Gramians and perform the balancing and reduction in the same way as for plain balanced reduction.

2.3 Performance configurations, optimal control and sensitivity

We use linear fractional transformations (Zhou et al., 1996) and refer to Fig. 2 for notation.

First let $K = O$, then we can use the upper feedback loop to define the upper linear fractional transformation:

$$N = \mathcal{F}_u(M, G) \triangleq M_{zw} + M_{z\bar{y}} G (I - M_{\bar{u}y} G)^{-1} M_{\bar{u}w} \quad (2)$$

Next, we use this N in conjunction with K in the lower feedback loop to define the lower linear fractional transformation:

$$F = \mathcal{F}_l(N, K) \triangleq N_{zw} + N_{zu} K (I - N_{yu} K)^{-1} N_{yw} \quad (3)$$

From linear fractional transformation theory (Zhou et al., 1996) we know that

$$\mathcal{I}(G(s), M(s), K(s)) = \mathcal{F}_l(\mathcal{F}_u(M(s), G(s)), K(s)) \\ = \mathcal{F}_u(\mathcal{F}_l(M(s), K(s)), G(s)).$$

The theory of optimal control in the H_2 and H_∞ case is rather complete now. We refer to (Zhou et al., 1996) for all details. The state-space approaches are coded for instance in the μ -tools for use with MATLAB (Balas et al., 1994). As our approach to low-order control design we use an order reduction procedure around standard optimal full-order

¹Matrix subscripts between parentheses are index vectors.

control synthesis. We introduce the following (full-order) control synthesis operations:

$$\begin{aligned} H_2 \mathcal{K}(\mathcal{I}(G, M, \underline{\cdot})) \\ H_\infty \mathcal{K}(\mathcal{I}(G, M, \underline{\cdot})) \end{aligned} \quad (4)$$

for computing the H_2 optimal controller, and the central H_∞ optimal controller respectively.

The output sensitivity function matrix and the input sensitivity function matrix are defined as

$$\begin{aligned} S_y &= (I - GK)^{-1} \\ S_u &= (I - KG)^{-1}. \end{aligned} \quad (5)$$

Define $S_u(G, K) = S_u$ and $S_y(G, K) = S_y$. Then

$$\begin{aligned} S_y(G, K) - S_y(G, K_0) &= \\ (I - GK)^{-1} - (I - GK_0)^{-1} &= \\ (I - GK)^{-1}G(K - K_0)(I - GK_0)^{-1} \end{aligned} \quad (7)$$

We can also write the twin feedback configuration in a sensitivity form,

$$\begin{aligned} \mathcal{I}(G, M, K) &= \begin{bmatrix} O_{z\bar{u}} & I_z & O_{zy} \end{bmatrix} M \cdot \\ &\cdot S_w(M, \Lambda) \begin{bmatrix} O_{\bar{y}w} \\ I_w \\ O_{\underline{u}w} \end{bmatrix}, \end{aligned} \quad (8)$$

with

$$\begin{aligned} \Lambda &= \text{diag}(G, O_{wz}, K) \\ S_w(M, \Lambda) &= (I - \Lambda M)^{-1}. \end{aligned}$$

3 Order reduction in a closed-loop setting

Our starting point is the twin feedback configuration with master weight M and system model G given. The aim is to find a low-order controller achieving a sufficient performance level. The approach is based on full-order optimal control combined with order reduction techniques.

First, we consider the objectives of closed-loop order reduction: in Section 3.1 the model reduction step is analysed, and in Section 3.2 we take a closer look at the controller reduction step. The algorithm that is used for both model and controller reduction, closed-loop balanced reduction, is explained in Section 3.3. Section 3.4 discusses the rationale for using the closed-loop balanced reduction algorithm, and finally Section 3.5 gives a procedure for using the new order reduction facility in connection with optimal controller synthesis.

3.1 Model reduction in closed-loop

As mentioned earlier, model reduction is only an auxiliary step in low-order control design. It is needed to enable an optimal controller synthesis step. The big issue is that there is not a clear measure, like the performance measure, to quantify the loss of information by model reduction. Minimizing some open-loop error $\|G - G_r\|$ is not appropriate due to the spill-over problem. A better approach is to minimize the closed-loop changes due to model reduction. This, however, requires a controller. Although we do not have the controller we are looking for yet, we often do have a preliminary stabilizing controller. Note that most servo-systems are designed for feedback operation with fairly simple feedback controllers. Our problem then is not mere closed-loop stability, but performance improvement (often with a limitation on the controller complexity). We state that model reduction for control design should make a trade-off between the model order and the change in closed-loop behaviour (both preferably low). The change in closed-loop behaviour can be measured by

$$c_m \triangleq \|\mathcal{I}(G_h, M, K) - \mathcal{I}(G_m, M, K)\| \quad (9)$$

with G_h and G_m the high order and moderate order model respectively. For ease of interpretation, we introduce a relative error: with

$$\gamma_h \triangleq \|\mathcal{I}(G_h, M, K)\|, \quad (10)$$

we define

$$\delta_m \triangleq c_m / \gamma_h = \frac{\|\mathcal{I}(G_h, M, K) - \mathcal{I}(G_m, M, K)\|}{\|\mathcal{I}(G_h, M, K)\|}. \quad (11)$$

Note that $\gamma_m = \|\mathcal{I}(G_m, M, K)\|$ itself is *not* a good measure, since minimization of γ_m yields the 'best-controllable' reduced-order system; a solution might even be a zeroth order model $G_0 = O$.

3.2 Controller reduction in closed-loop

Controller reduction in closed-loop has the same objective as the original control design problem. Given M and G , find a K_r that minimizes $\|\mathcal{I}(G, M, K_r)\|$. The only difference is that we have a high-order controller available. The assessment of the reduction result is much easier than in the model reduction case since we can use the performance criterion directly:

$$\gamma_r = \|\mathcal{I}(G, M, K_r)\| \quad (12)$$

Note that it is also possible to strive to reduced-order controllers that change the closed-loop minimally as in the model reduction case, but this may result in controllers that have worse performance.

3.3 Closed-loop balanced reduction

First we derive the algorithm and then we state some of its properties.

3.3.1 The algorithm

To define balanced reduction within the twin feedback configuration, we make a realization $F = \mathcal{I}(G_h, M, K_n)$ in which the state vector is built from the state vectors of G_h (length h), M , and K_n (length n) in that precise order: $\mathbf{x}_F^T = [\mathbf{x}_G^T \ \mathbf{x}_M^T \ \mathbf{x}_K^T]$. Balanced reduction of G within F follows the standard balanced reduction procedure with the difference that instead of taking the entire Gramians of F ,

$$\begin{aligned} P_F &= \mathcal{P}(F) \\ Q_F &= \mathcal{Q}(F), \end{aligned}$$

only specific parts of these Gramians are used. The scheme for G_h reduction within $\mathcal{I}(G_h, M, K)$ hinges on taking the left upper parts of the Gramians of the interconnected system realization and proceeds along classical lines from then on:²

$$\begin{aligned} P_G &= [\mathcal{P}(\mathcal{I}(G_h, M, K_n))]_{(1:h, 1:h)} \\ &\triangleq \mathcal{P}(\mathcal{I}(\underline{G}_h, M, K_n)) \\ Q_G &= [\mathcal{Q}(\mathcal{I}(G_h, M, K_n))]_{(1:h, 1:h)} \\ &\triangleq \mathcal{Q}(\mathcal{I}(\underline{G}_h, M, K_n)) \\ [\tilde{L}_h, \tilde{R}_h] &= \check{T}(P_G, Q_G) \\ \tilde{L}_m &= [\tilde{L}_h]_{(:, 1:m)} \\ \tilde{R}_m &= [\tilde{R}_h]_{(:, 1:m)} \\ \tilde{G}_m &= \mathcal{R}_{[\tilde{L}_m, \tilde{R}_m]}(G_h) \triangleq \text{bal}\mathcal{R}_m(\mathcal{I}(\underline{G}_h, M, K_n)) \end{aligned}$$

Exploiting the same notation, the procedure for closed-loop controller reduction is:

- Take the controller state part of the closed-loop Gramians, $P_K = \mathcal{P}(\mathcal{I}(G_h, M, \underline{K}_n))$ and $Q_K = \mathcal{Q}(\mathcal{I}(G_h, M, \underline{K}_n))$
- Extract a balancing transformation, $[\tilde{L}_n, \tilde{R}_n] = \check{T}(P_K, Q_K)$,
- Truncate the last $n - r$ columns of \tilde{L}_n and \tilde{R}_n yielding \tilde{L}_r and \tilde{R}_r , and
- apply a projection of K_n by means of \tilde{L}_r, \tilde{R}_r

This procedure is summarized as follows:

$$\tilde{K}_r = \text{bal}\mathcal{R}_r(\mathcal{I}(G_h, M, \underline{K}_n)).$$

²the reduction procedure is applied to the underlined system.

Since the closed-loop balancing of G and K are independent, we can also balance G and K simultaneously. In Fig. 3 the block-diagonal structure of the similarity transformation group that is allowed in closed-loop transformation is visualized. \check{G} and

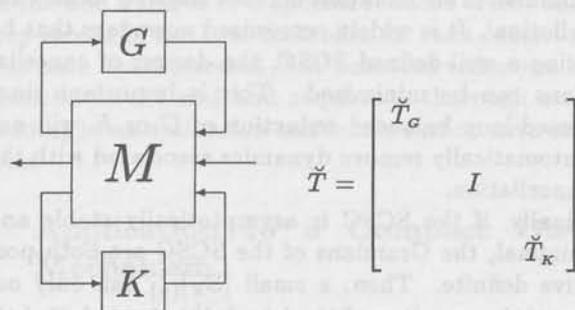


Fig. 3: Diagonal structure of similarity transformation group

\check{K} are closed-loop balanced if they induce diagonal blocks in the Gramians of $\mathcal{I}(\check{G}, M, \check{K})$:

$$\begin{aligned} \mathcal{P}(\mathcal{I}(\check{G}, M, \check{K})) &= \begin{bmatrix} \Sigma_G & \equiv & \equiv \\ \equiv & \equiv & \equiv \\ \equiv & \equiv & \Sigma_K \end{bmatrix} \\ \mathcal{Q}(\mathcal{I}(\check{G}, M, \check{K})) &= \begin{bmatrix} \Sigma_G & \equiv & \equiv \\ \equiv & \equiv & \equiv \\ \equiv & \equiv & \Sigma_K \end{bmatrix} \end{aligned}$$

with $\Sigma_G = \text{diag}(\sqrt{\lambda(P_G Q_G)})$ the $\mathcal{I}(G, M, K)$ -HSVs and $\Sigma_K = \text{diag}(\sqrt{\lambda(P_K Q_K)})$ the $\mathcal{I}(G, M, \underline{K})$ -HSVs. Note that $\mathcal{P}(\mathcal{I}(\check{G}, M, \check{K})) \neq \mathcal{Q}(\mathcal{I}(\check{G}, M, \check{K}))$ generically.

3.3.2 Properties

First we state a trivial property:

$$\text{bal}\mathcal{R}_m(\mathcal{I}(G_h, M, K) - \mathcal{I}(\underline{G}_h, M, K)) = \text{bal}\mathcal{R}_m(\mathcal{I}(\underline{G}_h, M, K))$$

This means that closed-loop balanced reduction does not discriminate between the relative and the absolute case.

It is important to stress that the result of closed-loop balanced reduction is not necessarily closed-loop balanced! This also implies that stepwise closed-loop balanced reduction may yield a different result than direct closed-loop balanced reduction.

Next we analyse the minimality properties. If K_n has $n - r$ uncontrollable state coordinates then $\mathcal{I}(G, M, K_n)$ has at least $n - r$ uncontrollable state

coordinates and $\mathcal{P}(\mathcal{I}(G, M, \underline{K}_n))$ has at most rank r . The same holds for G . Of course the observability case is completely dual.

In case G , M , and K are all minimal realizations, there may still be a chance that $\mathcal{I}(G, M, K)$ is not minimal. This situation is often referred to as 'cancellation'. It is widely recognized nowadays that by using a well-defined SCSC, the danger of cancellations can be minimized. This is important since closed-loop balanced reduction of G or K will not automatically remove dynamics associated with the cancellation.

Finally, if the SCSC is asymptotically stable and minimal, the Gramians of the SCSC are both positive definite. Then, a small $[\Sigma_K]_{(n)}$ can only occur if the coupling of the last (n^{th}) closed-loop balanced controller state coordinate with the other coordinates is sufficiently weak. And this means that truncation of the closed-loop balanced controller to order $n - 1$ in the SCSC gives almost the same result as truncation of the SCSC by plain balanced reduction.

3.4 Underlying weighted reduction problems

Next we investigate the relation between closed-loop balanced reduction and the objectives we had for closed-loop reduction of the model and controller. Also the relation with frequency-weighted reduction (Enns, 1984) is established, see also (Schelfhout, 1996).

Recall that for the controller reduction case we are interested in minimizing $\|\mathcal{I}(G, M, K_r)\|$, while the model reduction case requires small

$$\|\mathcal{I}(G_h, M, K) - \mathcal{I}(G_m, M, K)\|.$$

From (8) we know that closed-loop changes are fully due to changes in

$$\Lambda = \text{diag}(G, O_{\mathbf{wz}}, K).$$

For small perturbations $\delta G = \tilde{G} - G$ and $\delta K = \tilde{K} - K$ we have

$$S_{\mathbf{w}}(M, \tilde{\Lambda}) \approx S_{\mathbf{w}} + S_{\mathbf{w}} \delta \Lambda M S_{\mathbf{w}} \quad (13)$$

$$S_{\mathbf{w}} = (I - \Lambda M)^{-1} \quad (14)$$

$$\Lambda = \text{diag}(G, O_{\mathbf{wz}}, K) \quad (15)$$

$$\delta \Lambda = \text{diag}(\delta G, O_{\mathbf{wz}}, \delta K). \quad (16)$$

Define

$$\Phi_{\mathbf{z}} = \begin{bmatrix} O_{\mathbf{z}\bar{\mathbf{u}}} & I_{\mathbf{z}} & O_{\mathbf{z}\bar{\mathbf{y}}} \end{bmatrix} \quad (17)$$

$$\Phi_{\mathbf{w}} = \begin{bmatrix} O_{\bar{\mathbf{y}}\mathbf{w}} \\ I_{\mathbf{w}} \\ O_{\bar{\mathbf{u}}\mathbf{w}} \end{bmatrix} \quad (18)$$

$$\Phi_{\bar{\mathbf{y}}} = \begin{bmatrix} I_{\bar{\mathbf{y}}} \\ O_{\mathbf{w}\bar{\mathbf{y}}} \\ O_{\bar{\mathbf{u}}\bar{\mathbf{y}}} \end{bmatrix} \quad (19)$$

$$\Phi_{\bar{\mathbf{u}}} = \begin{bmatrix} I_{\bar{\mathbf{u}}} & O_{\bar{\mathbf{u}}\mathbf{z}} & O_{\bar{\mathbf{u}}\bar{\mathbf{y}}} \end{bmatrix} \quad (20)$$

$$\Phi_{\bar{\mathbf{u}}} = \begin{bmatrix} O_{\bar{\mathbf{y}}\bar{\mathbf{u}}} \\ O_{\mathbf{w}\bar{\mathbf{u}}} \\ I_{\bar{\mathbf{u}}} \end{bmatrix} \quad (21)$$

$$\Phi_{\bar{\mathbf{y}}} = \begin{bmatrix} O_{\bar{\mathbf{y}}\bar{\mathbf{u}}} & O_{\bar{\mathbf{y}}\mathbf{z}} & I_{\bar{\mathbf{y}}} \end{bmatrix}. \quad (22)$$

Then we can state that

$$\mathcal{I}(\tilde{G}, M, \tilde{K}) \approx \mathcal{I}(G, M, K) + W_{\delta G} \delta G V_{\delta G} + W_{\delta K} \delta K V_{\delta K} \quad (23)$$

with

$$W_{\delta G} = \Phi_{\mathbf{z}} M S_{\mathbf{w}} \Phi_{\bar{\mathbf{y}}}$$

$$V_{\delta G} = \Phi_{\bar{\mathbf{u}}} M S_{\mathbf{w}} \Phi_{\mathbf{w}}$$

$$W_{\delta K} = \Phi_{\mathbf{z}} M S_{\mathbf{w}} \Phi_{\bar{\mathbf{u}}}$$

$$V_{\delta K} = \Phi_{\bar{\mathbf{y}}} M S_{\mathbf{w}} \Phi_{\mathbf{w}}$$

Thus, if we want to minimize the influence of small model changes on the closed-loop system, we can try to minimize $\|W_{\delta G} \delta G V_{\delta G}\|$ with $\delta G = G_m - G_h$. We can also write the above as

$$\mathcal{I}(\tilde{G}, M, \tilde{K}) - \mathcal{I}(G, M, K) \approx W_{\delta \Lambda} \delta \Lambda V_{\delta \Lambda} \quad (24)$$

with

$$W_{\delta \Lambda} = \Phi_{\mathbf{z}} M (I - \Lambda M)^{-1} \quad (25)$$

$$V_{\delta \Lambda} = M (I - \Lambda M)^{-1} \Phi_{\mathbf{w}}. \quad (26)$$

In (Schelfhout, 1996) it is shown that the weights $V_{\delta \Lambda}$ and $W_{\delta \Lambda}$ are used implicitly in closed-loop balanced reduction, i.e.

$$\text{bal}\mathcal{R}(\mathcal{I}(G, M, \underline{K})) = \text{bal}\mathcal{R}(W_{\delta K} \underline{K} V_{\delta K}) \quad (27)$$

This equivalence does not depend on δG or δK . Note, however, that the computational scheme for closed-loop balanced reduction is more efficient and that unstable G and K can be reduced as long as the twin feedback configuration is strictly stable.

So far, we have only considered the case that the changes in G and K are sufficiently small for first-order approximations. We refer to (Wortelboer et al., 1997) for the results concerning larger changes.

3.5 A combined order reduction - control design strategy

Here we describe the main cycle in obtaining high-performance low-order control starting from a high-order model (Fig. 4). the model reduction step can be repeated a number of times to find an appropriate

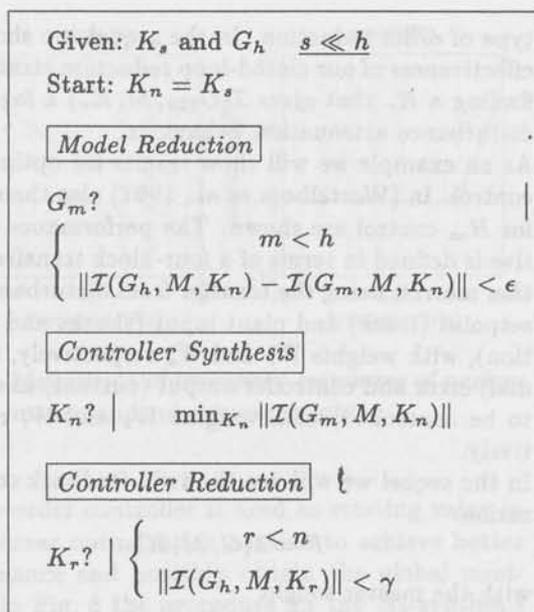


Fig. 4: The iteration scheme for low-order control design starting from a high-order model using interactive order reduction

m together with G_m . The evaluation of the order reduction effects is by means of $\|\mathcal{I}(G_h, M, K_n) - \mathcal{I}(G_m, M, K_n)\|$.

The control design step can be performed by H_2 or H_∞ optimal control if m is sufficiently small. With a new K_n we can go back to the model reduction phase and verify if G_m is still appropriate. If this iteration has converged we can proceed with the controller reduction iteration. After each step the performance is measured by means of $\mathcal{I}(G_h, M, K_r)$.

For the H_2 case we end with a search for the optimal fixed-order controller.

For the H_∞ case, we do not have a satisfactory algorithm to derive an optimal fixed-order H_∞ controller starting from the optimal full-order controller. Yet, we can find better performing low-order controllers by closed-loop balanced reduction of each controller that is generated by a bisection type H_∞ control algorithm. The idea behind this is to relax the performance requirement deliberately, thus hoping to find a full-order controller that can be reduced more easily. For the H_∞ -case, we can choose a γ that is somewhat higher than the optimal performance level that is attained in the full-order case. To be more specific, we can exploit the fact that in optimal H_∞ controller synthesis by bisection, a series of controllers of full order is generated that achieve progressively lower H_∞ -norms until the optimal γ is reached. One of these intermediate full-order controllers might induce an H_∞ -norm that is close to the optimal H_∞ -norm for a controller of order $r < n$.

Such a full-order controller might be a better starting point for reduction than the optimal full-order one, since the reduction does not necessarily imply a performance degradation. In computer codes for H_∞ controller design using the bisection principle we can add a search over specific reduced orders and keep record of the best controller for each order. It is once again stressed that balanced reduction is heuristic in the end and requires a thorough embedding in an evaluation and manipulation environment.

4 Application to a Compact Disc Mechanism

4.1 Tracking control problem

The control task of a Compact Disc mechanism is to achieve track following, which basically amounts to pointing the laser spot to the track of pits on the CD that is rotating. The reader is referred to (Steinbuch et al., 1994) and the references therein for details about the principles of CD-player control. The mechanism treated here, consists of a swing arm on which a lens is mounted, see Fig. 5.

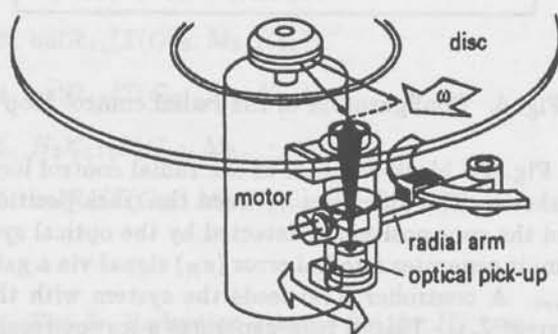


Fig. 5: Schematic view of a rotating arm Compact Disc mechanism.

The rotation of the arm in the horizontal plane enables reading of the spiral-shaped disc-tracks, and the suspended lens is used to focus the spot on the disc. Due to the fact that the disc is not perfectly flat, and due to irregularities in the spiral of pits on the disc, a feedback system is needed. The challenge is to find a low-cost controller that can make the servo-system faster and less sensitive to external shocks.

Performance improvements are sought via model-based control design in the frequency domain. The time domain specifications that underlie the frequency domain specifications are treated in (Steinbuch et al., 1994). Robustness is an important issue in practical control design, in fact, design for robustness against specific variations in the CD-system

has already been achieved (Steinbuch et al., 1994). Here, we concentrate on low-order control design for a single high-order model.

A detailed model is needed to describe the vibrational behaviour of the electro-mechanical system over a large frequency range in order to anticipate the interaction with a controller of possible high-bandwidth. A Finite Element Model was built, containing 60 vibration modes ($n = 120$), and has two inputs (actuation of arm and of focus lens), and two outputs (tracking error and focus error). The model is included in the WOR-toolbox (Wortelboer, 1994b). With respect to the disc, we can discern a radial (R) part and a focus (F) part of the feedback loop. Ideally, these parts have no interaction, but in practice there is (some) mutual interference. From control design view point, the radial loop poses a much more difficult problem compared to the focus loop. This is due to the more pronounced mechanical resonances. Therefore, in this paper, we concentrate our investigation on this loop only.

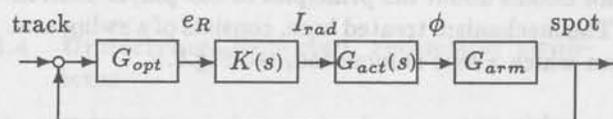


Fig. 6: Configuration of the radial control loop

In Fig. 6 a block-diagram of the radial control loop is shown. The difference between the track position and the spot position is detected by the optical system; it generates a radial error (e_R) signal via a gain G_{opt} . A controller $K(s)$ feeds the system with the current I_{rad} . This in turn generates a torque resulting in an angular acceleration. The transfer function from the current I_{rad} to the angular displacement ϕ of the arm is called $G_{act}(s)$. A (nonlinear) gain G_{arm} relates the angular displacement with the spot movement in the radial direction. Only the control error signal e_R is available for measurement.

In Fig. 13 the (1,1) element, i.e. the radial transfer function, of the (magnitude) frequency response of a Finite Element based 120th-order model is plotted. At low frequencies the actuator transfer function from current input I_{rad} to position error output e_R is a critically stable system with a phase lag of 180° (rigid body mode). At higher frequencies the plot shows parasitic dynamics.

Given the model $G = G_{opt}G_{arm}G_{act}$, the control design involves the definition of a configuration $\mathcal{I}(G, M, K)$, the choice for a suitable norm, the creation of relevant frequency weights in M , and finally, the synthesis of K . The design of a low-order controller via model-based control design requires some

type of order reduction. In the sequel, we show the effectiveness of our closed-loop reduction strategy in finding a K_r that gives $\mathcal{I}(G_{120}, M, K_r)$ a favorable disturbance attenuation behaviour.

As an example we will show results for optimal H_2 control. In (Wortelboer et al., 1997) also the results for H_∞ control are shown. The performance objective is defined in terms of a four-block transfer function matrix, being the transfer from disturbances on setpoint (track) and plant input (shocks and vibration), with weights W_r and W_d respectively, to (radial) error and controller output (current) as signals to be controlled, with weights W_s and W_t respectively.

In the sequel we will use the twin-feedback configuration:

$$F = \mathcal{I}(G, M, K)$$

with the master weight

$$M = \begin{bmatrix} O & O & W_d & I_{\bar{y}} \\ W_s & W_s W_r & O & O \\ O & O & O & W_t \\ I_{\bar{u}} & W_r & O & O \end{bmatrix} \quad (28)$$

Note that M can also be written in a matrix product form with each weight only occurring once.

4.2 Iterative model and controller reduction

The procedure for achieving high performance low-order controllers can be summarized as follows: we start with the 120th-order model G_{120} , a master weight M , including the weights $W_{s,t,r,d}$ and a stabilizing (PID) feedback K_3 . Fig. 7 shows the magnitudes of the frequency responses of the weights

$$W_s(s) = \frac{31417}{s + 62.832}$$

$$W_t(s) = \frac{78.467s^2 + 4.7330 \cdot 10^6 s + 1.9826 \cdot 10^{11}}{s^2 + 1.2566 \cdot 10^6 s + 3.9478 \cdot 10^{11}}$$

where W_s can be thought of as shaping the performance and W_t the robustness. Both W_r and W_d are chosen equal to 1.

The stabilizing controller K_3 is a PID controller with first-order low-pass at high frequencies:

$$K_3(s) = -7.6746 \cdot 10^4 \cdot \frac{(s + 3.1447 \cdot 10^2)(s + 8.6207 \cdot 10^2)}{s(s + 6.6225 \cdot 10^3)(s + 6.2832 \cdot 10^4)}$$

Applying closed-loop balanced reduction we can obtain a low-order model, for which an H_2 optimal controller can be calculated, which in turn can be reduced in closed-loop. With this low-order controller the procedure is restarted. As a final step

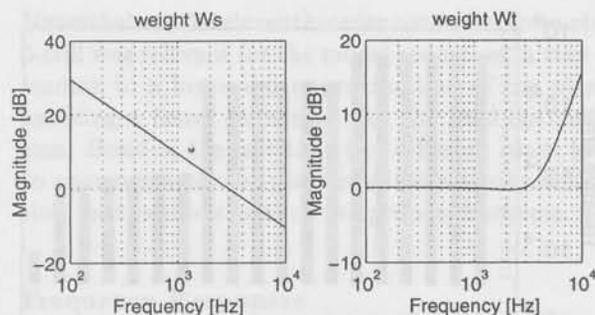


Fig. 7: Magnitude of frequency responses of performance and robustness weights

the low-order controller is used as starting value in a non-linear optimization routine to achieve better performance and possibly obtain the global minimum. In Fig. 8 the procedure for the H_2 -synthesis case is shown.

In the following section we will describe all the subsequent steps as mentioned in Fig. 8.

Step 1. Model Reduction:

$$G_{20} = \text{bal}\mathcal{R}_{20}(\mathcal{I}(G_{120}, M_3, K_3)) \quad (\text{aa})$$

The criterion in the model reduction step is the relative change of the closed-loop transfer function matrix measured in H_2 -norm:

$$\delta(r) = \frac{\|\mathcal{I}(G_{120}, M_3, K_3) - \mathcal{I}(G_r, M_3, K_3)\|_2}{\|\mathcal{I}(G_{120}, M_3, K_3)\|_2} \ll 1$$

Fig. 9 shows this relative error $\delta(r)$ as a function of the order r of the model.

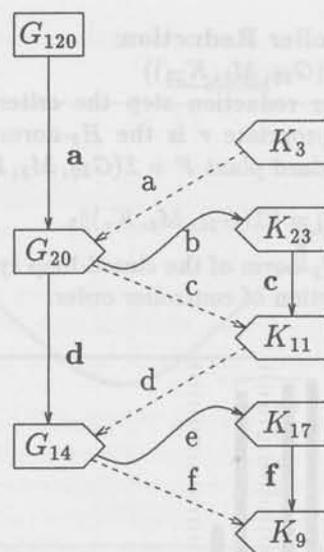
A good approximation is possible with $r = 20$. Of course, it is worthwhile to investigate even lower order approximations. However, in this stage we should be very careful about deleting any dynamics in the model which is not excited by the simple controller K_3 , but which might become important in the final stage. So here we will choose a model with a relatively high-order, and in a second iteration we will try to further reduce the model, using H_2 optimal controllers in the calculation of the closed-loop error transfer function.

The new model G_{20} enables the calculation of an H_2 optimal controller.

Step 2. Controller Calculation:

$$K_{23} = H_2\mathcal{K}_{23}(\mathcal{I}(G_{20}, M_3, \dots)) \quad (\text{b})$$

$$\gamma(23) = \|\mathcal{I}(G_{20}, M_3, K_{23})\|_2 = 913.76$$



The six steps of model reduction (1,4), controller design (2,5), and controller reduction (3,6):

1. $\text{bal}\mathcal{R}_{20}(\mathcal{I}(G_{120}, M_3, K_3))$ aa
2. $H_2\mathcal{K}_{23}(\mathcal{I}(G_{20}, M_3, \dots))$ b
3. $\text{bal}\mathcal{R}_{11}(\mathcal{I}(G_{20}, M_3, K_{23}))$ cc
4. $\text{bal}\mathcal{R}_{14}(\mathcal{I}(G_{20}, M_3, K_{11}))$ dd
5. $H_2\mathcal{K}_{17}(\mathcal{I}(G_{14}, M_3, \dots))$ e
6. $\text{bal}\mathcal{R}_9(\mathcal{I}(G_{12}, M_3, K_{17}))$ ff

Fig. 8: Reduction scheme for the H_2 case

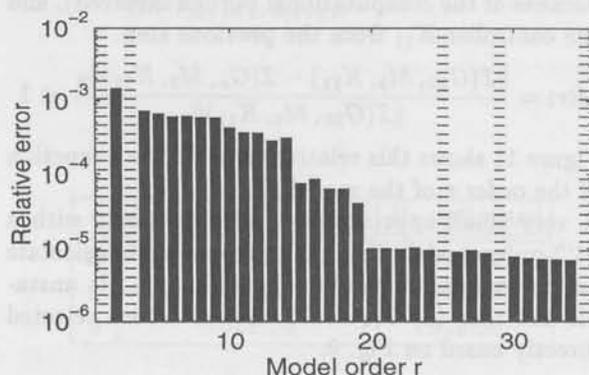


Fig. 9: H_2 -norm of the relative error for model reduction for G_{120}

Step 3. Controller Reduction:

$$K_{11} = \text{bal}\mathcal{R}_{11}(\mathcal{I}(G_{20}, M_3, K_{23})) \quad (\text{cc})$$

In the controller reduction step the criterion for choosing an appropriate r is the H_2 -norm of the closed-loop standard plant $F = \mathcal{I}(G_{20}, M_3, K_r)$:

$$\gamma(r) = \|\mathcal{I}(G_{20}, M_3, K_r)\|_2$$

In Fig. 10 the H_2 -norm of the closed-loop system is shown, as a function of controller order.

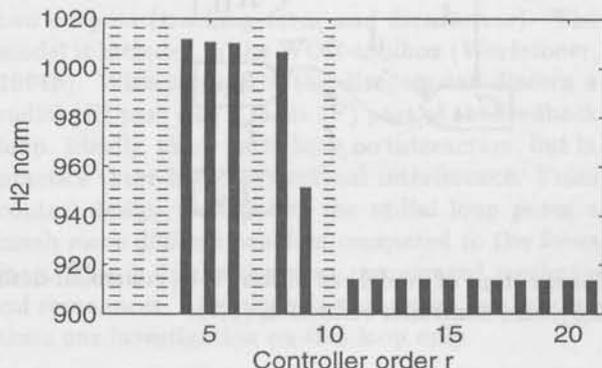


Fig. 10: H_2 -norm of the closed-loop system for reduction of K_{23}

The figure clearly shows that the reduced-order controllers for $r \geq 11$ are almost as good as the original K_{23} . The ninth-order controller would be appropriate and we could stop here. To see how the iteration proceeds we take the eleventh-order to do a further model reduction step.

Step 4. Model Reduction:

$$G_{14} = \text{bal}\mathcal{R}_{14}(\mathcal{I}(G_{20}, M_3, K_{11})) \quad (\text{dd})$$

Again in this model reduction step we use as criterion the H_2 -norm of the relative error, but now with the 20th-order model as starting point (and not G_{120} because of the computational burden involved), and the controller K_{11} from the previous step:

$$\delta(r) = \frac{\|\mathcal{I}(G_{20}, M_3, K_{11}) - \mathcal{I}(G_r, M_3, K_{11})\|_2}{\|\mathcal{I}(G_{20}, M_3, K_{11})\|_2} \ll 1$$

Figure 11 shows this relative error $\delta(r)$ as a function of the order r of the model.

A very small approximation error is made with a 14th-order model. Using G_{14} we can again calculate an H_2 optimal controller. Note that G_{14} is unstable and that (a) G_{14} could also have been selected directly based on Fig. 9.

Step 5. Controller Calculation:

$$K_{17} = H_2\mathcal{K}_{17}(\mathcal{I}(G_{14}, M_3, \dots)) \quad (\text{e})$$

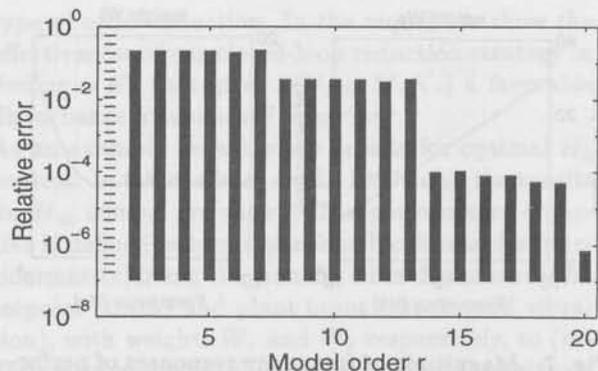


Fig. 11: H_2 -norm of the relative error for model reduction of G_{20}

$$\gamma(17) = \|\mathcal{I}(G_{14}, M_3, K_{17})\|_2 = 913.77$$

Step 6. Controller Reduction:

$$K_9 = \text{bal}\mathcal{R}_9(\mathcal{I}(G_{14}, M_3, K_{17})) \quad (\text{ff})$$

The H_2 -norm of the closed-loop standard plant F is:

$$\gamma(r) = \|\mathcal{I}(G_{14}, M_3, K_r)\|_2$$

In Fig. 12 the H_2 -norm of the closed-loop system is shown, as a function of controller order.

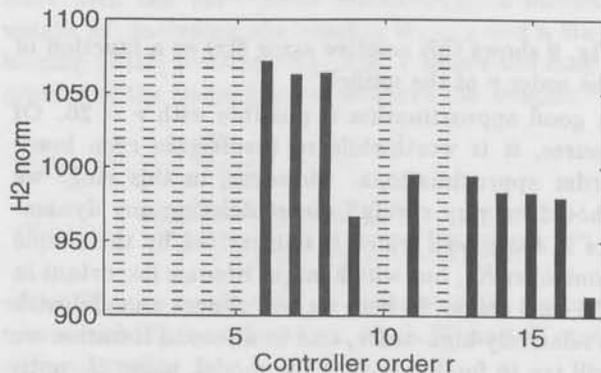


Fig. 12: H_2 -norm of the closed-loop system for reduction of K_{17}

From the figure the choice for $r = 9$ is evident. Notice that a similar norm would have been obtained if in step 3 (cc) a ninth-order controller would have been taken. Hence, steps 5 and 6 could have been skipped in this case. Since the final reduced-order controller will be further optimized using fixed-order H_2 optimization it is only relevant to have a reasonable starting value for the controller parameters.

Nevertheless, the eleventh-order controller from step 3 still was relevant for the model reduction in step 4, leading to a low-order approximation of the plant, enabling a faster calculation for the final optimization. Besides, the fact that two different roads lead to approximately the same result is a strong indication for the robustness of the proposed scheme.

Frequency Responses

In Fig. 13 the magnitude of the frequency responses of the model and model-errors are shown.

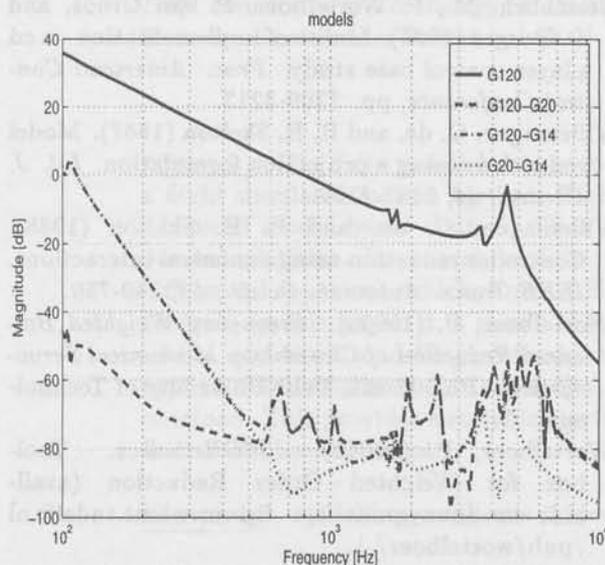


Fig. 13: Magnitude of the frequency responses of the models

Figure 14 shows the magnitude of the frequency responses of the controllers.

4.3 Fixed-order H_2 optimization

Using the reduced-order controllers as initial values, we are now able to start the fixed-order H_2 -optimization. The results are summarized in Tables 1 and 2 below.

Table 1 shows the results obtained using the reduced-order model (G_{20}), indicating that further optimization improves results significantly for $r < 10$. This also holds for the (computational much more involved) case with the full 120th-order model. It should be noted that the convergence of the algorithm is fast provided the closed-loop balanced low-order controllers are close to the minimizing reduced-order controllers. From Table 1 we see that for $r = 7$ the closed-loop balanced result is even worse than the fourth and fifth-order results and not surprisingly this case does not converge! It is emphasized that these results may be only local min-

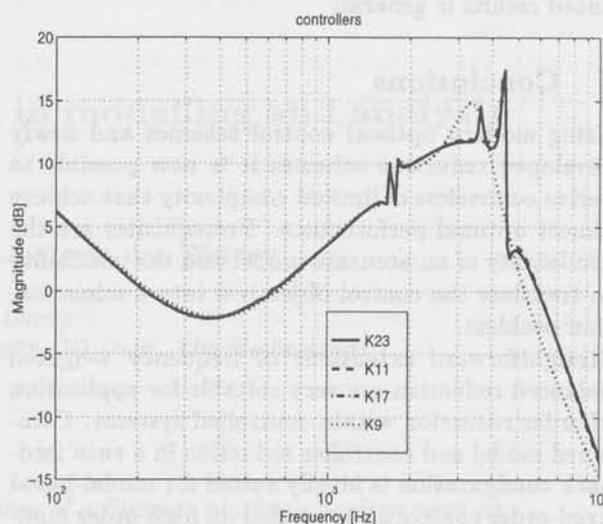


Fig. 14: Magnitude of the frequency responses of the H_2 controllers

r	closed-loop balanced	H_2 -optimized
23	913.76	913.76
17	913.77	913.77
11	914.27	914.17
9	946.61	916.63
7	1032.86	no conv.
5	1022.61	920.56
4	1023.43	979.24

Table 1: Performance of reduced-order controllers: $\|I(G_{20}, M_3, K_r)\|_2$.

r	closed-loop balanced	H_2 -optimized
17	913.80	913.80
9	946.61	916.65
5	1022.54	920.58

Table 2: Performance of reduced-order controllers: $\|I(G_{120}, M_3, K_r)\|_2$.

imizers, still the improvement over closed-loop balanced results is general.

5 Conclusions

Using modern optimal control schemes and newly developed reduction schemes it is now possible to derive controllers of limited complexity that achieve almost optimal performance. Prerequisites are the availability of an accurate model and the possibility to translate the control objectives into a minimization problem.

Straightforward extensions of frequency weighted balanced reduction are very suitable for application of order reduction within controlled systems. Combined model and controller reduction in a twin feedback configuration is ideally suited for model-based fixed-order control design based on high-order models since the interconnection structure incorporates the same weighting functions as the standard control design configuration. The key algorithm for closed-loop balanced reduction is almost as fast as standard balanced reduction. The iterative process for fixed-order control design starting from high-order models can be executed in an interactive way using a new toolbox for use with MATLAB.

References

- Anderson, B. and Y. Liu (1989). Controller reduction: concepts and approaches. *IEEE Trans. Automat. Contr.*, 34, 802-812.
- Balas, G., J. Doyle, K. Glover, A. Packard and R. Smith (1994). *μ -Analysis and Synthesis Toolbox*, The MathWorks, Inc.
- Balas, M. (1982). Trends in large space structure control theory: fondest hopes, wildest dreams. *IEEE Trans. Automat. Contr.*, AC-27, 522-535.
- Borghesani, C., Y. Chait, and O. Yaniv (1994). *The QFT Toolbox for MATLAB*, The MathWorks, Inc.
- Ceton, C., P. Wortelboer and O. Bosgra (1993). Frequency weighted closed-loop balanced reduction. *Proc. 2nd European Control Conference*, Groningen, The Netherlands, pp. 697-701.
- Enns, D. (1984). *Model Reduction for Control System Design*. PhD thesis, Stanford University, Stanford.
- Glover, K. (1984). All optimal Hankel-norm approximations of linear multivariable systems and their L_∞ -error bounds. *Int. J. Control*, 39, 1115-1193.
- Hyland, D. and D. Bernstein (1985). The optimal projection equations for model reduction and the relationships among the methods of wilson, skelton and moore. *IEEE Trans. Automat. Contr.*, AC-30, 1201-1211.
- Liu, K. and R. Skelton (1993). Integrated modeling and controller design with application to flexible structure control. *Automatica*, 29, 1291-1314.
- Moore, B. (1981). Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Automat. Contr.*, AC-26, 17-32.
- Schelfhout, G. (1996). *Model Reduction for Control Design*. PhD thesis, Katholieke Universiteit Leuven, Belgium.
- Skelton, R. (1989). Model error concepts in control design. *Int. J. Control*, 49, 1725-1753.
- Steinbuch, M., P. Wortelboer, P. van Groos, and O. Bosgra (1994). Limits of implementation - a cd player control case study. *Proc. American Control Conference*, pp. 3209-3213.
- Villemagne, C. de, and R. E. Skelton (1987). Model reductions using a projection formulation. *Int. J. Control*, 46, 2141-2169.
- Villemagne, C. de, and R. E. Skelton (1988). Controller reduction using canonical interactions. *IEEE Trans. Automat. Contr.*, 33, 740-750.
- Wortelboer, P. (1994a). *Frequency Weighted Balanced Reduction of Closed-loop Mechanical Servo-systems*. PhD thesis, Delft University of Technology.
- Wortelboer, P. (1994b). *WOR-toolbox*. Toolbox for Weighted Order Reduction (available via anonymous ftp: ftp-mr.wbmt.tudelft.nl/pub/wortelboer/).
- Wortelboer, P. and O. Bosgra (1994). Frequency weighted closed-loop order reduction in the control design configuration. *Proc. 33rd IEEE Conf. Dec. & Contr.*, Lake Buena Vista, FL, pp. 2714-2719.
- Wortelboer, P., M. Steinbuch and O. Bosgra (1997). Closed-loop balanced reduction in a twin feedback configuration with application to a Compact Disc Mechanism. Submitted to the Special Issue on Model Reduction of the *Int. Journal of Robust and Nonlinear Control*.
- Zhou, K., J.C. Doyle, and K. Glover (1996). *Robust and Optimal Control*. Prentice Hall, Englewood Cliffs, NJ.

Alternative parametrization in modelling and analysis of a Stewart platform

S.H. Koekebakker, P.C. Teerhuis and A.J.J. van der Weiden

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: s.koekebakker@wbmt.tudelft.nl*

Abstract. The fully parallel driven construction of a Stewart platform is often used as a flight simulator motion system. Higher standards in motion realism imply the use of advanced model based control. This paper considers the modelling and analysis of a Stewart platform. In general, models of parallel robots result in combined algebraic and differential equations, which causes difficulties with simulation, analysis and model based control. In this paper it is shown that by making the right choices in parametrization within a modern modelling method, these difficulties are circumvented. As a result an explicit differential model, in which the different model parameters are clearly separated, is obtained. This provides a suitable starting point for simulation, analysis, model reduction and model based control.

Keywords. modelling, nonlinear systems, multibody, mechanical systems.

1 Introduction

The use of robot manipulators is widely spread in industry nowadays. Most of these manipulators are constructed as a series connection of joints and links. The dual form of these robots, the parallel manipulator, is less often seen to be applied. In (flight)simulation motion systems however, the parallel construction is almost invariably in use. The Stewart platform (see Fig. 1), introduced by Stewart (1965) is a 6 degrees-of-freedom (d.o.f.) parallel manipulator which is applied in most of the current high fidelity flight simulators. These systems are the subject of this paper.

There are several advantages in applying a parallel construction. This kind of manipulators have higher rigidity and accuracy due to the parallel force path and averaged link to end-effector error. The inverse kinematics (from link to end-effector coordinates) which is a problem in path generation of serial manipulators is easily solved in parallel robots. There are also disadvantages. The dual forward kinematics is a complex algebraic problem and has in general more than one solution (Rahaven, 1991). Modelling

the dynamics is also less straight forward.

For several reasons, feedback control of these motion systems is still decentralized i.e. per actuator without taking mechanical coupling into account. Setting higher standards of motion realism in simulation will involve modern control strategies in order to fully benefit from recent constructional and computational improvements in flight simulators e.g. light weight, low center-of-gravity (c.o.g.) platforms, high frequency airplane dynamics simulation (Advani, 1993). This research has been done in direct interaction with this development within the Simona institute. Also use of a Stewart platform as a more general robot (Nguyen and Pooran, 1989) will require high performance control of motion.

By incorporating more structural system information i.e. a model into the controller, it is possible to achieve higher performance. Most modern control strategies are therefore model based in some way (directly, in design or evaluation). In this case the quality of motion depends on the fidelity of the model. Deriving a model of the mechanics of the Stewart platform manipulator for analysis, design and control will be the subject of this paper. To

arrive at a model with structure from which insight can be gained, modelling laws will be done based on physical laws.

Modelling the dynamics of a Stewart platform as a multibody system has been done in Lee and Geng (1993) who claim to be the first to present a complete model and with more simplifications in Do and Yang (1988) and Liu *et al.* (1991). Modelling the mechanics of this platform can be done in several ways and with various objectives in mind. The equations of motion can be derived by using the classical approach of Lagrange (Lee and Geng, 1993) or Newton-Euler (Do and Yang, 1988).

In general, deriving the equations of motion of a parallel manipulator results in combined differential and algebraic (constraint) equations (see e.g. Roberston and Schwertassek (1988)). In simulation and control this formulation can cause difficulties (index problems etc., Brenan *et al.* (1989)). In this paper it is shown that an explicit differential model for the Stewart platform results if one makes the right choices in parametrization. Dependent variables are explicit functions of the integrable differential equations. In this way index problems, etc. are circumvented.

By using a modern method like Kane's (Kane and Levinson, 1985), which have the advantages of both the Newton-Euler and Lagrange formulation but without the corresponding disadvantages (Huston, 1990), it will be shown that applying this approach can result in a model from which more insight can be gained.

Together with the alternative parametrization, this is advantageous over the models earlier presented in literature, if one wants to apply a model for both analysis, simulation and control. Model based feedback, however, is still more complex for parallel manipulators, since the dynamics are only described in end-effector coordinates and the measured signals are link related. Next to the fairly low performance requirements of flight simulator motion systems in the past, this will probably be the reason that in most of these motion systems, feedback controllers are decentralized (one siso loop per link i.e. actuator).

In this paper some of the disadvantages of decentralized control can be shown by analysis of the derived model. To apply a simple, but accurate model based controller in practise, one would like to quantify the errors made by undermodelling, to be able to do robust analysis of the control scheme. The modelling approach taken in this paper aims at a model from which the influences of different system parameters, like masses, inertias, velocity, gravity can be clearly separated.

In this paper only the mechanics of the system is

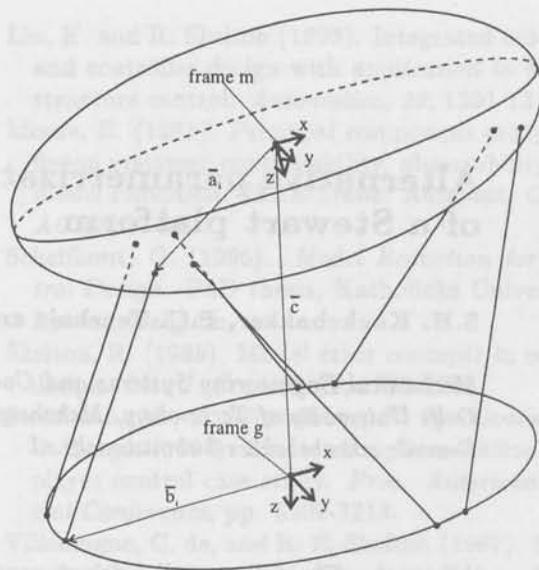


Fig. 1: A schematic view of the Stewart platform

modelled. The actuators are considered ideal force generators in the direction of the sliding joints. In practice the dynamics of the actuators e.g. hydraulic servo systems, has also to be taken into account in control design.

This paper is organized as follows. After stating some notation in Section 2 the fundamental formulas of mechanics to describe the kinematics and dynamics will be introduced in Section 3. Then in Section 4 the Stewart platform will be defined in order to derive a model of its kinematics and dynamics. After some model analysis in Section 5 with the control objective in mind, finally some conclusions will be given in Section 6.

2 Notation

Capital symbols, X are used for matrices, \bar{x} for vectors, x for scalars. With some scalar (energy) functions \mathcal{X} is used. $\bar{x} \times \bar{y}$ denotes the vector product which can also be written as $\tilde{X}\bar{y} = (\tilde{Y})^T\bar{x}$ where \tilde{X} is a skew symmetric ($X = -X^T$) matrix parametrized by the vector, $\bar{x}^T = [x_1 \ x_2 \ x_3]$, such that the result is the vector product.

$$\tilde{X} = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix} \quad (1)$$

$X \times Y$ denotes vector wise product of the columns stacked in the matrices.

The index \bar{x}_n is used for the normalizing operation $\bar{x}_n = \bar{x} / |\bar{x}|$ with $|\bar{x}| = \sqrt{\bar{x}^T \bar{x}}$. $P_{\bar{x}_n}$ denotes the projector to the (hyper)plane with normal vector \bar{x}_n and can be constructed from the vector product matrix $P_{\bar{x}_n} = (I - \bar{x}_n \bar{x}_n^T) = (\tilde{X}_n)^4 = \tilde{X}_n \tilde{X}_n^T = -(\tilde{X}_n)^2$.

Projection matrices have some nice properties like $P = P^T = P^n$.

Motion can be described w.r.t. various frames. A matrix or vector described in some frame can have a superscript referring to this frame. For the inertial frame or ground coordinates the index \bar{x}^g will be used. As a function of the moving end-effector or platform, vectors will be denoted \bar{x}^m . If a (rotation) matrix maps a vector into another frame it will be denoted as ${}^B R^A$ if R maps from A to B .

The subscript index like \bar{a}_i will be used to refer to the i^{th} -actuator if also non actuator dependent variables appear in the equation.

3 Fundamental mechanics

The aim of this section is to show how to derive a limited number of differential, and possibly also some algebraic, equations, which describe the motion of a rigid multi body mechanical system. Most theory described in this section can be found in Kane and Levinson (1985) as in many other textbooks on mechanics.

All mechanics treated here are based on the assumption of a semi-equilibrium given by Newton's second law $\bar{f} - m\ddot{\bar{p}} = 0$ in an inertial frame for any mass particle. To describe the acceleration, $\ddot{\bar{p}}$, of all parts in a system as a function of a limited number of variables some kinematics have to be introduced.

After defining the kinematics of a mechanical system, its dynamics can be specified. In the equations of motion the semi-equilibria are described in a compact form as a function of generalized velocities or coordinates. The integrals over the mass-particles of a body result in inertia matrices and the active forces are projected along the velocities by a virtual work argument.

3.1 Kinematics

The motion of a point (mass particle, joint, etc.) is usually most conveniently and invariantly defined w.r.t. the body frame to whom it's connected. The motion of a frame put in another frame generally consists of translation \bar{t} and rotation R . The orientation of a frame can be described by a rotation matrix. A rotation matrix consists of perpendicular unit vectors which describe the basis of the frame into the other frame. As a result a rotation matrix, R has the following property:

$$R^T R = I \quad (2)$$

Any 3×3 -matrix with this property and $\det(R) = 1$ is a rotation matrix. With $\det(R) = -1$ also the mirror operation is included (transformation of right hand frames to left hand frames and vice versa).

The position of a point \bar{p}^A in frame A can now be described in frame B by:

$$\bar{p}^B = \bar{t}^B + {}^B R^A \bar{p}^A \quad (3)$$

To describe the velocity of this point in the other frame one can simply differentiate this equation. Some properties of the time derivative of the rotation matrix can be derived by differentiating (2). This results in skew symmetric matrices which can be parametrized by the vector product matrix of the (thereby defined) angular velocity $\bar{\omega}$.

$$R^T \dot{R} = -\dot{R}^T R = \tilde{\Omega}^A \quad (4)$$

with

$$\tilde{\Omega}^A = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \quad (5)$$

Now with a vector which is rigidly attached to the frame A , $\dot{\bar{p}}^A = \bar{0}$, hence

$$\dot{\bar{p}}^B = \dot{\bar{t}}^B + {}^B R^A \tilde{\Omega}^A \bar{p}^A = \dot{\bar{t}}^B + \tilde{\Omega}^B \bar{p}^B \quad (6)$$

were the change of frame for the matrix $\tilde{\Omega}$, is given by $\tilde{\Omega}^B = {}^B R^A \tilde{\Omega}^A {}^A R^B$ and ${}^A R^B = ({}^B R^A)^T$.

If some variations or velocities can be described as a product of a (position-dependent) matrix and vector of other variations this matrix will be called a Jacobian matrix. In this case

$$\delta \dot{\bar{p}}^B = [I \quad {}^B R^A (\bar{P}^A)^T] \begin{bmatrix} \delta \bar{t}^B \\ \delta \bar{\omega}^A \end{bmatrix} = J_{p^B, x} \delta \bar{x} \quad (7)$$

with δ denoting the variation.

Although the rotation matrix consists of nine entries, its properties put constraints on these entries. Different parametrizations such as euler angles (three subsequent planar rotations) or euler parameters (four parameters with one normalizing constraint to describe one axis of rotation and the angle of rotation) are possible.

The three euler angles have the disadvantage of a highly non-linear appearance in both the rotation matrix and the euler angle velocity to angular velocity transformation. The latter can even become singular. The constraint equation with the euler parameters also imposes extra limitations. The choice of parametrization can however often be postponed till after the derivation of the equations of motion. By differentiating (6) the acceleration of the point rigidly attached to the frame ($\dot{\bar{p}}^A = \bar{0}$) can be calculated:

$$\begin{aligned} \ddot{\bar{p}}^B &= \ddot{\bar{t}}^B + {}^B R^A (\bar{P}^A)^T \dot{\bar{\omega}}^A + {}^B R^A (\tilde{\Omega}^A)^2 \bar{p}^A \\ &= \ddot{\bar{t}}^B + (\bar{P}^B)^T \dot{\bar{\omega}}^B + (\tilde{\Omega}^B)^2 \bar{p}^B \end{aligned} \quad (8)$$

If the point considered is already moving in the frame A ($\dot{\bar{p}}^A, \dot{\bar{v}}^A = \dot{\bar{p}}^A, \dot{\bar{a}}^A = \ddot{\bar{p}}^A$) by differentiation of

$$\dot{\bar{p}}^B = \dot{\bar{t}}^B + {}^B R^A(\tilde{\Omega})^A \bar{p}^A + {}^B R^A \dot{\bar{v}}^A \quad (9)$$

the coriolis acceleration appears as the third term in

$$\ddot{\bar{p}}^B = \ddot{\bar{a}}_{p^*}^B + {}^B R^A \ddot{\bar{a}}_{p^*}^A + 2 {}^B R^A \tilde{\Omega}^A \dot{\bar{p}}^A = \ddot{\bar{a}}_{p^*}^B + \ddot{\bar{a}}_{p^*}^B + 2 \tilde{\Omega}^B \dot{\bar{p}}^B \quad (10)$$

Where p^* is a point connected to A momentarily at the same position as \bar{p} . Its acceleration $\ddot{\bar{a}}_{p^*}$, is given by (8).

In stating the equations of motion, the state which describes the orientation, usually only appears in the rotation matrix. It is possible to parametrize the rotation by the unit vector pointing along the axis of rotation \bar{n}_μ , and the angle μ of rotation. Parametrization by the four euler parameters $\bar{\epsilon} = [\epsilon_0 \ \bar{\epsilon}_{13}^T]^T$, given by $\epsilon_0 = \cos(1/2 \mu)$ and $\bar{\epsilon}_{13} = \sin(1/2 \mu) \bar{n}_\mu$ results in very convenient (simple to calculate) relations of the rotation matrix and the angular velocity in which the euler parameters and its derivatives play an intermediate role. These relations are extensively dealt with in Nikravesh *et al.* (1985) without further derivation they will be given here.

The rotation matrix can be calculated by taking a product of two matrices which are linear in $\bar{\epsilon}$:

$$R(\bar{\epsilon}) = G(\bar{\epsilon})L(\bar{\epsilon})^T \quad (11)$$

With

$$G(\bar{\epsilon}) = [-\bar{\epsilon}_{13} \ \epsilon_0 I + \bar{\epsilon}_{13}] \quad (12)$$

and

$$L(\bar{\epsilon}) = [-\bar{\epsilon}_{13} \ \epsilon_0 I + (\bar{\epsilon}_{13})^T] \quad (13)$$

$\dot{\bar{\epsilon}}$ can be described as a product of $\dot{\bar{\omega}}$ and a matrix which is a linear function \bar{G} .

$$\dot{\bar{\epsilon}} = \frac{1}{2} G^T(\bar{\epsilon}) \dot{\bar{\omega}} \quad (14)$$

With angles $-\pi < 1/2\mu < \pi$, $\bar{\epsilon}_{13}$ can be used as the (orientation) state from which $\epsilon_0 = \sqrt{1 - \bar{\epsilon}_{13}^T \bar{\epsilon}_{13}}$ is solved. These relations are used to get from the angular velocity to the rotation matrix with help of an integration routine and initial conditions on $\bar{\epsilon}_{13}$. The kinematic relations provide means to state the motion of the system as a function of a limited number of (generalized) variables (coordinates or velocities). Together with the dynamics i.e. semi-equilibria of active and inertial forces stated in the next section the equations of motion result.

3.2 Dynamics

The equations of motion can be stated in several ways. First the Lagrange equations will be given. Then it will be shown that from these equations the more simple Newton-Euler equations follow if one rigid body is considered. Finally the method based on Kane and Levinson (1985) will be introduced.

With Lagrange the difference between the kinetic energy, \mathcal{K} , and potential energy, \mathcal{P} , of a system is called the Lagrangian, \mathcal{L} : $\mathcal{L} = \mathcal{K} - \mathcal{P}$. For an unconstrained system described by generalized coordinates \bar{x} the Lagrange equations are given by:

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\bar{x}}} \right) - \frac{\partial \mathcal{L}}{\partial \bar{x}} = \bar{f} \quad (15)$$

The driving moments/forces, \bar{f} , are all the working forces (non inertial or conservative) projected along the variations of the generalized coordinates. These are called the generalized forces.

By observing that the work δW , done by forces \bar{f} , does not change if a change of coordinates \bar{x} , is applied, it is easy to show that projecting the forces along other coordinates is equal to multiplying by the transpose jacobian.

$$\delta W = \bar{f}^T \delta \bar{x}_1 = \bar{f}^T J_{x_1, x_2} \delta \bar{x}_2 = (J_{x_1, x_2}^T \bar{f})^T \delta \bar{x}_2 \quad (16)$$

The kinetic energy can in general be described by

$$\mathcal{K} = \frac{1}{2} \dot{\bar{x}}^T M(\bar{x}) \dot{\bar{x}} \quad (17)$$

Where the mass matrix, $M(\bar{x})$, is a symmetric positive semi definite matrix. It can be shown that the partial derivative of the potential energy like gravity, with respect to the velocity is zero. In that case

$$M(\bar{x}) \ddot{\bar{x}} + \frac{d}{dt} (M(\bar{x}) \dot{\bar{x}} - \dot{\bar{x}}^T \left(\frac{\partial}{\partial \bar{x}_i} M(\bar{x}) \right) \dot{\bar{x}} + \frac{\partial}{\partial \bar{x}} \mathcal{P}(\bar{x})) = \bar{\tau} \quad (18)$$

compactly written as

$$M(\bar{x}) \ddot{\bar{x}} + C(\bar{x}, \dot{\bar{x}}) \dot{\bar{x}} + \bar{G}(\bar{x}) = \bar{\tau} \quad (19)$$

with a mass matrix M , a non linear coriolis/centripetal matrix C and a gravity vector \bar{G} . If a rigid body (with mass m , and inertia $I_{\bar{c}}$) is considered at its center of gravity \bar{c} , the Newton-Euler equations result. The mass matrix is block diagonal in this case.

$$\mathcal{K}_{body} = \frac{1}{2} \begin{bmatrix} \dot{\bar{c}} & \dot{\bar{\omega}} \end{bmatrix} \begin{bmatrix} mI & 0 \\ 0 & I_{\bar{c}} \end{bmatrix} \begin{bmatrix} \dot{\bar{c}} \\ \dot{\bar{\omega}} \end{bmatrix} \quad (20)$$

From the two blocks two independent equations result. The impulse law:

$$\Sigma \bar{f} = \frac{d}{dt} (m \frac{d}{dt} (\bar{c})) = m I \ddot{\bar{c}} \quad (21)$$

And the impulse moment law with the generalized forces \bar{f}_e (moments in this case):

$$\begin{aligned}\Sigma \bar{f}_e &= \frac{d}{dt}({}^B R^A I_e^A \bar{\omega}^A) = {}^B R^A I_e^A \dot{\bar{\omega}}^A + {}^B R^A \bar{\Omega}^A I_e^A \bar{\omega}^A \\ &= I_e^B \dot{\bar{\omega}}^B + \bar{\Omega}^B I_e^B \bar{\omega}^B\end{aligned}\quad (22)$$

The Newton-Euler equations are easily stated for each rigid body in a system. Extra equations with unknown internal forces result, however, in using this method to state the equations of motion for a multi body system. Applying Lagrange, results in taking partial derivatives of complex energy functions if the whole system is considered. These are disadvantages which can be circumvented by using Kane's method.

With Kane also generalized variations or velocities have to be specified. The general formula

$$\bar{f} + \bar{f}^* = \bar{0} \quad (23)$$

states the (semi-)equilibrium of the active forces, \bar{f} , and inertial forces, \bar{f}^* , projected along the directions of the generalized velocities. To calculate the over-all inertial forces, as in the Newton-Euler approach, the specific inertial forces generated in the frame of each body can be stated. As in the Lagrangian approach, a minimal number of equations results by writing the motion of the bodies as a function of the generalized velocities and projecting each specific force from its local coordinates to the generalized ones. Also the active forces can first be stated in an appropriate frame after which projection follows. The projection in general consists of a change of coordinates i.e. a multiplication with a jacobian. This procedure can also be automated (Kane and Levinson, 1996).

With this method it is possible to start with a strongly simplified system by calculating part of the (inertial) forces and separately adding other forces if a more accurate model has to be taken into account. The amount of generalized velocities can exceed the number of d.o.f. of a system. In that case, next to the differential equations given by the semi-equilibria of the forces (23) constraint equations of velocities and position, generally stated as

$$A(\bar{x}, t)\dot{\bar{x}} + \bar{b}(\bar{x}, t) = \bar{0}, \quad (24)$$

should be added. With a non-holonomic system these constraint equations cannot be integrated to constraint position equations.

A parallel manipulator, like the Stewart platform, is a holonomic system. Since there are kinematic chains, it is often easier to state the equations of motion of a parallel manipulator by using constraint equations. In that case the manipulator is described

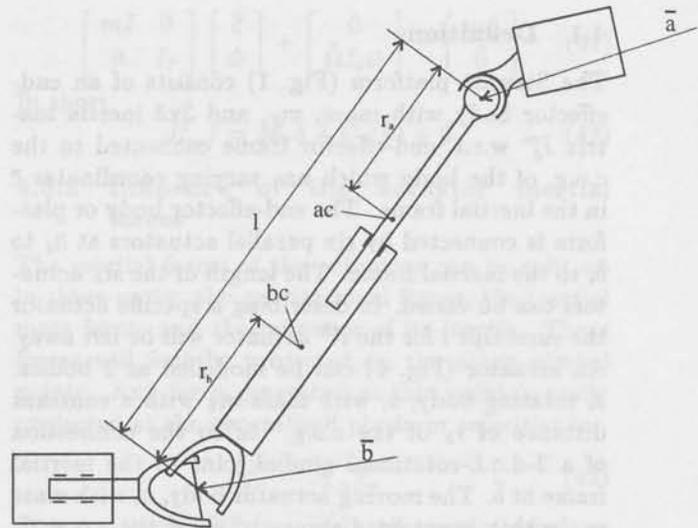


Fig. 2: Stewart platform actuator link construction

as a serial system (with some of the joints disconnected). The parallel connections are incorporated by adding constraints. A combined differential/algebraic description results. This kind of description causes difficulties (index problems etc., Brenan *et al.* (1989)) in simulation and model based control. With these goals in mind during modelling, it is more convenient to state the model in explicit differential equations if possible. In the next section it is shown that this can be done with the Stewart platform.

4 Modelling the mechanics of the Stewart platform

In the previous section the general procedure of stating the equations of motion was given. In this section this will be applied in modelling the Stewart platform. Kane's method of projecting local semi-equilibria (equations of motion) will be used to arrive at a compact description. To state the local equations, both Lagrange and Newton-Euler are applied wherever either one is most appropriate.

With the choice of platform position/orientation as the generalized coordinates, all equilibria can be written as explicit functions of these coordinates (and its derivatives).

The parallel manipulator construction of the Stewart platform is first defined. To derive the equations of motion the velocity and accelerations should be described w.r.t. a limited number of generalized variations. This defines the kinematics after which the semi-equilibrium equations of the active and inertial forces can be stated.

4.1 Definitions

The Stewart platform (Fig. 1) consists of an end-effector body with mass, m_e , and 3x3 inertia matrix I_e^m w.r.t. end-effector frame connected to the c.o.g. of the body which has varying coordinates \bar{c} in the inertial frame. The end-effector body or platform is connected by six parallel actuators at \bar{a}_i to \bar{b}_i to the inertial frame. The length of the six actuators can be varied. In describing a specific actuator the superscript i for the i^{th} actuator will be left away. An actuator (Fig. 4) can be modelled as 2 bodies. A rotating body, b , with mass m_b with a constant distance of r_b of the c.o.g. bc to the connection of a 2-d.o.f.-rotational gimbal joint to the inertial frame at \bar{b} . The moving actuator body, a , with mass m_a with a constant distance of r_a of the c.o.g. ac is connected with a 3-d.o.f.-rotational gimbal joint to the platform at \bar{a} . With a 1-d.o.f. controlled sliding joint between these two bodies the length of the actuator can be varied.

The inertia of the actuator bodies is neglected around the actuator axis. It is assumed to be uniform perpendicular to this axis. i_a is the inertia of the moving actuator body at \bar{ac} and any axis perpendicular to the actuator. i_b is the inertia of the rotating part of the actuator w.r.t. the connection to the inertial frame (\bar{b}) and any axis perpendicular to the actuator.

With this assumption also the case (often seen in practise) in which the moving part of the actuator both rotates and slides at the connection with the rotating part, and has only 2-d.o.f. rotation w.r.t. the platform, results in the same dynamics. In general (apart from singularities) with the 6-d.o.f. (one of which is controlled) of each actuator, the 6-d.o.f. of the platform, freely moving in the inertial frame, can be described and vice versa.

4.2 Kinematics

The kinematics of the Stewart platform will be described by first defining the transformation of the platform to actuator coordinates. Then by differentiation also velocity and acceleration of all relevant points can be calculated as a function of the platform motion, whose velocities will be taken as the generalized speeds.

Almost all vectors can be conveniently described in the inertial frame. Apart from \bar{a}_i^m whose time derivative in the moving frame is $\bar{0}$.

The vector, \bar{l}_i , between the two attachment points of an actuator can be described by

$$\bar{l}_i = \bar{c} + T\bar{a}_i^m - \bar{b}_i \quad (25)$$

Now the length of the actuator, $|\bar{l}_i|^2 = \bar{l}_i^T \bar{l}_i$, and the unit vector in direction of the actuator, $\bar{l}_{n,i} = \frac{\bar{l}_i}{|\bar{l}_i|}$

can be calculated from the platform variables \bar{c}^g and the orientation matrix $T = {}^gR^m$ which will be the only rotation matrix used.

The velocity of the length of the actuators can be calculated by projection of the velocity of the upper gimbal attachment point, \bar{v}_a , in the direction of the actuator, since $\frac{d}{dt} |\bar{l}| = \frac{d}{dt} \sqrt{\bar{l}^T \bar{l}} = \frac{\bar{l}^T \bar{v}_a}{|\bar{l}|} = \bar{l}_{n,i}^T \bar{v}_a$. The velocity of the upper gimbal points is given by

$$\bar{v}_{a_i} = \dot{\bar{c}} + \bar{\omega} \times T\bar{a}_i^m \quad (26)$$

By projecting this velocity, the velocity of the actuator appears.

$$\dot{l}_i = \bar{l}_{n,i}^T \bar{v}_i = \bar{l}_{n,i}^T \dot{\bar{c}} + \bar{l}_{n,i}^T (\bar{\omega} \times T\bar{a}_i^m) \quad (27)$$

With some reordering and written as matrix equation (e.g. \bar{v}_{a_i} stacked in V_a) for all the actuators the jacobian between the actuator and platform velocities is defined.

$$\dot{\bar{l}} = L_n^T \dot{\bar{c}} + (TA^m \times L_n)^T \bar{\omega} = J_{l,x} \dot{\bar{x}} = L_n^T V_a \quad (28)$$

This jacobian matrix is one of the most important variables in the Stewart platform. The jacobian between platform and gimbal point velocity is defined by

$$\bar{v}_{a_i} = [I \ T(\bar{A}_i^m)^T] \dot{\bar{x}} = J_{a_i,x} \dot{\bar{x}} \quad (29)$$

To determine the inertial forces of the actuators the jacobians from gimbal point to the c.o.g.'s of the actuators are also important. The angular velocity of the actuator perpendicular to the actuator $\bar{\omega}$, is defined by

$$\bar{\omega}_a = \bar{l}_n \times \frac{\bar{v}_a}{|\bar{l}|} \quad (30)$$

Now the velocities of the c.o.g.'s of the actuator bodies \bar{v}_{ac} and \bar{v}_{bc} can be stated as a function of \bar{v}_a :

$$\bar{v}_{ac} = \bar{v}_a + \bar{\omega}_a \times (-r_a \bar{l}_n) = (I - \frac{r_a}{|\bar{l}|} P_{l_n}) \bar{v}_a = J_{ac,a} \bar{v}_a, \quad (31)$$

and

$$\bar{v}_{bc} = \bar{\omega}_a \times r_b \bar{l}_n = \frac{r_b}{|\bar{l}|} P_{l_n} \bar{v}_a = J_{bc,a} \bar{v}_a. \quad (32)$$

The acceleration of the actuators can be calculated by differentiating (28).

$$\ddot{\bar{l}} = J_{l,x} \ddot{\bar{x}} + \dot{J}_{l,x} \dot{\bar{x}} = L_n^T \dot{V}_a + \dot{L}_n^T V_a \quad (33)$$

The derivative of the unit vectors $\bar{l}_{n,i}$ in the direction of each actuator can be calculated with:

$$\begin{aligned} \dot{\bar{l}}_n &= \frac{d}{dt} \frac{\bar{l}}{|\bar{l}|} = \frac{\dot{\bar{l}} |\bar{l}| - \bar{l} \frac{d}{dt} |\bar{l}|}{|\bar{l}|^2} \\ &= \frac{(I - \bar{l}_n \bar{l}_n^T)}{|\bar{l}|} \bar{v}_a = \frac{1}{|\bar{l}|} P_{l_n} \bar{v}_a \end{aligned} \quad (34)$$

The acceleration of the actuator length consists of a term which is the projection of the acceleration of the upper gimbal in the direction of the actuator and a positive quadratic term which is the centripetal acceleration of the actuator. So the acceleration of the actuator length is always positive if the platform is moving with constant speed in any direction. The acceleration of the upper gimbal can be derived directly with (8)

$$\ddot{\bar{a}}_i = \ddot{\bar{c}} + \dot{\bar{\omega}} \times \bar{a}_i + \bar{\omega} \times (\bar{\omega} \times \bar{a}_i) = J_{a,x} \ddot{x} - |\bar{\omega}|^2 P_{\omega} \bar{a}_i \quad (35)$$

The acceleration of the c.o.g. of the moving actuator part also generates inertial forces and can be written as a function of platform motion.

$$\dot{\bar{v}}_{ac} = \frac{d}{dt} \bar{v}_{ac} = \frac{d}{dt} (J_{ac,a} \bar{v}_a) = J_{ac,a} \dot{\bar{v}}_a + \dot{J}_{ac,a} \bar{v}_a \quad (36)$$

So the jacobian needs to be differentiated.

$$\begin{aligned} \dot{J}_{ac,a} &= \frac{d}{dt} \left(I - \frac{r_a}{|l|} P_{l_n} \right) \\ &= \frac{r_a}{|l|^2} \bar{v}_a^T \bar{l}_n P_{l_n} + \frac{r_a}{|l|^2} (P_{l_n} \bar{v}_a \bar{l}_n^T + \bar{l}_n \bar{v}_a^T P_{l_n}) \end{aligned} \quad (37)$$

Now

$$\dot{J}_{ac,a} \bar{v}_a = \frac{r_a}{|l|^2} (|P_{l_n} \bar{v}_a|^2 \bar{l}_n + 2(\bar{v}_a^T \bar{l}_n) P_{l_n} \bar{v}_a) \quad (38)$$

which clearly shows a centripetal and a coriolis term.

4.3 Dynamics

First a simplified model with the platform as the only (rigid) body, will be derived. Then the influence of the actuator inertial forces is quantified.

4.3.1 Dynamics of the platform alone

The basic structure of the Stewart platform model results if one considers the platform alone, not taking into account the inertial forces of the actuators. Since the system in this case consists of only one body, the equations of motion are easily derived with Newton-Euler taking the velocity of the platform coordinates as the generalized speed. With (21)

$$L_n \bar{f} + m_{\bar{z}} \bar{g} = m_{\bar{z}} \ddot{\bar{c}} \quad (39)$$

where \bar{f} are the forces generated by the actuators and \bar{g} is the gravity vector. And (22)

$$[TA^m \times L_n] \bar{f} = I_{\bar{z}} \dot{\bar{\omega}} + \bar{\Omega} I_{\bar{z}} \bar{\omega} \quad (40)$$

With $I_{\bar{z}} = T I_{\bar{z}}^m T^T$. Combining these two results in the simplified model of the Stewart platform.

$$\begin{bmatrix} L_n \\ TA^m \times L_n \end{bmatrix} \bar{f} = \dots$$

$$\begin{bmatrix} mI & 0 \\ 0 & I_{\bar{z}} \end{bmatrix} \begin{bmatrix} \ddot{\bar{c}} \\ \dot{\bar{\omega}} \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{\Omega} I_{\bar{z}} \bar{\omega} \end{bmatrix} = \begin{bmatrix} m\bar{g} \\ \bar{0} \end{bmatrix} \quad (41)$$

In short

$$J_{l,x}^T \bar{f} = M_{\bar{z}} \ddot{\bar{c}} + C_{\bar{z}}(\dot{\bar{c}}) + \bar{G}_{\bar{z}} \quad (42)$$

4.3.2 Influence of the actuator inertial forces

The inertial forces of the actuators can be split up in three parts: the gravitational forces, the inertial mass forces and the influence of its inertia. These forces will first be projected on the upper gimbal points. Any force generated at this point is easily projected at the generalized platform velocities by

$$\bar{f}_a^x = J_{a,x}^T \bar{f}_a^a \quad (43)$$

The gravitational forces are easily projected along the platform velocities, the rotating part of the actuator

$$\bar{f}_{bg}^a = J_{bc,a}^T m_b \bar{g} = \frac{m_b r_b}{|l|} P_{l_n} \bar{g} \quad (44)$$

In a position where the gravity vector is directed along the actuator (\bar{l}_n) this force would not contribute. The contribution of the moving part is in that case maximal at \bar{a} as is shown by

$$\bar{f}_{ag}^a = J_{ac,a}^T m_a \bar{g} = m_a \left(I - \frac{r_a}{|l|} P_{l_n} \right) \bar{g} \quad (45)$$

The inertial force generated by the mass at the c.o.g. of the moving part of the actuator is easily described in this point.

$$\bar{f}_{ac} = m_a \dot{\bar{v}}_{ac} \quad (46)$$

Projection of this force at the upper gimbal point results in

$$\begin{aligned} \bar{f}_{m_a}^a &= m_a \left(I - P_{l_n} + \frac{(|l| - r_a)^2}{|l|^2} P_{l_n} \right) \dot{\bar{v}}_a \dots (47) \\ &+ 2m_a \frac{(|l| - r_a)r_a}{|l|^3} (\bar{l}_n^T \bar{v}_a) P_{l_n} \bar{v}_a + \\ &+ m_a \frac{r_a}{|l|^2} \bar{l}_n \bar{v}_a^T P_{l_n} \bar{v}_a \end{aligned}$$

The first term consists of a part in the direction of the actuator where the mass directly acts on the gimbal point. Perpendicular to this direction the influence gets smaller with the squared ratio of distances to the lower gimbal point. The second and third term are the coriolis and centripetal force.

The inertial forces generated by the inertias of the lower and upper part of the actuator can be taken together since their contribution to the kinetic energy is equal:

$$\begin{aligned} \mathcal{K}_{i_a, i_b} &= \frac{1}{2} \bar{\omega}_a^T \bar{\omega}_a (i_a + i_b) \\ &= \frac{1}{2} \bar{v}_a^T \frac{(i_a + i_b)}{|l|^2} P_{l_n} \bar{v}_a = \frac{1}{2} \bar{v}_a^T M_{i_a, i_b}^a \bar{v}_a \end{aligned} \quad (48)$$

With Lagrange and

$$\frac{d}{dt}(M_{i_a, i_b}) = -\frac{(i_a + i_b)}{|l|^3} (2\bar{v}_a^T \bar{l}_n P_{l_n} + P_{l_n} \bar{v}_a \bar{l}_n^T + \bar{l}_n \bar{v}_a^T P_{l_n})$$

$$\frac{\delta \mathcal{K}_{i_a, i_b}}{\delta \bar{p}_a} = -\frac{(i_a + i_b)}{|l|^3} ((\bar{l}_n^T \bar{v}_a) P_{l_n} \bar{v}_a + \bar{l}_n \bar{v}_a^T P_{l_n} \bar{v}_a),$$

the inertial forces at \bar{a} result

$$\bar{f}_{i_a, i_b}^a = M_{i_a, i_b} \ddot{\bar{v}}_a - \frac{2(i_a + i_b)}{|l|^3} (\bar{l}_n^T \bar{v}_a) P_{l_n} \bar{v}_a. \quad (49)$$

The contribution to the mass matrix only exists at motion perpendicular to the direction of the actuator. Next to this, only coriolis and no centripetal terms appear. The coriolis force is generated as a result of the inertia points in the opposite direction as the one generated by the mass. This is due to the fact that the influence of the inertia decreases while that of the mass increases as the actuator gets longer.

4.4 The Stewart platform model

The equation of motion of the Stewart platform including the inertia of the actuators can still be put in form of

$$J_{l, z}^T \bar{f} = M_t \ddot{\bar{x}} + C_t(\dot{\bar{x}}, \bar{x}) + \bar{G}_t(\bar{x}) \quad (50)$$

Where M_t , C_t and \bar{G}_t are given by

$$M_t = M_{\bar{z}} + \sum_{i=1}^6 J_{a_i, z}^T (M_{m_{a,i}} + M_{i_{a,i} + i_{b,i}}) J_{a_i, z} \ddot{\bar{x}} \quad (51)$$

$$C_t = C_{\bar{z}} + \sum_{i=1}^6 J_{a_i, z}^T (C_{m_{a,i}} + C_{i_{a,i} + i_{b,i}}) J_{a_i, z} \dot{\bar{x}} \dots - |\bar{\omega}|^2 (M_{m_{a,i}} + M_{i_{a,i} + i_{b,i}}) P_{\omega} \bar{a}_i \quad (52)$$

$$\bar{G}_t = \bar{G}_{\bar{z}} + \sum_{i=1}^6 J_{a_i, z}^T (\bar{G}_{m_{a,i}} + \bar{G}_{m_{b,i}}) \quad (53)$$

This model is parametrized by the platform coordinates only and the effect of each term (mass, coriolis, centripetal, gravity, driving forces) and parameter (mass, inertia, centers of gravity, gimbal point) can be clearly distinguished.

5 Model analysis

In this section the model of the Stewart platform will be further analyzed. Three issues, with the control objective in mind, are treated. The jacobian matrix, $J_{l, z}$, given by (28), plays a central role in the system, as will be discussed first. Secondly analyzing a linearized version of the model reveals some of the drawbacks of decentralized control. Motivating the use of model based control. In that case one would like to use a simplified model as given by (41) instead of a complex model like (50). Quantification of the differences between these two models should be possible to justify this simplification. The influence of the parasitic actuator forces, which quantify this difference, will be discussed in the third subsection.

5.1 Interpretation and use of the jacobian matrix, $J_{l, z}$

If the system has to be controlled by the actuators, the jacobian specifies how the control inputs, the actuator forces, influence the platform (accelerations) which are, especially in flight simulation applications, often the variables to be controlled. Further the measurable outputs are often only the actuator lengths. The derivatives (actuator speed) of these outputs are given by the product of the jacobian and the platform speed.

There are two interpretations to the jacobian. In the force interpretation the rows of $J_{l, z}$ give the (generalized) forces in the platform coordinates given a unit force in an actuator. In the velocity interpretation the columns of $J_{l, z}$ specify the velocity of the actuators to have unit velocity of the platform.

In model based control, the inverse information is of interest. The measured variations of the actuator have to be put in platform variations to calculate corrections in a model specified in platform coordinates. Each column of the inverse jacobian, $J_{l, z}^{-1}$, specifies what velocity (angular velocity included) of the platform is necessary to have elongation of just one actuator while the others only rotate.

The correction forces in a model based controller are also calculated in platform coordinates. Each row of $J_{l, z}^{-1}$ specifies the forces necessary in the actuators to have unit force correction in platform coordinates. The inverse jacobian appears in feedback linearizing structures (like computed torque, etc.) which will be dealt with in forthcoming contributions.

Another problem of a parallel manipulator with only the link position measured are the forward kinematics. It is not known how to analytically calculate the platform position (without decision making about roots) from link measurements. The jacobian provides a way to apply a Newton-Raphson iteration to calculate the solution provided one starts in a point

sufficiently close to the solution and away from jacobian singularities.

$$\bar{x}_{j+1} = \bar{x}_j + J_{l,x_j}^{-1}(\bar{l}_{measured} - \bar{l}_j) \quad (54)$$

The condition number of $J_{l,x}$ also provides a measure for the controllability of the platform from the actuators which becomes uncontrollable at singularities of this matrix.

Further most of the constraints of the platform are caused by the characteristics of the actuators like limited stroke, maximum speed and force. The jacobian plays an important role in translating these limitations into platform coordinates.

5.2 Analysis of a linearized model

If the model of the Stewart platform is linearized at zero speed in some position with gravity assumed to be compensated for, the following equation results from actuator force to actuator acceleration:

$$\ddot{\bar{l}} = J_{l,x} M^{-1} J_{l,x}^T \bar{f} \quad (55)$$

Consider each actuator provided with a similar compensator, as is often the case in decentralized control. This compensator feeds back the difference between the desired and measured actuator length to the actuator which generates a force. Let the transfer function from the error length to the actuator force in the linearized case be given by $g(s)$. Now some interesting properties of the compensated system can be derived.

The mass matrix can be decomposed into a singular value or eigenvalue decomposition (which is the same for a positive definite symmetric matrix):

$$J M J^T = U \Lambda U^T \quad (56)$$

The unimodular matrix, U , can be interpreted as the interaction matrix. The Λ -matrix can be seen as the mass matrix in the decoupled direction. Each element λ_i now defines the mass seen in the direction specified by the i^{th} -column of U .

The properties of a closed loop transfer function like the sensitivity from reference length to the error length, $\bar{e}_l = S(s)\bar{r}_l$, are influenced by the system's mass matrix, $J M J^T$, in the following way.

$$S(s) = U \begin{pmatrix} \frac{s^2}{(s^2 + \lambda_1 g(s))} & & 0 \\ & \ddots & \\ 0 & & \frac{s^2}{(s^2 + \lambda_n g(s))} \end{pmatrix} U^T \quad (57)$$

With decentralized control the interaction directions are specified with the model. In a flight simulator motion system as described in Advani (1993) e.g. one typically finds interaction of surge/pitch

and sway/roll. Further the compensator should be able to deal with mass variations given by the singular values of the mass matrix. With a flight simulator the condition number of the mass matrix, which is a measure of these variations, is given by

$$\kappa(\Lambda) = \frac{\lambda_{max}}{\lambda_{min}} \quad (58)$$

It is already larger than 10 in the favourable neutral position. One cannot expect high performance from such a system.

With model based control, unlike decentralized control, decoupling of the mass matrix is possible and as a result of this, each 'reflected' mass, λ_i , can be compensated for separately.

5.3 Influence of parasitic actuator forces

Adding the inertial influence of the actuators to the simplified model did not change the compact form of six coupled second order differential equations. The equations, however, became much more complex which is not favourable in model based control in which the model has to be calculated at high speed.

In Ji (1994) it is claimed that the actuator inertial effects can be seen as a change of the platform mass, inertia and c.o.g. This claim should be carefully interpreted as this change is not only dependent on the position, but also on the direction of the motion i.e. not valid at one operating point. In case of the flight simulator e.g. the mass of the actuators add more to the mass matrix of simulator in heave than in the lateral directions of surge and pitch.

With the equations given, it is possible to give bounds on the forces not taken into account if the inertial forces of the actuators would be neglected. Although with conventional motion systems this is often justified, the tendency towards light weight platforms makes the actuator inertial forces more evidently come into play. Total neglect would result in a too rough approximation in that case. Approximation by a constant additive term would be more convenient.

6 Conclusion

In this paper the dynamics of the Stewart platform is stated as a set of differential equations without algebraic constraints resulting from the kinematic chains in the system. This is possible by writing the actuator motion explicitly as a function of platform motion.

By using Kane's method of projecting forces onto the generalized velocities, each contribution is quantified separately.

In linearizing the system, it is shown that with a decentralized controller the coupling through the joint mass matrix is not influenced. The singular values of this mass matrix quantify the mass variations a decentralized feedback controller has to deal with (moving in different directions). These variations appear to be considerable in practise which motivates the use of multivariable, possibly model based, control.

To apply known non-linear model based control techniques for the Stewart platform, one needs to deal with the forward kinematical problem of a parallel manipulator which appears in the feedback path. This will be dealt with in forthcoming contributions.

References

- Advani, S.K. (1993). The development of SIMONA: a simulator facility for advanced research into simulation techniques, motion system control and navigation systems technologies. *Proc. AIAA conference*, pp. 156-166.
- Brenan, K.E., S.L. Campbell and L.R. Petzold (1989). *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. Elsevier Science Publishers B.V.
- Do, W.Q.D. and D.C.H. Yang (1988). Inverse dynamics analysis and simulation of a platform type of robot. *Journal of Robotic Systems*, 5, 209-227.
- Huston, R.L. (1990). Multibody dynamics formulations via Kane's equations. In J.L. Junkins (Ed.), *Mechanics and Control of Large Flexible Structures*, Chap. 4, pp. 71-86. Number 129 In *Progress in Astronautics and Aeronautics*. AIAA.
- Ji, Zhiming (1994). Dynamics decomposition for stewart platforms. *Journal of Mechanical Design*, 116, 67-69.
- Kane, Thomas R. and David A. Levinson (1985). *Dynamics: Theory and Applications*. McGraw-Hill Book Company.
- Kane, Thomas R. and David A. Levinson (1996). *Dynamics Online, Theory and Implementation with Autolev*. OnLine Dynamics, Inc.
- Lee, J.D. and Z. Geng (1993). A dynamic model of a flexible stewart platform. *Computers and Structures*, 48, 367-374.
- Liu, Kai, Mick Fitzgerald, Darren M. Dawson and Frank L. Lewis (1991). Modeling and control of a stewart platform manipulator. In N. Sadegh and Ye-Hwa Chen (Eds.), *Control of Systems with Inexact Dynamic Models*, ASME, New York, pp. 83-89. Presented at Winter Annual Meeting ASME, Atlanta, GA, USA.
- Nguyen, Charles C. and Farhad J. Pooran (1989). Dynamic analysis of a 6 dof ckm robot end-effector for dual-arm telerobot systems. *Robotics and Autonomous Systems*, 5, 377-394.
- Nikravesh, P.E., R.A. Wehage and O.K. Kwon (1985). Euler parameters in computational kinematics and dynamics. *Trans. of the ASME*, 107, 358-369.
- Rahaven, M. (1991). The stewart platform of general geometry has 40 configurations. *Proc. ASME Des. Tech. Conf.-Advances in design automation-DE Vol. 32-2*.
- Roberson, R.E. and R. Schwertassek (1988). *Dynamics of Multibody Systems*. Springer-Verlag.
- Stewart, D. (1965). A platform with six degrees of freedom. *Proc. Inst. Mech. Engrs.*, pp. 371-386.

High-performance motion control of a flexible mechanical servomechanism[‡]

Dick de Roover[§], Okko H. Bosgra, Frank B. Sperl[¶] and Maarten Steinbuch[‡]

Mechanical Engineering Systems and Control Group

Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.

E-mail: D.deRoover@wbmt.tudelft.nl

Abstract. This paper shows how, using modern control theory, a high-performance motion control system is designed for a flexible mechanical servosystem, which shows improved performance compared to a standard motion control system. The design comprises multivariable feedback control taking modelled dynamic interaction into account, together with model-based design of reference and command signals which minimize residual vibration at the end of a movement. As a consequence of repeating the applied reference and command signals, robustness of the scheme against modelling errors can be improved by iteratively learning the reference and command signals.

Keywords. flexible mechanical servomechanism, multivariable feedback control, point-to-point control, model-based feedforward control, iterative learning control.

1 Introduction

In order to be competitive, modern mechanical positioning devices, are required to perform both fast and accurately. Due to the inherent flexibility of the mechanical construction of most positioning devices, these performance requirements are conflicting, i.e. the faster the system moves, the less accurate it will be, due to large vibrations induced by fast movements and large acceleration forces.

One possible solution for this flexibility problem, is to redesign the positioning mechanism, for example by enlarging the stiffness of the flexible components. However, besides the fact that this solution may be rather expensive, it is often impossible to change only some components without altering the construction of the machine. Therefore, the goal of the research reported in this paper, is to find a solution for the flexibility problem by designing a

high-performance motion control system.

In general, a motion control system of a servo mechanism has three degrees of freedom, see Figure 1. The first degree of freedom is the choice of an out-

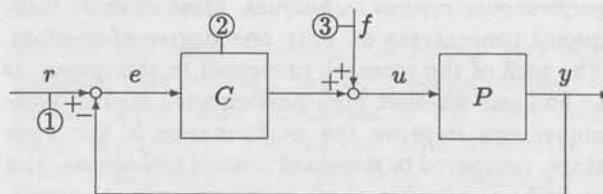


Fig. 1: General 3 degree-of-freedom motion control system; P, C, f, r, u, y, e , denote the plant, feedback compensator, force input, output reference, system input, system output, and tracking error, respectively.

put reference signal r , i.e. a desired trajectory which the measured position output y is assumed to follow. The second degree of freedom is the choice of a feedback compensator C , applied to stabilize the system at any desired position, to suppress disturbance signals acting on the system, and to force the

[‡]This paper is presented at the Philips Conference on Applications of Control Technology, PACT'96, 29-30 October 1996, Epe, The Netherlands.

[§]The work of Dick de Roover is financially supported by Philips Research Laboratories, Eindhoven, The Netherlands.

[¶]Philips Research, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands.

system output y to follow the desired trajectory r . The third degree of freedom is the choice of a force profile f applied to the system input, i.e. a force generated by an actuator, most times applied as a feedforward signal of r , in order to speed up the tracking of r .

In a recent report, the motion control system of a flexible $xy\phi$ -stage is described, which fits well into the general configuration of Figure 1, see Bartelings *et al.* (1996). In this motion control system, henceforth denoted as the *standard control system*, the force profile f is chosen as a time-optimal 'bang-bang' profile, i.e. first the stage is maximally accelerated, whereafter it is maximally decelerated, so as to reach the desired end position. In practical situations, also the derivative of the force profile (actuator jerk) and the derivative of the position output (stage velocity) are limited by physical constraints. The output reference signal r is obtained by integrating the force profile f twice, and dividing it by a constant gain corresponding to the mass of the stage, according to Newton's law; in this case, the signal r is fed forward by the signal f . The feedback compensator in the standard control system is obtained by statically decoupling the actuator inputs from the position outputs, and thereafter placing the poles and zeros of each single loop at desired locations. Conform the notion of standard control system, the techniques used for design of this system will be denoted as *standard control techniques*.

At present, modern theory on systems and control offers a large number of (different) techniques for designing a high-performance motion control system, like H_∞ feedback control, point-to-point control, two-degree-of-freedom control, and many others. These techniques will be denoted as *high-performance control techniques*. Most of these techniques concentrate on only one degree-of-freedom. The goal of the research presented in this paper, is to find out whether high-performance control techniques can improve the performance of the $xy\phi$ -stage, compared to standard control techniques, and to find out whether there exists an optimal combination of the three degrees of control freedom, so that a maximum performance of the stage can be obtained. An important feature of almost any high-performance control technique, is the use of system knowledge, for example the use of a linear time-invariant model of a system. Therefore this aspect will get much attention.

This paper is organized as follows. In the next section the experimental setup, used for the experiments shown in this paper, is described, and system knowledge is obtained by modelling this setup. In Section 3, a high-performance multivariable feedback controller is presented. Section 4 continues

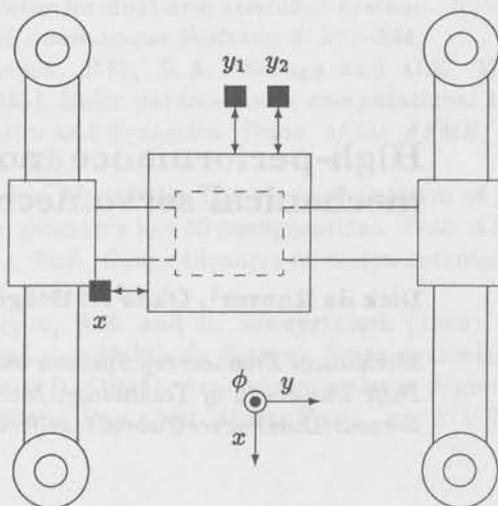


Fig. 2: Schematic top-view of an $xy\phi$ -stage; the average value of the measurements y_1 and y_2 is taken as position in y -direction, and the difference between y_1 and y_2 is a measure for rotation ϕ .

with high-performance design of the force and reference signals f and r , respectively, known as *point-to-point control*. In Section 5 a special type of feedback scheme is presented, known as *iterative learning control*, which iteratively updates the force profile in case the output of the system has to follow the reference signal repeatedly. Finally, in Section 6 some conclusions are drawn.

2 Modelling the experimental setup

Figure 2 shows a schematic top-view of a prototype $xy\phi$ -stage experimental setup, used for the experiments shown in this paper. The stage, consisting of airfoot, translator (translating part of a linear motor), and mirror block, is driven by a linear motor in x -direction. The stator part of that motor is fixed to the translator parts of two other linear motors, which drive the stage in y -direction; by driving these two motors independently, also a slight rotation ϕ of the stage is possible. The position of the stage in the horizontal plane is measured with three laser interferometers, one in x -direction and two in y -direction. Therefore, the positioning system is multivariable, having three actuators, henceforth denoted as *inputs*, and three sensors, henceforth denoted as *outputs*. Since the dynamics in rotational direction are of less importance, in this paper only results in x and y direction are shown. To gain insight in the dynamic system behaviour, two different ways of modelling have been followed: *analytic* modelling and *experimental* modelling. An-

alytic modelling, also called *white box* or *physical* modelling, concerns the modelling of a system on the basis of first principles, like the laws of Newton, explicitly taking into account the physical structure of the system, see for example Kane and Levinson (1985), Führer and Schwertassek (1990). In de Roover and van Marrewijk (1995) an analytic model was derived for the stage, which describes the most relevant dynamic behaviour of the stage in the horizontal plane. The use of this model is to understand the physical system behaviour in a qualitative way, and not to give an exact quantitative description. A quantitative description of the dynamic system behaviour can be obtained by performing experimental modelling, also called *black box* modelling or *system identification*. The idea of experimental modelling, is to excite the system dynamics with some suitable force profile and/or reference signal, see Figure 1, and to measure some of the input and output signals during excitation. Using some realization or optimization technique, a model can be derived which explains the measured data, without a direct physical interpretation, see for example Ljung (1987), Söderström and Stoica (1989).

In de Callafon *et al.* (1996) an experimental model was derived for the stage, using frequency-domain identification techniques; a linearly parametrized time-invariant model was fitted to a frequency response of the system, computed from sets of time-domain data. Figure 3 shows the computed frequency response, together with a 30th order model fitted to this response. In the fit procedure, extra weights were applied which emphasized the mid-frequency range, important for control design. This figure shows a typical response of a general mechanical servomechanism: at low frequencies, the response has the shape of a double integrator, according to the law of Newton, and at middle and high frequencies some resonances turn up, due to the flexible components. This model is used for all control design methods in the remainder of this paper.

3 Multivariable feedback control

In the standard control system, the inputs are decoupled from the outputs at low frequencies, by multiplying the system with a static pre-compensator. In Figure 3 it is seen that at frequencies above ≈ 1250 rad/s (≈ 200 Hz), the dynamic interaction between inputs and outputs cannot be neglected anymore. The surplus value of modern high-performance feedback control techniques like H_∞ feedback control, μ -synthesis, or Quantitative Feedback Theory (QFT), is the ability to explicitly cope with the dynamic interaction of a system.

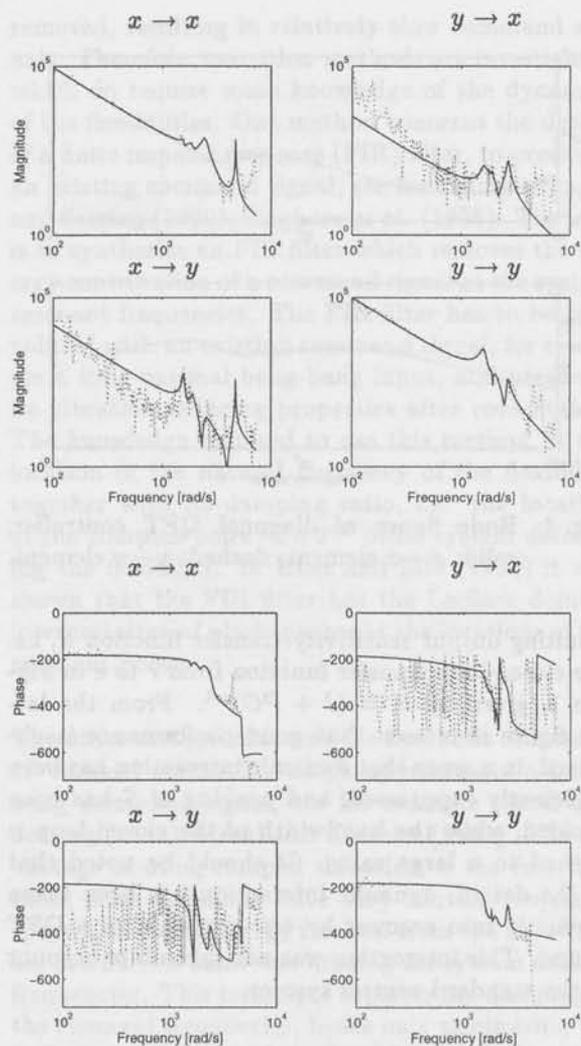


Fig. 3: Bode figure of computed frequency response $xy\phi$ -stage (dashed), together with 30th order model (solid), from inputs x, y to outputs x, y .

For the $xy\phi$ -stage, feedback controllers have been designed both with H_∞ and QFT design techniques. In this paper, only the results of the latter technique are shown. One of the main objectives of QFT, is to design simple, low order controllers with minimum bandwidth, that satisfy a number of performance specifications in the presence of uncertain system knowledge, see for example Horowitz (1963), Borghesani *et al.* (1995). In a first step, performance specifications are translated to so-called QFT bounds in a Nichols chart. In a second step, semi-automatic shaping of a systems frequency response is performed using a graphical user interface, to satisfy the QFT bounds. The outcome is a diagonal controller of any desired order, specified by the user.

A complete QFT design has been performed for the $xy\phi$ -stage. Figure 4 shows a Bode plot of the resulting controller, and Figure 5 a Bode plot of the

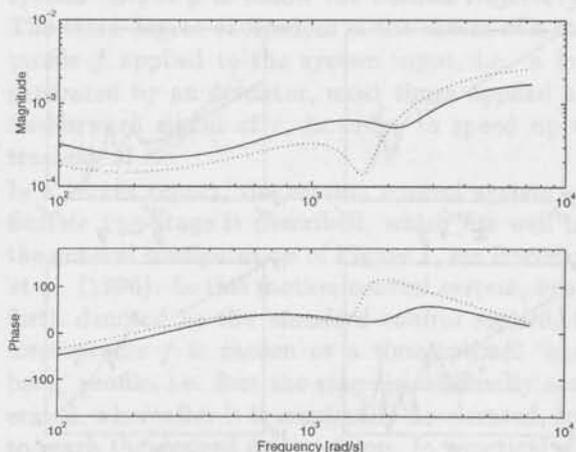


Fig. 4: Bode figure of diagonal QFT controller; solid: $x-x$ element, dashed: $y-y$ element.

resulting output sensitivity transfer function S , i.e. the closed-loop transfer function from r to e in Figure 1, given by $S = (I + PC)^{-1}$. From the latter figure it is seen that good performance is obtained, in a sense that dynamic interaction has been sufficiently suppressed, and peaking of S has been avoided, while the bandwidth of the closed loop is pushed to a large value. It should be noted that in the design, dynamic interaction has been taken explicitly into account by translating it to a QFT bound. This interaction was not taken into account in the standard control system.

4 Point-to-point control

4.1 Input shaping

To make a comparison between standard and high-performance control techniques more transparent, in this and the next section, only results are shown in one direction of the stage. One of the main topics of the research presented in this paper, is the design of force (acceleration) profiles and position reference trajectories, that minimize vibration of the system at the end of a transient from one chip position to another. From optimal control theory, it is well known that the shortest transient time is generated by a 'bang-bang' force profile, i.e. a profile which first maximally accelerates the system, whereafter it maximally decelerates the system, so as to bring it to rest in the desired end-position, see Figure 6 (a). However, when applying this signal to the stage, large vibrations were induced during the transient, originating from the flexible elements, see Figure 6 (c). Although the *step time*, defined as the duration of the force profile, see Figure 6 (a), is the shortest possible, the *settling time*, defined as

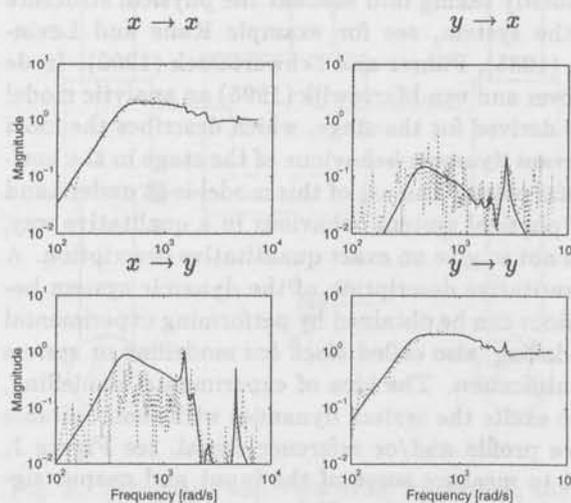


Fig. 5: Magnitude Bode figure of output sensitivity S , from inputs x, y to outputs x, y ; solid: model, dashed: frequency response data.

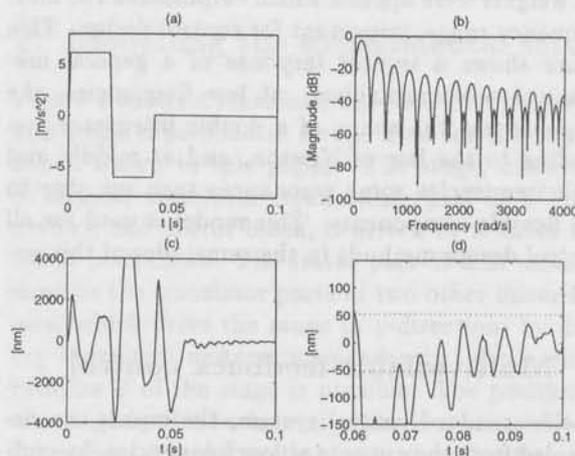


Fig. 6: Command response bang-bang force profile for 2.5 mm step in x -direction; (a) force profile, (b) normalized spectrum of force profile, (c) tracking error, (d) close up tracking error with performance bounds.

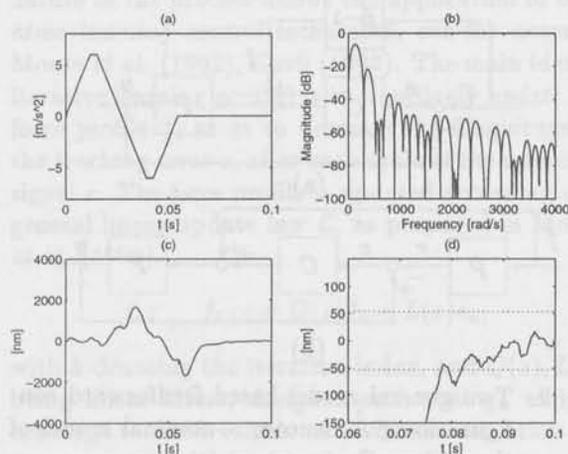


Fig. 7: Command response limited slope force profile for 2.5 mm step in x -direction; (a) force profile, (b) normalized spectrum of force profile, (c) tracking error, (d) close up tracking error with performance bounds.

the time between the end of the step and the moment at which the tracking error has settled within a narrow band surrounding the desired end-position, is even longer than the step time, see Figure 6 (d). Hence the resulting *cycle time*, defined as the sum of step time and settling time, is rather large. The reason for the excessive settling time is easily explained from the spectrum of the force profile, shown in Figure 6 (b). Comparing this figure with Figure 3, it is seen that the bang-bang force profile contains relatively much energy at the frequency range where the flexible dynamics are located ($\approx 1500 - 3000$ rad/s), and has even noticeable energy content for frequencies above 3000 rad/s.

To reduce the high frequency energy content of the force profile, the standard control system applies bang-bang acceleration profiles with limited jerk (time derivative of acceleration), see Figure 7 (a). By limiting the slope of the profile, the step time of the profile increases, but, in general the settling time decreases, as a result of reduced energy content of the profile at the locations of the resonant system poles, see Figure 7 (b). Therefore, an optimal slope can be determined as that value at which the cycle time, i.e. the sum of step and settling time, is the shortest. For a 2.5 mm step, a jerk of 500 m/s^3 resulted in a minimal cycle time, see Figures 7 (c) and (d). Clearly, the spectral content at the system resonant poles has been sufficiently reduced.

The standard technique for minimizing residual vibration is easy to apply, since no knowledge of the flexible dynamics is required at all. However, a disadvantage of this approach might be the fact that *all* of the high frequency content of the input signal is

removed, resulting in relatively slow command signals. Therefore, two other methods are investigated which do require some knowledge of the dynamics of the flexibilities. One method concerns the design of a finite impulse response (FIR) filter, to preshape an existing command signal, see for example Singer and Seering (1990), Singhose *et al.* (1995). The idea is to synthesize an FIR filter which removes the energy contribution of a command signal at the system resonant frequencies. The FIR filter has to be convoluted with an existing command signal, for example a time optimal bang-bang input, and preserves its vibration reducing properties after convolution. The knowledge required to use this method, is the location of the natural frequency of the flexibility together with its damping ratio, i.e. the location of the complex poles of a 2^{nd} order system describing the flexibility. In Bhat and Miu (1990) it was shown that the FIR filter has the Laplace domain interpretation of placing zeros at the locations of the resonant poles.

The other method concerns the synthesis of a series of 'ramped sinusoids', which approximate a bang-bang command signal, see for example Meckl and Seering (1985), Meckl and Kinceler (1994). The advantage of using ramped sinusoids, is the fact that these basis functions have very narrow frequency spectra, allowing energy removal from the input signal in a narrow band surrounding the system natural frequencies. This technique neglects the damping of the resonant frequencies, hence only minimizing the energy contribution of the input signal at the natural frequencies in the Fourier domain; consequently, the only knowledge of the system required, are the locations of the natural frequencies of the flexibilities.

Since both methods perform equally well, only the results of the latter method are shown in Figure 8. The results are quite appealing. Especially Figure 8 (b) shows that fine control over the spectrum of the force profile is possible; by removing energy from the command signal only at those frequencies where this is necessary, a relatively fast command signal is preserved. Note that robustness to uncertainty in the location of the system resonant frequencies, can be tuned by varying the width and the depth of the notch(es) in the frequency spectrum.

Table 1 quantitatively summarizes the obtained results, showing that high-performance command generating techniques allow for an extra 6% reduction of cycle time, compared to the standard command generating techniques.

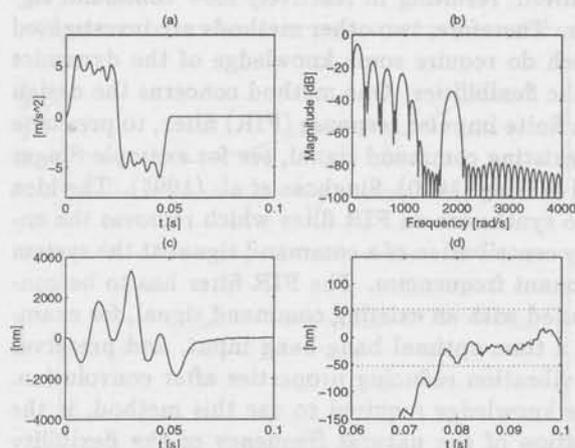


Fig. 8: Command response ramped sinusoid force profile for 2.5 mm step in x -direction; (a) force profile, (b) normalized spectrum of force profile, (c) tracking error, (d) close up tracking error with performance bounds.

	bang-bang	lim. slope	sinusoids
T_{step} [s]	$4.26e^{-2}$	$5.46e^{-2}$	$5.07e^{-2}$
T_{settle} [s]	$4.90e^{-2}$	$2.73e^{-2}$	$2.59e^{-2}$
T_{cycle} [s]	$9.16e^{-2}$	$8.19e^{-2}$	$7.66e^{-2}$
T_{step} [%]	100	128	119
T_{cycle} [%]	100	89.4	83.6

Table 1: Experimental results of bang-bang command signal and two vibration reducing command signals.

4.2 Closed-loop implementation

Since the $xy\phi$ -stage is marginally stable, the force profiles had to be implemented in a closed loop, according to Figure 1, i.e. besides suitable force profiles, also suitable reference trajectories had to be generated. As mentioned in the introduction, the reference signal in the standard control system was obtained by integrating the limited slope force profile twice, hence making f a feedforward signal of r . For the high-performance command generating technique, a reference signal could also have been obtained by integrating f twice, see for example Meckl and Kinceler (1994). However, we prefer to choose a more general *model based* feedforward setting. Figure 9 shows two general model based feedforward configurations. In Figure 9 (a), a reference trajectory r is generated, and fed forward by filtering it with the inverse of a nominal model \hat{P} of the system P . In Figure 9 (b), the force profile f is generated, and a corresponding reference trajectory is obtained by filtering f with a nominal model

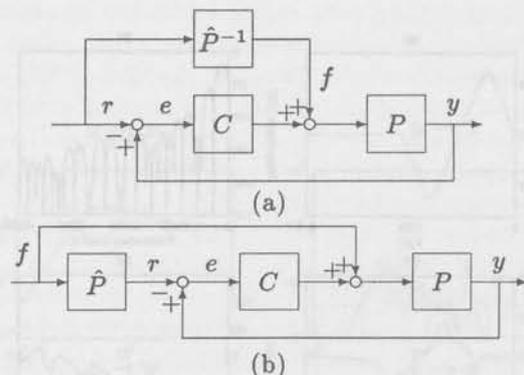


Fig. 9: Two general model based feedforward configurations; \hat{P} denotes a nominal model of the system P .

\hat{P} . This latter scheme was used for implementation of the model-based command generating technique. Since f is fed both to the system and the nominal model, the feedback controller C operates only if there is a difference between the system P and the nominal model \hat{P} , or if there are external disturbances acting on the system. That is the reason why the resulting closed loop system is called a general *two-degree-of-freedom nominal tracking system*, see for example Vidyasagar (1985), Hara and Sugie (1988).

5 Iterative learning control

Although high-performance point-to-point control shows promising results, there is still a noticeable difference between the reference trajectory and the system output, according to Figure 8 (c). The main reason for this, is a discrepancy between the nominal model \hat{P} and the system P in the two-degree-of-freedom *nominal tracking* structure of Figure 9. From the theory on systems and control, it is well-known that *robust tracking* can be achieved, for reference and command signals which are *persistent* in nature, i.e. signals which are non-decaying in time, like steps, ramps, undamped sinusoids, etc, see for example Davison (1972), Francis and Wonham (1976). Two devices are found in literature, which solve this so-called robust tracking problem, namely the servo compensator, see for example Davison (1975), and the disturbance observer, see for example Johnson (1976). In de Roover and Bosgra (1996) the dual nature of both concepts is explained, and guidelines are given for a proper choice between the servo compensator and the disturbance observer in a multivariable system.

The reference and command signals applied to the $xy\phi$ -stage, are also persistent, because they are *repeated* an indefinite number of times. This repetitive

nature of the process allows the application of *iterative learning control* techniques, see for example Moore *et al.* (1992), Kavli (1992). The main idea of iterative learning control is to iteratively *update* the force profile f , so as to *decrease* the magnitude of the tracking error e , after each cycle of the reference signal r . The force profile is updated according to a general linear update law \mathcal{L} , as proposed in Moore *et al.* (1992):

$$\mathcal{L}: f_{k+1} = Q(z)f_k + L(z)e_k, \quad (1)$$

with k denoting the iteration index, and $Q(z), L(z)$ being linear filters, designed such that f_k and e_k converge to fixed values. A sufficient condition for convergence of (1) is given by:

$$\|Q - L(I + PC)^{-1}P\|_p < 1, \quad (2)$$

with $\|X\|_p$ denoting the gain of X , measured in some induced p -norm. With update law (1), the control system of Figure 9 (b) changes to the one shown in Figure 10.

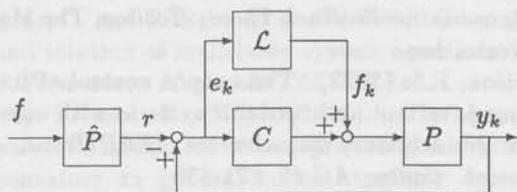


Fig. 10: Modified control structure with learning algorithm \mathcal{L}

It can be shown that the best choice for the filter Q would be $Q = I$. However, it is seen from equation (2) that, if $Q = I$, L should be equal to $(P^{-1} + C)$ for guaranteeing convergence, i.e. the inverse of the system P should be exactly known, which is impossible in most practical applications. Therefore, the filter Q is chosen as a low-pass filter with magnitude equal to 1 at low frequencies, and cut-off frequency near that frequency point where the knowledge of P does not allow the determination of an inverse anymore. In de Roover (1996), a systematic analysis is proposed for design of the filters Q and L . Using the model and the controller shown in Figure 3 and 4, respectively, filters Q and L have been designed for the x -direction of the $xy\phi$ -stage, shown in Figure 11. With these filters, a learning iteration was performed for all experiments shown in the previous section. For each command signal, 10 iterations were performed, and each experiment was repeated 3 times, so as to average out the effect of random noise. The results are shown in Figures 12, 13 and 14 for the bang-bang, the limited slope, and the ramped sinusoid command signal, respectively.

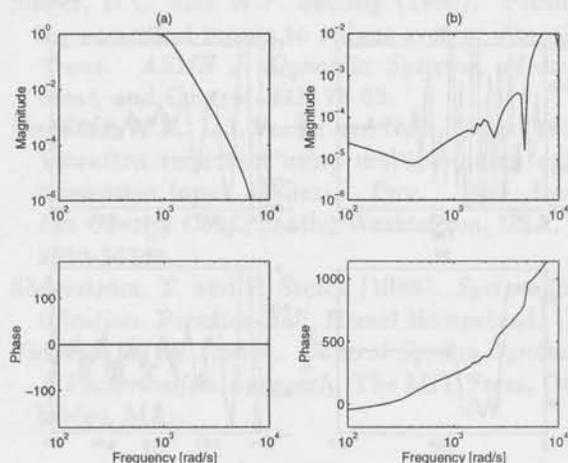


Fig. 11: Bode plot of (a): filter Q and (b): filter L . The cut-off frequency of Q is approximately 1000 rad/s, and anti-causal filtering with Q was applied with net-phase zero.

The results are quite impressive! It is seen that for each command signal, the tracking error has been tremendously reduced. Two important comments have to be made with respect to these results. First, when comparing Figures 13 and 14, the best results are obtained for learning control applied to the standard reference and command signals, see Figure 13; the tracking error is reduced within a band of 150 nm, and the cycle time has been reduced to $6.23e^{-2}$ s. The cycle time for the bang-bang and ramped sinusoid command signal, are $7.45e^{-2}$ s and $6.50e^{-2}$ s, respectively. The reason for this, is the fact that the filter Q has a cut-off frequency at 1000 rad/s, hence removing energy from the command signals above 1000 rad/s. However, the nominal ramped sinusoid reference and command signal were designed to have energy content above 1000 rad/s, see Figure 8 (b). Hence, the ramped sinusoid reference signal is too fast for the learned force profile, resulting in the showing up of residual vibration in the tracking error, see Figure 14 (a),(b).

Second, although the residual vibration in the tracking error has been removed significantly, the learned force profiles show unwanted residual behaviour. It can be shown that this residual behaviour is caused by badly damped *zeros* of the system. Especially if the reference signal is too fast for the learned command signal, like the bang-bang and the ramped sinusoid reference signal, the residual vibration in the learned force profile is simply a result of excitation of the *inverse* system dynamics with a reference signal that has too much energy content at the locations of the system zeros, see Figures 12,14 (c),(d).

Regarding these comments, it is worthwhile to put more effort in the design of learning filters Q and L , in order to increase the cut-off frequency of Q ,

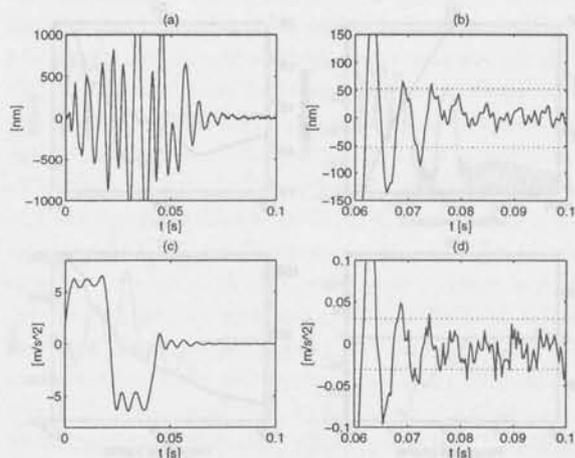


Fig. 12: Learning iteration for bang-bang command and reference signal; (a) tracking error after 10 iterations, (b) close up of (a) with performance bounds, (c), force profile after 10 iterations, (d) close up of (c) with bounds on regulation level.

without destroying the convergence of the learning iteration, so that fast nominal reference and command signals can be correctly learned.

6 Conclusions

In this paper, a high-performance motion control system for a flexible mechanical servosystem has been designed and implemented, and compared to a standard motion control system. It was shown that accurate modelling of the multivariable system, enables the designer to reduce both dynamic interaction, and the effect of flexible components, which limit the performance of the control system. The high-performance control system showed improved performance with respect to a standard control system, in a sense that motions were performed more fast and accurate. Exploiting the repetitive nature of the motion, robustness of the scheme against modelling errors, could be improved, resulting in even more fast and accurate movements, which are close to the maximum obtainable performance for the system at hand.

References

- Bartelings, H.J., P. Dirksen, G.A.J. de Fockert, H.W.A. Janssen, F.B. Sperling and J. v.d Werf (1996). *Sire-4 Scanning lithography test rig*, Nat.Lab. Technical Note, 212/96.
- Bhat, S.P. and D.K. Miu (1990). Precise point-to-point positioning control of flexible structures. *Trans. ASME J. Dynamic Systems, Measurement, and Control*, 112, 667-674.

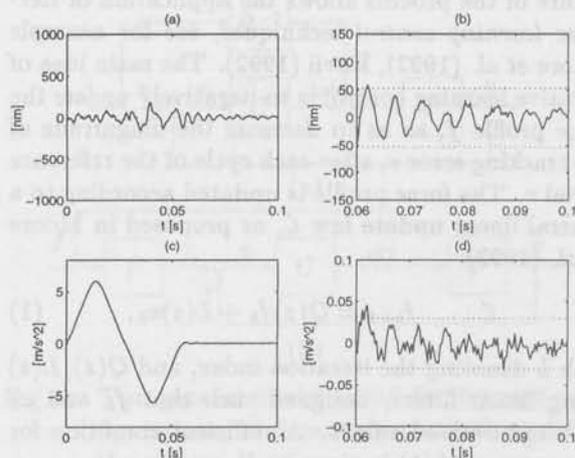


Fig. 13: Learning iteration for limited slope command and reference signal; (a) tracking error after 10 iterations, (b) close up of (a) with performance bounds, (c), force profile after 10 iterations, (d) close up of (c) with bounds on regulation level.

- Borghesani, C., Y. Chait and O. Yaniv (1995). *Quantitative Feedback Theory Toolbox*, The MathWorks, Inc.
- Davison, E.J. (1972). The output control of linear time-invariant multivariable systems with unmeasurable arbitrary disturbances. *IEEE Trans. Automat. Contr.*, AC-17, 621-630.
- Davison, E.J. (1975). Robust control of a general servomechanism problem: the servo compensator. *Automatica*, 11, 461-471.
- de Callafon, R.A., D. de Roover and P.M.J. Van den Hof (1996). Multivariable least squares frequency domain identification using polynomial matrix fraction descriptions. *Proc. 35th IEEE Conf. Decis. Control*, Kobe, Japan. (See also this issue pp. 31-37).
- de Roover, D. and R.N. van Marrewijk (1995). Systematic analysis of the dynamic behaviour of an electro-mechanical servosystem. In: O.H. Bosgra et al. (Eds.), *Sel. Topics Identif., Modelling and Control*. Delft Univ. Press, Vol. 8, pp. 75-84.
- de Roover, D. (1996). Synthesis of a robust iterative learning controller using an H_∞ approach. *Proc. 35th IEEE Conf. on Decision and Control*, Kobe, Japan. (See also this issue).
- de Roover, D. and O.H. Bosgra (1996). *Dualization of the Internal Model Principle in Compensator and Observer Theory*. Internal Report N-492, Mechan. Engin. Systems and Control Group, Delft Univ. Technology.
- Francis, B.A. and W.M. Wonham (1976). The internal model principle of control theory. *Automatica*, 12, 457-465.

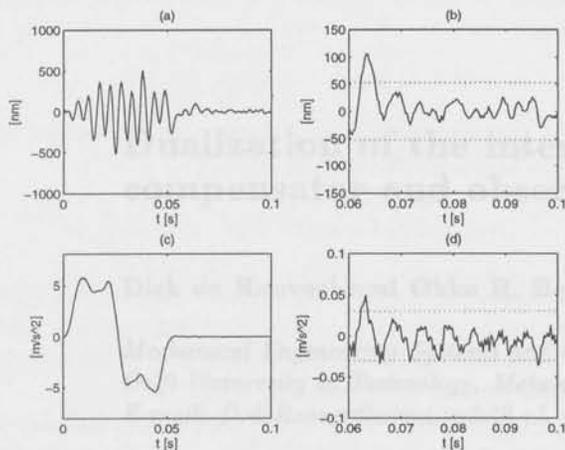


Fig. 14: Learning iteration for ramped sinusoid command and reference signal; (a) tracking error after 10 iterations, (b) close up of (a) with performance bounds, (c), force profile after 10 iterations, (d) close up of (c) with bounds on regulation level.

Führer, C. and R. Schwertassek (1990). Generation and solution of multibody system equations. *Int. J. Non-linear Mechanics*, 25, 127-141.

Hara, S. and T. Sugie (1988). Independent parametrization of two-degree-of-freedom compensators in general robust tracking systems. *IEEE Trans. Automat. Contr.*, AC-33, 59-67.

Horowitz, I.M. (1963). *Synthesis of Feedback Systems*. Academic Press, New York.

Johnson, C.D. (1976). Theory of disturbance-accommodating controllers. In: C.T. Leondes (Ed.), *Control and Dynamic Systems*, Ac. Press, New York, Vol. 12, pp. 387-489.

Kane, T.R. and D.A. Levinson (1985). *Dynamics: Theory and Applications*. McGraw-Hill, New York.

Kavli, T. (1992). Frequency domain synthesis of trajectory learning controllers for robot manipulators. *J. Robotic Systems*, 9, 663-680.

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

Meckl, P.H. and R. Kinceler (1994). Robust motion control of flexible systems using feedforward forcing functions. *IEEE Trans. Control Systems Techn.*, CST-2, 245-254.

Meckl, P.H. and W.P. Seering (1985). Minimizing residual vibration for point-to-point motion. *Trans. ASME J. Vibration, Acoustics, Stress, and Reliability in Design*, 107, 378-382.

Moore, K.L., M. Dahleh and S.P. Bhattacharyya (1992). Iterative learning control: a survey and new results. *J. Robotic Systems*, 9, 563-594.

Singer, N.C. and W.P. Seering (1990). Preshaping command inputs to reduce system vibration. *Trans. ASME J. Dynamic Systems, Measurement, and Control*, 112, 76-82.

Singhose, W.E., L.J. Porter and N.C. Singer (1995). Vibration reduction using multiple-hump extra-insensitive input shapers. *Proc. 1995 American Control Conf.*, Seattle, Washington, USA, pp. 3830-3834.

Söderström, T. and P. Stoica (1989). *System Identification*. Prentice-Hall, Hemel Hempstead, UK.

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

Stinger, M.C. and W.F. Steing (1990) *Prosp...*
 ing computer models to reduce system vibration
 Trans. ASME & Dynamic Systems Division
 and Control 113 74-81

Stinger, M.C., J.J. Evers and M.C. Steing (1992)
 Vibration reduction using nonlinear control
 Innovative Technology in Design, Proc. 1992 Amer.
 Institute of Aeronautics and Astronautics, Inc., 1992
 San Diego, Calif., USA, pp. 1530-1534

Stinger, M.C. and P. Stiles (1989) *Structural...*
 Efficient Finite Element Method Simulation
 Vibration, M. (1982) Control System Simulation
 Applications Symposium, The MIT Press, Cam-
 bridge, MA

Fig. 11. Comparison of control strategies for
 the system. (a) Control strategy (b) Control strategy
 (c) Control strategy (d) Control strategy
 (e) Control strategy (f) Control strategy

without developing any special control
 strategy, so that the control system can be
 used for any system as long as the control
 strategy is known.

6. Conclusions

In this paper, a high-order nonlinear control
 system has been designed to reduce the
 vibration of a flexible structure. The control
 system is designed to reduce the vibration
 of the structure by using a nonlinear control
 strategy. The control system is designed to
 reduce the vibration of the structure by
 using a nonlinear control strategy. The
 control system is designed to reduce the
 vibration of the structure by using a
 nonlinear control strategy. The control
 system is designed to reduce the vibration
 of the structure by using a nonlinear
 control strategy. The control system is
 designed to reduce the vibration of the
 structure by using a nonlinear control
 strategy. The control system is designed
 to reduce the vibration of the structure
 by using a nonlinear control strategy.

References

Stinger, M.C., J.J. Evers and M.C. Steing
 (1992) *Innovative Technology in Design*,
 Proc. 1992 Amer. Institute of Aeronautics
 and Astronautics, Inc., 1992, San Diego,
 Calif., USA, pp. 1530-1534

Stinger, M.C. and W.F. Steing (1990)
Prosp... ing computer models to reduce
 system vibration Trans. ASME & Dynamic
 Systems Division and Control 113 74-81

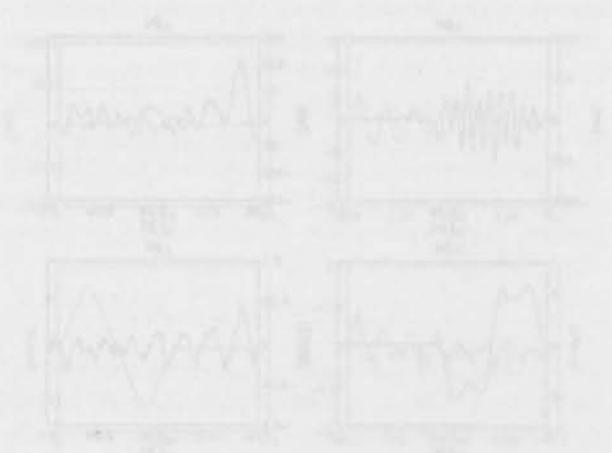


Fig. 12. Comparison of control strategies for
 the system. (a) Control strategy (b) Control strategy
 (c) Control strategy (d) Control strategy

The control system is designed to reduce
 the vibration of the structure by using a
 nonlinear control strategy. The control
 system is designed to reduce the vibration
 of the structure by using a nonlinear
 control strategy. The control system is
 designed to reduce the vibration of the
 structure by using a nonlinear control
 strategy. The control system is designed
 to reduce the vibration of the structure
 by using a nonlinear control strategy. The
 control system is designed to reduce the
 vibration of the structure by using a
 nonlinear control strategy. The control
 system is designed to reduce the vibration
 of the structure by using a nonlinear
 control strategy. The control system is
 designed to reduce the vibration of the
 structure by using a nonlinear control
 strategy. The control system is designed
 to reduce the vibration of the structure
 by using a nonlinear control strategy.

Dualization of the internal model principle in compensator and observer theory

Dick de Roover[†] and Okko H. Bosgra

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekekeweg 2, 2628 CD Delft, The Netherlands.
E-mail: D.deRoover@wbmt.tudelft.nl, O.H.Bosgra@wbmt.tudelft.nl*

Abstract. This paper reconsiders the concepts of *servo compensator* and *disturbance observer*. Both concepts make use of an internal model representing a general persistent disturbance. The use of such a model, also known as the *Internal Model Principle (IMP)*, has been well recognized to be necessary within the servo compensator concept, leading to general necessary and sufficient conditions for a servo compensator to asymptotically compensate the disturbance. However, for the disturbance observer concept, no general existence conditions are present in the literature. In this paper these conditions will be derived by reformulating the servo compensator and disturbance observer problem in a general standard plant. Moreover, it is shown that a robust servo compensator, stabilized with state feedback, is dual to a disturbance observer stabilized with output injection. Equally important, it is shown that a robust servo compensator, stabilized by an observer based controller using output feedback, is dual to a disturbance observer in combination with state feedback and disturbance compensation.

Keywords. asymptotic disturbance rejection, Internal Model Principle (IMP), robust servo compensator, robust disturbance observer, dualization

1 Introduction

One of the main reasons for adding feedback compensation to a system is the rejection of undesired signals that disturb the performance of that system. In practical situations these *disturbance signals* range in nature from purely stochastic to purely deterministic signals. In this paper we restrict our attention to purely deterministic signals.

The compensation of deterministic signals has been explored and developed by several researchers in the late sixties and early seventies, see Johnson (1971), Davison (1972a, 1975), Francis and Wonham (1976). A breakthrough was the notion of the *Internal Model Principle (IMP)*, which states that, in order to compensate a general persistent disturbance, a feedback compensator has to include an in-

ternal model of the dynamics of that disturbance¹, see Francis and Wonham (1976). The IMP led to the development of the rather celebrated *robust servo compensator*, see Davison (1975), which asymptotically rejects a general persistent disturbance in the face of perturbations in the parameters of the uncompensated system.

Simultaneous to the development of the servo compensator, a Luenberger type of observer was developed for systems subject to the same general class of persistent disturbances, by including an internal model of the disturbance in the observer, see Johnson (1971, 1976), Meditch and Hostetter (1974). In the most general case, this *disturbance observer* estimates both the system state and the disturbance state. If the disturbance is assumed to act on the input of the system, the estimated disturbance can

[†]The work of Dick de Roover is financially supported by Philips' Research Laboratories, Eindhoven, The Netherlands.

¹A persistent disturbance is a disturbance generated by an autonomous dynamical system having poles in the closed right complex half plane.

be used to compensate the real disturbance, see for example Johnson (1971, 1976) Profeta *et al.* (1990), Nagasawa and Yokamada (1993).

Since both the servo compensator and the disturbance observer have been used for compensating the same class of persistent disturbances, several researchers have tried to compare both concepts as if they were equal to each other, see Johnson (1972), Davison (1972b), Kwatny and Kalnitsky (1978), Desoer and Wang (1980). This led to a misunderstanding of, in particular, the disturbance observer concept. Although it was recognized that both concepts are not equal but dual, neither general conditions for a disturbance observer to exist, nor general conditions for such an observer to compensate a persistent disturbance are available. In this paper general existence conditions will be derived that turn out to be dual to conditions obtained in servo compensator theory, by reformulating the servo compensator and disturbance observer problem in a general standard plant.

The next section starts with an explanation of the notation used throughout this paper, together with some fundamental aspects from linear system theory. Then Section 3 describes the theory of the servo compensator within a general standard plant. In Section 4 this general theory is dualized to that of disturbance observer theory. Finally Section 5 ends up with some conclusions.

2 Notation and fundamentals

Let \mathbb{R} (\mathbb{C}) denote the field of real (complex) numbers. Let \mathbb{C}_- (\mathbb{C}_+) denote the open left (closed right) complex half-plane. Let $\mathbb{R}^{n \times m}$ (\mathbb{R}^n) be the set of all $n \times m$ matrices (n -vectors) with elements in \mathbb{R} . Let $\mathbb{R}(s)$ denote the set of all rational functions with real coefficients in s , with s denoting the Laplace operator. Let $\mathbb{R}(s)^{n \times m}$ be the set of all $n \times m$ matrices with elements in $\mathbb{R}(s)$. Let $A \in \mathbb{R}^{n \times n}$ then ϕ_A denotes the minimal polynomial of A and $\sigma(A)$ denotes the spectrum of A . Let $M \in \mathbb{R}^{n \times m}$ then $\rho(M)$ denotes the rank of M , and $\rho(M) \leq \min\{n, m\}$. The normal rank ρ_n of a matrix $M(s) \in \mathbb{R}(s)^{n \times m}$ is defined as:

$$\rho_n(M(s)) := \max_{s \in \mathbb{C}} \rho(M(s)).$$

Consider a linear time invariant (LTI) system \mathcal{G} represented by:

$$\mathcal{G}: \begin{cases} \dot{x}(t) = Ax(t) + Bu(t), & x(t_0) = x_0, \\ y(t) = Cx(t) + Du(t), \end{cases} \quad (1)$$

with x, x_0, u and y denoting the state, initial state, input and output of the system, respectively. The

system (1) can be written in a more compact matrix form:

$$\mathcal{G}: \begin{bmatrix} \dot{x} \\ y \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}.$$

The transfer matrix of \mathcal{G} from u to y is given by (zero initial conditions):

$$G(s) = C(sI - A)^{-1} + D =: \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

The system \mathcal{G} is said to be *exponentially stable*, if for any initial state x_0 the zero-input response will tend to zero exponentially. The system \mathcal{G} is exponentially stable for any x_0 if and only if $\sigma(A) \in \mathbb{C}_-$.

The system \mathcal{G} is said to be (state) *controllable* if there exists an input $u(t)$ that will transfer the state x_0 to any state x_f in any finite time interval $[t_0, t_f]$. Otherwise \mathcal{G} is said to be *uncontrollable*. The system \mathcal{G} is said to be (state) *observable* if for any finite time interval $[t_0, t_f]$, the state x_0 (at time t_0) can be determined with knowledge of the input $u(t)$ and the output $y(t)$ over the interval $[t_0, t_f]$. Otherwise \mathcal{G} is said to be *unobservable*. Let $A \in \mathbb{R}^{n_s \times n_s}$, $B \in \mathbb{R}^{n_s \times n_u}$, $C \in \mathbb{R}^{n_y \times n_s}$, then the pair $\{A, B\}$ ($\{A, C\}$) is controllable (observable) if and only if

$$\rho([\lambda I - A \ B]) = n_s, \quad \left(\rho \begin{bmatrix} \lambda I - A \\ C \end{bmatrix} \right) = n_s \quad \forall \lambda \in \mathbb{C}.$$

The system \mathcal{G} is said to be *stabilizable* if there exists a state feedback $u = Kx$ such that \mathcal{G} is stable, i.e. $A - BK \in \mathbb{C}_-$. The system \mathcal{G} is said to be *detectable* if there exists a constant matrix L such that $A - LC \in \mathbb{C}_-$.

Let $G(s) \in \mathbb{R}(s)^{n_y \times n_u}$ be proper and have normal rank n , and let $\{A, B, C, D\}$ be a *minimal* state space realization of $G(s)$, then a real or complex number λ is called a *transmission zero* of $G(s)$ if²:

$$\rho \left(\begin{bmatrix} \lambda I - A & B \\ -C & D \end{bmatrix} \right) < \rho_n \left(\begin{bmatrix} sI - A & B \\ -C & D \end{bmatrix} \right) = n_s + n,$$

i.e. if the Rosenbrock system matrix Rosenbrock (1970) w.r.t. u and y loses rank.

Let $D^{-1}(s)N(s)$ ($\bar{N}(s)\bar{D}^{-1}$) be a left (right) coprime factorization of $G(s)$. Let $G(s)$ and, consequently, $N(s)$ ($\bar{N}(s)$) have rank n , then λ is a transmission zero of $G(s)$ if and only if $\rho N(\lambda) < n$ ($\rho \bar{N}(\lambda) < n$).

3 Theory of the servo compensator

3.1 The robust servomechanism problem (RSP)

Consider the feedback configuration of Figure 1. The continuous-time LTI generalized system \mathcal{G} is

²This definition is due to Davison and Wang (1974)

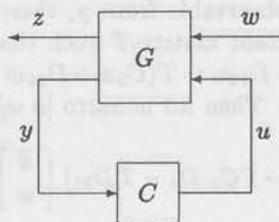


Fig. 1: Standard feedback configuration.

given by:

$$G: \begin{bmatrix} \dot{x} \\ z \\ y \end{bmatrix} = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_1 & D_{12} \\ C_2 & D_{21} & D_2 \end{bmatrix} \begin{bmatrix} x \\ w \\ u \end{bmatrix}, \quad (2)$$

$$x(t_0) = x_0,$$

where $x(t) \in \mathbb{R}^{n_x}$ describes the state of the system, $w(t) \in \mathbb{R}^{n_w}$ represents an exogenous system input, $u(t) \in \mathbb{R}^{n_u}$ is the control input to the system, $z(t) \in \mathbb{R}^{n_z}$ is the output to be regulated, e.g. the difference between a measured and a desired output, and $y(t) \in \mathbb{R}^{n_y}$ is the measured output of the system. The matrices $\{A, B_1, B_2, C_1, C_2, D_1, D_{12}, D_{21}, D_2\}$ are assumed to have appropriate dimensions, i.e. $A \in \mathbb{R}^{n_x \times n_x}$, $B_1 \in \mathbb{R}^{n_x \times n_w}$, $B_2 \in \mathbb{R}^{n_x \times n_u}$, $C_1 \in \mathbb{R}^{n_z \times n_x}$, $C_2 \in \mathbb{R}^{n_y \times n_x}$, $D_1 \in \mathbb{R}^{n_z \times n_w}$, $D_{12} \in \mathbb{R}^{n_z \times n_u}$, $D_{21} \in \mathbb{R}^{n_y \times n_w}$, and $D_2 \in \mathbb{R}^{n_y \times n_u}$. The corresponding transfer matrix of G is given by:

$$G(s) = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_1 & D_{12} \\ C_2 & D_{21} & D_2 \end{bmatrix}.$$

The measured system output y is fed to a dynamic compensator C with transfer matrix:

$$C(s) = \begin{bmatrix} A_c & B_c \\ C_c & D_c \end{bmatrix},$$

and the output of this compensator is fed back to the control input u of G .

In this paper we consider the situation where $w(t)$ is assumed to be a *persistent signal*, generated by an autonomous system³:

$$\mathcal{W}: \begin{bmatrix} \dot{x}_w \\ w \end{bmatrix} = \begin{bmatrix} A_w \\ C_w \end{bmatrix} x_w, \quad x_w(t_0) = x_{w0}, \quad (3)$$

where $x_w(t) \in \mathbb{R}^{n_w}$, $A_w \in \mathbb{R}^{n_w \times n_w}$, and $C_w \in \mathbb{R}^{n_w \times n_w}$. Without loss of generality it is assumed

³Note that many common signals can be described in this setup, like steps, ramps, sinusoids, etc.

that the pair $\{A_w, C_w\}$ is observable, and $\sigma(A_w) \subset \mathbb{C}_+$.

Furthermore, we want to construct the compensator C such that the closed loop is stable, and the regulated system output z vanishes to zero asymptotically, in the presence of w . Besides, this rejection property should be *robust*, i.e. should be maintained even in case the dynamics of G vary.

More formal, we want to solve the robust servomechanism problem, which is defined as:

Definition 3.1 *The robust servomechanism problem (RSP) is to find a feedback compensator C for the system G such that:*

1. *The resulting compensated system is exponentially stable.*
2. *The system output $z(t)$ tends to zero asymptotically, for all $x_0 \in \mathbb{R}^{n_x}$ and for all exogenous signals $w(t)$ satisfying (3).*
3. *Properties 1. and 2. are robust, i.e. they also hold in case the dynamics of the system G are perturbed.*

It was Davison (1972a, 1975) who first solved this problem, and at present a general solution can be found in many textbooks. In this paper the solution to this problem is presented within the general setup of Figure 1.

3.2 The internal model principle

Before presenting a general solution to the RSP, some comments have to be made on the compensator structure which, in fact, is not completely free to choose. In Francis and Wonham (1975) it was shown that for robust regulation of z in the presence of w generated by (3), the compensator necessarily has to incorporate a model of the dynamic system \mathcal{W} . This is known as the *Internal Model Principle* (IMP), see Francis and Wonham (1975, 1976), Conant and Ashby (1970). More formal, the matrix A_c of the controller incorporates an internal model of the dynamic system \mathcal{W} , if the minimal polynomial of A_w divides at least n_z invariant factors of A_c , with n_z the number of independent outputs z_i to be regulated, (Francis and Wonham, 1976).

In Wolovich and Ferreira (1979) it is shown that such an internal model must make itself present at the junction where w enters the closed loop from w to z . Therefore, since the internal model is inside the compensator C , the structure of C must be such that it commutes with the open-loop transfer matrix from w to z . Besides, the internal model must be observable from the control input u and has to be

controllable from the regulated output z , (Francis and Wonham, 1976).

A feedback compensator which obeys this specific structure is given by:

$$C_{sc}: \begin{bmatrix} \dot{x}_{sc} \\ u \end{bmatrix} = \begin{bmatrix} A_{sc} & B_{sc} \\ K_{sc} & 0 \end{bmatrix} \begin{bmatrix} x_{sc} \\ z \end{bmatrix}, \quad (4)$$

$$x_{sc}(t_0) = x_{sc0}$$

where⁴

$$A_{sc} = \text{block diag} \underbrace{[\Gamma, \Gamma, \dots, \Gamma]}_{n_x\text{-tuple}} \in \mathbb{R}^{n_{sc} \times n_{sc}}$$

$$B_{sc} = \text{block diag} \underbrace{[\gamma, \gamma, \dots, \gamma]}_{n_x\text{-tuple}} \in \mathbb{R}^{n_{sc} \times n_x}$$

$$K_{sc} = \text{a stabilizing gain matrix} \in \mathbb{R}^{n_u \times n_{sc}},$$

with $\{\Gamma, \gamma\}$ any controllable pair, such that $\sigma(\Gamma)$ equals the roots of $\phi_w(s)$, with $\phi_w(s)$ being the minimal polynomial of A_w . For example, the pair $\{\Gamma, \gamma\}$ can be chosen as a controllable canonical pair:

$$\Gamma = \begin{bmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ -\alpha_q & -\alpha_{q-1} & \dots & -\alpha_1 \end{bmatrix}, \quad \gamma = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

with $\phi_w(s) = s^q + \alpha_1 s^{q-1} + \dots + \alpha_{q-1} s + \alpha_q$. Since the above compensator is fed by the regulated output z , while the measured output y is available, an extra restriction is posed on the problem: it is necessary that z has to be observable from y , i.e. there exists a nonsingular constant matrix T such that $z = Ty$, (Francis and Wonham, 1975, Davison, 1976). If this is the case, z can be reconstructed from y . The existence of such a matrix T is guaranteed by:

Lemma 3.2 Consider the system (2). The regulated output z is observable from the measured output y , if and only if:

$$\rho \left(\begin{bmatrix} C_1 & D_1 \\ C_2 & D_{21} \end{bmatrix} \right) = \rho \left(\begin{bmatrix} C_2 & D_{21} \end{bmatrix} \right). \quad (5)$$

Proof: (\Leftarrow) If condition (5) holds, there exists a nonsingular matrix T^* such that for nonzero $[x \ w]^T$:

$$[T^*[C_1 \ D_1] - [C_2 \ D_{21}]] \begin{bmatrix} x \\ w \end{bmatrix} = 0.$$

Thus $T^*(C_1 x + D_1 w) = C_2 x + D_{21} w$, and hence z can be reconstructed from: $z = T^{*-1}(y - D_2 u) + D_{12} u$.

⁴Note that by construction the pair $\{A_{sc}, B_{sc}\}$ is controllable.

(\Rightarrow) If z is observable from y , there exists a nonsingular constant matrix T such that $z = Ty$, i.e. $C_1 x + D_1 w + D_{12} u = T(C_2 x + D_{21} w + D_2 u)$. W.l.g. choose $u = 0$. Then for nonzero $[x \ w]^T$ we have:

$$[C_1 - TC_2 \ D_1 - TD_{21}] \begin{bmatrix} x \\ w \end{bmatrix} = 0,$$

i.e. $C_1 = TC_2$ and $D_1 = TD_{21}$. Consequently

$$\rho \left(\begin{bmatrix} C_1 & D_1 \\ C_2 & D_{21} \end{bmatrix} \right) = \rho \left(\begin{bmatrix} TC_2 & TD_{21} \\ C_2 & D_{21} \end{bmatrix} \right) = \rho \left(\begin{bmatrix} C_2 & D_{21} \end{bmatrix} \right).$$

□

3.3 Solution of the RSP using state feedback

If the compensator (4) is put inside the feedback loop of Figure 1, it is likely that the closed loop is unstable, that is, we need some additional feedback to stabilize the system \mathcal{G} extended with the unstable dynamic system C_{sc} given by (4). In this subsection we assume that the state x of \mathcal{G} is available for feedback. In this case, the problem is reduced to finding a constant compensator:

$$\bar{C} = [K \ K_{sc}], \quad (6)$$

which stabilizes the extended system \mathcal{G}_e given by the series connection of \mathcal{G} and C_{sc} :

$$\mathcal{G}_e: \begin{bmatrix} \dot{x} \\ \dot{x}_{sc} \\ z \\ x \\ x_{sc} \end{bmatrix} = \bar{G}_e \begin{bmatrix} x \\ x_{sc} \\ w \\ u \end{bmatrix}, \quad (7)$$

with

$$\bar{G}_e = \begin{bmatrix} \begin{bmatrix} A & 0 \\ B_{sc}C_2 & A_{sc} \end{bmatrix} & \begin{bmatrix} B_1 \\ B_{sc}D_{21} \end{bmatrix} & \begin{bmatrix} B_2 \\ B_{sc}D_2 \end{bmatrix} \\ \begin{bmatrix} C_1 & 0 \\ I_{n_x} & 0 \end{bmatrix} & D_1 & D_{12} \\ \begin{bmatrix} 0 & I_{n_{sc}} \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{bmatrix}.$$

Figure 2 shows the resulting closed loop. The existence of a constant \bar{C} which solves the RSP is given by:

Theorem 3.3 Given the system (2) and suppose the signal $w(t)$ is generated according to (3). The RSP, defined by Definition 3.1, is solvable using state feedback, if and only if:

i. C includes an internal model of w ,

ii. $\rho \left(\begin{bmatrix} C_1 & D_1 \\ C_2 & D_{21} \end{bmatrix} \right) = \rho \left(\begin{bmatrix} C_2 & D_{21} \end{bmatrix} \right),$

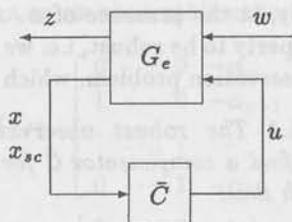


Fig. 2: Modified feedback configuration which solves the RSP.

iii. The pair $\{A, B_2\}$ is stabilizable,

iv.

$$\rho \left(\begin{bmatrix} \lambda I - A & B_2 \\ -C_1 & D_{12} \end{bmatrix} \right) = n_x + n_z, \quad \forall \lambda \in \sigma(A_w). \quad (8)$$

Proof: See Appendix. \square

Two remarks regarding this theorem are of importance:

Remark 3.4 The rank condition (8) implies that $n_u \geq n_z$, i.e. \mathcal{G} has at least as many control inputs as there are outputs to be regulated, and the subsystem $\{A, B_2, C_1, D_{12}\}$ has no transmission zeros located at the spectrum of A_w .

Remark 3.5 The closed loop system of Figure 2 is robust to parameter variations in the system parameters $\{A, B_1, B_2, C_1, C_2, D_1, D_{12}, D_{21}, D_2\}$, which do not destabilize the closed loop. However, neither variations in the parameters of $\{A_w, A_{sc}\}$, nor in the structure of $\{A_{sc}, B_{sc}\}$ are tolerated, as they might destroy the property of asymptotic disturbance rejection, or might even result in an unstable closed loop system. Besides, z must remain observable from y .

The resulting compensator for the original system \mathcal{G} , given by:

$$C(s) = \left[\begin{array}{c|c} A_{sc} & \begin{bmatrix} B_{sc}^T & 0 \end{bmatrix} \\ \hline K_{sc} & \begin{bmatrix} 0 & K \end{bmatrix} \end{array} \right], \quad (9)$$

is often referred to as *robust servo compensator*, (Davison, 1975). Robust asymptotic disturbance rejection of w is achieved by duplicating the dynamics of w inside the feedback loop, i.e. the servo compensator constitutes an *internal model* of the unstable dynamics of w ; in the appendix it is shown that this duplication produces transmission zeros in the closed loop from w to z , at the spectrum of A_w , which completely block the transmission from $w(t)$ to $z(t)$.

3.4 Solution of the RSP using output feedback

If the state x of \mathcal{G} is not available, the output y can be used for stabilization. The existence of a compensator which solves the RSP using output feedback, is given by:

Corollary 3.6 Given the system (2) and suppose the signal $w(t)$ is generated according to (3). The RSP, defined by Definition 3.1, is solvable using output feedback, if and only if:

i. Conditions i. – iv. of Theorem 3.3 hold,

ii. The pair $\{A, C_2\}$ is detectable.

Proof: Condition ii. guarantees the detectability of the series connection of \mathcal{G} and C_{sc} ; therefore the resulting closed loop system can be made exponentially stable. The asymptotic rejection and robustness property are guaranteed by the separation principle, applied to the servo compensator and any stabilizing compensator. \square

If Corollary 3.6 is true, a control input u can be designed which solves the RSP, according to:

$$u(t) = K_{sc}x_{sc}(t) + Kx_c(t),$$

with $x_c \in \mathbb{R}^{n_c}$ the state-vector of any stabilizing compensator. For example, an observer based compensator can be designed, which stabilizes

$$\begin{bmatrix} \dot{x} \\ y \end{bmatrix} = \begin{bmatrix} A & B_2 \\ C_2 & D_2 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}.$$

In this case, the resulting compensator, which solves the RSP, is given by:

$$C(s) = \left[\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array} \right],$$

with

$$\begin{aligned} A_c &= \begin{bmatrix} A_{sc} & 0 \\ (\tilde{B}_2 + L\tilde{D}_2)K_{sc} & \tilde{A} + \tilde{B}_2K + L(\tilde{C}_2 - \tilde{D}_2K) \end{bmatrix} \\ B_c &= \begin{bmatrix} B_{sc}^T \\ K \end{bmatrix}, \quad C_c = [K_{sc} \quad -L], \quad D_c = 0, \end{aligned} \quad (10)$$

and L denoting an observer gain, designed such that $\tilde{A} + L\tilde{C}_2 \in \mathbb{C}_-$, and $\{\tilde{A}, \tilde{B}_2, \tilde{C}_2, \tilde{D}_2\}$ is a duplication of the system $\{A, B_2, C_2, D_2\}$. According to the separation principle, L and $[K \quad K_{sc}]$ can be designed independently.

4 Dualization of the IMP in the theory of disturbance observers

4.1 The robust observation problem

Again consider the system (2):

$$\mathcal{G}: \begin{bmatrix} \dot{x} \\ z \\ y \end{bmatrix} = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_1 & D_{12} \\ C_2 & D_{21} & D_2 \end{bmatrix} \begin{bmatrix} x \\ w \\ u \end{bmatrix},$$

where the signals x, w, u, z, y and the tuple $\{A, B_1, B_2, C_1, C_2, D_1, D_{12}, D_{21}, D_2\}$ have appropriate dimensions, and again assume $w(t)$ to be generated according to (3):

$$\mathcal{W}: \begin{bmatrix} \dot{x}_w \\ w \end{bmatrix} = \begin{bmatrix} A_w \\ C_w \end{bmatrix} x_w, \quad x_w(t_0) = x_{w0},$$

where x_w, w, A_w and C_w also have the appropriate dimensions. Again it is assumed that $\sigma(A_w) \in \mathbb{C}_+$ and the pair $\{A_w, C_w\}$ is observable.

In this section we are considered with finding an observer for the system \mathcal{G} , in the presence of w generated by (3). Consider the following general observer system:

$$\tilde{\mathcal{G}}: \begin{bmatrix} \dot{\tilde{x}} \\ \tilde{z} \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} \tilde{A} & \tilde{B}_1 & \tilde{B}_2 \\ \tilde{C}_1 & \tilde{D}_1 & \tilde{D}_{12} \\ \tilde{C}_2 & \tilde{D}_{21} & \tilde{D}_2 \end{bmatrix} \begin{bmatrix} \tilde{x} \\ w \\ \tilde{u} \end{bmatrix}, \quad (11)$$

$$\tilde{x}(t_0) = \tilde{x}_0,$$

where $\tilde{x} \in \mathbb{R}^{n_s}$ denotes the difference between the state and the observed state, $w \in \mathbb{R}^{n_w}$ represents the exogenous system input, $\tilde{u} \in \mathbb{R}^{n_u}$ denotes the observer input, $\tilde{z} \in \mathbb{R}^{n_z}$ is the difference between the measured system output y and a reconstructed system output, i.e. the observer output to be regulated, and $\tilde{y} \in \mathbb{R}^{n_y}$ is the output of the observer system, available for compensation of the observer dynamics. The matrices $\{\tilde{A}, \tilde{B}_1, \tilde{B}_2, \tilde{C}_1, \tilde{C}_2, \tilde{D}_1, \tilde{D}_{12}, \tilde{D}_{21}, \tilde{D}_2\}$ are assumed to have appropriate dimensions.

Next, according to Figure 3, we want to construct a compensator \tilde{C} , which stabilizes the observer sys-

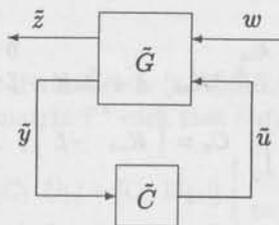


Fig. 3: Standard observer configuration

tem $\tilde{\mathcal{G}}$, and brings the reconstruction error \tilde{z} to zero

asymptotically, in the presence of w . Again, we require this property to be robust, i.e. we want to solve the robust observation problem, which is defined as:

Definition 4.1 The robust observation problem (ROP) is to find a compensator \tilde{C} for the observer system $\tilde{\mathcal{G}}$ such that:

1. The resulting compensated observer system is exponentially stable.
2. The reconstruction error $\tilde{z}(t)$ tends to zero asymptotically, for all $\tilde{x}_0 \in \mathbb{R}^{n_s}$ and all $w(t)$ satisfying (3).
3. Properties 1. and 2. are robust, i.e. they also hold in case the dynamics of the observer system $\tilde{\mathcal{G}}$ are perturbed

In the literature, the ROP can be found under the name *disturbance observer*, see for example Johnson (1971, 1976), Meditch and Hostetter (1974), Levin and Kreindler (1976), Kwatny and Kalnitsky (1978), Sievers and von Flotow (1989), Profeta *et al.* (1990), Nagasawa and Yokamada (1993). However, non of these references have considered, and solved, the ROP in its full general form as we stated it in Definition 4.1.

4.2 Use of the IMP

It can be shown that, in order to solve the ROP, the compensator \tilde{C} necessarily has to include an internal model of (3), which is observable from w , and controllable from \tilde{y} . Besides, this internal model must make itself present at the junction where w enters the closed loop from w to \tilde{z} , in order to observe the actual disturbance w . A compensator with these commuting properties is given by:

$$\tilde{C}_{do}: \begin{bmatrix} \dot{x}_{do} \\ w \end{bmatrix} = \begin{bmatrix} A_{do} & L_{do} \\ C_{do} & 0 \end{bmatrix} \begin{bmatrix} x_{do} \\ \tilde{y} \end{bmatrix}, \quad (12)$$

$$x_{do}(t_0) = x_{do0}$$

where⁵

$$A_{do} = \text{block diag} \underbrace{[\Upsilon, \Upsilon, \dots, \Upsilon]}_{n_w \text{-tuple}} \in \mathbb{R}^{n_{do} \times n_{do}}$$

$$L_{do} = \text{a stabilizing gain matrix} \in \mathbb{R}^{n_{do} \times n_y}$$

$$C_{do} = \text{block diag} \underbrace{[v, v, \dots, v]}_{n_w \text{-tuple}} \in \mathbb{R}^{n_w \times n_{do}},$$

with $\{\Upsilon, v\}$ any observable pair, such that $\sigma(\Upsilon)$ equals the roots of $\phi_w(s)$, with $\phi_w(s)$ being the minimal polynomial of A_w . For example, the pair $\{\Upsilon, v\}$

⁵Note that by construction the pair $\{A_{do}, C_{do}\}$ is observable.

can be chosen as an observable canonical pair:

$$\Upsilon = \begin{bmatrix} 0 & \cdots & 0 & -\alpha_p \\ 1 & \cdots & 0 & -\alpha_{p-1} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & -\alpha_1 \end{bmatrix},$$

$$v = [0 \ \cdots \ 0 \ 1],$$

with $\phi_w(s) = s^q + \alpha_1 s^{p-1} + \cdots + \alpha_{p-1} s + \alpha_p$. Since the above compensator has as output w , while the best we can do is to make a reconstruction \tilde{u} of w , an extra restriction is posed to the problem: it is necessary that w is controllable from \tilde{u} , i.e. there exists a nonsingular matrix \tilde{T} such that $w = \tilde{T}\tilde{u}$. The existence of such a matrix is guaranteed by:

Lemma 4.2 Consider the observer system (11). The exogenous input w is controllable from the reconstructed input \tilde{u} , if and only if:

$$\rho \left(\begin{bmatrix} \tilde{B}_1 & \tilde{B}_2 \\ \tilde{D}_1 & \tilde{D}_{12} \end{bmatrix} \right) = \rho \left(\begin{bmatrix} \tilde{B}_2 \\ \tilde{D}_{12} \end{bmatrix} \right). \quad (13)$$

Proof: Follows by similar reasoning from the proof of Lemma 3.2 \square

4.3 Solution of the ROP using output injection

In this subsection we assume that the observer system (11) can be stabilized using output injection, i.e. we assume to have full access to the observer state with the measured output y . In this case, we can reduce the ROP to finding a constant compensator:

$$\tilde{C} = \begin{bmatrix} L \\ L_{do} \end{bmatrix}, \quad (14)$$

which stabilizes the extended system \tilde{G}_e given by the series connection of \tilde{G} and \tilde{C}_{do} :

$$\tilde{G}_e : \begin{bmatrix} \dot{\tilde{x}} \\ \dot{x}_{do} \\ \tilde{z} \\ \tilde{y} \end{bmatrix} = \tilde{G}_e \begin{bmatrix} \tilde{x} \\ x_{do} \\ w \\ v_1 \\ v_2 \end{bmatrix}, \quad (15)$$

with

$$\tilde{G}_e = \begin{bmatrix} \begin{bmatrix} \tilde{A} & \tilde{B}_2 C_{do} \\ 0 & A_{do} \end{bmatrix} & \begin{bmatrix} \tilde{B}_1 \\ 0 \end{bmatrix} & \begin{bmatrix} I_{n_x} & 0 \\ 0 & I_{n_{do}} \end{bmatrix} \\ \begin{bmatrix} \tilde{C}_1 & \tilde{D}_{12} C_{do} \\ \tilde{C}_2 & \tilde{D}_2 C_{do} \end{bmatrix} & \begin{bmatrix} \tilde{D}_1 \\ \tilde{D}_{21} \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix},$$

and v_1, v_2 being intermediate state variables. Figure 4 shows the resulting closed loop. The existence of a constant \tilde{C} which solves the ROP is given by:

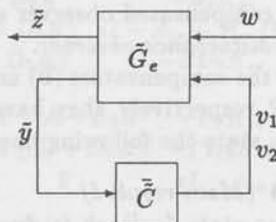


Fig. 4: Modified observer configuration which solves the ROP.

Theorem 4.3 Given the observer system (11), and suppose the signal $w(t)$ is generated according to (3). The ROP, defined by Definition 4.1, is solvable using output injection, if and only if:

- \tilde{C} includes an internal model of w ,
- $\rho \left(\begin{bmatrix} \tilde{B}_1 & \tilde{B}_2 \\ \tilde{D}_1 & \tilde{D}_{12} \end{bmatrix} \right) = \rho \left(\begin{bmatrix} \tilde{B}_2 \\ \tilde{D}_{12} \end{bmatrix} \right)$,
- The pair $\{\tilde{A}, \tilde{C}_2\}$ is detectable,
-

$$\rho \left(\begin{bmatrix} \lambda I - \tilde{A} & \tilde{B}_1 \\ -\tilde{C}_2 & \tilde{D}_{21} \end{bmatrix} \right) = n_x + n_w, \quad \forall \lambda \in \sigma(A_w). \quad (16)$$

Proof: See Appendix. \square

Remark 4.4 The rank condition (16) implies that $n_{\tilde{y}} \geq n_w$, i.e. the number of measured outputs should be greater than or at least equal to the number of exogenous inputs, which was already intuitively noticed in Meditch and Hostetter (1974), pp.478. Besides, condition (16) implies that the subsystem $\{\tilde{A}, \tilde{B}_1, \tilde{C}_2, \tilde{D}_{21}\}$ has no transmission zeros located at the spectrum of A_w .

Remark 4.5 The resulting compensated observer system is robust to parameter variations in the sense that the disturbance state is correctly reconstructed despite perturbations in the system parameters $\{\tilde{A}, \tilde{B}_1, \tilde{B}_2, \tilde{C}_1, \tilde{C}_2, \tilde{D}_1, \tilde{D}_{12}, \tilde{D}_{21}, \tilde{D}_2\}$, as long as the compensated observer system remains asymptotically stable. Again variations neither in the structure of the observer system, nor in the parameters $\{A_{do}, C_{do}\}$ are tolerated, and w has to remain controllable from \tilde{u} .

The resulting compensator for the observer system \tilde{G} is given by:

$$\tilde{C}(s) = \left[\begin{array}{c|c} A_{do} & L_{do} \\ \hline \begin{bmatrix} \tilde{T} C_{do} \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ L \end{bmatrix} \end{array} \right]. \quad (17)$$

The resulting compensated observer system will be referred to as *disturbance observer*.

If we compare the compensators (9) and (17) for the RSP and ROP, respectively, they have a dual form. In fact, we can state the following important result:

Theorem 4.6 (Main result I)

The RSP using state feedback is dual to the ROP using output injection.

Proof: This can be verified by solving an RSP using state feedback for the transposed system \tilde{G}^T , and transposing the resulting compensator. \square

Remark 4.7 Note that conditions i. – iv. of Theorem 3.3 are indeed dual to conditions i. – iv. of Theorem 4.3, and that (6), (7) are indeed dual to (14), (15), respectively.

4.4 Compensation using a disturbance observer

In most cases, the reconstructed state of \tilde{G} is used to compensate the system \mathcal{G} . Moreover, the reconstruction of w is always used to cancel the actual disturbance w , i.e. the observer system is fed by the system output y , and the control input u is chosen as:

$$u(t) = K\hat{x}(t) + C_{do}x_{do}(t),$$

with K denoting a state feedback gain, designed such that $A + B_2K \in \mathbb{C}_-$, and \hat{x} denoting the reconstructed system state. In this case, the resulting compensator for \mathcal{G} has the following transfer matrix:

$$C(s) = \begin{bmatrix} A_c & B_c \\ C_c & D_c \end{bmatrix},$$

with

$$A_c = \begin{bmatrix} A_{do} & L_{do}(\tilde{C}_2 + \tilde{D}_2K) \\ 0 & \tilde{A} + \tilde{B}_2K + L(\tilde{C}_2 - \tilde{D}_2K) \end{bmatrix}, \quad (18)$$

$$B_c = \begin{bmatrix} L_{do} \\ -K \end{bmatrix}, \quad C_c = [TC_{do} \quad L], \quad D_c = 0.$$

According to the separation principle, K and $[L \quad L_{do}]^T$ can be designed independently. The existence of compensator (18), which stabilizes the system \mathcal{G} , and cancels the disturbance w asymptotically, is given by:

Corollary 4.8 Given the system (2) and suppose the signal $w(t)$ is generated according to (3). The compensator (18) can stabilize \mathcal{G} and cancel w asymptotically, if and only if:

- i. Conditions i. – iv. of Theorem 4.3 hold,
- ii. The pair $\{A, B_2\}$ is stabilizable.

The following result is equally important as Theorem 4.6:

Theorem 4.9 (Main result II)

The RSP using output feedback with an observer based controller according to (10), is dual to the ROP with compensation according to (18).

Proof: This can again be verified by solving an RSP with output feedback, for the transposed ROP with compensation, and transposing the resulting compensator; compare (10) with (18). \square

5 Conclusions

In this paper the concepts of *servo compensator* and *disturbance observer* have been reconsidered in a general standard plant framework. Both concepts use an internal model of a persistent disturbance to asymptotically attenuate this disturbance. It has been shown that both concepts are not equal, but dual to each other, i.e. the servo compensator uses the internal model to asymptotically compensate the persistent disturbance, while the disturbance observer uses the internal model to asymptotically observe the disturbance state, enabling, under certain conditions, to asymptotically compensate the disturbance using this observation.

Exploiting this duality, it has been shown that a necessary and sufficient condition for the existence of a general disturbance observer can be derived, that is dual to the existence condition obtained in servo compensator theory. The main part of this condition is a simple rank condition on the Rosenbrock system matrix with respect to the disturbance inputs and system outputs. Besides, a necessary and sufficient condition was given for a disturbance observer to be able to compensate an observed disturbance. These simple conditions may allow a control system designer to decide upon when to use the servo compensator concept and when to use the disturbance observer concept.

Appendix

Proof of Theorem 3.3 To proof this theorem, we need to show that the compensator C_{sc} satisfies the three properties of Definition 3.1.

1. exponential stability

The compensated system can be made exponentially stable, if and only if the extended system (7) has a minimal state-space realization, i.e. the pair $\left\{ \begin{bmatrix} A & 0 \\ B_{sc}C_2 & A_{sc} \end{bmatrix}, \begin{bmatrix} B_2 \\ B_{sc}D_2 \end{bmatrix} \right\}$ is stabilizable

and the pair $\left\{ \begin{bmatrix} I_{n_x} & 0 \\ 0 & I_{n_{sc}} \end{bmatrix}, \begin{bmatrix} A & 0 \\ B_{sc}C_2 & A_{sc} \end{bmatrix} \right\}$ is detectable. Clearly the latter condition on detectability is always met, since we solve the problem with state feedback. Hence the extended system can be made exponentially stable if and only if

$$\rho \left(\begin{bmatrix} sI - A & 0 & B_2 \\ -B_{sc}C_2 & sI - A_{sc} & B_{sc}D_2 \end{bmatrix} \right) = n_x + n_{sc} \quad \forall s \in \mathbb{C}. \quad (19)$$

Now for $s \notin \sigma(A_{sc})$, condition (19) is true, since the pair $\{A, B_2\}$ is stabilizable by assumption *i.* of Theorem 3.3. Now for $s \in \sigma(A_{sc})$, rewrite:

$$\begin{bmatrix} sI - A & 0 & B_2 \\ -B_{sc}C_2 & sI - A_{sc} & B_{sc}D_2 \end{bmatrix} = \begin{bmatrix} I_{n_x} & 0 & 0 \\ 0 & -B_{sc} & sI - A_{sc} \end{bmatrix} \\ \times \begin{bmatrix} I_{n_x} & 0 & 0 \\ 0 & T^{-1} & 0 \\ 0 & 0 & I_{n_{sc}} \end{bmatrix} \begin{bmatrix} sI - A & 0 & B_2 \\ C_1 & 0 & -D_{12} \\ 0 & I_{n_{sc}} & 0 \end{bmatrix}, \quad (20)$$

where we used the fact that $z = Ty$ by assumption *ii.* of Theorem 3.3 and Lemma 3.2. Since the pair $\{A_{sc}, B_{sc}\}$ is controllable, the first factor of (20) has always rank $n_x + n_{sc}$. Clearly the second factor has full rank, and by assumption *iii.* of Theorem 3.3, the third factor has rank $n_x + n_z + n_{sc}$, $\forall s \in \sigma(A_{sc})$. Hence, by Sylvester's inequality⁶:

$$\rho \left(\begin{bmatrix} sI - A & 0 & B_2 \\ -B_{sc}C_2 & sI - A_{sc} & B_{sc}D_2 \end{bmatrix} \right) = n_x + n_{sc} \quad \forall s \in \sigma(A_{sc}),$$

i.e. the extended system (7) is stabilizable.

2. asymptotic rejection of w

Let the extended system (7) be stabilized by the state feedback (6). Consider the closed-loop transfer matrix from w to z , given by:

$$T_{zw}(s) = \left[\begin{array}{c|c} \begin{bmatrix} A + B_2K & B_2K_{sc} \\ B_{sc}(C_2 + D_2K) & A_{sc} + B_{sc}D_2K_{sc} \end{bmatrix} & \begin{bmatrix} B_1 \\ B_{sc}D_{21} \end{bmatrix} \\ \hline \begin{bmatrix} C_1 + D_{12}K & D_{12}K_{sc} \end{bmatrix} & D_1 \end{array} \right] \quad (21)$$

According to Lemma 3.2, there exists a nonsingular T such that $z = Ty$, i.e. $C_1 = TC_2$, $D_1 = TD_{21}$ and $D_{12} = TD_2$. Substituting this result in (21), and premultiplying the Rosenbrock system matrix w.r.t. w and z with a unimodular matrix yields:

$$\begin{bmatrix} I_{n_x} & 0 & 0 \\ 0 & 0 & I_{n_z} \\ 0 & I_{n_{sc}} & -B_{sc}T^{-1} \end{bmatrix}$$

⁶Sylvester's ineq.: Let $A \in \mathbb{R}^{q \times n}$, $B \in \mathbb{R}^{n \times p}$, then $\rho(A) + \rho(B) - n \leq \rho(AB) \leq \min(\rho(A), \rho(B))$.

$$\begin{bmatrix} sI - A - B_2K & -B_2K_{sc} & B_1 \\ -B_{sc}(C_2 + D_2K) & sI - A_{sc} - B_{sc}D_2K_{sc} & B_{sc}D_{21} \\ -T(C_2 + D_2K) & -TD_2K_{sc} & TD_{21} \end{bmatrix} \\ = \begin{bmatrix} sI - A - B_2K & -B_2K_{sc} & B_1 \\ -T(C_2 + D_2K) & -TD_2K_{sc} & TD_{21} \\ 0 & sI - A_{sc} & 0 \end{bmatrix}.$$

Clearly, this matrix loses rank for all $s \in \sigma(A_{sc})$, i.e. the roots of the minimal polynomial of A_{sc} appear as *transmission zeros* in the closed loop from w to z . Since these transmission zeros appear in every element of $T_{zw}(s)$, the transmission of all unstable dynamics of w to z is blocked in the closed loop, for any $x_{w0} \neq 0$. Hence, what remains in z is only due to stable dynamics, which goes to zero asymptotically.

3. robustness

Let the parameters of the system \mathcal{G} be perturbed, i.e. $\hat{A} = A + \delta A$, $\hat{B}_1 = B_1 + \delta B_1$, etc., and let $\hat{G}(s)$ be the transfer matrix of the perturbed system:

$$\hat{G}(s) = \left[\begin{array}{c|cc} \hat{A} & \hat{B}_1 & \hat{B}_2 \\ \hline \hat{C}_1 & \hat{D}_1 & \hat{D}_{12} \\ \hat{C}_2 & \hat{D}_{21} & \hat{D}_2 \end{array} \right].$$

If the perturbations $\delta A, \delta B_1$, etc. are sufficiently small (not equal to zero), the perturbed closed loop system (\hat{G} closed with C) remains stable, i.e. property *i.* of Definition 3.1 is always robust. Suppose that the perturbed closed loop remains stable, and that the regulated output \hat{z} of the perturbed system remains observable from the measured output \hat{y} of the perturbed system, then it can be readily deduced that the roots of the minimal polynomial of A_{sc} appear also as transmission zeros in the perturbed closed loop from w to z , i.e.

$$\begin{bmatrix} sI - \hat{A} - \hat{B}_2K & -\hat{B}_2K_{sc} & \hat{B}_1 \\ -B_{sc}(\hat{C}_2 + \hat{D}_2K) & sI - A_{sc} - B_{sc}\hat{D}_2K_{sc} & B_{sc}\hat{D}_{21} \\ -\hat{C}_1 - \hat{D}_{12}K & -\hat{D}_{12}K_{sc} & \hat{D}_1 \end{bmatrix}$$

loses rank $\forall s \in \sigma(A_{sc})$. Hence z of the perturbed system also goes to zero asymptotically, as long as the closed loop remains stable. Note that also perturbations in K and K_{sc} are allowed, as long as the system remains stable. \square

Proof of Theorem 4.3 To proof this theorem, we need to show that the compensator \hat{C}_{do} satisfies the three properties of Definition 4.1.

1. exponential stability

This follows by similar reasoning from part 1. of the proof of Theorem 3.3, where now stabilizability is always met, and detectability is guaranteed by conditions *iii.* and *iv.*

2. asymptotic rejection of w

This follows by similar reasoning from part 2. of the proof of Theorem 3.3, where now T_{zw} is given by:

$$\left[\begin{array}{cc|c} \tilde{A} + L\tilde{C}_2 & (\tilde{B}_2 + L\tilde{D}_2)C_{do} & \tilde{B}_1 + L\tilde{D}_{21} \\ L_{do}\tilde{C}_2 & A_{do} + L_{do}\tilde{D}_2C_{do} & L_{do}\tilde{D}_{21} \\ \hline C_1 & \tilde{D}_{12}C_{do} & \tilde{D}_1 \end{array} \right].$$

The result now follows by substituting $\tilde{B}_1 = \tilde{B}_2\tilde{T}$, $\tilde{D}_1 = \tilde{D}_{12}\tilde{T}$, $\tilde{D}_{21} = \tilde{D}_2\tilde{T}$, and postmultiplying the Rosenbrock system matrix w.r.t. w and \tilde{z} by the unimodular matrix:

$$\begin{bmatrix} I_{n_s} & 0 & 0 \\ 0 & 0 & I_{n_w} \\ 0 & I_{n_{do}} & \tilde{T}^{-1}C_{do} \end{bmatrix}.$$

3. robustness

This follows by similar reasoning from part 3. of the proof of Theorem 3.3. \square

References

- Conant, R.C. and W.R. Ashby (1970). Every good regulator of a system must be a model of that system. *Int. J. Systems Sci.*, 1, 89-97.
- Davison, E.J. (1972a). The output control of linear time-invariant multivariable systems with unmeasurable arbitrary disturbances. *IEEE Trans. Automat. Contr.*, AC-17, 621-630.
- Davison, E.J. (1972b). Author's reply to comments on "Output control of linear time-invariant multivariable systems with unmeasurable arbitrary disturbances". *IEEE Trans. Automat. Contr.*, AC-17, 838.
- Davison, E.J. (1975). Robust control of a general servomechanism problem: the servo compensator. *Automatica*, 11, 461-471.
- Davison, E.J. (1976). The robust control of a servomechanism problem for linear time-invariant multivariable systems. *IEEE Trans. Automat. Contr.*, AC-21, 25-34.
- Davison, E.J. and S.H. Wang (1974). Properties and calculation of transmission zeros of linear multivariable systems. *Automatica*, 10, 643-658.
- 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, Vol. 16, pp. 81-129.
- Francis, B.A. and W.M. Wonham (1975). The internal model principle for linear multivariable regulators. *J. Applied Mathematics and Optimization*, 2, 175-194.
- Francis, B.A. and W.M. Wonham (1976). The internal model principle of control theory. *Automatica*, 12, 457-465.
- Johnson, C.D. (1971). Accommodation of external disturbances in linear regulator and servomechanism problems. *IEEE Trans. Automat. Contr.*, AC-16, 635-644.
- Johnson, C.D. (1972). Comments on "The output control of linear time-invariant multivariable systems with unmeasurable arbitrary disturbances". *IEEE Trans. Automat. Contr.*, AC-17, 836-837.
- Johnson, C.D. (1976). Theory of disturbance-accommodating controllers. In: C.T. Leondes (Ed.), *Control and Dynamic Systems*. Academic Press, New York, Vol. 12, pp. 387-489.
- Kwatny, H.G. and K.C. Kalnitsky (1978). On alternative methodologies for the design of robust linear multivariable regulators. *IEEE Trans. Automat. Contr.*, AC-23, 930-933.
- Levin, V. and E. Kreindler (1976). Use of disturbance estimator for disturbance suppression. *IEEE Trans. Automat. Contr.*, AC-21, 776-778.
- Meditch, J.S. and G.H. Hostetter (1974). Observers for systems with unknown and inaccessible inputs. *Int. J. Control*, 19, 473-480.
- Nagasawa, M. and E. Yokamada (1993). Precision motor control system for VCR using disturbance observer. *IEEE Trans. J. Magnetics in Japan*, 8, 461-467.
- Profeta, J.A., W.G. Vogt and M.H. Mickle (1990). Disturbance estimation and compensation in linear systems. *IEEE Trans. Aerospace and Electronic Syst.*, 26, 225-231.
- Rosenbrock, H.H. (1970). *State-space and Multivariable Theory - Studies in Dynamical Systems*. Thomas Nelson and Sons LTD, London, UK.
- Sievers, L.A. and A.H. von Flotow (1989). Comparison of two LQG-based methods for disturbance rejection. *Proc. 28th IEEE Conf. Decision and Control*, Tampa, FL, pp. 483-485.
- Wolovich, W.A. and P. Ferreira (1979). Output regulation and tracking in linear multivariable systems. *IEEE Trans. Automat. Contr.*, AC-24, 460-465.

Synthesis of a robust iterative learning controller using an H_∞ approach[†]

Dick de Roover[‡]

Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: D.deRoover@wbmt.tudelft.nl

Abstract. Iterative Learning Control (ILC) is a powerful feedback methodology that iteratively improves the transient behaviour of processes that are repetitive in nature. Although most of the published ILC schemes are heuristic in nature, some initial research has been performed on the formulation of the ILC problem in the H_∞ mathematical framework. However, so far only the performance and robustness *analysis* of the ILC schemes has been performed for a given (heuristically designed) learning controller. In this paper it is shown how the *synthesis* of an iterative learning controller can be generalized to the synthesis of an H_∞ (sub)optimal controller. It is shown how a general learning control problem can be reformulated in the so-called 'standard plant' format, by choosing an appropriate weighting function for learning performance. Moreover, process uncertainty can be included explicitly in the ILC design, by choosing appropriate weighting functions related to this uncertainty. It turns out that convergence and learning performance of this ILC scheme can be obtained for all systems in the uncertainty set, by solving a μ -synthesis problem. The practical usefulness of the scheme is verified on an $xy\phi$ -stage experimental setup.

Keywords. Iterative Learning Control (ILC), H_∞ control theory, robust learning performance, μ -synthesis, $xy\phi$ -stage experimental setup.

1 Introduction

For control systems that have to perform their tasks repeatedly, Iterative Learning Control (ILC) has turned out to be an effective tool for improving the transient performance. After its introduction in the systems and control community by Arimoto *et al.* (1984), the number of 'newly' proposed learning schemes has become almost as large as the number of practitioners, see for example Bondi *et al.* (1988), Kavli (1992), Liang and Looze (1993), Moore *et al.* (1992); the reader is referred to Moore *et al.* (1992) for an extensive list of references. Almost all publi-

cations concern the design of a *convergent* scheme. In fact, ILC also has to deal with the well known trade-off between performance and robustness of the controller, i.e. the theoretically superb performance of tracking a transient without error, may have to be sacrificed in practice, due to the more severe demand of convergence of the scheme in the face of uncertain system dynamic knowledge. Although most of the published ILC schemes are heuristic in nature, some initial research has been performed on the formulation of the ILC problem in the H_∞ mathematical framework, references Padiou and Su (1990), Moore *et al.* (1992), Liang and Looze (1993). However, the cited references only considered the performance and robustness *analysis* of ILC schemes, based on heuristically designed learning controllers.

In this paper it is shown that the *synthesis* of an iterative learning controller can be generalized to the

[†]This paper is presented at the 35th IEEE Conference on Decision and Control, 11-13 December 1996, Kobe, Japan. Copyright of this paper remains with IEEE.

[‡]The work of Dick de Roover is financially supported by Philips' Research Laboratories, Eindhoven, the Netherlands.

synthesis of an H_∞ (sub)optimal controller, by reformulating the ILC problem in the so-called 'standard plant' format. The key issue of this solution to the ILC problem, is the specification of an appropriate weighting function for learning performance, and the fact that the delay line, between successive updates of the forcing function, has magnitude equal to 1, hence allowing effective use of the small gain theorem. Moreover, solving the ILC problem in this format, allows the designer to explicitly trade-off learning performance and robustness, by incorporating process uncertainty weighting functions into the synthesis problem. It is shown that this *robust performance* problem can be solved using a μ -synthesis approach.

The remainder of this paper is organized as follows. Section 2 describes the general ILC problem, and an heuristic solution to it. Section 3 shows how the ILC problem can be reformulated and solved in the standard plant format. Furthermore, Section 4 extends the results of Section 3 to the formulation and solution of a *robust* ILC problem, by including process uncertainty descriptions into the design. Section 5 verifies the practical usefulness of the proposed scheme on a real life experimental setup of an $xy\phi$ -stage, and finally Section 6 ends up with some conclusions.

2 Heuristic approach to ILC

2.1 Problem formulation

Consider the feedback configuration depicted in Figure 1. It is assumed that the plant P and the con-

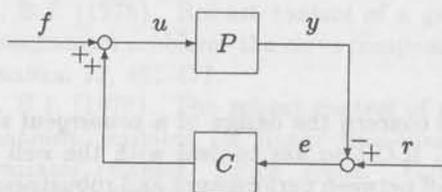


Fig. 1: Feedback configuration of plant P and controller C ; the signals f, r, u, y , and e denote a forcing function, a reference signal, the plant input, the plant output, and the servo error, respectively.

troller C are discrete time, linear, and time invariant. The map from $\text{col}(r, f)$ to $\text{col}(y, e)$ can be obtained from simple block diagram manipulation:

$$(I + PC)^{-1} \begin{bmatrix} PC & P \\ I & -P \end{bmatrix} =: \begin{bmatrix} T & R \\ S & -R \end{bmatrix}, \quad (1)$$

thus: $T = (I + PC)^{-1}PC$, $R = (I + PC)^{-1}P$, and $S = (I + PC)^{-1}$. Suppose the reference signal $r(t)$,

defined on a finite interval $[0, \Delta T, \dots, T_N]$ with ΔT being the sampling time and $T_N = N\Delta T$ with N the number of samples, is repeated an indefinite number of times, each time starting at the same initial condition; if so, the servo error $e(t)$ will be repeated too, apart from random noise. The main idea of iterative learning control is to iteratively update the forcing function f , so as to decrease the magnitude of e , after each cycle of the reference signal r . More formal, let k denote the number of iterations, then ILC is about to find an update \mathcal{U} of the forcing function at the k -th iteration, based on the servo error at the k -th iteration, i.e.:

$$f_{k+1}(t) = \mathcal{U}(f_k(t), e_k(t)), \quad t \in [0, T_N], \quad k \in \mathbb{N}, \quad (2)$$

such that

$$\lim_{k \rightarrow \infty} f_k(t) = f_*(t) \quad \text{and} \quad \lim_{k \rightarrow \infty} \|e_k(t)\| = e_*, \quad (3)$$

with $f_*(t)$ and e_* being fixed points, and e_* is minimal over the interval $[0, T_N]$, measured in some signal norm $\|\cdot\|$. Obviously, convergence of an ILC scheme to fixed points f_* and e_* , depends on the choice of the update law (2); the great body of literature on ILC is mainly concerned with 'newly' proposed update laws. Roughly speaking, for linear systems two different types of update laws can be distinguished: a *PID-type* of update law and a *model based* update law. The most general PID-type of update law can be found in Arimoto (1985), which updates the forcing function by:

$$f_{k+1}(t) = f_k(t) + \alpha e_k(t) + \beta \dot{e}_k(t) + \gamma \int e_k(t) dt. \quad (4)$$

For this type of learning rule, convergence conditions are derived, so as to obtain the gains $\{\alpha, \beta, \gamma\}$. The most general model based update law is proposed in Moore *et al.* (1992), and reads as follows:

$$f_{k+1}(t) = Q(q)f_k(t) + L(q)e_k(t), \quad (5)$$

General convergence conditions on the filters Q and L are derived, based on knowledge of the plant P and the feedback controller C . In fact, update law (5) can be seen as a generalization of update law (4), by making specific choices of the filters Q and L . Hence, learning rule (5) will be the starting point for the research presented in this report.

Figure 2 shows how this general learning rule can be implemented as an *offline ad-on device*. During execution of the reference signal r , together with forcing function f_k at the k -th trial (within the setup of Figure 1), the servo error e_k at the k -th trial is measured and logged in a memory table. After this reference trial, the logged error is filtered with the

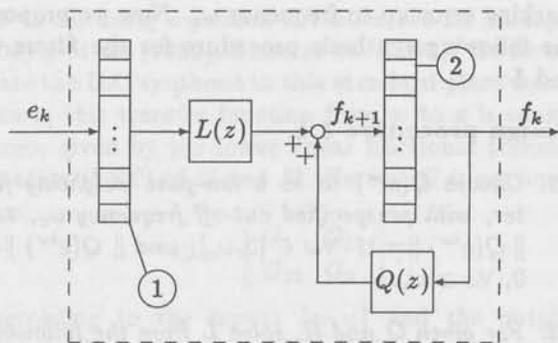


Fig. 2: Offline learning algorithm; 1: logged error table, 2: feedforward table

L -filter, and added to the Q -filtered forcing function f_k of the k -th trial. The sum of both signals is stored in another memory table, and constitutes the forcing function f_{k+1} , to be applied in the next ($(k+1)$ th) trial.

2.2 Convergence criterion

In this paper, contraction mapping theory is used in convergence analysis and synthesis (Moore *et al.*, 1992). The following convergence result is obtained for the general update law (5):

Theorem 2.1 (L_2 convergence) Consider the feedback configuration of Figure 1, and suppose $f, e \in L_2(0, \infty)$, then the learning iteration (5) converges to a fixed point $f_*(t), t \in [0, T_N]$, if

$$\|Q(z) - L(z)R(z)\|_\infty < 1, \quad (6)$$

with $R(z)$ as defined in (1), and $\|\cdot\|_\infty$ denoting the matrix ∞ -norm (equal to the induced 2-norm).

Proof: Can be found either in Padieu and Su (1990) or in Moore *et al.* (1992). \square

If (6) is satisfied, the fixed points $f_*(t)$ and e_* can be obtained by substituting the lower part of equation (1) into equation (5), and using the fact that in the limit $f_{k+1} = f_k = f_*$:

$$f_*(t) = (I - Q + LR)^{-1}LSr(t),$$

and

$$e_* = \|(I - R(I - Q + LR)^{-1}L)Sr(t)\|_2 \quad (7)$$

2.3 Design of filters $Q(z)$ and $L(z)$

Whereas convergence analysis results like Theorem 2.1 have received considerable attention in the literature, practical guidelines for synthesis of the filters $Q(z)$ and $L(z)$ are hard to find. In making an appropriate choice for these filters, the following result is very important:

Theorem 2.2 Suppose $R, L \neq 0$, then for the feedback configuration of Figure 1, and the learning iteration (5), the fixed point e_* is zero, if and only if (6) is true, and $Q(z) = I$.

Proof: Trivial, (6) has to be met in order for a fixed point e_* to exist.

(\Leftarrow) This simply follows by substituting $Q = I$ into equation (7).

(\Rightarrow) Suppose $Q(z) = (1 + \epsilon)I$, and substitute this into equation (7):

$$\begin{aligned} e_* &= \|(I - R(I - (1 + \epsilon)I + LR)^{-1}L)Sr(t)\|_2 \\ &= \|(I - R(-\epsilon I + LR)^{-1}L)Sr(t)\|_2 \\ &= \|(I + R1/\epsilon IL)^{-1}Sr(t)\|_2 \\ &= \left\| \left(- \sum_{j=1}^{\infty} (R1/\epsilon L)^{-j} \right) Sr(t) \right\|_2 \end{aligned}$$

Trivial, for $R, L \neq 0$, this last term is zero if and only if $\epsilon = 0$, i.e. $Q(z) = I$. \square

This result is the main reason why most of the proposed ILC schemes operate with $Q(z) = I$. In this case, the convergence criterion of Theorem 2.1 simplifies to:

$$\|I - L(z)R(z)\|_\infty < 1. \quad (8)$$

For this criterion to hold true $\forall z$, it can be easily seen that (8) implies invertibility of the mapping R , i.e. $L(z)$ should be chosen equal to $R^{-1}(z)$, (Moore *et al.*, 1992). However, most times in practical situations it is not possible to obtain an exact description of the inverse (closed loop) system, either because it does not exist¹ due to strict properness or nonminimum phase behaviour, or because it is too complex to describe. It was well recognized that the use of the filter Q in the update law, although destroying the perfect tracking property described in Theorem 2.2, robustifies the learning algorithm with respect to uncertainty in the dynamics of R , (Moore *et al.*, 1992, Kavli, 1992).

On the basis of the abovementioned arguments, in Kavli (1992) the following (heuristic) frequency domain design procedure was proposed for the filters Q and L :

Design procedure 2.1

1. Choose $L(e^{i\omega}) \approx R^{-1}(i\omega)$, $\omega \in [0, \omega_c]$, i.e. choose L to be the best possible (approximate) inverse of R , up to some frequency ω_c .

¹Note that in the MIMO case, R should be left invertible, i.e. the plant P should have at least as many outputs as inputs!

- Choose $Q(e^{i\omega})$ to be a low-pass filter with cut-off frequency near ω_c , with $\|Q(e^{i\omega})\|=1$, $\forall \omega \in [0, \omega_c]$, and $\|Q(e^{i\omega})\| \approx 0$, $\forall \omega > \omega_c$.

The idea is, that most real life systems can be described rather well at low frequencies, and that the uncertainty in the description shows up at high frequencies. Now for all frequencies $\omega \in [0, \omega_c]$ with $\|Q(e^{i\omega})\|=1$, the frequency content of e_* will be zero, according to Theorem 2.2, and for $\omega > \omega_c$ the frequency content of e_* will be equal to the frequency content of e_0 , i.e. the servo error without learned forcing function, (Kavli, 1992).

Although good practical results can be obtained using this approach, the design procedure is rather ad lib, in a sense that it is hard to decide at which frequency ω_c the uncertainty starts to play a role, i.e. the trade-off between *performance* on the one hand ($e_* = 0$ for $Q(z) = I$) and *robustness* on the other hand ($Q(z) \approx 0 \forall z$ for which $L(z) \neq R^{-1}(z)$), is made in a rather heuristic way. In the next two sections, a novel synthesis procedure for the filters Q and L is proposed, providing explicit control over the trade-off between learning performance and robustness to plant uncertainty.

3 ILC synthesis using an H_∞ approach

3.1 Novel design procedure

In this section we propose a slight change to learning rule (5), without altering its generality:

$$f_{k+1}(t) = Q(f_k(t) + Le_k(t)), \quad (9)$$

for two reasons. First, due to the strict properness of most real life systems, in practical applications the filter L will behave as a differentiator at high frequencies. Therefore, to avoid differentiating of high-frequency signals, L is also cut off at some high frequency.

Second, by choosing the update law equal to (9), the convergence criterion (6) changes to:

$$\|Q(z)(I - L(z)R(z))\|_\infty < 1. \quad (10)$$

From an H_∞ -controller design point of view, equation (10) motivates to interpret the filter $Q(z)$ as a *weighting function* for *learning performance*, i.e.:

$$\|I - L(z)R(z)\|_\infty < \|Q^{-1}(z)\|_\infty.$$

Consistent with Theorem 2.2, it seems fairly natural to view the filter $Q(z)$ as a measure for learning performance: the cut-off frequency ω_c has to be chosen as large as possible, in order to guarantee zero

tracking error up to frequency ω_c . Now we propose the following synthesis procedure for the filters Q and L :

Design procedure 3.1

- Choose $Q(e^{i\omega})$ to be a low-pass weighting filter, with prespecified cut-off frequency ω_c , s.t. $\|Q(e^{i\omega})\|=1$, $\forall \omega \in [0, \omega_c]$, and $\|Q(e^{i\omega})\| \approx 0$, $\forall \omega > \omega_c$.
- For given Q and R , solve L from the following (sub)optimal H_∞ -synthesis problem:

$$L(z) = \arg \min_{L \in H_\infty} \|Q(z)(I - L(z)R(z))\|_\infty \quad (11)$$

In fact, (11) describes the well-known 'model matching problem', i.e. for given Q , L is *matched* to the inverse of R . Obviously, for convergence of the proposed ILC scheme, the minimizing argument $L_*(z)$ of the proposed synthesis procedure should result in

$$\|Q(z)(I - L_*(z)R(z))\|_\infty = \gamma_* < 1,$$

according to (10). Note that the smaller γ_* , the faster f and e converge to their fixed points f_* and e_* , which can be easily seen from:

$$\begin{aligned} \|f_{k+1} - f_k\|_2 &\leq \gamma_* \|f_k - f_{k-1}\|_2 \leq \dots \\ &\leq \gamma_*^k \|f_1 - f_0\|_2, \end{aligned} \quad (12)$$

i.e. by minimizing $\|Q(I - LR)\|_\infty$, the highest convergence rate in L_2 is obtained.

3.2 Workable solution to proposed synthesis

To solve the problem described in Equation (11) for practical situations, we adopt the approach suggested in Balas *et al.* (1991), based on Doyle *et al.* (1989). Therefore, the ILC synthesis problem is reformulated in the standard plant format, depicted in Figure 3. Within this framework, tools are avail-

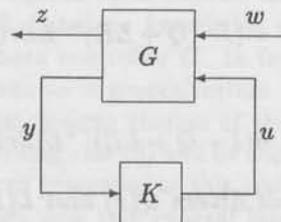


Fig. 3: 'Standard plant' configuration of generalized plant G and controller K , see Balas *et al.* (1991).

able for computing a stabilizing K that minimizes $\|T_{zw}\|_\infty$ (with T_{zw} being the transfer function from

w to z), using a *gamma iteration*, see for example, Doyle *et al.* (1989), Zhou *et al.* (1996). To formulate the ILC synthesis in this standard plant framework, this transfer function from w to z is considered, given by the lower linear fractional transformation (LFT) of G and K . Hereto, G is partitioned as:

$$G := \begin{bmatrix} G_1 & G_{12} \\ G_{21} & G_2 \end{bmatrix},$$

according to the inputs $\{w, u\}$ and the outputs $\{y, z\}$. Now the lower LFT of G and K , denoted $\mathcal{F}_l(G, K)$, is defined as:

$$\mathcal{F}_l(G, K) := G_1 + G_{12}K(I - G_2K)^{-1}G_{21} = T_{zw}. \quad (13)$$

Clearly, (11) can be described in this format, by taking:

$$K = L \quad \text{and} \quad G := \begin{bmatrix} G_1 & G_{12} \\ G_{21} & G_2 \end{bmatrix} = \begin{bmatrix} Q & Q \\ -R & 0 \end{bmatrix}. \quad (14)$$

According to (13), this choice for G and K results in the computation of a stabilizing L , such that $\|Q(I - LR)\|_\infty$ is minimized. In fact, the signals z and w represent the signals f_{k+1} and f_k respectively. Moreover, due to the fact that the pure delay line of length N , between f_{k+1} and f_k , has magnitude equal to 1, the small gain theorem, requires $\|T_{zw}\|_\infty = \|Q(I - LR)\|_\infty$ to be less than 1 for guaranteeing stability of the closed loop; not remarkable, this requirement is precisely the convergence condition (10). Figure 4 shows the ILC design problem, described in the format of Figure 3.

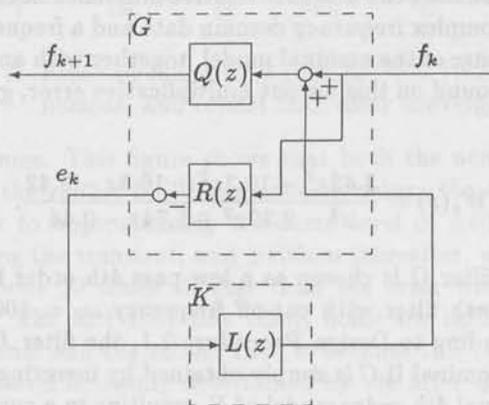


Fig. 4: 'Standard plant' configuration for ILC synthesis problem.

It should be noted that for solution of the synthesis problem using the standard plant format, the plant G is described in state space coordinates:

$$\begin{aligned} \dot{x} &= Ax + B_1w + B_2u \\ y &= C_1x + D_1w + D_{12}u \end{aligned}$$

$$z = C_2x + D_{21}w + D_2u.$$

Now solving the ILC synthesis problem using the 2 coupled Riccati equations, see Doyle *et al.* (1989), will in general result in a singular H_∞ synthesis problem, because the matrices D_{12} and D_{21} do not have full column and row rank, respectively; this is due to the fact that the transfer functions $Q(z)$ and $R(z)$ are not proper, respectively. However, this problem can be easily cured, by slightly perturbing the original problem, (Stoorvogel, 1990).

4 Synthesis of a robust learning controller

One advantage of Design Procedure 3.1 with respect to Design Procedure 2.1, is the fact that for a given filter Q , the best possible L_2 -convergence is obtained, since (10) is actually *minimized*, remind equation (12); hence there is no other filter L that can obtain a higher convergence rate w.r.t. a given Q . Another advantage of the newly proposed synthesis procedure, is the fact that this setup provides a suitable framework for designing real *multivariable* learning controllers; interaction is explicitly taken into account in the design of the filters Q and L , and hence learning performance is guaranteed for the multivariable system.

Another, even more important, advantage, is the fact that formulating the ILC synthesis problem in the H_∞ framework, allows the designer to explicitly take uncertainty, with respect to the transfer function R , into account. For example, suppose the real closed loop system can be described by some nominal transfer function $R_0(z)$ and some output multiplicative uncertainty, specified by a stable and stably invertible weighting function $W_o(z)$:

$$R(z) := \{(I + W_o(z)\Delta(z))R_0(z) \mid \|\Delta(z)\|_\infty \leq 1\}, \quad (15)$$

then the newly proposed ILC synthesis procedure can be easily extended by choosing:

$$K = L \quad \text{and} \quad G = \begin{bmatrix} Q & 0 & Q \\ R_0 & 0 & 0 \\ -R_0 & W_o & 0 \end{bmatrix}, \quad (16)$$

and solving K by minimizing $T_{zw} = \mathcal{F}_l(G, K)$. However, taking the lower LFT of these extended G and K gives:

$$\begin{aligned} \mathcal{F}_l(G, K) &= \begin{bmatrix} Q & 0 \\ R_0 & 0 \end{bmatrix} + \begin{bmatrix} Q \\ 0 \end{bmatrix} L[-R_0 \ W_o] \\ &= \begin{bmatrix} Q(I - LR_0) & QLW_o \\ R_0 & 0 \end{bmatrix}, \quad (17) \end{aligned}$$

and hence:

$$\|T_{zw}\|_\infty \geq \|R_0\|_\infty.$$

Since in practical situations it is likely that $\|R_0\|_\infty$ is greater than one, this problem can in general not be solved using the standard H_∞ synthesis procedure. This is not surprising, since taking robustness against specified uncertainty into account, turns this ILC synthesis problem into a *robust performance* design problem, which cannot be solved in general. Still it might be possible to solve the synthesis problem, using the *structured singular value* (μ), see for example Balas *et al.* (1991); exploiting the diagonal structure in the mapping from z to w , μ can be computed of the matrix M , the lower LFT of G and K given in (17), for a finite number of frequencies. Now a successive iteration can be performed, of scaling the off-diagonal elements of M (D-scaling by means of frequency domain curve fitting on $\mu(M)$) and solving the (sub)optimal H_∞ -synthesis problem for the scaled system. Although convergence of this so-called *D-K iteration* is not guaranteed, in a large number of cases good results are reported, (Zhou *et al.*, 1996).

Now it is quite natural to proceed with maximizing the learning performance specified in Theorem 2.2, by iterating over the bandwidth ω_c of the filter Q . Hereto, we propose the following design procedure:

Design procedure 4.1

1. Model the transfer function R as a nominal model together with an upper bound on the (output multiplicative) model uncertainty, according to (15).
2. Choose $Q(e^{i\omega})$ to be a low-pass weighting filter, with cut-off frequency ω_c , s.t. $\|Q(e^{i\omega})\| = 1$, $\forall \omega \in [0, \omega_c]$, and $\|Q(e^{i\omega})\| \approx 0$, $\forall \omega > \omega_c$.
3. For given Q and R , find an L that minimizes (17) using the proposed μ -synthesis.
4. If an L can be found, resulting in $\|T_{zw}\|_\infty < 1$, increase the bandwidth ω_c of the filter Q and again perform step 3; else decrease ω_c . Perform steps 2 and 3 iteratively, until the maximum obtainable ω_c has been reached.

5 Application to an $xy\phi$ -stage

In this section, the proposed synthesis procedure is applied to an experimental setup of an $xy\phi$ -stage, a high accuracy positioning mechanism. The stage is moved repeatedly, according to some smooth trajectory; for the experiments in this paper, a 3rd order polynomial was used, making a 1cm step in approximately 0.12 seconds, with $\Delta T = 3e^{-4}s$.

For this system, two learning controllers have been designed and implemented: one designed according

to Design Procedure 2.1, henceforth denoted as the *nominal* ILC, and the other designed according to Design Procedure 4.1, denoted as the *robust* ILC.

Design of both learning controllers requires knowledge of the transfer function R from f to e . A common approach for mechanical servo systems to obtain this knowledge is identification in the frequency domain, see de Callafon *et al.* (1996). Figure 5 shows a magnitude Bode plot of frequency domain data, obtained with a Hewlett-Packard signal an-

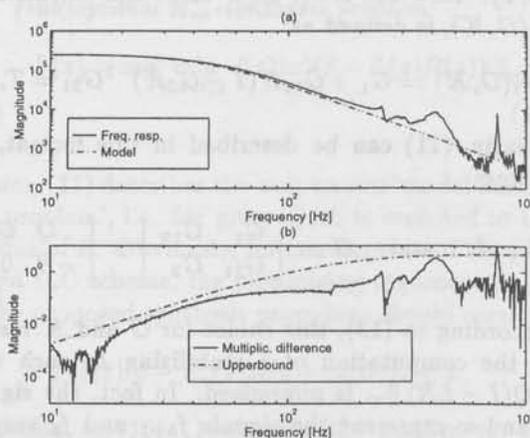


Fig. 5: (a) Magnitude Bode plot of frequency response data of R with a 4th order nominal model (b) relative difference between the frequency response data of R and nominal model, together with an upper bound.

alyzer, together with a simple 4th order nominal model fitted to this curve; the lower part of this figure shows the absolute relative difference between the complex frequency domain data and a frequency response of the nominal model, together with an upper bound on this output multiplicative error, given by:

$$W_o(z) = \frac{3.42z^3 - 10.3z^2 + 10.3z - 3.42}{z^3 - 2.29z^2 + 1.74z - 0.44}$$

The filter Q is chosen as a low-pass 4th order Butterworth filter with cut-off frequency $\omega_c = 100\text{Hz}$. According to Design Procedure 2.1, the filter L for the nominal ILC is simply obtained by inverting the nominal 4th order model of R , resulting in a convergence rate of $\gamma_* = \|Q(I - LR)\|_\infty = 0.12$; Figure 6 shows a Bode plot of this filter. Also shown in this figure, is a Bode plot of the robust filter L , obtained via Design Procedure 4.1; the performed μ -synthesis started with a value $\gamma_* = 1.38$, and converged to a value $\gamma_* = 0.98$ after 5 iterations. It can be seen from this Figure that the robust learning filter is somewhat more 'cautious' at all frequencies, in a sense that it has a smaller gain.

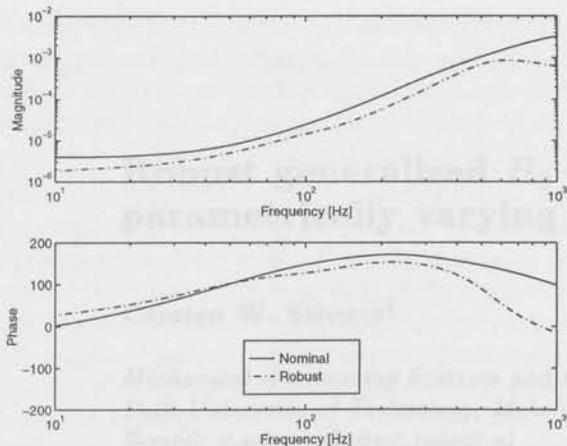


Fig. 6: Bode plot of nominal learning filter and robust learning filter L , designed according to Design Procedures 2.1 and 4.1 respectively.

With these two filters, a learning iteration was performed, according to (5). Figure 7 shows the resulting error signals after convergence of the ILC

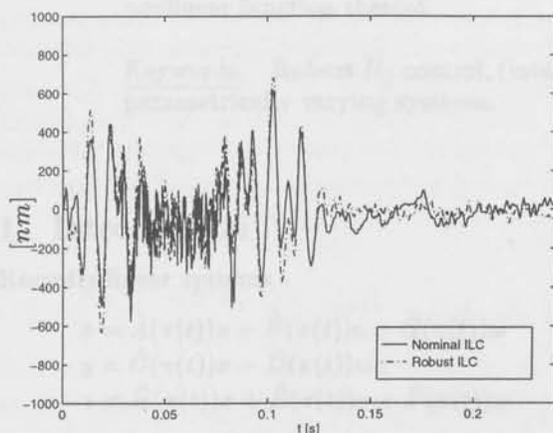


Fig. 7: Error signals $e_*(t)$, $t \in [0, 3e^{-4}, \dots, 0.24]$, of nominal and robust ILC, after convergence.

schemes. This figure shows that both the nominal and the robust ILC tremendously reduce the servo error to approximately the same level of $\pm 600\text{nm}$ during the transient, and $\pm 100\text{nm}$ thereafter, which is about 20 times smaller than the error without ILC. The fact that this result holds for both the nominal and the robust ILC, is because this reduction level is mainly determined by the filter Q , according to Theorem 2.2; apparently, the bandwidth $\omega_c = 100\text{Hz}$ of the filter Q is high enough to give this large error reduction.

The main difference between the nominal and robust ILC is the convergence rate of both schemes ($\gamma_* = 0.12$ v $\gamma_* = 0.98$); whereas the nominal ILC converged within 3 iterations, the robust ILC needed 10 iterations to converge to its fixed point, clearly showing the trade-off between convergence

rate and robustness.

Despite its slower convergence, an important advantage of the robust ILC is the fact that it has been designed for all systems characterized by the 4th order nominal model and the upper bound on the multiplicative model uncertainty. Hence, it is likely that the robust ILC will give the same learning performance for other systems that can be described by the same class of systems, for example other $xy\phi$ -stages of the same type as the one used for the experiments shown in this paper.

Finally, it was investigated that the bandwidth of the filter Q could be increased up to $\omega_c \approx 170\text{Hz}$, to obtain a final value of $\gamma_* = 0.99$, using Design Procedure 4.1; for higher values of ω_c , the μ -synthesis did not converge to a value below 1 for the specified uncertainty set. Not surprisingly, the nominal ILC diverged for values $\omega_c > 170\text{Hz}$, indicating the reliability of the chosen uncertainty set W_o , and the usefulness of the proposed robust synthesis procedure.

6 Conclusions

Since most schemes on iterative learning control analyze their convergence in an L_2 sense, this paper has shown how the design of an iterative learning controller (ILC) can be generalized to the design of an L_2 induced (i.e. an H_∞), (sub)optimal controller, by choosing an appropriate weighting function for learning performance, and reformulating the ILC synthesis problem in the standard plant format. Besides the advantages of maximizing the L_2 convergence rate and allowing a real multivariable ILC design, an important advantage is the fact that uncertain system knowledge can be incorporated explicitly into the design procedure, by specifying appropriate uncertainty weighting functions, turning the nominal ILC synthesis problem into a robust performance synthesis problem, which is hard to solve in general. However, using a μ -synthesis approach, most times this robust ILC synthesis problem can be solved, by performing a so-called D-K iteration. This allows the designer to maximize the learning performance for a specified class of systems. Application of this synthesis procedure on a real life experimental setup of an $xy\phi$ -stage, has shown the practical ability of the proposed method.

Acknowledgement

The author greatly acknowledges the support of an $xy\phi$ -stage experimental setup by the Philips' Research Laboratories, Eindhoven, The Netherlands. He also wishes to thank Frank Sperling and Maarten Steinbuch, members of the Philips' Lab., for useful

discussions on the ILC subject.

References

- Arimoto, S. (1985). Mathematical theory of learning with applications to robot control. In *Proc. 4th. Yale Workshop on Applications of Adaptive Systems*, New Have, C.T., pp. 379-388.
- Balas, G., J.C. Doyle, K. Glover, A. Packard and R. Smith (1991). *μ -Analysis and Synthesis Toolbox*. MUSYN Inc. and The MathWorks, Inc.
- Bondi, P., G. Casalino and L. Gambardella (1988). On the iterative learning control theory for robotic manipulators. *IEEE J. Robotics and Automation*, 4, 14-22.
- de Callafon, R.A., D. de Roover and P.M.J. Van den Hof (1996). Multivariable least squares frequency domain identification using polynomial matrix fraction descriptions. To appear in *IEEE Proc. 35th Conf. Decision and Control*, Kobe, Japan, 1996. (See also this issue, pp. 31-37).
- Doyle, J.C., K. Glover, P.P. Khargonekar and B.A. Francis (1989). State-space solutions to standard H_2 and H_∞ control problems. *IEEE Trans. Automat. Contr.*, AC-34, 831-847.
- Kavli, T. (1992). Frequency domain synthesis of trajectory learning controllers for robot manipulators. *J. Robotic Systems*, 9, 663-680.
- Liang, Y-J. and D.P. Looze (1993). Performance and robustness issues in iterative learning control. *Proc. 32nd IEEE Conf. Decision and Control*, San Antonio, TX, pp. 1990-1995.
- Moore, K.L., M. Dahleh and S.P. Bhattacharyya (1992). Iterative learning control: a survey and new results. *J. Robotic Systems*, 9, 563-594.
- Padiou, F. and R. Su (1990). An H_∞ approach to learning control systems. *Int. J. Adapt. Contr. and Signal Proc.*, 4, 465-474.
- Stoorvogel, A.A. (1990). *The H_∞ Control Problem: A State Space Approach*. Ph.D. Thesis, Eindhoven Univ. Techn., The Netherlands.
- Zhou, K., J.C. Doyle and K. Glover (1996). *Robust and Optimal Control*, Prentice Hall, Englewood Cliffs, NJ.

Robust generalized H_2 control for uncertain and linear parametrically varying systems with full block scalings[†]

Carsten W. Scherer[‡]

*Mechanical Engineering Systems and Control Group
 Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands
 E-mail: c.scherer@wbmt.tudelft.nl*

Abstract. In this paper we extend known analysis and synthesis results for robust generalized H_2 performance to general full block scalings and provide a full solution to the corresponding linear parametrically varying control problem. This is only possible if scheduling the controller not just with a copy of the plant uncertainties but with a nonlinear function thereof.

Keywords. Robust H_2 control, (integral) quadratic constraints, full block scalings, linear parametrically varying systems.

1 Introduction

Recently linear systems

$$\begin{aligned} \dot{x} &= \hat{A}(\pi(t))x + \hat{B}(\pi(t))u + \hat{G}(\pi(t))w \\ y &= \hat{C}(\pi(t))x + \hat{D}(\pi(t))w \\ z &= \hat{H}(\pi(t))x + \hat{E}(\pi(t))u + \hat{F}(\pi(t))w \end{aligned} \quad (1)$$

which depend on a time-varying a priori unknown but on-line measurable parameter $\pi(t)$ which is contained in some given set Π have gained a lot of interest (Apkarian and Gahinet, 1995; Apkarian et al., 1994; Becker et al., 1993; Becker and Packard, 1994; Packard, 1994; Scorletti and El Ghaoui, 1995; Scherer, 1995). These so-called linear parametrically-varying (LPV) systems appear in robustness problems, in gain-scheduling techniques for nonlinear systems, or in synthesis problems for nonlinear systems that can be described by a differential inclusion (Boyd et al., 1994). Given an LPV system, the goal is to construct a controller

$$\begin{aligned} \dot{x}_c &= \hat{A}_c(\pi(t))x_c + \hat{B}_c(\pi(t))y \\ u &= \hat{C}_c(\pi(t))x_c + \hat{D}_c(\pi(t))y \end{aligned} \quad (2)$$

[†]This paper is presented at the 35th IEEE Conference on Decision and Control, 11-13 December 1996, Kobe, Japan. Copyright of this paper remains with IEEE.

[‡]The author would like to thank Samir Bennani, Dehlia Willemsen and Edwin Njio (all from Delft University of Technology) for many helpful discussions.

which not only uses the measured output y but, in addition, the on-line measured actual parameter $\pi(t)$ as its information in order to exponentially stabilize the system (1) and to provide good performance properties expressed as a property of the channel $w \rightarrow z$. Up to now, the performance objective was mostly specified as reducing the L_2 -gain of the channel $w \rightarrow z$ below a certain a priori given level γ , the so-called H_∞ -problem for LPV systems (Apkarian and Gahinet, 1995; Apkarian et al., 1994; Becker et al., 1993; Becker and Packard, 1994; Helmersson, 1995; Packard, 1994; Scorletti and El Ghaoui, 1995; Wu, 1995). The corresponding problem for the LQG criterion has been discussed in Wu (1995) and Wu and Packard (1995) and the mixed H_2/H_∞ problem is considered in Scherer (1995).

One of the approaches to solve this problem proceeds as follows: Represent (1) using a linear fractional transformation as

$$\begin{bmatrix} \dot{x} \\ y \\ z \end{bmatrix} = \begin{bmatrix} A & B & G & G_2 \\ C & 0 & D & D_2 \\ H & E & F & F_{12} \\ H_2 & E_2 & F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} x \\ u \\ w \\ w_2 \end{bmatrix}, \quad (3)$$

$$w_2 = \Delta(\pi(t))z_2$$

where $\Delta : \Pi \rightarrow \mathcal{R}^{k \times l}$ is a possibly nonlinear known

function. Typically, if $\hat{A}(\pi)$, $\hat{B}(\pi)$, etc. in (1) are rational functions of $\pi \in \mathcal{R}^m$ without pole in 0, one can obtain this representation with

$$\Delta(\pi) = \text{diag}(\pi_1 I_1, \dots, \pi_m I_m) \quad (4)$$

where I_j denote identity matrices of varying size. The description (3) then resembles an uncertain system as considered in μ -theory with real time-varying parametric uncertainty. In this case Π is often assumed to be a box in \mathcal{R}^m centered around 0.

The controller is then assumed to admit the same structure: It consists of an LTI system

$$\dot{x}_c = A_c x_c + B_c \begin{bmatrix} y \\ w_c \end{bmatrix}, \quad \begin{bmatrix} u \\ z_c \end{bmatrix} = C_c x_c + D_c \begin{bmatrix} y \\ w_c \end{bmatrix} \quad (5)$$

that is scheduled as

$$w_c = \Delta_c(\pi(t)) z_c \quad (6)$$

with a possibly nonlinear function $\Delta_c: \Pi \rightarrow \mathcal{R}^{k_c \times l_c}$. The LPV synthesis problem is posed as follows: Find A_c, B_c, C_c, D_c and a function

$$\Delta_c: \Pi \rightarrow \mathcal{R}^{k_c \times l_c}$$

such that, for all parameter curves $\pi: [0, \infty) \rightarrow \Pi$, the closed-loop system is exponentially stable and has a certain desired performance property described using the channel $w \rightarrow z$.

Note that the system (3) controlled by (5)-(6) can be also obtained by connecting

$$\begin{bmatrix} \dot{x} \\ y \\ w_c \\ z \\ z_2 \\ z_c \end{bmatrix} = \begin{bmatrix} A & B & 0 & G & G_2 & 0 \\ C & 0 & 0 & D & D_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{k_c} \\ H & E & 0 & F & F_{12} & 0 \\ H_2 & E_2 & 0 & F_{21} & F_{22} & 0 \\ 0 & 0 & I_{l_c} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ u \\ z_c \\ w \\ w_2 \\ w_c \end{bmatrix} \quad (7)$$

$$\begin{bmatrix} w_2 \\ w_c \end{bmatrix} = \begin{bmatrix} \Delta(\pi(t)) & 0 \\ 0 & \Delta_c(\pi(t)) \end{bmatrix} \begin{bmatrix} z_2 \\ z_c \end{bmatrix}$$

with the LTI controller (5).

If dealing with the L_2 gain as a performance measure, if $\Delta(\pi)$ admits the special structure (4), and if choosing $\Delta_c(\pi) = \Delta(\pi)$, the resulting LPV problem hence turns out to be a robust H_∞ problem for static time-varying uncertainties. The first approaches to this problem were restricted to solve the corresponding upper bound μ -synthesis problem with block-diagonal constant D -scales (Packard, 1994; Apkarian and Gahinet, 1995). Recently it has been shown how to include the block-diagonal G -scales to reflect the fact that the parameters are real-valued and to reduce conservatism (Fan et al., 1991; Helmersson, 1995; Scorletti and El Ghaoui, 1995). However, for

the robust stabilization problem, it has been pointed out (Rantzer and Megretski, 1994; Iwasaki et al., 1995) that it is possible to use a much larger class of scalings defined via (integral) quadratic constraints.

The purpose of this paper is twofold. Firstly we want to show that it is possible to approach many other robust performance problems using this general class of scalings, including robust H_2 -criteria (Iwasaki, 1993; El Ghaoui and Folcher, 1996). Along the lines as described e.g. in Boyd et al. (1994), one just needs to combine the Lyapunov shaping design technique with an uncertainty description using quadratic constraints. This leads to less conservative results than those derived in Boyd et al. (1994) via the S-procedure. As a paradigm example we choose the $L_2 \rightarrow L_\infty$ gain of the channel $w \rightarrow z$ as a performance measure which has been called generalized H_2 norm (Rotea, 1993). However, the techniques easily extend to the other problems considered in Boyd et al. (1994) and Scherer et al. (1995). Similarly as in Iwasaki et al. (1995) for the robust stabilization problem we derive the corresponding analysis and synthesis results for uncertain systems.

Secondly, in the spirit of Packard (1994), Apkarian and Gahinet (1995), Helmersson (1995), and Scorletti and El Ghaoui (1995), we will fully solve the LPV control problem for the general class of scalings Q, R, S that are simply described by

$$Q < 0, R > 0, \begin{bmatrix} \Delta(\pi) \\ I \end{bmatrix}^T \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} \Delta(\pi) \\ I \end{bmatrix} > 0$$

for all $\pi \in \Pi$, and for the generalized H_2 performance measure. We do not need to assume any additional specific structure on the function $\Delta(\pi)$ (such as linearity) or on the scalings Q, R, S (such as being block-diagonal) as done before. It is not difficult to see that this set of scalings admits a nice description in terms of finitely many LMIs if $\Delta(\Pi)$ is a convex polytope with finitely many extreme points. (Note that this amounts to a specific structure of the image of Π under Δ and not necessarily of the parameter set Π itself.) As an essential new ingredient it will turn out that the controller scheduling function $\Delta_c(\pi)$ cannot be chosen equal to $\Delta(\pi)$ but, even in the simple case (4) and if Π is a box, it has to be taken as a nonlinear function of $\Delta(\pi)$ that can be explicitly constructed.

Throughout the paper system are considered on the time-interval $[0, \infty)$. L_2 denotes $L_2^n[0, \infty)$ (for some n) with norm $\|x\|_2^2 = \int_0^\infty x(t)^T x(t) dt$ and L_∞ denotes $L_\infty^n[0, \infty)$ (for some n) with norm $\|x\|_\infty^2 = \text{ess sup}_{t \geq 0} x(t)^T x(t)$. The function $\text{sym}(X) = X + X^T$ is used to shorten the layout.

2 Analysis for uncertain systems

Suppose that the uncertain system is described as

$$\begin{bmatrix} \dot{x} \\ z \\ z_2 \end{bmatrix} = \begin{bmatrix} A & G & G_2 \\ H & 0 & 0 \\ H_2 & F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} x \\ w \\ w_2 \end{bmatrix}, \quad w_2 = \Delta(t)z_2 \quad (8)$$

with the time-varying (continuous) perturbation $\Delta(t)$. Here $w \rightarrow z$ is the performance channel and $w_2 \rightarrow z_2$ the uncertainty channel. The class of uncertainties is specified through a family of quadratic constraints (Rantzer and Megretski, 1994) defined by a set \mathcal{P} of scalings P that are tacitly assumed to have the structure

$$P = P^T = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}, \quad Q < 0, \quad R > 0$$

with Q/R of the size of w_2/z_2 respectively. Indeed, we just assume that all uncertainties affecting the plant satisfy the quadratic constraints

$$\begin{bmatrix} \Delta(t) \\ I \end{bmatrix}^T P \begin{bmatrix} \Delta(t) \\ I \end{bmatrix} > 0 \quad (9)$$

for all $P \in \mathcal{P}$ and for all $t \geq 0$.

As a typical example we mention *polytopic uncertainty*. Suppose $\Delta(t)$ is known to be contained in a convex polytope with finitely many extreme points:

$$\Delta(t) \in \text{conv}\{\Delta_1, \dots, \Delta_\delta\} \quad \text{for all } t \geq 0.$$

Then \mathcal{P} is simply described as $Q < 0, R > 0$ and

$$\begin{bmatrix} \Delta_j \\ I \end{bmatrix}^T P \begin{bmatrix} \Delta_j \\ I \end{bmatrix} > 0 \quad \text{for all } j = 1, \dots, \delta.$$

Due to $Q < 0$, it is easily seen that (9) indeed holds for all uncertainties in the convex hull. Hence, in this case, the set of scalings \mathcal{P} admits a nice parametrization in terms of finitely many LMIs.

The goal of this section is as follows: Characterize whether (8) is robustly exponentially stable and whether the gain of $L_2 \ni w \rightarrow z \in L_\infty$ is robustly smaller than γ . If both properties hold we say that the system has *robust generalized H_2 performance level γ* .

It is very easy to derive the analogue of the constantly scaled bounded real lemma (Rantzer and Megretski, 1994; Iwasaki, 1993) for the robust generalized H_2 -criterion.

Theorem 1 *If there exist $X > 0$ and $P \in \mathcal{P}$ such that*

$$\begin{bmatrix} A^T X + X A & X G & X G_2 \\ G^T X & -I & 0 \\ G_2^T X & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & H_2^T \\ 0 & F_{21}^T \\ I & F_{22}^T \end{bmatrix} P \begin{bmatrix} 0 & H_2^T \\ 0 & F_{21}^T \\ I & F_{22}^T \end{bmatrix}^T < 0, \\ \begin{bmatrix} X & H^T \\ H & \gamma I \end{bmatrix} > 0, \quad (10)$$

then the system (8) has robust generalized H_2 performance level γ .

To obtain the results in Section 4 it is crucial to reformulate the first inequality in (10) in a symmetric fashion that seems new and that reveals the relation of the general scalings to the standard D, G scalings used in μ -theory. Based on the formulas in the appendix we transform the scalings as

$$P \rightarrow T = \begin{bmatrix} U & W \\ W^T & V \end{bmatrix} = \begin{bmatrix} Q - SR^{-1}S^T & SR^{-1} \\ R^{-1}S^T & -R^{-1} \end{bmatrix} \quad (11)$$

and denote the image of \mathcal{P} under this bijective transformation as \mathcal{T} . Then (9) is equivalent to

$$\begin{bmatrix} U & W \\ W^T & V \end{bmatrix}^{-1} - \begin{bmatrix} 0 & \Delta(t) \\ \Delta(t)^T & 0 \end{bmatrix} < 0 \quad (12)$$

and (10) is equivalent to

$$\begin{bmatrix} A^T X + X A & X G & X G_2 & H_2^T \\ G^T X & -I & 0 & F_{21}^T \\ G_2^T X & 0 & U & F_{22}^T + W \\ H_2 & F_{21} & F_{22} + W^T & V \end{bmatrix} < 0, \\ \begin{bmatrix} X & H^T \\ H & \gamma I \end{bmatrix} > 0. \quad (13)$$

Corollary 2 *Suppose there exists an $X > 0$ and some $\begin{bmatrix} U & W \\ W^T & V \end{bmatrix} \in \mathcal{T}$ with (13). Then the system (8) has a robust generalized H_2 performance level γ .*

Note that the inequality (12) describing the uncertainty is affine in T^{-1} and the inequality (13) is affine in T itself. Hence the proposed scaling transformation has a linearizing effect on these inequalities. Numerical benefits of this transformation for synthesis (Section 3) remain to be explored.

Similar results can be obtained for the following performance criteria (Boyd et al., 1994; Scherer et al., 1995):

- General quadratic constraints

$$\int_0^\infty \begin{bmatrix} w(t) \\ z(t) \end{bmatrix} \begin{bmatrix} Q_p & S_p \\ S_p^T & R_p \end{bmatrix} \begin{bmatrix} w(t) \\ z(t) \end{bmatrix} dt \leq 0$$

with $Q_p \leq 0, R_p \geq 0$ which includes L_2 -gain and dissipativity requirements.

- A robust bound on the H_2 norm. Here the H_2 norm is defined as follows: Let the disturbance w be absent. Moreover, let z^j denote the output of the system for the initial condition

$x(0) = Ge_j$ with the standard unit vectors e_j . Then the squared H_2 norm is defined as

$$\sum_j \int_0^\infty z^j(t)^T z^j(t) dt.$$

- A robust bound on the $L_\infty \ni w \rightarrow z \in L_\infty$ gain of the system.
- Mixed criteria as presented in Boyd et al. (1994), El Ghaoui and Følcher (1996) and Scherer et al. (1995).

We have extended the results of Rantzer and Megretski (1994) and Iwasaki et al. (1995) to robust performance problems that can be expressed in terms of time-invariant quadratic Lyapunov functions. Moreover, although in spirit quite similar to Iwasaki et al. (1995), Boyd et al. (1994), El Ghaoui and Følcher (1996), it generalizes these results in two respects: Firstly, we allow for quite general uncertainty structures not restricted to block-diagonal matrices. Secondly, even if the uncertainty is block-diagonal, we allow for full block scalings P . This avoidance of the channel-wise application of the S-procedure leads to less conservative analysis results.

3 Synthesis for uncertain systems

Let the system be described by (3) with $\Delta(\pi(t))$ replaced by $\Delta(t)$. In the synthesis problem we search for a controller

$$\dot{x}_c = A_c x_c + B_c y, \quad u = C_c x_c + D_c y$$

to achieve robust generalized H_2 performance of level γ for the resulting closed-loop system. Only for notational simplicity we assume that

$$D_c = 0 \text{ and } F = 0, \quad F_{12} = 0, \quad F_{21} = 0, \quad F_2 = 0.$$

Through a suitable change of the controller parameters it is immediate to obtain from Corollary 2 the required synthesis inequalities and a recipe for constructing a controller (Masubuchi et al., 1995; Scherer et al., 1995).

Theorem 3 Suppose there exist X, Y, K, L, M and $\begin{bmatrix} U & W \\ W^T & V \end{bmatrix} \in \mathcal{T}$ satisfying the two inequalities

$$\begin{bmatrix} \text{sym}(XA + LC) & * & * & * & * \\ K^T + A & \text{sym}(AY + BM) & * & * & * \\ (XG + LD)^T & G^T & -I & 0 & 0 \\ (XG_2 + LD_2)^T & G_2^T & 0 & U & W^T \\ H_2 & H_2 Y + E_2 M & 0 & W & V \end{bmatrix} < 0,$$

$$\begin{bmatrix} X & I & H^T \\ I & Y & (HY + EM)^T \\ H & HY + EM & \gamma I \end{bmatrix} > 0.$$

With nonsingular M_1 and M_2 such that $I - XY = M_1 M_2$, the controller

$$C_c = M M_2^{-1}, \quad B_c = M_1^{-1} L$$

$$A_c = M_1^{-1} [K - XAY - M_1 B_c C Y - X B C_c M_2] M_2^{-1}$$

renders the robust generalized H_2 performance level of the closed-loop system smaller than γ .

Contrary to previous results in Iwasaki et al. (1995), we end up with synthesis inequalities that are affine in all variables, including the scalings. Nonconvexity enters the problem via the constraint set \mathcal{T} which does not admit a nice LMI parametrization.

Due to the specific structure of the synthesis LMIs one can easily eliminate the parameters K and L (Scherer, 1995). Let us, instead, briefly turn to the state-feedback problem $C = I, D = 0, D_2 = 0$. Since, for any X , the left-upper block of the first synthesis LMI can be arbitrarily assigned, the inequalities reduce to

$$\begin{bmatrix} \text{sym}(AY + BM) & G & G_2 (H_2 Y + E_2 M)^T \\ G^T & -I & 0 & 0 \\ \hline G_2^T & 0 & U & W^T \\ H_2 Y + E_2 M & 0 & W & V \end{bmatrix} < 0, \quad (14)$$

$$\begin{bmatrix} Y & (HY + EM)^T \\ HY + EM & \gamma I \end{bmatrix} > 0. \quad (15)$$

Note that, in this case, the controller can be chosen as $u = MY^{-1}x$ and is, hence, static.

Dualization (appendix) leads to the inequality

$$\begin{bmatrix} \text{sym}(AY + BM) & G & * \\ G^T & -I & 0 \\ \hline H_2 Y + E_2 M & 0 & 0 \end{bmatrix} + \begin{bmatrix} G_2 & 0 \\ 0 & 0 \\ 0 & I \end{bmatrix} \tilde{P} \begin{bmatrix} G_2 & 0 \\ 0 & 0 \\ 0 & I \end{bmatrix}^T < 0 \quad (16)$$

for $\tilde{P} = \begin{bmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{bmatrix}$ which is equivalent to (14). For polytopic uncertainties, the set of transformed scalings \tilde{P} admits the LMI description $\tilde{Q} > 0, \tilde{R} < 0$,

$$\begin{bmatrix} I \\ \Delta_j^T \end{bmatrix}^T \tilde{P} \begin{bmatrix} I \\ \Delta_j^T \end{bmatrix} > 0 \text{ for all } j = 1, \dots, \delta.$$

Hence using (15) and (16), the state-feedback generalized H_2 control problem for uncertain systems is reduced to solving a genuine LMI problem. This extends Iwasaki (1993) and El Ghaoui and Følcher (1996) to general scales and Iwasaki et al. (1995) to the generalized H_2 -criterion.

4 Synthesis for LPV systems

With a (compact) set Π and a (continuous) function $\Delta : \Pi \rightarrow \mathcal{R}^{k \times l}$, the underlying system is described as (7) where $\pi(t)$ is an arbitrary (continuous) curve with $\pi(t) \in \Pi$. The LPV synthesis problem is posed as follows: Find an LTI controller (2) and a (continuous) function $\Delta_c : \Pi \rightarrow \mathcal{R}^{k_c \times l_c}$ such that the system

$$\begin{bmatrix} \dot{x} \\ z \\ z_e \end{bmatrix} = \begin{bmatrix} A & G & G_e \\ \mathcal{H} & 0 & 0 \\ \mathcal{H}_e & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ w \\ w_e \end{bmatrix},$$

$$w_e = \begin{bmatrix} \Delta(\pi(t)) & 0 \\ 0 & \Delta_c(\pi(t)) \end{bmatrix} z_e$$

that results from (7) controlled with (5) has robust generalized H_2 performance level γ for all $\pi(t) \in \Pi$. We are now ready to formulate the core result of this paper whose proof cannot be included for reasons of space.

Theorem 4 *The next two statements are equivalent.*

1) *There exist a controller (2), a function $\Delta_c : \Pi \rightarrow \mathcal{R}^{k_c \times l_c}$, a scaling $P_e = \begin{bmatrix} Q_e & S_e \\ S_e^T & R_e \end{bmatrix}$, $Q_e < 0$, $R_e > 0$, and a matrix $\mathcal{X} > 0$ such that*

$$\begin{bmatrix} \Delta(\pi) & 0 \\ 0 & \Delta_c(\pi) \\ I & 0 \\ 0 & I \end{bmatrix}^T P_e \begin{bmatrix} \Delta(\pi) & 0 \\ 0 & \Delta_c(\pi) \\ I & 0 \\ 0 & I \end{bmatrix} > 0 \text{ for all } \pi \in \Pi$$

(17)

and

$$\begin{bmatrix} A^T \mathcal{X} + \mathcal{X} A & \mathcal{X} G & \mathcal{X} G_e \\ G^T \mathcal{X} & -I & 0 \\ G_e^T \mathcal{X} & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \mathcal{H}_e^T \\ 0 & 0 \\ I & 0 \end{bmatrix} P_e \begin{bmatrix} 0 & \mathcal{H}_e^T \\ 0 & 0 \\ I & 0 \end{bmatrix}^T < 0,$$

$$\begin{bmatrix} \mathcal{X} & \mathcal{H}^T \\ \mathcal{H} & \gamma I \end{bmatrix} > 0.$$

(18)

2) *There exist matrices X, Y, L, M and scalings $P = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}$, $\tilde{P} = \begin{bmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{bmatrix}$, $Q, \tilde{R} < 0$, $R, \tilde{Q} > 0$, with*

$$\begin{bmatrix} \Delta(\pi) \\ I \end{bmatrix}^T P \begin{bmatrix} \Delta(\pi) \\ I \end{bmatrix} > 0,$$

(19)

$$\begin{bmatrix} I \\ \Delta(\pi)^T \end{bmatrix}^T \tilde{P} \begin{bmatrix} I \\ \Delta(\pi)^T \end{bmatrix} > 0$$

for all $\pi \in \Pi$ and

$$\begin{bmatrix} \text{sym}(XA + LC) & * & * \\ (XG + LD)^T & -I & 0 \\ (XG_2 + LD_2)^T & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & H_2^T \\ 0 & 0 \\ I & 0 \end{bmatrix} P \begin{bmatrix} 0 & H_2^T \\ 0 & 0 \\ I & 0 \end{bmatrix}^T < 0,$$

$$\begin{bmatrix} \text{sym}(AY + BM) & G & * \\ G^T & -I & 0 \\ H_2 Y + E_2 M & 0 & 0 \end{bmatrix} + \begin{bmatrix} G_2 & 0 \\ 0 & 0 \\ 0 & I \end{bmatrix} \tilde{P} \begin{bmatrix} G_2 & 0 \\ 0 & 0 \\ 0 & I \end{bmatrix}^T < 0,$$

$$\begin{bmatrix} X & I & H^T \\ I & Y & (HY + EM)^T \\ H & HY + EM & \gamma I \end{bmatrix} > 0.$$

Let us now comment on how to apply this result for solving the LPV control problem. Trivially, the set of all scalings with (19) is convex. In a first step one has to solve the three synthesis LMIs for X, Y, L, M, P, \tilde{P} over these convex constraints. This is indeed possible with standard algorithms if $\Delta(\Pi)$ is a polytope since then the sets of scalings P, \tilde{P} satisfying (19) admit descriptions in terms of finitely many LMIs (Section 2). Note that one can even directly minimize the performance level γ (Gahinet et al., 1994). In particular, this is assured if Π itself is a polytope and $\Delta(\pi)$ is affine in π . A special case is (4) where π is contained in a box (containing 0). Compared to Helmersson (1995) and Scorletti and El Ghaoui (1995), we even then allow for a larger class of not necessarily block-diagonal scalings. This reduces conservatism at the expense of increasing the number of synthesis LMIs due to the implicit description of the scalings.

Hence one can determine the achievable performance level, and it remains to construct a suitable controller. For this purpose we assume that the synthesis LMIs have been solved for X, Y, L, M, P, \tilde{P} . Due to $Q < 0$ and $R > 0$, P is nonsingular, and the same is true of \tilde{P} . We can as well assume w.l.o.g. that $P - \tilde{P}^{-1}$ is nonsingular. Let

$$k_c \text{ and } l_c$$

be the number of negative/positive eigenvalues of $P - \tilde{P}^{-1}$ respectively. It is possible to prove that one can construct an extension

$$P_e = \begin{bmatrix} Q_e & S_e \\ S_e^T & R_e \end{bmatrix} = \begin{bmatrix} Q & Q_{12} & S & S_{12} \\ Q_{21} & Q_{22} & S_{21} & S_{22} \\ S^T & S_{21}^T & R & R_{12} \\ S_{21}^T & S_{22}^T & R_{21} & R_{22} \end{bmatrix}$$

with dimension $(k + k_c) \times (l + l_c)$ that has the three properties $\begin{bmatrix} Q & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} < 0$, $\begin{bmatrix} R & R_{12} \\ R_{21} & R_{22} \end{bmatrix} > 0$, and

$$P_e^{-1} = \begin{bmatrix} -\tilde{Q} & * & \tilde{S} & * \\ * & * & * & * \\ \tilde{S}^T & * & -\tilde{R} & * \\ * & * & * & * \end{bmatrix}.$$

(The proof of this fact is nontrivial but constructive. It shows that k_c and l_c are indeed minimal;

there is no smaller extension which has all the desired properties. If the scales are block-diagonal, this extension can be performed block-wise and is easier.)

Due to (19), it is then possible to determine a function $\Delta_c(\pi)$ that satisfies (17). With

$$\begin{bmatrix} U_{11} & U_{12} & W_{11} & W_{12} \\ U_{21} & U_{22} & W_{21} & W_{22} \\ W_{11}^T & W_{21}^T & V_{11} & V_{12} \\ W_{21}^T & W_{22}^T & V_{21} & V_{22} \end{bmatrix} = \begin{bmatrix} -Q_e^{-1} & Q_e^{-1} S_e \\ S_e^T Q_e^{-1} R_e - S_e^T Q_e^{-1} S_e \end{bmatrix}$$

(partitioned in the same way as P_e), the function

$$\begin{bmatrix} U_{21} & W_{21} \end{bmatrix} \begin{bmatrix} U_{11} & W_{11} + \Delta(\pi) \\ * & V_{11} \end{bmatrix}^{-1} \begin{bmatrix} W_{12} \\ V_{12} \end{bmatrix} - W_{22}$$

is a possible choice. Note that, in general, and contrary to what is done in Packard (1994), Apkarian and Gahinet (1995), Helmersson (1995), Scorletti and El Ghaoui (1995), $\Delta_c(\pi)$ differs from $\Delta(\pi)$, even for the standard structure (4)! Once the scaling P_e is constructed it remains to apply Theorem 3 for the system (7) and the transformed version T_e of the fixed scaling P_e to determine A_c, B_c, C_c by solving an LMI.

This controller is guaranteed to satisfy (18) for some $\mathcal{X} > 0$. Since P_e satisfies (17), we can finally infer from Theorem 1 that the closed-loop system has a robust generalized H_2 performance level γ - the LPV problem is solved.

5 Conclusions

The analysis and synthesis results for achieving robust generalized H_2 performance with general scalings are pretty straightforward extensions of well-established ideas. In this paper we provide a full solution of the corresponding problem for LPV systems what extends previous specialized results in a nontrivial fashion. As a crucial and structurally interesting step one has to schedule the controller with a function that is generally different from, and actually a nonlinear function of, the parameters of the plant.

References

Apkarian, P. and P. Gahinet (1995). A convex characterization of gain-scheduled H_∞ -controllers, *IEEE Trans. Automatic Control*, 40, 853-864.
 Apkarian, P., P. Gahinet, and G. Becker (1994). Self-scheduled \mathcal{H}_∞ control of linear parameter-varying systems, *Proc. Amer. Contr. Conf.*, pp. 856-860.

Becker, G., A. Packard, D. Philbrick, and G. Balas (1993). Control of parametrically-dependent linear systems: A single quadratic Lyapunov approach, *Proc. Amer. Contr. Conf.*, San Francisco, CA, pp. 2795-2799.
 Becker, G. and A. Packard (1994). Robust performance of linear parametrically varying systems using parametrically dependent linear feedback, *Systems Control Lett.*, 23, 205-215.
 Boyd, S., L. El Ghaoui, E. Feron, and V. Balakrishnan (1994). *Linear Matrix Inequalities in System and Control Theory*. SIAM Studies in Applied Mathematics 15, SIAM, Philadelphia.
 Fan, M., A. Tits and J.C. Doyle (1991). Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics. *IEEE Trans. Automat. Control*, 36, 25-38.
 Gahinet, P., A. Nemirovski, A. Laub, and M. Chilali (1994). *LMI Control Toolbox*. The Mathworks, Inc.
 El Ghaoui, L. and J.P. Folcher (1996). Multiobjective robust control of LTI systems subject to unstructured perturbations, *Systems Control Lett.*, 28, 23-30.
 Helmersson, A. (1995). *Methods for robust gain-scheduling*. PhD thesis, Linköping University, Sweden.
 Iwasaki, T. (1993). *A Unified Matrix Inequality Approach to Linear Control Design*. Ph.D. Thesis, Purdue University, USA.
 Iwasaki, T., S. Hara, and T. Asai (1995). Well-posedness theorem: a classification of LMI/BMI-reducible robust control problems, *Preprint*.
 Lind, R., G. Balas, and A. Packard (1995). Optimal scaled H_∞ full information synthesis with real parametric uncertainty, *Proc. Amer. Contr. Conf.*, pp. 3463-3467.
 Masubuchi, I., N. Suda, and A. Ohara (1995). LMI-based controller synthesis: a unified formulation and solution, *Proc. Amer. Control Conf.*, pp. 3473-3477.
 Packard, A. (1994). Gain scheduling via linear fractional transformations, *System and Control Letters*, 22, 79-92.
 Rantzer, A. and A. Megretski (1994). System analysis via integral quadratic constraints, *Proc. IEEE Conf. Decision Contr.*, pp. 3062-3067.
 Rotea, M.A. (1993). The generalized \mathcal{H}_2 control, *Automatica*, 29, 373-385.
 Scherer, C.W. (1995). Mixed H_2/H_∞ control. In: A. Isidori (Ed.), *Trends in Control, A European Perspective*, Springer-Verlag, Berlin, pp. 173-216.
 Scherer, C.W., P. Gahinet, and M. Chilali (1995). Multi-objective output-feedback control. *Preprint*.

there is no smaller extension, which has all the desired properties. If the zeros are block-diagonal, this extension can be performed block-wise and it is easier.)

Due to (19), it is then possible to determine a function $\Delta_1(s)$ that satisfies (17). With

$$\begin{bmatrix} U_{11} & U_{12} & W_{11} & W_{12} \\ U_{21} & U_{22} & W_{21} & W_{22} \\ V_{11} & V_{12} & V_{21} & V_{22} \\ W_{11} & W_{21} & V_{11} & V_{21} \end{bmatrix} = \begin{bmatrix} -Q_1^{-1} & Q_1^{-1}x \\ Q_1^{-1}Q_2^{-1}R & -Q_1^{-1}x \end{bmatrix}$$

(partitioned in the same way as P_2), the function

$$[U_{21} \quad W_{21}] \begin{bmatrix} W_{11} & W_{12} + \Delta_1(s) \\ V_{11} & V_{12} \end{bmatrix}^{-1} \begin{bmatrix} W_{11} \\ V_{11} \end{bmatrix} = W_{22}$$

is a possible choice. Note that, in general, and contrary to what is done in Pappas (1984), Spharim and Gahleitner (1988), Holmstrom (1992), Sridhar and El Ghoul (1993), $\Delta_1(s)$ differs from $\Delta(s)$, even for the standard structure (4). Once the scaling P_2 is constructed it remains to apply Theorem 3 for the system (1) and the transformed version P_2 of the feed scaling P , to determine A_1, B_1, C_1 , by solving an LMI.

This controller is guaranteed to satisfy (3) for most $\alpha > 0$. Since P_2 satisfies (17), by our final step from Theorem 1 that the closed-loop system has a robust gain and H_2 performance level γ close to γ^* if α is small.

5. Conclusions

The analysis and synthesis results for achieving robust gain-scheduled H_2 performance with general scaling are given. State-invariant extensions of well-established ideas. In this paper we provide a full solution of the corresponding problem for LTV systems, where extended previous specified results in a convexified fashion. As a control and structure learning step one has to choose the controller with a function that is generally different from, and actually a nonlinear function of, the parameters of the plant.

References

Apharim, P. and P. Gahleitner (1985). A convex characterization of gain-scheduled H_2 controllers. *IEEE Trans. Automatic Control*, 30, 853-854.

Apharim, P., P. Gahleitner, and G. Neuen (1986). Self-scheduled H_2 control of linear parameter-varying systems. *Proc. Amer. Control Conf.*, pp. 480-483.

Chen, Y. (1995). *Robust Control: A Practical Approach*. Wiley, New York.

Chen, Y. and J.P. Young (1995). Robust control of linear parameter-varying systems. *IEEE Trans. Automatic Control*, 40, 1000-1004.

Chen, Y., J.P. Young, and J. Spong (1995). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 40, 1005-1009.

Chen, Y., J.P. Young, and J. Spong (1996). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 41, 1010-1014.

Chen, Y., J.P. Young, and J. Spong (1997). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 42, 1015-1019.

Chen, Y., J.P. Young, and J. Spong (1998). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 43, 1020-1024.

Chen, Y., J.P. Young, and J. Spong (1999). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 44, 1025-1029.

Chen, Y., J.P. Young, and J. Spong (2000). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 45, 1030-1034.

Chen, Y., J.P. Young, and J. Spong (2001). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 46, 1035-1039.

Chen, Y., J.P. Young, and J. Spong (2002). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 47, 1040-1044.

Chen, Y., J.P. Young, and J. Spong (2003). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 48, 1045-1049.

Chen, Y., J.P. Young, and J. Spong (2004). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 49, 1050-1054.

Chen, Y., J.P. Young, and J. Spong (2005). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 50, 1055-1059.

Chen, Y., J.P. Young, and J. Spong (2006). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 51, 1060-1064.

Chen, Y., J.P. Young, and J. Spong (2007). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 52, 1065-1069.

Chen, Y., J.P. Young, and J. Spong (2008). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 53, 1070-1074.

Chen, Y., J.P. Young, and J. Spong (2009). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 54, 1075-1079.

Chen, Y., J.P. Young, and J. Spong (2010). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 55, 1080-1084.

Chen, Y., J.P. Young, and J. Spong (2011). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 56, 1085-1089.

Chen, Y., J.P. Young, and J. Spong (2012). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 57, 1090-1094.

Chen, Y., J.P. Young, and J. Spong (2013). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 58, 1095-1099.

Chen, Y., J.P. Young, and J. Spong (2014). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 59, 1100-1104.

Chen, Y., J.P. Young, and J. Spong (2015). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 60, 1105-1109.

Chen, Y., J.P. Young, and J. Spong (2016). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 61, 1110-1114.

Chen, Y., J.P. Young, and J. Spong (2017). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 62, 1115-1119.

Chen, Y., J.P. Young, and J. Spong (2018). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 63, 1120-1124.

Chen, Y., J.P. Young, and J. Spong (2019). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 64, 1125-1129.

Chen, Y., J.P. Young, and J. Spong (2020). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 65, 1130-1134.

Chen, Y., J.P. Young, and J. Spong (2021). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 66, 1135-1139.

Chen, Y., J.P. Young, and J. Spong (2022). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 67, 1140-1144.

Chen, Y., J.P. Young, and J. Spong (2023). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 68, 1145-1149.

Chen, Y., J.P. Young, and J. Spong (2024). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 69, 1150-1154.

Chen, Y., J.P. Young, and J. Spong (2025). Robust control of linear parameter-varying systems with state-invariant extensions. *IEEE Trans. Automatic Control*, 70, 1155-1159.

Robust linear parametrically varying flight control system design with bounded rates

Dehlia M.C. Willemsen, Samir Bennani[†] and Carsten W. Scherer[‡]

*Faculty of Aerospace Engineering, Stability and Control Group
Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands
E-mail: s.bennani@lr.tudelft.nl*

Abstract. This paper discusses an approach to the control of Linear Parametrically Varying (LPV) systems that can take the rate of parameter variations into account and also guarantees robustness against parametric and dynamic uncertainties. To illustrate the technique we consider a missile control problem that has been extensively studied in the literature. For this highly nonlinear model, the objective is to design a controller with guaranteed performance robustness over a given operating range.

Keywords. H_∞ control, linear parametrically varying systems, integral quadratic constraints, robust LPV control, linear matrix inequalities.

1 Introduction

The classical approach to gain scheduling relies on the interpolation of controllers designed for frozen parameters as, e.g., the operating conditions. This procedure, even if seemingly working well in practice, does not take the time-variations of the involved parameters into account. In particular, one cannot provide a priori stability and performance guarantees, as shown in Athans and Shamma (1992) for linear systems that depend on a time-varying parameter. Such systems are called linear parametrically varying (LPV). In the early nineties (Packard, 1994; Apkarian and Gahinet, 1995) it has been observed that the techniques of robust control can be generalized to arriving at a systematic design procedure for such LPV systems. These solutions, however, lead to the desired performance guarantees even if the rate of change of the parameters is unbounded. A refined approach which will be pursued in this paper takes bounds on this rate of change into account (Wu, 1995; Apkarian and Adams, 1995; Willemsen, 1996; Scherer, 1995). In ad-

dition, our technique allows to incorporate robustness properties into the design procedure.

The paper is structured as follows. We first provide the description of the uncertain LPV system. Then it is shown how the structural knowledge about the uncertainties is reflected in suitable classes of so-called multipliers or scalings. This leads to the analysis characterization of stability and robustness of the LPV systems in terms of a scaled differential Bounded Real Lemma. For LPV controller synthesis, we apply a linearizing transformation of the controller parameters (Masubuchi et al., 1995; Scherer, 1995). Introducing basis functions and gridding the parameter set will result, for fixed scalings, in finitely many linear matrix inequalities that can be readily solved (Gahinet et al., 1994). If optimizing as well over the scalings, we have to resort to a D - K -like iteration (Balas et al., 1993).

The theory is illustrated on a missile benchmark problem as studied in Rugh et al. (1993) and Wu (1995). In contrast to Wu (1995), which considers the nominal performance LPV problem, we also address robust performance issues. Starting from a given nonlinear model, we obtain an uncertain LPV representation accessible for design. Then the design specifications are translated into suitable

[†]Author to whom correspondence should be addressed.

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

weighting functions as in the H_∞ -approach. We end up with an interconnection structure and perform the iteration that is comparable to the D - K procedure in μ -synthesis (Balas et al., 1993). Finally, we validate the robust LPV controller by nonlinear simulations.

2 LPV design

2.1 LPV systems

The uncertain LPV system is described by

$$\begin{aligned} \dot{x} &= A(p)x + G(p)w + B(p)u \\ z &= H(p)x + F(p)w + E(p)u \\ y &= C(p)x + D(p)w \end{aligned} \quad (1)$$

where, with a suitable partition of the signals $w = [w_0^T, w_1^T, \dots, w_k^T]^T$ and $z = [z_0^T, z_1^T, \dots, z_k^T]^T$, the uncertainty enters as

$$w_i = \Delta_i z_i, \quad i = 1, \dots, k, \quad (2)$$

and $w_0 \mapsto z_0$ is the channel to describe the performance specification (Figure 1). The parameter $p(t)$ and its rate of variation $\dot{p}(t)$ are assumed to be contained in the a priori given compact sets P and P_r respectively. For controller design the parameter

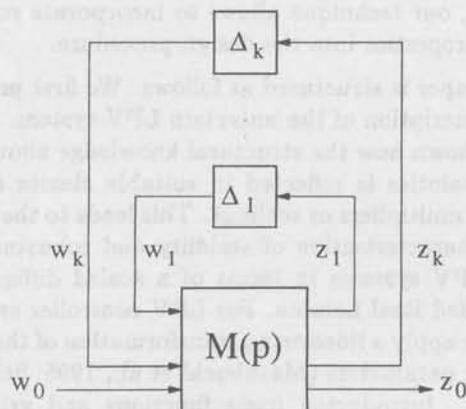


Fig. 1: LPV system with uncertainty and a performance channel

$p(t)$ is assumed to be on-line measurable. Hence, LPV controllers take the form

$$\begin{aligned} \dot{x}_c &= A_c(p)x_c + B_c(p)y \\ u &= C_c(p)x_c + D_c(p)y \end{aligned} \quad (3)$$

such that the resulting closed-loop system is described by

$$\begin{aligned} \dot{\xi} &= \mathcal{A}(p)\xi + \mathcal{G}(p)w \\ z &= \mathcal{H}(p)\xi + \mathcal{F}(p)w \end{aligned} \quad (4)$$

together with (2), where

$$\begin{aligned} \mathcal{A}(p) &= \begin{bmatrix} A(p) + B(p)D_c(p)C(p) & B(p)C_c(p) \\ B_c(p)C(p) & A_c(p) \end{bmatrix}, \\ \mathcal{G}(p) &= \begin{bmatrix} G(p) + B(p)D_c(p)D(p) \\ B_c(p)D(p) \end{bmatrix}, \\ \mathcal{H}(p) &= [H(p) + E(p)D_c(p)C(p) \quad E(p)C_c(p)], \\ \mathcal{F}(p) &= [F(p) + E(p)D_c(p)D(p)]. \end{aligned}$$

2.2 Analysis of uncertain LPV systems

In this section we will provide an analysis result that characterizes robust stability and robust performance for the uncertain LPV system (4), (2). For that purpose we need to introduce scalings that characterize the nature of the uncertainties Δ_i affecting the plant in terms of integral quadratic constraints (IQCs) (Rantzer and Megretski, 1994). Hence, for each channel, we define a collection of matrices Q_i, S_i, R_i such that the IQC

$$\int_0^T \begin{bmatrix} w_i \\ z_i \end{bmatrix}^T \begin{bmatrix} Q_i & S_i^T \\ S_i & R_i \end{bmatrix} \begin{bmatrix} w_i \\ z_i \end{bmatrix} dt \geq 0 \quad (5)$$

holds for $T \geq 0$ and for all signals w_i, z_i of finite energy that are related by $w_i = \Delta_i z_i$. As examples, we mention

- time-varying parametric uncertainties $w_i(t) = [\delta_i(t)I]z_i(t)$, $|\delta_i(t)| \leq 1$:

$$Q_i < 0 \quad R_i = -Q_i, \quad S_i + S_i^T = 0.$$

- dynamic uncertainty $\Delta_j : L_2[0, \infty) \mapsto L_2[0, \infty)$ with gain not larger than 1:

$$Q_i = q_i I < 0, \quad R_i = -Q_i, \quad S_i = 0.$$

We take the L_2 -gain of the channel $w_0 \rightarrow z_0$ as a measure for performance. The L_2 -gain of this channel is bounded by the value γ if the IQC (5) holds with the fixed scaling

$$Q_0 = -\frac{1}{\gamma}I, \quad R_0 = \gamma I, \quad S_0 = 0.$$

Finally, we collect the scalings into block-diagonal matrices as $Q = \text{diag}(Q_0, Q_1, \dots, Q_k)$, $R = \text{diag}(R_0, R_1, \dots, R_k)$, and $S = \text{diag}(S_0, S_1, \dots, S_k)$.

Now we are ready to provide the characterization of robust stability and robust performance in terms of the solvability of a so-called scaled differential Bounded Real Lemma whose proof is straightforward (Helmersson, 1995; Scorletti and El Ghaoui, 1995; Scherer, 1995; Willemsen, 1996). For notational convenience we define $\text{sy}(M) := M + M^T$.

Theorem 1 Suppose there exist smooth and bounded functions $\mathcal{X}(p)$, $Q(p)$, $R(p)$, $S(p)$ on P such that

$$\begin{aligned} & \mathcal{X}(p) > 0 \\ & \begin{bmatrix} \text{sy}(\mathcal{X}(p)\mathcal{A}(p)) + \mathcal{X}'(p, \dot{p}) & \mathcal{X}(p)\mathcal{G}(p) + \mathcal{H}^T(p)S(p) & * \\ \mathcal{G}^T(p)\mathcal{X}(p) + S^T(p)\mathcal{H}(p) & \text{sy}(\mathcal{F}^T(p)S(p)) + R(p) & * \\ Q(p)\mathcal{H}(p) & Q(p)\mathcal{F}(p) & Q(p) \end{bmatrix} < 0 \end{aligned} \quad (6)$$

holds for all $p \in P$ and $\dot{p} \in P_r$. Then, for all parameter curves $(p(t), \dot{p}(t)) \in P \times P_r$ and for all uncertainties (2), the system (4) remains stable and the L_2 -gain of the performance channel is bounded by γ .

Here $\mathcal{X}'(p, \dot{p})$ is defined as

$$\mathcal{X}'(p, \dot{p}) = \sum_{j=1}^m \frac{\partial \mathcal{X}}{\partial p_j}(p) \dot{p}_j.$$

Hence we have to find a parameter dependent Lyapunov function and parameter dependent scalings to satisfy a differential linear matrix inequality (Apkarian and Addams, 1995; Scherer, 1995). This generalizes the idea of using a constant Lyapunov function (Becker, 1993; Apkarian and Gahinet, 1995) and constant scalings for arbitrarily fast varying parameters.

To solve the inequalities numerically we choose continuously differentiable functions $f_1(p) \dots f_l(p)$ and search for the coefficients in the expansion

$$[\mathcal{X}(p) \quad Q(p) \quad R(p) \quad S(p)] = \sum_{j=1}^l f_j(p) [\mathcal{X}_j \quad Q_j \quad R_j \quad S_j].$$

The resulting infinitely many LMIs are reduced to finitely many inequalities by picking a finite number of points in P and P_r . If P_r is described as a convex combination of finitely many vertices, it suffices to choose the extreme points since the parameter \dot{p} appears linearly in (6) (Wu, 1995; Apkarian and Addams, 1995).

2.3 Controller synthesis

The synthesis problem consists of designing a controller (3) that minimizes the robust performance level γ as characterized in Theorem 1. However, the inequalities (6) are not linear in all the unknowns, the Lyapunov function, the scalings, and the controller parameters. It has been shown in Masubuchi et al. (1995) and Scherer et al. (1995) how the inequalities can be linearized, for fixed scalings, by a suitable nonlinear transformation of the controller parameters as follows: If denoting the first

block rows of $\mathcal{X}(p)$ and $\mathcal{X}(p)^{-1}$ as $[X(p) \ U(p)]$ and $[Y(p) \ V(p)]$ respectively, introduce the new controller parameters

$$\begin{aligned} K(p, \dot{p}) &= X(p)[A(p) + B(p)D_c(p)C(p)]Y(p) + \\ &+ U(p)B_c(p)C(p)Y(p) + X(p)B(p)C_c(p)V^T(p) + \\ &+ U(p)A_c(p)V^T(p) + X'(p, \dot{p})Y(p) + U'(p, \dot{p})V^T(p) \\ L(p) &= X(p)B(p)D_c(p) + U(p)B_c(p) \\ M(p) &= D_c(p)C(p)Y(p) + C_c(p)V^T(p) \\ N(p) &= D_c(p) \end{aligned} \quad (7)$$

and transform the blocks in (6) as

$$\mathcal{X}\mathcal{A} \rightarrow \begin{bmatrix} XA + LC & K \\ A + BNC & AY + BM \end{bmatrix}$$

$$\mathcal{X}\mathcal{G} \rightarrow \begin{bmatrix} XG + LD \\ G + BND \end{bmatrix}, \quad \mathcal{H}^T \rightarrow \begin{bmatrix} (H + ENC)^T \\ (HY + EM)^T \end{bmatrix}$$

$$\mathcal{X} \rightarrow \begin{bmatrix} X & I \\ I & Y \end{bmatrix}, \quad \mathcal{X}' \rightarrow \begin{bmatrix} X' & 0 \\ 0 & -Y' \end{bmatrix}$$

where we dropped the dependence on p and \dot{p} .

As explained for analysis, we can introduce basis functions and grid the parameter set to end up with finitely many inequalities. However, the resulting inequalities are still nonlinear in the new variables and the scalings together. Hence we have to resort to a D - K -like iteration scheme that proceeds as follows: Start with the uncertainty scalings $Q_i(p) = -I$, $R_i(p) = I$, $S_i(p) = 0$ and iterate the following two steps until the performance level cannot be improved:

1. Fix the scalings and minimize γ over $X(p)$, $Y(p)$ and the transformed controller parameters $K(p, \dot{p})$, $L(p)$, $M(p)$, $N(p)$.
2. Fix the controller parameters $K(p, \dot{p})$, $L(p)$, $M(p)$, $N(p)$ and minimize the performance level γ over $\mathcal{X}(p)$ and the uncertainty scalings $Q_i(p)$, $R_i(p)$, $S_i(p)$ as described for analysis.

Suppose the iteration stops with $X(p)$, $Y(p)$, $K(p, \dot{p})$, $L(p)$, $M(p)$, $N(p)$. Then one simply needs to choose nonsingular smooth and bounded functions $U(p)$, $V(p)$ satisfying

$$Y(p)X(p) + V(p)U^T(p) = I \quad (8)$$

in order to calculate the controller by solving (7) for $A_c(p, \dot{p})$, $B_c(p)$, $C_c(p)$, $D_c(p)$.

Since $A_c(p, \dot{p})$ depends on \dot{p} , one needs to measure not only the parameter value $p(t)$ itself but also its

rate of variation $\dot{p}(t)$ to implement the resulting controller. To avoid this undesired structure, we choose $K(p)$ independent of \dot{p} and either $X(p)$ or $Y(p)$ independent of p . Exploiting the freedom in the choice of $U(p)$ and $V(p)$ allows to construct a controller that depends on p only:

- If $X(p)$ is parameter dependent and Y is constant, choose $U(p) = I - X(p)Y$ and $V = I$. Taking derivatives in (8) reveals

$$YX'(p, \dot{p}) + VU'(p, \dot{p})^T = 0 \quad (9)$$

such that the terms in (7) that depend on \dot{p} indeed drop out.

- If X is constant and $Y(p)$ is parameter dependent, choose $U = I$ and $V(p) = I - Y(p)X$. This implies $X'(p, \dot{p}) = 0$, $U'(p, \dot{p}) = 0$ such that, again, the variable \dot{p} disappears in (7).

Note that this restriction to constant $X(p)$ or $Y(p)$ certainly introduces conservatism, with the benefit of a simpler controller implementation.

To speed up the computation, we finally remark that we performed all calculations after eliminating the transformed controller parameters along standard lines by using the projection lemma and by fixing $S(p) = 0$ (Apkarian and Addams, 1995).

3 The missile control problem

For the application we have chosen a missile benchmark problem that has been extensively studied in Packard and Balas (1992), Rugh et al. (1993), Wu (1995), Helmersson (1995) and is particularly suited for addressing gain scheduling as well as robustness issues. The problem is to design a longitudinal autopilot for a tail-fin controlled missile providing normal acceleration tracking over a large range of speed and angle-of-attack. In order to arrive at a design model (in section 3.3), the exact problem specifications are first given in section 3.2. These are based on the missile model which is defined in section 3.1.

3.1 The missile model

The non-linear state equations of the missile are

$$\begin{aligned} \dot{\alpha} &= f_1(\alpha, q, \delta, M) = \frac{\cos(\alpha)^2}{mu} F_z(\alpha, \delta, M) + q(10) \\ \dot{q} &= f_2(\alpha, q, \delta, M) = \frac{M_y}{I_y}(\alpha, \delta, M) \quad (11) \end{aligned}$$

with

α	angle-of-attack (rad)
q	pitch rate (rad/s)
F_z	$C_n(\alpha, \delta, M)0.7p_0M^2S$ (lbs)
M_y	$C_m(\alpha, \delta, M)0.7p_0M^2Sd$ (ft - lbs)
δ	tail fin deflection (rad)
p_0	973.3 lb/ft ² (static pressure at 20,000 ft)
S	0.44 ft ² (reference area)
d	0.75 ft (diameter)
m	13.98 slugs (mass of missile)
u	$V \cos(\alpha)$ ft/s (speed along missile center line)
V	Mss ft/s (velocity of the missile)
M	2 - 4 (Mach number of the missile)
ss	1036.4 ft/s (speed of sound at 20,000 ft)
I_y	182.5 slug - ft ² (pitch moment of inertia)
g	32.2 ft/s ² (acceleration due to gravity)
n_z	normal acceleration of the missile (per g).

The aerodynamic nonlinearity and parameter dependence in the missile model are reflected in the normal force and moment coefficients $C_n(\alpha, \delta, M)$ and $C_m(\alpha, \delta, M)$ respectively. Taking the missile symmetry into account it suffices to consider the positive values of the angle-of-attack. The aerodynamic coefficients are then given by

$$\begin{aligned} C_n(\alpha, \delta, M) &= a_n\alpha^3 + b_n\alpha^2 + c_n(2 + \frac{M}{3})\alpha + d_n\delta \\ C_m(\alpha, \delta, M) &= a_m\alpha^3 + b_m\alpha^2 - c_m(7 - \frac{8M}{3})\alpha + d_m\delta \end{aligned}$$

where the polynomial coefficients are

$$\begin{aligned} a_n &= +0.000103 \text{ deg}^{-3} & a_m &= +0.000215 \text{ deg}^{-3} \\ b_n &= -0.009450 \text{ deg}^{-2} & b_m &= -0.019500 \text{ deg}^{-2} \\ c_n &= -0.169600 \text{ deg}^{-1} & c_m &= +0.051000 \text{ deg}^{-1} \\ d_n &= -0.034000 \text{ deg}^{-1} & d_m &= -0.206000 \text{ deg}^{-1}. \end{aligned}$$

These coefficients are valid for the missile traveling between Mach = 2 and Mach = 4 at an altitude of 20,000 ft. Typical maneuvers for this missile result in angle-of-attack values ranging between -20 and +20 degrees. Hence the approximation $\cos(\alpha) \approx 1$ is legitimate. Then (10) simplifies to

$$\dot{\alpha} = \frac{F_z(\alpha, \delta, M)}{mV} + q. \quad (12)$$

One way to obtain an LPV model for the missile is to parametrize the set of all equilibrium models. For any angle-of-attack $\alpha \in [0, 20]$ and Mach number $M \in [2, 4]$, the fin deflection and pitch rate

$$\begin{aligned} \delta(\alpha, M) &= -\frac{1}{d_m} \left[a_m\alpha^3 + b_m\alpha^2 - c_m(7 - \frac{8M}{3})\alpha \right] \\ q(\alpha, M) &= -\frac{F_z(\alpha, \delta, M)}{mV} \end{aligned}$$

lead to an equilibrium of (10)-(11). The specific normal force n_z is measured by an accelerometer

placed at the center of gravity of the missile. It is defined as $n_z = \frac{F_x}{W}$ where $W = mg$. For convenience we use the shorthands $K_\alpha = \frac{0.7p_0 S}{m s s}$, $K_q = \frac{0.7p_0 S d}{I_y}$, $K_n = \frac{\pi}{180} \frac{0.7p_0 S}{W}$. The Jacobi linearization of the missile dynamics is then given as

$$\begin{aligned}\dot{x} &= A(\alpha, M)x + B(\alpha, M)u \\ y &= C(\alpha, M)x + D(\alpha, M)u\end{aligned}$$

where

$$\begin{aligned}A(\alpha, M) &= \begin{pmatrix} K_\alpha M \partial C_n(\alpha, 0, M) / \partial \alpha & 1 \\ K_q M^2 \partial C_m(\alpha, 0, M) / \partial \alpha & 0 \end{pmatrix} \\ B(\alpha, M) &= \begin{pmatrix} K_\alpha M \partial C_n(0, \delta, M) / \partial \delta \\ K_q M^2 \partial C_m(0, \delta, M) / \partial \delta \end{pmatrix} \\ C(\alpha, M) &= \begin{pmatrix} 0 & 1 \\ K_n M^2 \partial C_n(\alpha, 0, M) / \partial \alpha & 0 \end{pmatrix} \\ D(\alpha, M) &= \begin{pmatrix} 0 \\ K_n M^2 \partial C_n(0, \delta, M) / \partial \alpha \end{pmatrix}.\end{aligned}$$

We end up with a family of linearized system that are parametrized by $p = (\alpha, M)$. For a particular parameter value p in the allowable parameter set, the LPV dynamics are called frozen and reflect a local linearization of the missile dynamics.

3.2 The uncertainty description and performance specifications

The specifications to be achieved by the controller have to hold over the whole Mach range [2, 4]. Therefore, the system should globally provide normal acceleration command tracking features, with rise-time not greater than 0.35 s, overshoot not greater than 10 %, and steady state error not greater than 1 %. The measurements available for control are the normal acceleration n_z , the pitch rate q and the Mach number M . During a maneuver, the angle-of-attack should satisfy $|\alpha| \leq 20$ degrees while the tail-fin deflection rate should not exceed 25 deg /s per commanded g-level.

As very strong simplifications in the missile modeling have been made, we take the robustness issue originating from the uncertainty in the aerodynamic coefficients C_n and C_m into account. The uncertainty levels considered are $\Delta C_n = \pm 10$ % and $\Delta C_m = \pm 25$ %.

The controller provides fin commands δ_c that are processed through second order actuator dynamics given by $G_{act}(s) = \frac{\omega_a^2}{s^2 + 2\zeta\omega_a s + \omega_a^2}$, with natural frequency $\omega = 150$ rad/s and damping $\zeta = 0.7$. To avoid exciting unmodeled high frequency dynamics, the multiplicative input uncertainty weighting $W_{in} = 1.5 \frac{s+2}{s+80}$ is placed at the actuator.

3.3 Control strategy

To realize the specifications over the prescribed operating range, the missile dynamics are reformulated into an uncertain parameter varying system representation as used in section 2.1. The LPV system with parameter $p = (\alpha, M)$ has now an uncertain part arising from the perturbations in the aerodynamic coefficients. A further strategy is to view the angle-of-attack in the parameter vector p as uncertain. Hence, the angle-of-attack α and the uncertainties in the aerodynamic coefficients C_m and C_n are pulled out of the system and rescaled to $[-1, +1]$. The resulting uncertainty structure in the missile dynamics is then $\Delta_u = \text{diag}(\delta_\alpha I_2, \Delta_{C_n}, \Delta_{C_m})$.

The control architecture for the missile problem is depicted in figure 2. The tracking specification has been translated into an ideal acceleration model that the closed loop system should match. The ideal model comes from Wu (1995) and is $W_{id}(s) := \frac{144(-0.05s+1)}{s^2+19.2s+144}$ for which the allowable error is weighted as $W_{perf}(s) := \frac{0.5s+17.321}{s+0.0577}$. The low frequency gain of W_{perf} is 300 to bound the tracking error by 0.33 %. The high frequency gain is chosen to be 0.5 in order to limit the overshoot to be less than 5 %. To reflect the tail-fin deflection and deflection rate limits of 20 degrees and 25 degrees/s per g respectively, the filters $W_\delta = \frac{1}{20}$ and $W_{\dot{\delta}} = \frac{1}{25}$ have been chosen accordingly. Finally, noise filters $W_{n1} = 0.001$, and $W_{n2} = 0.001$ are used to reflect the measurement imperfections in pitch rate and normal acceleration.

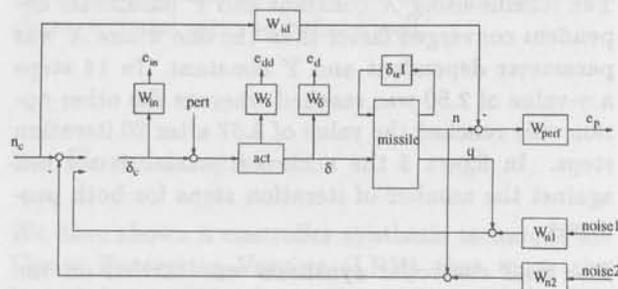


Fig. 2: The controller synthesis interconnection structure

4 Results

4.1 The design of the LPV controller

In this section we will use the synthesis LMIs (6) as derived in section 2. Solving the LMIs is done via basis functions and gridding of the parameter space. For the missile control problem, the Mach

number M is the remaining parameter for scheduling. The angle-of-attack α , the aerodynamic coefficients C_n and C_m and the actuator were all assumed uncertain. As the Mach number M can be placed in an LFT linearly, we choose a basis function as $f_1(M) = M$ (see also Apkarian and Addams (1995)). For the function $X(M)$, e.g., we thus have

$$X(M) = X_0 + X_1 M.$$

The other functions depending on the parameter M are $Y(M)$, $Q(M)$, $K(M)$, $L(M)$, $\hat{M}(M)$, $N(M)$, and have the same structure as $X(M)$.

The grid of the parameter set $M \in [2, 4]$ consisted of five points equally spaced between Mach = 2 and Mach = 4. As the Mach number will decrease from $M = 4$ to $M = 2$ in five seconds in the non-linear simulations, the parameter rate was taken $|\dot{M}| < 0.5/s$. Further, we use a block diagonal scaling matrix $Q = \text{diag}(Q_0, \dots, Q_4)$ arising from the uncertainty and the performance channels: the matrix Q_1 of dimension 2×2 for the uncertain α , the two scalar blocks Q_2, Q_3 for the uncertainty in C_n, C_m , the scalar block Q_4 for the dynamic actuator uncertainty, and the 3×3 block Q_0 for the performance specification. In the first iteration the scalings are set to unity. Once convergence is achieved, the large LMIs (6) are solved for the last scaling Q that was found in the iteration.

In section 2.2 it was explained that choosing either X or Y constant led to a controller that does not need a measurement of the parameter rate. For the missile control problem, both options were tested. The scheme using X constant and Y parameter dependent converged faster than the one where X was parameter dependent and Y constant. In 14 steps a γ -value of 2.50 was reached whereas the other option only reached the value of 3.87 after 20 iteration steps. In figure 3 the achieved γ -value is set out against the number of iteration steps for both processes.

The final controller synthesis was carried on the choice $(X_0, Y(M))$ where $\gamma = 2.5$ was achieved. The scaling matrix Q of the last iteration is used to synthesize the controller using the full LMIs (6). The achieved γ -value was 2.51 and the test on a denser parameter grid (with twice the density) gave γ values between 2.22 and 2.45. Comparing the achieved performance level with Wu (1995) and Helmersson (1995), we concluded that the iteration procedure was successful.

4.2 Simulations

The non-linear simulations of the LPV controlled missile are depicted in the figures 4 to 7. The ma-

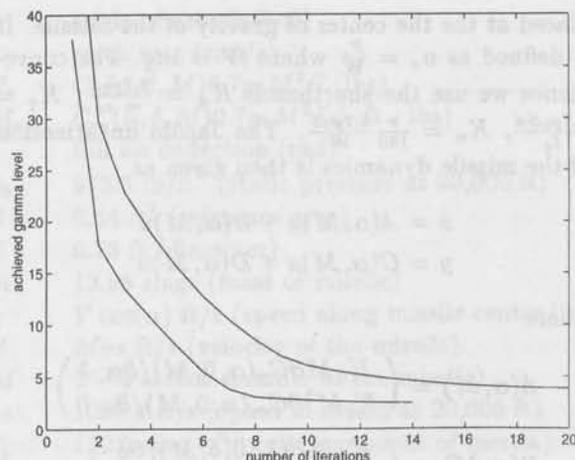


Fig. 3: Comparing the two options for solving the synthesis LMIs

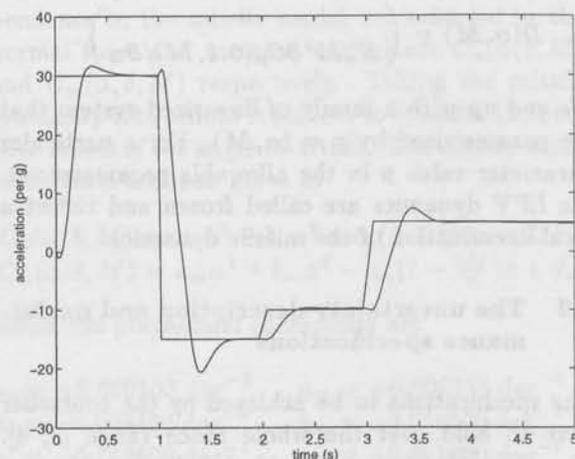


Fig. 4: Normal acceleration n_z for the commanded acceleration scenario n_c of the LPV controlled missile.

neuver, during which the Mach number varies as shown in figure 7, consists of a series of acceleration step commands as depicted in figure 4. The acceleration command response of the LPV controlled missile has a rise time that is less than the prescribed 0.35 s. The steady state error is within the required bounds. Overshoot characteristics are also within the limits. Only the step command from 30 g to -15 g causes a 3% overshoot violation. As a remedy one could try to redesign the weightings. A possible choice to enhance damping of the acceleration response is to increase the weight on the fin rate filter $W_{\dot{\delta}}$ since maximum fin rate is by far not reached in the non-linear simulation. Also the performance filter could be adjusted to further punish the overshoot (increase high frequency gain of the filter). However, we left the filters W_{id} and W_{perf} the same as in Wu (1995) in order to be able to com-

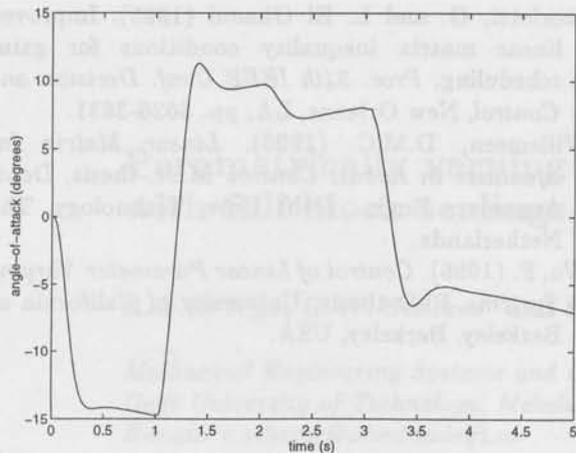


Fig. 5: Angle-of-attack α for the commanded acceleration scenario n_c of the LPV controlled missile.

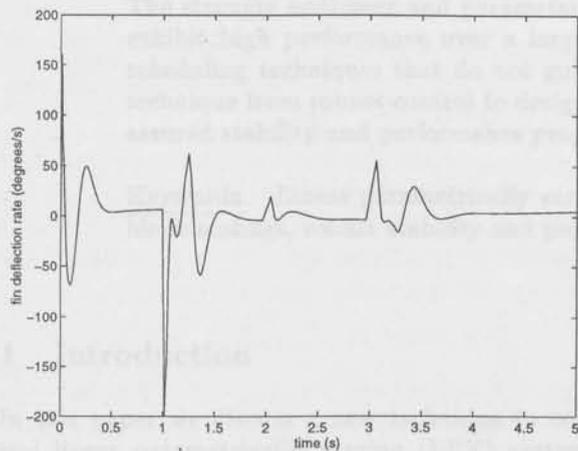


Fig. 6: Fin deflection rate $\dot{\delta}$ for the commanded acceleration scenario n_c of the LPV controlled missile.

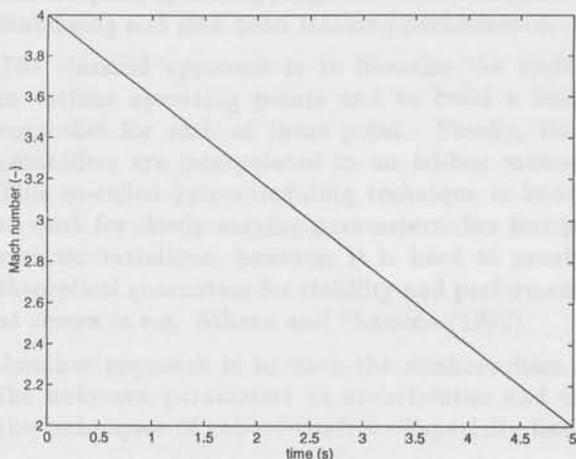


Fig. 7: The parameter trajectory $M(t)$.

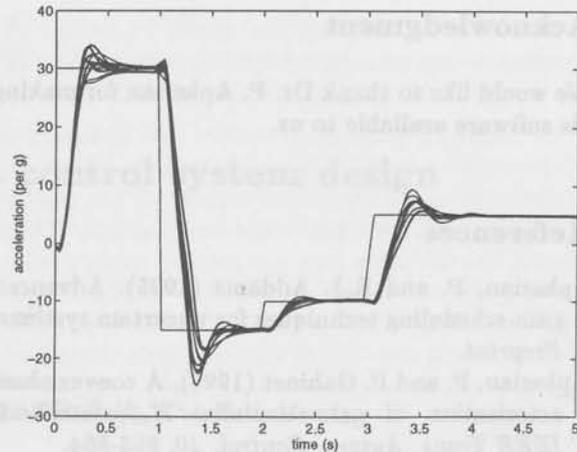


Fig. 8: Command response for all combination of perturbed aerodynamics

pare the results. The LPV controller synthesized here has a larger overshoot, but it is faster than the one in Wu (1995). Moreover the LPV controller of Wu (1995) exceeds the angle-of-attack limit of -20 degrees while also needing much more fin deflection rate. Figure 5 and 6 shows that our controller remains within the limits. It should be noted that the missile in Wu (1995) runs along a slightly different Mach trajectory. Finally, to demonstrate the robustness properties, figure 8 shows the acceleration command responses for all combinations of the aerodynamic uncertainties. As can be seen from the figure, overshoot in the 45 g step is the most sensitive to uncertainties, while the other performance characteristics seem to behave well.

5 Conclusions

We have shown a controller synthesis technique for Linear Parameter Varying (LPV) that takes the boundedness of the parameter variation rates into account. This technique gives guaranteed stability and performance levels. Moreover, robustness against uncertainties has been incorporated via the use of scalings. Using basis functions and gridding, the synthesis problem is reduced to an iteration of solving finitely many linear matrix inequalities. The method was applied and tested on a missile benchmark problem. The non-linear simulations have shown that the proposed method is successful in achieving the desired performance and robustness goals. Further research should be directed towards finding systematic procedures to choose the basis functions in the presented approach.

Acknowledgment

We would like to thank Dr. P. Apkarian for making his software available to us.

References

- Apkarian, P. and R.J. Addams (1995). Advanced gain-scheduling techniques for uncertain systems. *Preprint*.
- Apkarian, P. and P. Gahinet (1995). A convex characterization of gain-scheduled H_∞ -controllers. *IEEE Trans. Autom. Control*, 40, 853-864.
- Athans, M. and J.S. Shamma (1992). Gain-scheduling: potential hazards and possible remedies. *IEEE Control Systems Magazine*, 12, 101-106.
- Balas, G.J., J.C. Doyle, K. Glover, A. Packard, and R. Smith (1993). μ -Analysis and Synthesis TOOLBOX, For Use with MATLAB. The Mathworks, Inc.
- Becker, G. (1993). *Quadratic Stability and Performance of Linear Parameter Dependent Systems*. PhD thesis, University of California at Berkeley, Berkeley, USA.
- Gahinet, P., A. Nemirovski, A. Laub, and M. Chilali (1994). *LMI Control Toolbox*. The Mathworks, Inc.
- Helmersson, A. (1995). *Methods for Robust Gain-Scheduling*. PhD thesis, Linköping University, Sweden.
- Masubuchi, I., N. Suda and A. Ohara (1995). LMI-based controller synthesis: a unified formulation and solution. *Proc. American Control Conf.*, pp. 3473-3477.
- Packard, A. and G.J. Balas (1992). Time-varying controllers for missile autopilots. *Proc. 1st IEEE Conf. Control Applications*, Dayton, OH.
- Packard, A. (1994). Gain scheduling via linear fractional transformations. *System and Control Letters*, 22, 79-92.
- Rantzer, A. and A. Megretski (1994). System analysis via integral quadratic constraints. *Proc. 33rd IEEE Conf. Decision Contr.*, Lake Buena Vista, FL, pp. 3062-3067.
- Rugh, W.J., R.A. Nichols and R.T. Reichert (1993). Gain scheduling for H_∞ controllers: a flight control example. *IEEE Trans. Control Syst. Techn.*, CST-1, 69-78.
- Scherer, C.W. (1995). Mixed H_2/H_∞ control. In: A. Isidori (Ed.), *Trends in Control, A European Perspective*, Springer-Verlag, Berlin, pp. 173-216.
- Scherer, C.W., P. Gahinet and M. Chilali (1995). Multi-objective output-feedback control. *Preprint*.
- Scorletti, G. and L. El Ghaoui (1995). Improved linear matrix inequality conditions for gain-scheduling. *Proc. 34th IEEE Conf. Decision and Control*, New Orleans, LA, pp. 3626-3631.
- Willemsen, D.M.C. (1996). *Linear Matrix Inequalities in Robust Control*. M.Sc.-thesis, Dept. Aerospace Engin., Delft Univ. Technology, The Netherlands.
- Wu, F. (1995). *Control of Linear Parameter Varying Systems*. PhD thesis, University of California at Berkeley, Berkeley, USA.

Parametrically varying flight control system design with full block scalings

R.G.E. Njio, C.W. Scherer[†] and S. Bennani[§]

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: c.scherer@wbmt.tudelft.nl*

Abstract. A systematic procedure for synthesizing a parametrically varying controller for nonlinear systems is presented and applied to a high performance flight control system. The strongly nonlinear and parameter dependent system has to be stabilized and must exhibit high performance over a large operating range. In contrast to classical gain-scheduling techniques that do not guarantee stability or performance, we apply a new technique from robust control to design a parameter dependent nonlinear controller with assured stability and performance properties.

Keywords. Linear parametrically varying systems, (integral) quadratic constraints, full block scalings, robust stability and performance, linear matrix inequalities.

1 Introduction

In this paper we discuss a new technique to control linear parametrically varying (LPV) systems and apply it to a tail-fin controlled missile. The model of this system is strongly nonlinear and parameter dependent as it must operate over a large range of Mach numbers and angles of attack. Over the complete operating range the controller must be stabilizing and give good tracking performance.

The classical approach is to linearize the system in various operating points and to build a linear controller for each of these point. Finally, these controllers are interpolated in an ad-hoc manner. This so-called gain-scheduling technique is known to work for slowly varying parameters. For fast parameter variations, however, it is hard to provide theoretical guarantees for stability and performance as shown in e.g. Athans and Shamma (1992).

Another approach is to view the nonlinearities or the unknown parameters as uncertainties and use the techniques of robust control. Especially linear

matrix inequality (LMI) techniques (Apkarian and Gahinet, 1995; Becker and Packard, 1994; Boyd et al., 1994; Helmersson, 1995; Packard, 1994; Scherer, 1995; Scherer, 1996; Scorletti and El Ghaoui, 1995; Wu, 1995) where stability is guaranteed by searching for a suitable quadratic Lyapunov function, and uncertainties as well as performance are described by (integral) quadratic constraints (Rantzer and Megretski, 1994) are suitable as they can cope with time-varying uncertainties. The resulting design scheme involves a nonconvex optimization problem that is usually approached by a *D-K*-like iteration.

In gain-scheduling, however, the operating point and the parameters affecting the plant can be measured on-line. In designing a robust controller, this essential information is not taken into account what might introduce considerable conservatism. This leads to the idea to design a parametrically varying controller with similar guarantees on stability and performance. It turns out that the LMI techniques for uncertain systems not only extend to the design of such controllers but, in addition, the underlying optimization problem becomes even convex (Apkarian and Gahinet, 1995; Becker and Packard, 1994; Packard, 1994; Helmersson, 1995; Scorletti

[†] Author to whom correspondence should be addressed.

[§] Faculty of Aerospace Engineering, Stability and Control Group, Delft University of Technology

and El Ghaoui, 1995). In this paper we extend the previous techniques based on block-diagonal scalings to full block scalings (Scherer, 1996).

In the approach presented here the rate of change of the parameters is not limited what might lead to unnecessary conservatism since bounds on the parameter derivatives are often known in practice. These bounds can be taken into account by using parameter dependent Lyapunov functions (Apkarian and Addams, 1995; Scherer, 1995). For an application of one of these alternative techniques to the missile control problem we refer to Willemsen (1996).

2 Analysis of uncertain systems

Suppose the system is described as

$$\begin{bmatrix} \dot{x} \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} A & G_1 & G_2 \\ \mathcal{H}_1 & \mathcal{F}_1 & \mathcal{F}_{12} \\ \mathcal{H}_2 & \mathcal{F}_{21} & \mathcal{F}_2 \end{bmatrix} \begin{bmatrix} x \\ w_1 \\ w_2 \end{bmatrix}, w_2 = \Delta(t)z_2. \quad (1)$$

The channel $w_1 \mapsto z_1$ is used to describe performance measured in terms of the L_2 -gain. Moreover, $w_2 \mapsto z_2$ is the uncertainty channel and $\Delta(t)$ is a time-varying parametric uncertainty of which we only know that it satisfies

$$\Delta(t) \in \text{conv}\{\Delta^1, \dots, \Delta^\delta\}. \quad (2)$$

Hence the values of the uncertainties are simply specified as the convex hull of finitely many given extreme points.

The goal is to characterize whether, for this class of uncertainties, the perturbed system remains stable and the L_2 -gain of $w_1 \mapsto z_1$ does not exceed a given value $\gamma > 0$.

The information about the uncertainties is coded in a set of scalings \mathcal{P} that consists of symmetric matrices

$$P = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \quad \text{with } Q < 0, R > 0$$

that satisfy the constraints

$$\begin{bmatrix} \Delta^j \\ I \end{bmatrix}^T \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} \Delta^j \\ I \end{bmatrix} > 0 \quad \text{for all } j = 1, \dots, \delta. \quad (3)$$

Since the uncertainty is contained in a convex set with finitely many extreme points (2), we are able to describe the set of scalings by finitely many linear matrix inequalities.

The performance specification can be cast in an integral quadratic constraint (Rantzer and Megretski, 1994) as

$$\int_0^\infty \begin{bmatrix} w_1 \\ z_1 \end{bmatrix}^T \begin{bmatrix} -\gamma & 0 \\ 0 & \frac{1}{\gamma} \end{bmatrix} \begin{bmatrix} w_1 \\ z_1 \end{bmatrix} dt \leq 0.$$

Now we are ready to state the desired analysis result to guarantee robust stability and robust performance.

Theorem 2.1 (Analysis Result) *If there exists $\mathcal{X} > 0$ and $P \in \mathcal{P}$ satisfying*

$$\begin{bmatrix} A^T \mathcal{X} + \mathcal{X} A & \mathcal{X} G_1 & \mathcal{X} G_2 \\ G_1^T \mathcal{X} & 0 & 0 \\ G_2^T \mathcal{X} & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \mathcal{H}_1^T \\ I & \mathcal{F}_1^T \\ 0 & \mathcal{F}_{12}^T \end{bmatrix} \begin{bmatrix} -\gamma & 0 \\ 0 & \frac{1}{\gamma} \end{bmatrix} \begin{bmatrix} 0 & I & 0 \\ \mathcal{H}_1 & \mathcal{F}_1 & \mathcal{F}_{12} \end{bmatrix} + \begin{bmatrix} 0 & \mathcal{H}_2^T \\ 0 & \mathcal{F}_{21}^T \\ I & \mathcal{F}_2^T \end{bmatrix} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} 0 & 0 & I \\ \mathcal{H}_2 & \mathcal{F}_{21} & \mathcal{F}_2 \end{bmatrix} < 0 \quad (4)$$

then the system (1) is exponentially stable and has a robust L_2 -gain level of at most γ .

This analysis result involves only finitely many LMIs whose solvability can be verified by standard software (Gahinet et al., 1994).

Note that we use full block scalings that are only indirectly described by the inequalities (3). In the literature, analysis results with scalings are usually provided for block diagonal real repeated uncertainties

$$\Delta = \text{diag}(\delta_1 I_1, \dots, \delta_m I_m), \quad |\delta_j| \leq 1.$$

Then the scalings are usually restricted to have the block diagonal structure $Q = \text{diag}(Q_1, \dots, Q_m)$, $R = -Q$, $S = \text{diag}(S_1, \dots, S_m)$ and to satisfy $Q < 0$, $S + S^T = 0$ (Scorletti and El Ghaoui, 1995) or even $Q < 0$, $S = 0$ (Apkarian and Gahinet, 1995). Simple examples reveal that this unnecessary restriction of the scalings lead to a larger infimal bound on the robust performance level as guaranteed by Theorem 2.1. Note also that the fixed scalings $Q = I$, $R = -I$, $S = 0$ lead to the Bounded Real Lemma.

Finally, we stress that we can extend this result to parametric or dynamic uncertainties that admit a description in terms of integral quadratic constraints (Rantzer and Megretski, 1994).

3 LPV systems

Consider a nonlinear system that is represented as

$$\begin{aligned} \dot{x} &= f(x, p) + g(x, p)u \\ y &= h(x, p) + k(x, p)u \end{aligned} \quad (5)$$

where $x(t)$ is the state, $u(t)$ is the input, $y(t)$ is the output, and $p(t)$ is a time-varying parameter. Let

us assume that both the states and the parameters are on-line measurable and that they are known to be contained in the given sets \mathbf{Q} and \mathbf{P} respectively.

If $x = 0$ is an equilibrium of the system for all parameters, one can rewrite it as

$$\begin{aligned} \dot{x} &= \hat{A}(x, p)x + \hat{B}(x, p)u \\ y &= \hat{C}(x, p)x + \hat{D}(x, p)u. \end{aligned} \quad (6)$$

In order to apply the techniques from linear robust control theory to analyze and synthesize controllers, the system (6) is replaced with

$$\begin{aligned} \dot{x} &= \hat{A}(q, p)x + \hat{B}(q, p)u \\ y &= \hat{C}(q, p)x + \hat{D}(q, p)u \end{aligned} \quad (7)$$

where $q(t) \in \mathbf{Q}$ and $p(t) \in \mathbf{P}$ are viewed as time-varying parameters. We arrive at a so-called linear parametrically varying (LPV) system. Due to the decoupling of q and x , the LPV system (7) describes a larger set of trajectories than the original nonlinear system (6) and is, therefore, potentially harder to control. This might introduce conservatism. Yet this is done for two reasons. Firstly, for the LPV system we can use linear design techniques to build a gain-scheduled controller, a controller that depends on the on-line measurable time-varying parameters. Secondly, if we have found a controller with the desired properties for the LPV system (7), the controller guarantees the same properties for the nonlinear system (6) as well.

Let us now abbreviate $\pi = (q, p)$ and $\Pi = \mathbf{Q} \times \mathbf{P} \subset \mathbb{R}^m$. Moreover, we assume that the system (7) can be represented as a linear fractional transformation

$$\begin{bmatrix} \dot{x} \\ z_1 \\ z_2 \\ y \end{bmatrix} = \begin{bmatrix} A & G_1 & G_2 & B \\ H_1 & F_1 & F_{12} & E_1 \\ H_2 & F_{21} & F_2 & E_2 \\ C & D_1 & D_2 & 0 \end{bmatrix} \begin{bmatrix} x \\ w_1 \\ w_2 \\ u \end{bmatrix}, w_2 = \Delta(\pi)z_2 \quad (8)$$

where $\Delta : \Pi \mapsto \mathbb{R}^{k_c \times l_c}$ is a possibly nonlinear (continuous) function. Similarly as in Section 2, we assume that the possible values of the parameter $\Delta(\pi)$ are contained in a convex set with finitely many extreme points:

$$\Delta(\Pi) \subset \text{conv}\{\Delta^1, \dots, \Delta^\delta\}.$$

Moreover, we have specified an extra channel $w_1 \mapsto z_1$ to characterize performance.

If $\hat{A}(\pi), \dots$ in (7) are rational functions of $\pi \in \Pi$, it is well-known that one can rescale the parameters to

$$\Pi = \{\pi \mid |\pi_j| \leq 1 \text{ for all } j = 1, \dots, m\}$$

and that one can obtain this representation with the linear function

$$\Delta(\pi) = \text{diag}(\pi_1 I_1, \dots, \pi_m I_m)$$

where I_j denotes an identity matrix of suitable size.

The controller is assumed to take the same structure, an LTI part K

$$\begin{bmatrix} \dot{x}_c \\ u \\ z_c \end{bmatrix} = \begin{bmatrix} A_c & B_{c1} & B_{c2} \\ C_{c1} & D_{c11} & D_{c12} \\ C_{c2} & D_{c21} & D_{c22} \end{bmatrix} \begin{bmatrix} x_c \\ y \\ w_c \end{bmatrix} \quad (9)$$

that is scheduled with feedback

$$w_c = \Delta_c(\pi)z_c \quad (10)$$

where $\Delta_c : \Pi \mapsto \mathbb{R}^{k_c \times l_c}$ is possibly nonlinear. Due to this structure we can rewrite the controlled system as an augmented LTI system P_e

$$\begin{bmatrix} \dot{x} \\ z_1 \\ z_2 \\ z_c \\ y \\ w_c \end{bmatrix} = \begin{bmatrix} A & G_1 & G_2 & 0 & B & 0 \\ H_1 & F_1 & F_{12} & 0 & E_1 & 0 \\ H_2 & F_{21} & F_2 & 0 & E_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & I \\ C & D_1 & D_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ w_1 \\ w_2 \\ w_c \\ u \\ z_c \end{bmatrix} \quad (11)$$

interconnected with the LTI controller (9) and scheduled with the feedback

$$\begin{bmatrix} w_2 \\ w_c \end{bmatrix} = \begin{bmatrix} \Delta(\pi) & 0 \\ 0 & \Delta_c(\pi) \end{bmatrix} \begin{bmatrix} z_2 \\ z_c \end{bmatrix}. \quad (12)$$

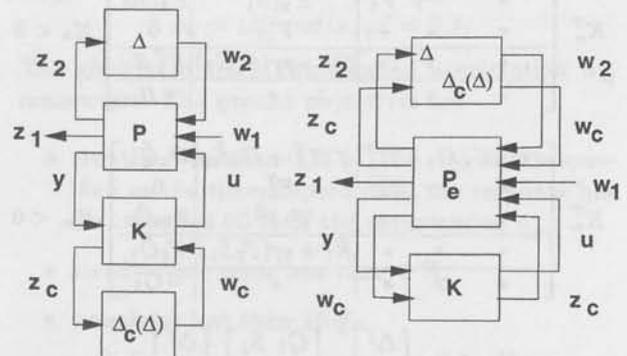


Fig. 1: LFT-system with LFT-scheduled controller

4 The LPV-synthesis problem

The LPV synthesis problem is posed as follows: Find an LTI controller K and a function

$$\Delta_c : \Pi \mapsto \mathbb{R}^{k_c \times l_c}$$

such that, for all parameter curves $\pi : [0, \infty) \mapsto \Pi$, the closed-loop system is exponentially stable and the L_2 -gain of $w_1 \mapsto z_1$ is not larger than γ .

If we interconnect the extended system P_e and the LTI controller K (Figure 1), the resulting LTI system has the structure of (1) with state $[x^T x_c^T]^T$, performance channel $w_1 \mapsto z_1$, and uncertainty channel $[z_2 \mapsto [w_2]$
 $[z_c \mapsto [w_c]$. Let us partition the scalings as

$$Q = \begin{bmatrix} Q_1 & Q_{12} \\ Q_{12}^T & Q_2 \end{bmatrix}, R = \begin{bmatrix} R_1 & R_{12} \\ R_{12}^T & R_2 \end{bmatrix}, S = \begin{bmatrix} S_1 & S_{12} \\ S_{21} & S_2 \end{bmatrix} \quad (13)$$

according to this uncertainty channel. We can now apply the analysis inequality (4) to the closed-loop system and dualize it according to the relations given in the appendix. If we exploit the special structure of the system (11) and if we eliminate the controller parameters in the analysis inequalities, we arrive at the following LMI-solution of the LPV-synthesis problem. As an abbreviation we use $\text{sy}(M) := M + M^T$.

Theorem 4.1 (Synthesis Result) *There exists a controller K (9) and a scheduling function Δ_c (10) that solve the LPV problem if there exist X, Y with $\begin{bmatrix} X & I \\ I & Y \end{bmatrix} > 0$ and scalings $Q_1, R_1, S_1, \tilde{Q}_1, \tilde{R}_1, \tilde{S}_1$ such that*

$$K_n^T \begin{array}{c|ccc|ccc} \text{sy}(XA) & XG_1 & H_1^T & XG_2 + H_2^T S_1^T & H_2^T R_1 & & & \\ * & -\gamma & F_1^T & F_{21}^T S_1^T & F_{21}^T R_1 & & & \\ * & * & -\gamma & F_{12} & 0 & & & \\ \hline * & * & * & Q_1 + \text{sy}(S_1 F_1) & F_2^T R_1 & & & \\ * & * & * & * & -R_1 & & & \end{array} K_n < 0$$

$$K_m^T \begin{array}{c|ccc|ccc} \text{sy}(AY) & G_1 & YH_1^T & YH_2^T + G_2 \tilde{S}_1 & G_2 \tilde{Q}_1 & & & \\ * & -\gamma & F_1^T & F_{21}^T & 0 & & & \\ * & * & -\gamma & F_{12} \tilde{S}_1 & F_{12} \tilde{Q}_1 & & & \\ \hline * & * & * & \tilde{R}_1 + \text{sy}(F_2 \tilde{S}_1) & F_2 \tilde{Q}_1 & & & \\ * & * & * & * & -\tilde{Q}_1 & & & \end{array} K_m < 0$$

$$Q_1 < 0, \begin{bmatrix} \Delta^j \\ I \end{bmatrix}^T \begin{bmatrix} Q_1 & S_1 \\ S_1^T & R_1 \end{bmatrix} \begin{bmatrix} \Delta^j \\ I \end{bmatrix} > 0$$

$$\tilde{R}_1 < 0, \begin{bmatrix} I \\ (\Delta^j)^T \end{bmatrix}^T \begin{bmatrix} \tilde{Q}_1 & \tilde{S}_1 \\ \tilde{S}_1^T & \tilde{R}_1 \end{bmatrix} \begin{bmatrix} I \\ (\Delta^j)^T \end{bmatrix} > 0$$

hold for all $j = 1, \dots, \delta$, where K_m and K_n are basis matrices of the kernels of

$$[B^T \ 0 \ E_1^T \ 0 \ E_2^T], [C \ D_1 \ 0 \ D_2 \ 0].$$

Note that the conditions to be verified take the form of linear matrix inequalities that can be readily validated. It is even possible to directly minimize this performance level since γ enters the inequalities linearly.

5 Controller construction

Once we have solved the synthesis LMIs in Theorem 4.1 for X, Y and for the scalings $Q_1, R_1, S_1, \tilde{Q}_1, \tilde{R}_1, \tilde{S}_1$, we are left with the task to construct the controller. As a first step we determine an extension of the scalings to (13) such that $Q < 0, R > 0, \tilde{Q} > 0, \tilde{R} < 0$ and such that the duality relation (20) holds.

Hence we try to find the unspecified matrices in

$$\begin{bmatrix} M & N \\ N^T & L \end{bmatrix} = \begin{bmatrix} -Q_1 & S_1 & -Q_{12} & S_{12} \\ S_1^T & -R_1 & S_{21}^T & -R_{12} \\ -Q_{21} & S_{21} & -Q_2 & S_2 \\ S_{12}^T & -R_{21} & S_2^T & -R_2 \end{bmatrix}$$

such that

$$\begin{bmatrix} M & N \\ N^T & L \end{bmatrix}^{-1} = \begin{bmatrix} \tilde{M} & \tilde{N} \\ \tilde{N}^T & \tilde{L} \end{bmatrix} := \begin{bmatrix} \tilde{Q}_1 & \tilde{S}_1 & * & * \\ \tilde{S}_1^T & \tilde{R}_1 & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \quad (14)$$

Since the synthesis inequalities are strict, we can perturb the scalings such that, in addition to all inequalities, M, \tilde{M} and $(M - \tilde{M}^{-1})$ are nonsingular.

With a decomposition $(M - \tilde{M}^{-1})^{-1} = V\Lambda V^T$ where V is orthogonal and $\Lambda = \text{diag}(\Lambda_1, -\Lambda_2)$ such that $\Lambda_1 > 0, \Lambda_2 > 0$, we define

$$\begin{bmatrix} M & N \\ N^T & L \end{bmatrix} := \begin{bmatrix} M & VT \\ T^T VT & T^T \Lambda T \end{bmatrix} \quad (15)$$

for some quadratic nonsingular T . It is easy to see that relation (14) holds for any T . If we introduce

$$Z := \begin{bmatrix} 0_{r_1 \times c_1} \\ I_{c_1} \end{bmatrix} \quad \text{and} \quad Z_\perp := \begin{bmatrix} I_{r_1} \\ 0_{c_1 \times r_1} \end{bmatrix}$$

where r_1/c_1 equal the number of rows/columns of Δ , it can be shown that the matrices

$$(\Lambda - V^T Z_\perp (Z_\perp^T M Z_\perp)^{-1} Z_\perp^T V) \quad (16)$$

$$(\Lambda - V^T Z (Z^T M Z)^{-1} Z^T V) \quad (17)$$

have c_2 positive and r_2 negative eigenvalues respectively, where c_2 is the dimension of Λ_1 and r_2 is that of Λ_2 . Hence there exists a nonsingular $T := [T_1 \ T_2]$ where T_1/T_2 have c_2/r_2 columns respectively such that $T_1^T (16) T_1 > 0$ and $T_2^T (17) T_2 < 0$. Using this matrix T in (15) then leads to all the required properties for the full scalings.

Let us then construct Δ_c . With the transformed scalings U, V, W as in (19) we define

$$\begin{bmatrix} \hat{U} & \hat{W} \\ \hat{W}^T & \hat{V} \end{bmatrix} = \begin{bmatrix} U & W \\ W^T & V \end{bmatrix}^{-1} \quad (18)$$

and partition them as inherited by (15). Then a suitable $\Delta_c(\pi)$ can be chosen as

$$\Delta_c(\pi) := -\hat{W}_2 + \begin{bmatrix} \hat{U}_{21} & \hat{W}_{21} \end{bmatrix} \begin{bmatrix} \hat{U}_1 & \hat{W}_1 + \Delta(\pi) \\ \hat{W}_1^T + \Delta^T(\pi) & \hat{V}_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{W}_{12} \\ \hat{V}_{12} \end{bmatrix}$$

Once we have constructed the scalings, the LTI part of the controller K can be computed by solving a standard H_∞ -like problem: just determine \mathcal{X} and a controller such that (4) holds as described in Scherer (1996) and Njio (1996). The full parameter dependent nonlinear controller is then described by (9)-(10). For the justification of Theorem 4.1 and the steps in the controller computation the reader is referred to Scherer (1996) and Njio (1996).

In practice, the procedure for controller construction is found to be numerically delicate. It is improved by fixing γ to a larger than the infimal value and by putting extra constraints on the scalings. Further research is required to systematically improve the numerical conditioning of all the steps involved in the controller design.

6 The missile control problem

The above described gain-scheduling technique has been applied to a missile control problem. The model and its objectives are found in many publications (Rugh et al., 1993; Packard and Balas, 1992; Helmersson, 1995; Wu, 1995; Willemsen, 1996) and are restated here for completeness.

6.1 The missile model

The nonlinear state equations for the control problem are

$$\begin{aligned} \dot{\alpha} &= f_1(\alpha, q, \delta, M) = \frac{\cos(\alpha)^2}{mu} F_z(\alpha, \delta, M) + q \\ \dot{q} &= f_2(\alpha, q, \delta, M) = \frac{M_y(\alpha, \delta, M)}{I_y} \\ n_z &= \frac{F_z(\alpha, \delta, M)}{mg} \end{aligned}$$

where

α	angle of attack (rad)
q	pitch rate (rad/s)
F_z	$C_n(\alpha, \delta) 0.7 p_1 M^2 S$ (lbs)
M_y	$C_m(\alpha, \delta) 0.7 p_1 S d$ (ft-lbs)
δ	tail fin deflection (rad)
p_1	973.3 lb/ft ² (static pressure at 20,000 ft)
S	0.44 ft ² (reference area)
d	0.75 ft (diameter)

m	13.98 slugs (mass of missile)
u	$V \cos(\alpha)$ ft/s (speed along missile center line)
V	$M s s$ ft/s (velocity of the missile)
M	2-4 (Mach number of the missile)
$s s$	1036.4 ft/s (speed of sound at 20,000 ft)
I_y	182.5 slug-ft ² (pitch moment of inertia)
g	32.2 ft/s ² (acceleration due to gravity)
n_z	normal acceleration of the missile (per g).

The aerodynamic coefficients C_n and C_m are given by the polynomial expressions

$$C_n = \text{sign}(\alpha) [a_n |\alpha|^3 + b_n |\alpha|^2 + c_n (2 + \frac{M}{3}) |\alpha| + d_n \delta]$$

$$C_m = \text{sign}(\alpha) [a_m |\alpha|^3 + b_m |\alpha|^2 - c_m (7 - \frac{8M}{3}) |\alpha| + d_m \delta].$$

Typical values for the missile operating between Mach 2 and Mach 4 with an angle-of-attack between -20 and +20 degrees at 20,000 ft are

$$\begin{aligned} a_n &= 0.000103 \text{ deg}^{-3} & a_m &= 0.000215 \text{ deg}^{-3} \\ b_n &= -0.00945 \text{ deg}^{-2} & b_m &= -0.0195 \text{ deg}^{-2} \\ c_n &= -0.1696 \text{ deg}^{-1} & c_m &= 0.051 \text{ deg}^{-1} \\ d_n &= -0.034 \text{ deg}^{-1} & d_m &= -0.206 \text{ deg}^{-1}. \end{aligned}$$

The fin is driven by an actuator of second order

$$\delta = G \delta_c, \quad G(s) = \frac{\omega_a^2}{s^2 + 2\zeta \omega_a s + \omega_a^2}$$

with

$$\omega_a = 150 \text{ rad/s}, \quad \zeta = 0.7.$$

The goal is to track commanded acceleration n_{z_c} maneuvers. The precise objectives are

- rise-time less than 0.35 s. The rise-time is specified as the time elapsed until the response has first reached 90 % of the commanded n_{z_c} .
- steady state error less than 1 %.
- overshoot less than 10 %.

Because of the physical limitations of the fin actuator, there are extra constraints:

- tail-fin deflection less than 25 degrees.
- tail-fin deflection rate less than 25 degrees/s per commanded g .

6.2 Control strategy

Because of the symmetry in the aerodynamic model, synthesis is done for positive α only (Packard and Balas, 1992). The state equations are then given by

$$\dot{\alpha} = K_\alpha M \left[\left(a_n \alpha^2 + b_n \alpha + c_n \left(2 - \frac{M}{3} \right) \right) \alpha + d_n \delta \right] + q$$

$$\dot{q} = K_q M^2 \left[\left(a_m \alpha^2 + b_m \alpha + c_m \left(-7 + \frac{8M}{3} \right) \right) \alpha + d_m \delta \right]$$

$$\dot{n}_z = K_n M^2 \left[\left(a_n \alpha^2 + b_n \alpha + c_n \left(2 - \frac{M}{3} \right) \right) \alpha + d_n \delta \right]$$

From these nonlinear equations we arrive at the LFT form (Figure 2) by 'pulling out the deltas'. The rise-

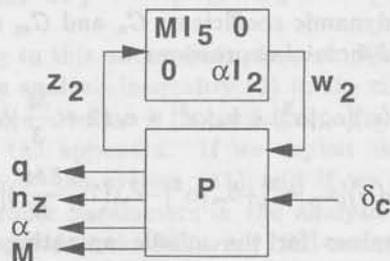


Fig. 2: LFT model of the missile

time and overshoot specifications are captured in an ideal model W_{id} , and the response of the closed-loop model of the missile is compared with the ideal response (Figure 3). The error e is weighted with W_{perf} and the L_2 -induced gain $n_{z_c} \mapsto e$ is minimized to force the controlled system to resemble the ideal model as closely as possible.

In order to comply with the extra constraints on the fin actuator, weighting filters are placed on the actuator outputs fin deflection (W_δ) and fin deflection rate ($W_{\dot{\delta}}$). As a first approach, we used the weighting filters as found in Wu (1995). The ideal model and weighting filters are given by

$$W_{id}(s) = \frac{144(-0.05s+1)}{s^2+19.2s+144}, \quad W_{perf}(s) = \frac{0.5s+17.321}{s+0.0577}$$

$$W_\delta(s) = \frac{1}{19}, \quad W_{\dot{\delta}}(s) = \frac{1}{25}$$

In order to compare the results to Apkarian and Gahinet (1995), the number of external inputs in the interconnection structure (Figure 3) is made equal to the number of external outputs by the definition of extra noise inputs on the measurements q and n_z with weighting filters $W_{n_1} = W_{n_2} = 0.001$.

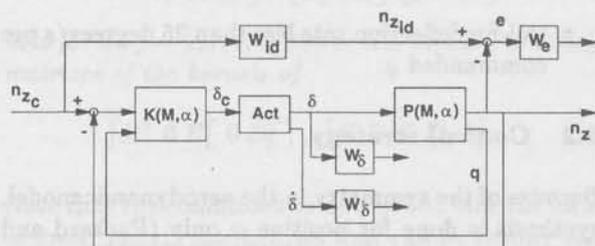


Fig. 3: The synthesis interconnection structure

7 Results

The controller is tested in a nonlinear simulation. The missile is decelerated in 5 s from Mach 4 to Mach 2 (Figure 4). To excite the nonlinearities of

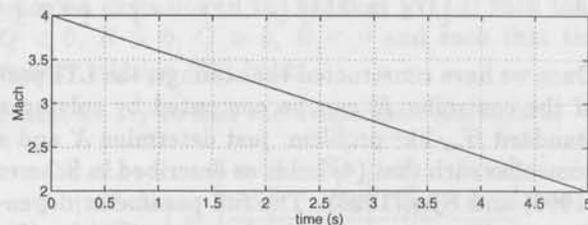


Fig. 4: Mach number as a function of time

the system, we use a normal acceleration command profile n_{z_c} as shown in Figure 5. The response to this command is good, although not all specifications are met. Note that the overshoot decreases for decreasing Mach numbers (Figure 5). It is suggested

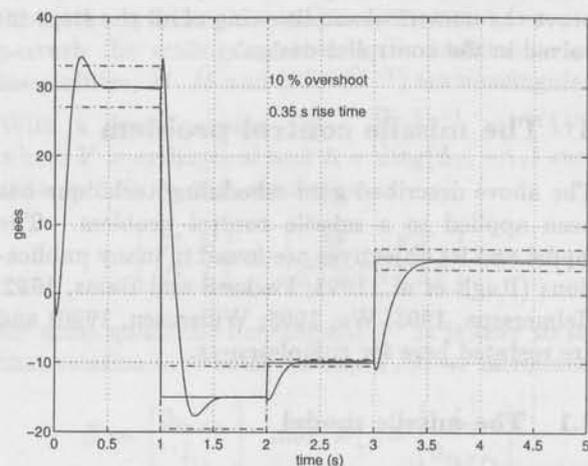


Fig. 5: Acceleration command and response

to use a parameter dependent ideal model in order to improve this acceleration response. The angle of attack (Figure 6) remains within its bounds (± 20 degrees). The fin deflection (Figure 7) remains well within its bounds, such that the weighting W_δ can be chosen less restrictive, while the fin rate (Figure 7) is too high, such that the weighting $W_{\dot{\delta}}$ should be increased.

The controller is compared to the technique with 'partial' scalings (block-diagonal and $S = 0$) (Apkarian and Gahinet, 1995) in Figure 8. Although the improvement in performance is small, it is expected to be better when more parameters and uncertainties are taken into account since then the conservatism due to using partial instead of full block scalings is larger.

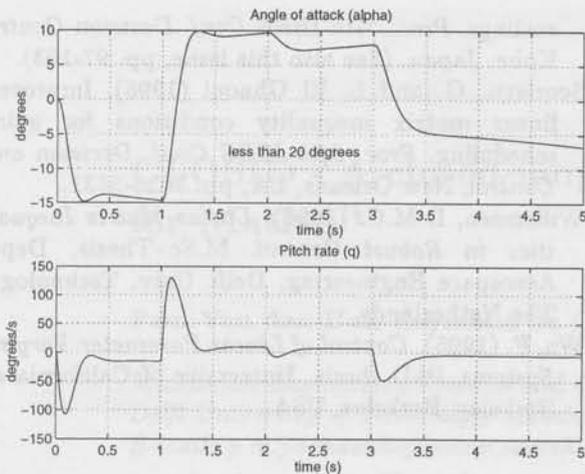


Fig. 6: Angle of attack and pitch rate

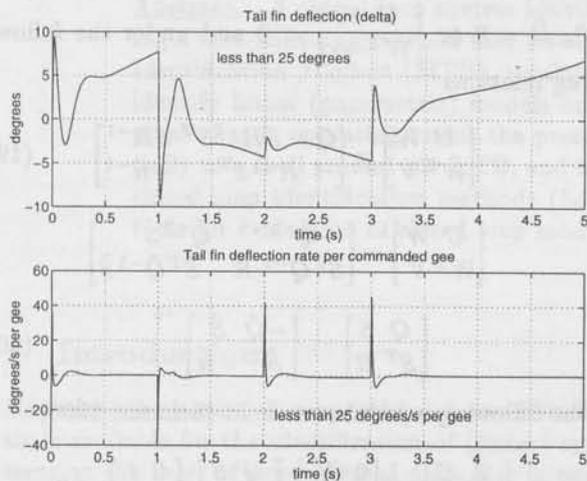


Fig. 7: Tail fin deflection and tail fin deflection rate per commanded gee

Although easily possible, the presented design procedure does not account explicitly for dynamic uncertainty in the actuator or parametric uncertainties in C_n ($\pm 10\%$) and C_m ($\pm 25\%$). Nevertheless, we found that the performance hardly degrades under perturbations in C_n, C_m (Figure 9). Even if the controller is implemented if keeping the parameter π in $\Delta_c(\pi)$ at a fixed value, the performance is still quite good (Figure 9).

8 Conclusions

In this paper we have extended an existing design technique for LPV systems that is based on block-diagonal scalings to full block scalings described by suitable linear matrix inequalities. Using a richer class of scalings not only reduces conservatism but allows to apply the technique to parameter struc-

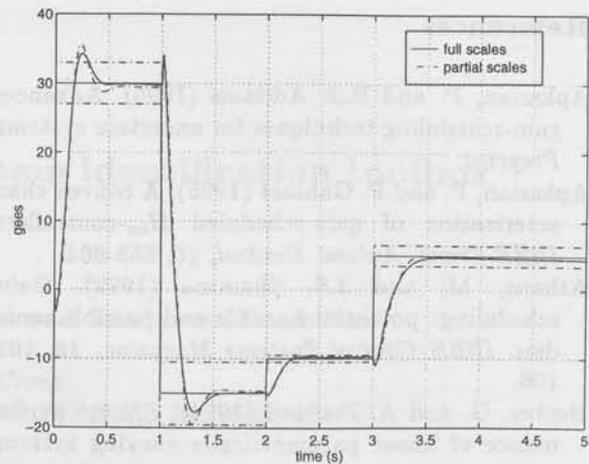


Fig. 8: Acceleration for a controller with full and partial scalings

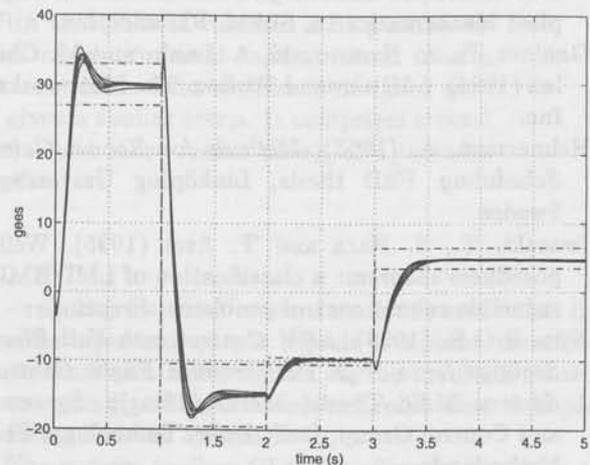


Fig. 9: Acceleration with perturbed C_m and C_n and controller implemented with fixed parameter

tures that are not necessarily block-diagonal. As a new structural ingredient, the controller scheduling function cannot be chosen identical to but has to be nonlinear function of the parameter structure of the LPV system.

The approach has been successfully applied to a nonlinear control problem for a missile model as confirmed by nonlinear simulations.

Acknowledgment

We are grateful to Dr. P. Apkarian for making his software available to us.

References

- Apkarian, P. and R.J. Addams (1995). Advanced gain-scheduling techniques for uncertain systems. *Preprint*.
- Apkarian, P. and P. Gahinet (1995). A convex characterization of gain-scheduled H_∞ -controllers. *IEEE Trans. Autom. Control*, 40, 853-864.
- Athans, M. and J.S. Shamma (1992). Gain-scheduling: potential hazards and possible remedies. *IEEE Control Systems Magazine*, 12, 101-106.
- Becker, G. and A. Packard (1994). Robust performance of linear parametrically varying systems using parametrically dependent linear feedback. *Systems Control Lett.*, 23, 205-215.
- Boyd, S., L. El Ghaoui, E. Feron, and V. Balakrishnan (1994). *Linear Matrix Inequalities in System and Control Theory*. SIAM Studies in Applied Mathematics 15, SIAM, Philadelphia.
- Gahinet, P., A. Nemirovski, A. Laub, and M. Chilali (1994). *LMI Control Toolbox*. The Mathworks, Inc.
- Helmersson, A. (1995). *Methods for Robust Gain-Scheduling*. PhD thesis, Linköping University, Sweden.
- Iwasaki, T., S. Hara and T. Asai (1995). Well-posedness theorem: a classification of LMI/BMI-reducible robust control problems. *Preprint*.
- Njio, R.G.E. (1996). *LPV Control with Full Block Scalings for a High Performance Flight Control System*. M.Sc.-Thesis, Mechan. Engin. Systems and Control Group, Delft Univ. Technology, The Netherlands.
- Packard, A. and G.J. Balas (1992). Time-varying controllers for missile autopilots. *Proc. 1st IEEE Conf. Control Applications*, Dayton, OH.
- Packard, A. (1994). Gain scheduling via linear fractional transformations. *System and Control Letters*, 22, 79-92.
- Rantzer, A. and A. Megretski (1994). System analysis via integral quadratic constraints. *Proc. 33rd IEEE Conf. Decision Contr.*, Lake Buena Vista, FL, pp. 3062-3067.
- Rugh, W.J., R.A. Nichols and R.T. Reichert (1993). Gain scheduling for H_∞ controllers: a flight control example, *IEEE Trans. Control Systems Techn.*, CST-1, 69-78.
- Scherer, C.W. (1995). Mixed H_2/H_∞ control. In: A. Isidori (Ed.), *Trends in Control, A European Perspective*, Springer-Verlag, Berlin, pp. 173-216.
- Scherer, C.W., P. Gahinet and M. Chilali (1995). Multi-objective output-feedback control. *Preprint*.
- Scherer, C.W. (1996). Robust generalized H_2 control for uncertain and LPV systems with general scalings. *Proc. 35th IEEE Conf. Decision Contr.*, Kobe, Japan. (See also this issue, pp. 97-103).
- Scorletti, G. and L. El Ghaoui (1995). Improved linear matrix inequality conditions for gain-scheduling. *Proc. 34th IEEE Conf. Decision and Control*, New Orleans, LA, pp. 3626-3631.
- Willemsen, D.M.C. (1996). *Linear Matrix Inequalities in Robust Control*. M.Sc.-Thesis, Dept. Aerospace Engineering, Delft Univ. Technology, The Netherlands.
- Wu, F. (1995). *Control of Linear Parameter Varying Systems*. PhD thesis, University of California at Berkeley, Berkeley, USA.

Appendix: Duality relations

Under the assumptions $Q < 0$, $R > 0$ or $\tilde{Q} > 0$, $\tilde{R} < 0$ or $\begin{bmatrix} U & W \\ W^T & V \end{bmatrix} < 0$ and under the following relations

$$\begin{bmatrix} U & W \\ W^T & V \end{bmatrix} = \begin{bmatrix} Q - SR^{-1}S^T & SR^{-1} \\ R^{-1}S^T & -R^{-1} \end{bmatrix} \quad (19)$$

$$\begin{bmatrix} U & W \\ W^T & V \end{bmatrix} = \begin{bmatrix} -\tilde{Q}^{-1} & \tilde{Q}^{-1}\tilde{S} \\ \tilde{S}^T\tilde{Q}^{-1} & \tilde{R} - \tilde{S}^T\tilde{Q}^{-1}\tilde{S} \end{bmatrix}$$

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} = \begin{bmatrix} -\tilde{Q} & \tilde{S} \\ \tilde{S}^T & -\tilde{R} \end{bmatrix}^{-1}, \quad (20)$$

the following equivalences hold (Scherer, 1996):

$$\begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} + \begin{bmatrix} 0 & C^T \\ I & D^T \end{bmatrix} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} 0 & I \\ C & D \end{bmatrix} < 0$$

$$\begin{bmatrix} \Delta \\ I \end{bmatrix}^T \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} \Delta \\ I \end{bmatrix} > 0$$

\Leftrightarrow

$$\begin{bmatrix} A & B & C^T \\ B^T & U & D^T + W \\ C & D + W^T & V \end{bmatrix} < 0$$

$$\begin{bmatrix} U & W \\ W^T & V \end{bmatrix}^{-1} + \begin{bmatrix} 0 & \Delta \\ \Delta^T & 0 \end{bmatrix} < 0$$

\Leftrightarrow

$$\begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix} + \begin{bmatrix} B & 0 \\ D & I \end{bmatrix} \begin{bmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{bmatrix} \begin{bmatrix} B^T & D^T \\ 0 & I \end{bmatrix} < 0$$

$$\begin{bmatrix} I \\ \Delta^T \end{bmatrix}^T \begin{bmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{bmatrix} \begin{bmatrix} I \\ \Delta^T \end{bmatrix} > 0.$$

CLOSID - A closed-loop system identification toolbox for Matlab^{‡§}

Paul Van den Hof, Raymond de Callafon[‡] and Edwin van Donkelaar[§]

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: p.m.j.vandenhof@wbmt.tudelft.nl*

Abstract. A closed-loop system identification toolbox for MATLAB is presented, including a user-friendly graphical user interface, that communicates with MathWork's System Identification Toolbox (SITB), version 4.0. With the CLOSID toolbox it is possible to identify linear (parametric) models on the basis of experimental data obtained from a plant that is operating under the presence of a controller. The toolbox is designed partially as a shell around the SITB, and has been given a similar setup. It comprises several closed-loop identification methods (both classical and more recent ones), and includes tools for evaluation of closed-loop model properties.

1 Introduction

Nowadays there are well-supported and user friendly tools available for the identification of (linear) systems on the basis of experimental data. See in particular the Mathwork's System Identification Toolbox SITB, version 4.0, which is equipped with a graphical user interface. This enables the user to identify and validate models in different types of model structures by mouse-clicking, rather than by entering (complex) commands. Additionally there is users' support in terms of graphical tools for model evaluation as well as support for e.g. bookkeeping of identified models.

In the tools that are currently available, there are only limited possibilities to identify models on the basis of data that is obtained under closed-loop experimental conditions. This particular experimental situation - which often occurs in practical situations

- requires a special treatment, in the sense that besides input and output signals of a plant, measured external excitation signals can be involved, as well as some (possibly known) controller that is implemented on the system.

The current toolbox CLOSID offers an extension to the open-loop toolbox SITB, in the sense that

- It provides a graphical user interface supported tool for identification of models from closed-loop observations;
- It enables the use of external excitation signals as well as a (possibly) known controller in the loop;
- It communicates with the SITB, meaning that for the actual estimation part of the closed-loop identification methods, SITB is automatically opened and applied, while in the CLOSID tool the data processing and the (closed-loop) model processing is performed. Therefore full performance and flexibility of the estimation methods in SITB is retained.
- It provides evaluation of models in terms of their closed-loop properties, as e.g. sensitivity functions, complementary sensitivities, closed-loop poles, etcetera.

[‡]MATLAB is a registered trademark of the Mathworks, Inc.

[§]The software described in this paper is available through anonymous ftp at: <ftp-mesc.wbmt.tudelft.nl>, directory <pub/matlab/closid>.

[‡]The work of Raymond de Callafon is financially supported by the Dutch Institute of Systems and Control (DISC).

[§]The work of Edwin van Donkelaar is financially supported by the Dutch Technology Foundation (STW) under contract DWT55.3618

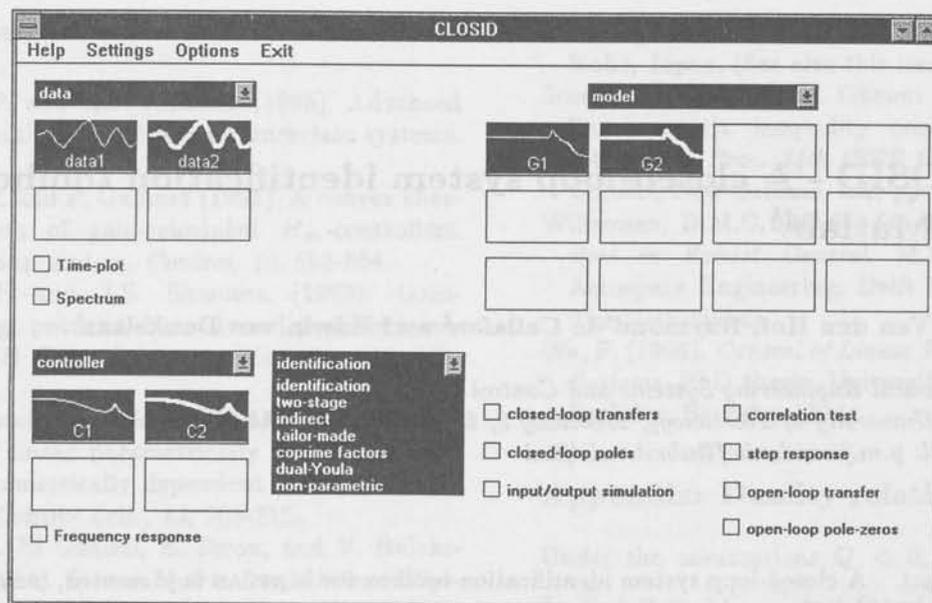


Fig. 1: CLOSID main window

In the current version, the graphical user interface of CLOSID is able to deal with SISO models only.

2 Main CLOSID-window

The graphical user interface of the CLOSID toolbox is opened by entering `closid` in the MATLAB command window. This opens the main window as shown in Figure 1.

The main window shows the following basic parts:

- a **data board** on the left upper part, where imported data sets are represented by colored line-icons, that can be selected by a mouse action.
- a **controller board** on the left lower part, where imported controllers are represented by colored line-icons, with similar selection options.
- an **identification menu** in the middle; this pop-up menu provides the user with a list of identification methods that can be applied.
- a **model board** on the right upper part, showing identified or imported models of the plant to be identified.
- a **model evaluation area**, containing check boxes for the application of several (closed-loop) evaluation procedures for the models on the model board.

Besides the controller board, the composition of the CLOSID main window is very similar to the

main window of the SITB. This controller board is required, as some of the closed-loop identification methods need the a priori knowledge of the controller.

Additionally, this enables the user to evaluate the models in the presence of a (user-chosen) feedback controller.

Data sets, controllers and models can be imported from the MATLAB workspace, through selecting the respective pop-up menus for data, controller and model.

The closed-loop configuration that is considered all through the toolbox is depicted in Figure 2. It is also displayed in the **data import** window.

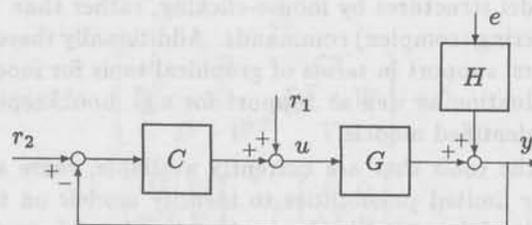


Fig. 2: Closed-loop system configuration

A data set is composed of experimental data $\{u, y\}$ over a given time horizon, together with either one of the external excitation signals r_1 and/or r_2 . Data sets can be viewed on screen in terms of time sequences and power spectra, by clicking on the corresponding check boxes under the data board. Models, as well as controllers, can be imported from and exported to the MATLAB workspace, in different formats:

- $[num; den]$: polynomial coefficients of numerator and denominator, in descending powers of z , stacked in a matrix with height 2.
- $[A \ B; C \ D]$: state space matrices (A,B,C,D) placed in a system matrix.
- $theta$: $theta$ -format as used in the SITB.

The particular model import window is depicted in Figure 3.

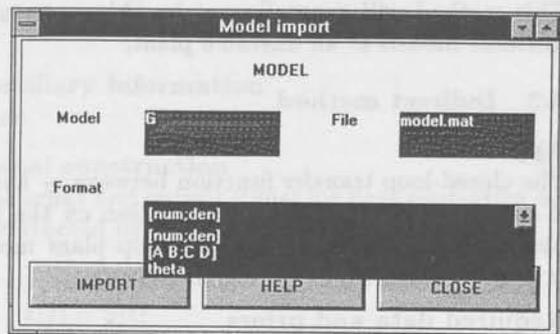


Fig. 3: Model import window

3 Closed-loop identification

The CLOSID toolbox contains five identification methods for parametric model identification, and one nonparametric method. The methods are denoted by

1. **two-stage** method,
2. **indirect** identification,
3. identification with a **taylor-made** parametrization,
4. **coprime factor** identification,
5. identification in the **dual-Youla** parametrization,
6. **non-parametric** (spectral) estimation.

For details on the different methods, one is referred to the references, in particular the survey paper Van den Hof and Schrama (1995).

The methods are all characterized by three steps, focussed on a specific closed-loop object that is going to be identified. E.g. in the indirect method, this closed-loop object is the plant-times-sensitivity $G/(1+CG)$, that is identified on the basis of measured signals r_1 and y . The three steps are clearly indicated in the several identification windows and are characterized as follows.

- **Construction of auxiliary i/o signals.**

A first step of choosing/constructing appropriate auxiliary input and output signals, that are going to be used to identify a particular transfer function object.

- **Identification.**

A second step of actual identification of the considered object, by estimating parameters through a least-squares identification criterion.

- **Calculate plant model.**

From the identified object a plant model is constructed and this plant model is copied to the CLOSID model board.

By choosing one of the identification methods from the **identification** pop-up menu, a particular window is opened, displaying the three steps mentioned above.

The first step is trivial for some methods, but requires a separate identification for some others, as e.g. the identification of the sensitivity function for the two-stage method. In these latter cases, quick-start options provide a simple means to construct the appropriate signals.

Apart from the "tailor-made" approach, all identification methods will perform the second step by opening MATLAB's SITB automatically, copying the appropriate signals from the CLOSID tool to the data board of SITB, allowing the user to identify the required transfer function object in the open-loop toolbox. In all of these situations, the second step is an identification problem that can be handled by the (open-loop) tools in the SITB.

When an appropriate model is identified and validated in SITB, it can be copied to the CLOSID tool, by pushing the **Calculate and copy plant model** in the CLOSID identification window. This third step then transfers the plant model to the CLOSID model board.

As an illustration the **coprime factor** identification window is shown in Figure 4

The nonparametric identification method identifies spectral models for the one input, two output transfer from r to $col(y, u)$, and constructs a plant model by taking the quotient of the two scalar nonparametric estimates.

4 Parametric methods

A brief overview is given of the characteristics of the different parametric methods. In the descriptions it is specified which data and priors are required (measured signal and/or knowledge of the controller), and which auxiliary information needs to be specified before the actual identification in step 2 can be performed.

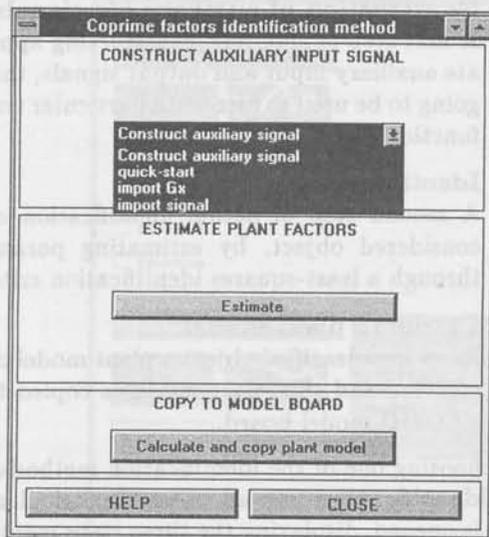


Fig. 4: Window for coprime factor identification

4.1 Two-stage method

Approach

In the first stage the transfer function between reference signal r_1 and input signal u (sensitivity function) is estimated, possibly with a high-order model. With this estimate a noise-free input signal is simulated, which is used in the second stage together with the measured output signal, to identify a plant model.

Required data and priors

- r_1, u, y

Auxiliary information

An estimate is required of the sensitivity function S_0 , i.e. the transfer between r_1 and u . This is obtained in the first stage of the identification procedure, by

$$\hat{\beta}_N = \arg \min_{\beta} \frac{1}{N} \sum_{t=1}^N [u(t) - S(q, \beta)r(t)]^2$$

An accurate (high-order) model is obtained and denoted as

$$\hat{S}(q) = S(q, \hat{\beta}_N).$$

A quick-start option for this estimation is available.

Signal construction

The input and output signal for final estimation are constructed by

$$\begin{aligned} x(t) &= \hat{S}(q)r(t) \\ z(t) &= y(t) \end{aligned}$$

Estimation (in SITB)

Parameters are estimated according to (e.g.)

$$\hat{\theta}_N = \arg \min_{\theta} \frac{1}{N} \sum_{t=1}^N [z(t) - G(q, \theta)x(t)]^2.$$

Plant model

A model of the plant is obtained as

$$\hat{G}(q) = G(q, \hat{\theta}_N).$$

Comments

This method will generally not be able to provide unstable models of an unstable plant.

4.2 Indirect method

Approach

The closed-loop transfer function between r_1 and y is estimated, and by using information on the implemented controller C , an open-loop plant model is reconstructed from this estimate.

Required data and priors

- r_1, y
- C

Auxiliary information

none.

Signal construction

The input and output signal for final estimation are constructed by

$$\begin{aligned} x(t) &= r_1(t) \\ z(t) &= y(t) \end{aligned}$$

Estimation (in SITB)

The exact transfer function between x and z , i.e. the object of identification, is given by

$$R_o = \frac{G_o}{1 + CG_o}$$

Parameters are estimated according to e.g.

$$\hat{\theta}_N = \arg \min_{\theta} \frac{1}{N} \sum_{t=1}^N [z(t) - R(q, \theta)x(t)]^2,$$

leading to the identified transfer function

$$\hat{R}(q) = R(q, \hat{\theta}_N).$$

Plant model

A model of the plant is obtained as

$$\hat{G}(q) = \frac{\hat{R}}{1 - C\hat{R}}.$$

Comments

If the controller is stable, then \hat{G} is guaranteed to be stabilized by C . The model order of \hat{G} will generically be equal to model order of \hat{R} plus order of C .

4.3 Identification with tailor-made parametrization

Approach

The closed-loop transfer function between r_1 and y is estimated, using a dedicated parametrization in terms of the parameters of the open-loop plant model and the known controller C .

Required data and priors

- r_1, y
- C

Auxiliary information

none.

Signal construction

The input and output signal for final estimation are constructed by

$$\begin{aligned} x(t) &= r_1(t) \\ z(t) &= y(t) \end{aligned}$$

Estimation (in CLOSID)

The exact transfer function between x and z , i.e. the object of identification, is given by

$$R_o = \frac{G_o}{1 + CG_o}$$

Parameters are estimated according to

$$\hat{\theta}_N = \arg \min_{\theta} \frac{1}{N} \sum_{t=1}^N [z(t) - \frac{G(q, \theta)}{1 + C(q)G(q, \theta)} x(t)]^2,$$

leading to the identified transfer function

$$\hat{G}(q) = G(q, \hat{\theta}_N).$$

Plant model

A model of the plant is obtained as

$$\hat{G}(q).$$

Comments

The parameter set that corresponds to stable closed-loop systems may be disconnected in the case that the model order of $G(q, \theta)$ is smaller than the order of C . In this case inaccurate models can result.

4.4 Coprime factor method

Approach

The closed-loop transfer functions between r (as input) and (y, u) are estimated, and an open-loop plant model is obtained by taking the quotient of the two estimates.

Required data and priors

- u, y
- C, r_1 and/or r_2

Auxiliary information

Any auxiliary system G_x with a factorization

$$G_x = \frac{N_x}{D_x}$$

that is stabilized by C .

Signal construction

The input and output signals for final estimation are constructed by

$$\begin{aligned} x(t) &= r_1(t) + C(q)r_2(t) \\ z(t) &= \begin{pmatrix} y(t) \\ u(t) \end{pmatrix} \end{aligned}$$

Estimation (in SITB)

The exact transfer function between x and z , i.e. the object of identification, is given by

$$\begin{pmatrix} N_o \\ D_o \end{pmatrix} = \begin{pmatrix} \frac{G_o F^{-1}}{1 + CG_o} \\ \frac{F^{-1}}{1 + CG_o} \end{pmatrix}$$

with $F^{-1} = D_x + CN_x$.

Parameters are estimated according to

$$\hat{\theta}_N = \arg \min_{\theta} \frac{1}{N} \sum_{t=1}^N \text{tr} [z(t) - \begin{pmatrix} N(q, \theta) \\ D(q, \theta) \end{pmatrix} x(t)][\cdot]^T$$

leading to the identified transfer functions

$$\begin{pmatrix} \hat{N} \\ \hat{D} \end{pmatrix} = \begin{pmatrix} N(q, \hat{\theta}_N) \\ D(q, \hat{\theta}_N) \end{pmatrix}.$$

Plant model

A model of the plant is obtained as

$$\hat{G}(q) = \frac{\hat{N}(q)}{\hat{D}(q)}.$$

Comments

By using a normalization procedure, and a common denominator parametrization in the identification, the model order of \hat{G} will be equal to the maximum model order of \hat{N} and \hat{D} .

Identification method	Data	Auxiliary information	Signals for estimation (x, z)	Estimated object $x \rightarrow z$	Exported model
Two-step	r_1, u, y	\hat{S}	$x = \hat{S}r$ $z = y$	G_o	\hat{G}
Indirect	r_1, y C		$x = r$ $z = y$	$R_o = \frac{G_o}{1 + CG_o}$	$\hat{G} = \frac{\hat{R}}{1 - C\hat{R}}$
Tailor-made	r_1, y C		$x = S(\theta)r$ $z = y$	G_o	\hat{G}
Coprime factors	r, u, y C	$G_x = N_x D_x^{-1}$	$x = \frac{r}{D_x + CN_x}$ $z = (y, u)$	(N_o, D_o)	$\hat{G} = \hat{N}\hat{D}^{-1}$
Dual Youla	r, u, y C	$C = N_c D_c^{-1}$ $G_x = N_x D_x^{-1}$	$x = \frac{r}{D_x + CN_x}$ $z = \frac{y - G_x u}{D_c + G_x N_c}$	R_o	$\hat{G} = \frac{N_x + D_c \hat{R}}{D_x - N_c \hat{R}}$

Table 1: Synopsis of closed-loop identification methods

4.5 Dual-Youla method

Approach

A particular closed-loop transfer function is estimated, and by using knowledge of the controller an open-loop plant model is reconstructed. The plant model is guaranteed to be stabilized by the implemented controller. This method is a generalization of the *Indirect method*.

Required data and priors

- u, y
- C, r_1 and/or r_2

Auxiliary information

The controller C is required to be known in a coprime factor representation

$$C = \frac{N_c}{D_c},$$

as well as any auxiliary system G_x with a factorization

$$G_x = \frac{N_x}{D_x}$$

that is stabilized by C .

Signal construction

The input and output signals for final estimation are constructed by

$$x(t) = r_1(t) + C(q)r_2(t)$$

$$z(t) = \frac{1}{D_c + G_x N_c} [y(t) - G_x(q)u(t)].$$

Estimation (in SITB)

The exact transfer function between x and z , i.e. the object of identification, is given by

$$R_o = \frac{(G_o - G_x)D_x}{D_c(1 + CG_o)}.$$

Parameters are estimated according to

$$\hat{\theta}_N = \arg \min_{\theta} \frac{1}{N} \sum_{t=1}^N [z(t) - R(q, \theta)x(t)]^2,$$

leading to the identified transfer function

$$\hat{R}(q) = R(q, \hat{\theta}_N).$$

Plant model

A model of the plant is obtained as

$$\hat{G}(q) = \frac{N_x + D_c \hat{R}}{D_x - N_c \hat{R}}$$

Comments

The model order of \hat{G} will generically be equal to the sum of the model orders of G_x , C and \hat{R} .

Synopsis of parametric methods

In Table 1 a synopsis is given of the parametric identification methods. In this table the signal r is used as an abbreviation for $r_1 + Cr_2$.

5 Model evaluation

Once a model is estimated and made available on the model board, several open-loop and closed-loop model properties can be evaluated. This is done using the seven **Model evaluation** options at the bottom of the main **Closid** window:

1. **closed-loop transfer functions.** The frequency responses of the four transfer functions from $col(r_2, r_1)$ to $col(y, u)$, are shown in a separate window, using the current models from the model board and the current controller C from the controller board. In the window the amplitude of the frequency responses are shown, see Figure 5.

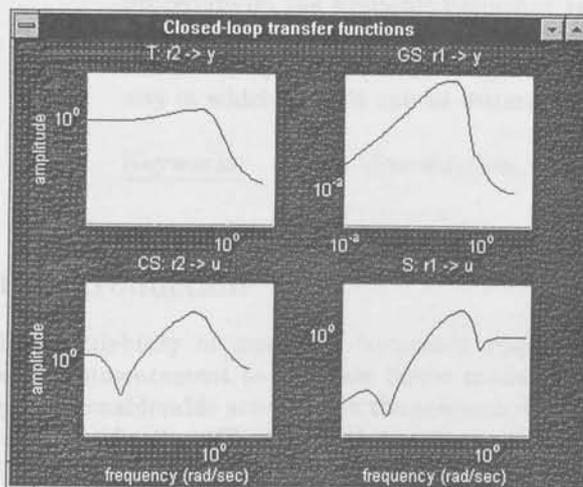


Fig. 5: Closed-loop frequency responses

2. **closed-loop poles.** When clicking this option, the poles of the closed-loop transfer functions are plotted in a separate window, also showing the stability region (unit circle). Thus a simple check is executed showing the (in)stability of the closed-loop system.
3. **input/output simulation.** Using the available reference signal(s) in the current data set, a plant input signal u and plant output signal y are simulated (noise-free), employing the current model and controller. These simulated signals are plotted together with the actual (measured) input and output signals from the current data set.

4. **correlation test.** The sample cross-covariance function is shown between the external reference signal r in the current data set, and the output simulation error (top) and the input simulation error (bottom). This test indicates whether there is still reference signal information in the output and/or input residual, see Figure 6.

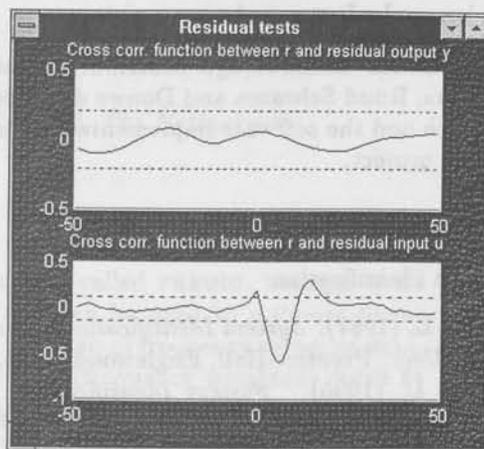


Fig. 6: Closed-loop correlation test

5. **step responses.** This option displays the step responses of the four closed-loop transfer functions from $col(r_2, r_1)$ to $col(y, u)$, for the current models on the model board and the current controller on the controller board.
6. **open-loop transfer.** The (open-loop) Bode diagram is displayed of the current plant models on the model board. This reflects the estimated transfer function between plant input u and output y .
7. **pole-zero plot** of the estimated transfer function between the plant input u and output y .

Selecting one or several of these evaluation tools will open a figure with a plot of the evaluation result for the current models from the model board; where appropriate the current data and current controller will also be employed. A zoom option is available in each figure. By selecting multiple models from the model board, evaluation results of several models can be compared in one figure.

6 Summary

A MATLAB toolbox has been presented for closed-loop system identification on the basis of time domain data. It has been designed as a "shell"

around Mathworks' "open-loop" System Identification Toolbox (SITB). A graphical user interface constructed similar to the SITB supports the user, and facilitates exchange of models between the SITB and the current tool. In its current version the graphical user interface supports the identification of SISO models; the provided MATLAB m-files are implemented to handle also multivariable models.

Acknowledgements

The authors acknowledge contributions of Peter Bongers, Ruud Schrama and Douwe de Vries to the research and the software implementations involved in this project.

References

System identification:

- Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.
- Ljung, L. (1995). *System Identification Toolbox User's Guide*, Version 4.0, The Mathworks, Inc., Natick, Mass., 1995.

Survey papers on closed-loop identification:

- Gevers, M. (1993). Towards a joint design of identification and control? In: *Essays on control: Perspectives on the Theory and its Applications*, Birkhäuser, Boston, pp. 111-151.
- Van den Hof, P.M.J. and R.J.P. Schrama (1995). Identification and control - closed-loop issues. *Automatica*, 31, 1751-1770.

Two-stage method:

- Van den Hof, P.M.J. and R.J.P. Schrama (1993). An indirect method for transfer function estimation from closed-loop data. *Automatica*, 29, 1523-1527.

Indirect method:

- Söderström, T. and P. Stoïca (1989). *System Identification*. Prentice-Hall, Hemel Hempstead, U.K.
- Van den Hof, P.M.J. and R.A. de Callafon (1996). Multivariable closed-loop identification: from indirect identification to dual-Youla parametrization. *Proc. 35th IEEE Conf. on Dec. and Control*, Kobe, Japan. (See also this issue, pp. 1-7).

Tailor-made approach:

- Van Donkelaar, E.T. and P.M.J. Van den Hof (1996). Analysis of closed-loop identification with a tailor-made parametrization. In: O.H. Bosgra et al. (Eds.), *Selected Topics in Identif., Modelling and Control*, Vol. 9, Delft University Press, 1996, pp. 17-24.

Coprime factor identification:

- Van den Hof, P.M.J. R.J.P. Schrama, R.A. de Callafon and O.H. Bosgra (1995). Identification of normalised coprime plant factors from closed-loop experimental data. *Europ. J. Control*, 1, 62-74.

Dual-Youla identification:

- Lee, W.S., B.D.O. Anderdson, R.L. Kosut and I.M.Y. Mareels (1993). On robust performance improvement through the windsurfer approach to adaptive robust control. *Proc. 32nd IEEE Conf. Decision and Control*, San Antonio, TX, pp. 2821-2827.
- Schrama, R.J.P. (1992). *Approximate Identification and Control Design*. Ph.D. Thesis, Mech.Eng. Sys. and Contr. Group, Delft Univ. of Techn., The Netherlands.

FREQID - Frequency domain identification toolbox for use with Matlab^{†§}

Raymond A. de Callafon[‡] and Paul M.J. Van den Hof

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands
E-mail: r.a.decalfon@wbmt.tudelft.nl*

Abstract. This paper presents a new MATLAB toolbox, called FREQID. FREQID is an abbreviation of FREQUENCY domain IDENTIFICATION and can be used to estimate linear (multivariable) discrete or continuous time models on the basis of frequency response data. Within FREQID, a model is being estimated by applying a (frequency weighted) curve fit procedure on the available frequency response. To simplify the operations involved with choosing frequency dependent weightings, model order selection and model evaluation, FREQID is equipped with a Graphical User Interface (GUI). The usage of the GUI and the way in which models can be estimated within FREQID is the core of this paper.

Keywords. system identification; frequency domain; multivariable systems; curve fitting

1 Introduction

The availability of measured frequency responses as a commencement to estimate linear models has gained considerable attention in the research on system identification. First of all this is due to the fact that estimating a model on the basis of a frequency response has several advantages compared to a time domain approach, see e.g. Ljung (1993) or Pintelon *et al.* (1994). Additionally, many engineers have a strong inclination towards a frequency domain related identification procedure, as the "shape" or quality of the model can be influenced directly in the frequency domain by the usage of so-called frequency dependent weightings.

This paper describes the usage of FREQID, a toolbox for use with MATLAB (version 4.2c) for performing identification on the basis of frequency response measurements. FREQID is an abbreviation of FREQUENCY domain IDENTIFICATION, which is supposed to

cover the main purpose of this software: it can handle the estimation of both discrete and continuous time (multivariable) models on the basis of a frequency response in which the frequency vector can be arbitrarily spaced.

Estimating a model in FREQID is done by a curve fitting procedure. In such a procedure, a model is being estimated by fitting the frequency response of the model on a measured frequency response. Within the curve fitting a frequency dependent weighting can be used to emphasize specific parts of the frequency response, so as to influence the quality or "shape" of the model being estimated.

Compared to the frequency domain identification toolbox of Kollár (1994), FREQID focuses solely on frequency response curve fitting. To simplify the operations involved with the estimation and validation of a model, FREQID is equipped with a Graphical User Interface (GUI). This GUI is meant to simplify both the manipulation of frequency domain measurements, the shaping of frequency dependent weightings and the model order selection during the estimation of a model. Furthermore, the GUI serves as a bookkeeper of the models being estimated and enables the user to validate and compare various

[†]MATLAB is a registered trademark of the MathWorks, Inc.

[§]The software described in this paper is available through anonymous ftp at: <ftp://mesc.wbmt.tudelft.nl>, directory [pub/matlab/freqid](ftp://pub/matlab/freqid).

[‡]The work of Raymond de Callafon is sponsored by the Dutch Institute of Systems and Control (DISC)

models relatively easily. As the GUI is designed to be user friendly, most of the information described in this paper is apparent from the GUI of FREQID. By pushing the various help-buttons present in the GUI of FREQID, additional information is displayed. Therefore, this paper will only focus on the main elements present in the GUI of FREQID.

For notational convenience and reasons of clarity, different fonts are used to indicate different objects in this paper. Text in various windows of the GUI of FREQID like titles and text on buttons are typeset in this font. Names of files or directories, commands to be typed and editable text in the GUI of FREQID are typeset in this font. Finally, most abbreviations will be typeset IN THIS FONT. In this way, the difference between *freqid* in a title, the command *freqid* to be typed and FREQID as an abbreviation will be unambiguous.

First in section 2, the main window of FREQID will be discussed. This section also shows the possibilities and bookkeeping facilities of the main window and how frequency responses (the data) and models can be imported. Section 3 describes the possibility to estimate a model on the basis of a frequency response using the GUI of FREQID. Subsequently, section 4 presents the available procedures to evaluate the models being estimated. Finally, the paper is ended by a short summary.

2 The main window

2.1 Overview

If the FREQID toolbox has been installed properly, typing the command

```
>> freqid
```

in the MATLAB command window, will invoke the GUI of FREQID. First a small message window will be opened, that contains information on FREQID and the authors who wrote the software. By clicking the continue-button, the main window of FREQID will be opened. This main window is depicted in Figure 1 and consists of the following distinguishable parts.

- At the top of the window a menu bar can be found. Via the options on the menu bar, session files can be loaded or saved, different MATLAB windows can be accessed and the layout of the FREQID windows can be modified.
- At the left top part of the window one can find the Data Board. This is used to store and manipulate the frequency domain measurements and/or frequency dependent weights used for estimating a model. It also contains a data-popup menu.

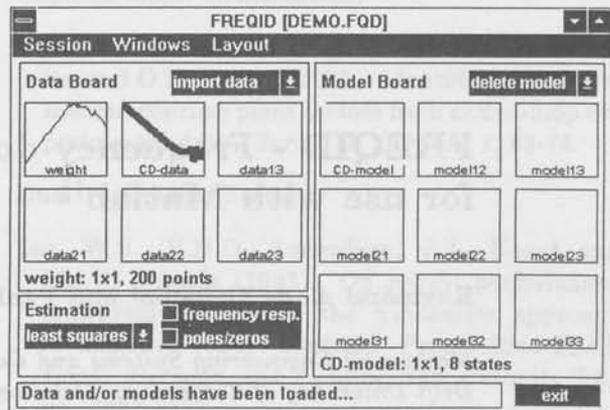


Fig. 1: Main window of FREQID

- The right part of the window contains the Model Board. Similar to the Data Board, this is used to store and manipulate models and contains a model-popup menu.
- The left bottom part of the main window is reserved for Estimation (and evaluation) of models.
- Finally, at the bottom of the window the status line is displayed. This is used to display all kinds of messages to the user. In Figure 1 the user is notified of the fact that Data and/or models have been loaded..., just after a session file called DEMO.FQD was loaded.

A more detailed description on importing data, models and mouse actions defined within the main window of FREQID can be found in the following sections.

2.2 Importing data

The starting point for estimating models within FREQID is the availability of a frequency response that needs to be fitted. Subsequently, the Data Board can be used to store and manipulate the frequency response and/or frequency domain weights used for estimating a model. For this purposes, a data-popup menu and specific mouse actions (clicking, drag & drop) are defined within the Data Board depicted in Figure 1.

Frequency responses (or frequency dependent weightings) can be imported from a file or from the MATLAB workspace onto the Data Board in three different formats:

- MVFR matrix (MFD tools) In such a MultiVariable Frequency Response (MVFR) matrix, a

frequency domain measurement (single- or multivariable) is stacked columnwise for each frequency point separately. The frequency vector (always in [rad/s]) corresponding to it, must be specified separately. This format is also supported by the Multivariable Frequency Domain (MFD) toolbox, (Maciejowski, 1990).

- **FREQFUNC matrix (IDENT tools)** This is a format to store frequency domain data supported by the System Identification ToolBox (SITB), (Ljung, 1995). Such a matrix already contains the corresponding frequency vector.
- **Varying matrix (MU tools)** This is the format supported by the μ -analysis and synthesis toolbox, (Balas *et al.*, 1995). Such a matrix already contains the frequency vector.

To import frequency responses on the Data Board, the import data option of the data-popup menu can be used. Invoking this menu option yields the window depicted in Figure 2. From Figure 2 the three

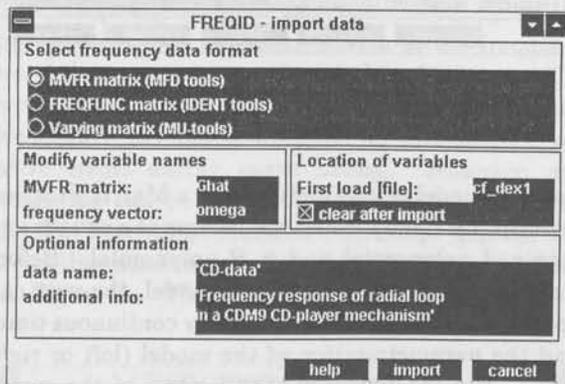


Fig. 2: Import data window.

different formats discussed above can be recognized, whereas for bookkeeping purposes, the name of the data and some additional information can be specified.

2.3 Obtaining models

Quite similar to the Data Board, the Model Board has been defined on the main window of FREQID. Importing a model on the Model Board can be done by estimating a model on the basis of a frequency response available on the Data Board. However, the discussion of this option is postponed until section 3. Additionally, a model can be imported by the import model option of the model-popup menu. Invoking this menu option yields the window depicted in Figure 3. Three different formats are supported to

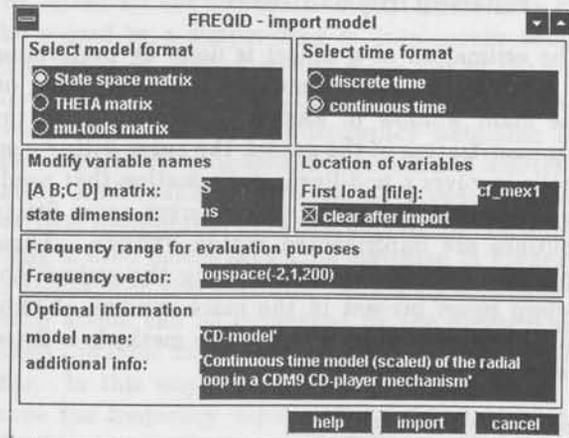


Fig. 3: Import model window.

import a model (discrete or continuous time) from a file or the MATLAB workspace onto the Model Board of FREQID:

- **State space matrix** In this format the state space matrices A, B, C and D of a model (single- or multivariable) are stacked in a system matrix $S=[A \ B;C \ D]$. In order to be able to re-extract the state space, the size of the matrix A (state space dimension) must be given too.
- **THETA matrix.** This is a format to store a model as supported by the System Identification toolbox, (Ljung, 1995).
- **mu-tools matrix.** This is the format supported by the μ -analysis and synthesis toolbox, (Balas *et al.*, 1995).

To import a model, a frequency vector must be added in order to evaluate the frequency response of the model. Finally, the name of the model and some additional information can be specified for bookkeeping purposes, see also Figure 3.

2.4 Mouse actions

Once some frequency response (data) or a model has been loaded successfully, it will appear as an icon in one of the boxes present in the main window depicted in Figure 1. This icon is formed by plotting the Bode amplitude diagram of the first element of the (multivariable) frequency response of either the data or the model. The icon can now be selected simply by clicking on the corresponding box. If the icon has been selected, a fat line will be drawn in the corresponding box, see e.g. the second box in Figure 1. Information on the icon can also be accessed by a simple click on the name. By a simple drag & drop action, an icon can also be copied.

3 Estimating a model

The estimation of a model is done by performing a curve fit on a frequency response available on the main window of FREQID. Depending on the parametrization of the model, the curve fitting generally involves a non-linear optimization that needs to be solved. Currently, two different curve fitting routines are implemented within FREQID. These routines are available by invoking the estimation-popup menu present in the main window of FREQID. A short summary of the two methods is listed below.

least squares estimation

The least-squares estimation routine implemented aims to minimize the 2-norm of a (weighted) difference between the frequency response of the model and the data. A frequency dependent weighting is a so-called Schur-weighting in which the weighting is specified for each transfer function separately. The (multivariable) model is parametrized by either a left or right Matrix Fraction Description (MFD), which reduces to a simple numerator/denominator representation for estimating scalar models. For a more detailed discussion on the procedure, one is referred to de Callafon *et al.* (1996).

maximum amplitude

The maximum amplitude routine implemented aims at minimizing the (weighted) maximum difference between the frequency response of the model and the data, element wise. Again the weighting can be specified for each transfer function separately. The (multivariable) model is parametrized by a combined diagonal left and right Matrix Fraction Description (MFD), which reduces to a simple numerator/denominator representation for estimating scalar models. For a more detailed discussion on the procedure, one is referred to Hakvoort and Van den Hof (1994).

For both the methods discussed above, an iteration based on the Sanathanan-Koerner procedure (Sanathanan and Koerner, 1963) is used to tackle the non-linear minimization involved. Although there is no direct guarantee of convergence, the method generally leads to useful models. Furthermore it is reasonably fast and due to the subsequent convex optimization steps it supports the estimation of relatively high order models. The procedure to estimate a model using the GUI of FREQID is nearly the same for both methods. Furthermore, the least squares estimation routine is included in FREQID, whereas for the maximum amplitude cri-

terion the installation of the MATLAB optimization toolbox, version 1.0c, is a prerequisite (Hakvoort, 1994). Therefore, only the least squares estimation routine will be illustrated here.

3.1 Least squares estimation

Once data has been loaded and selected, invoking the least squares option from the estimation-popup menu in the main window of FREQID will present the least squares estimation window on the screen. An overview of this window is depicted in Figure 4. In the least squares estimation algorithm, a multi-

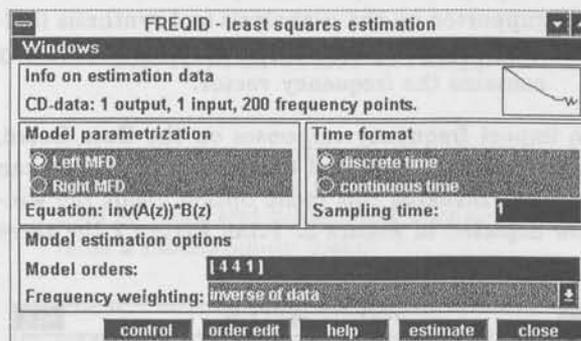


Fig. 4: Least squares estimation window

variable model is parametrized by a Matrix Fraction Description (MFD), using the inverse of a square and monic A -polynomial and a B -polynomial. Before starting up the estimation of a model, the user can specify the time format (discrete or continuous time) and the parametrization of the model (left or right MFD). For the left MFD, the inverse of the monic A -polynomial appears at the output of the model, whereas for the right MFD, the inverse appears at the input. For a scalar system, both parametrizations are the same and reflect an ordinary numerator/denominator parametrization. Subsequently, the model orders or number of parameters to be estimated can be specified, for which a separate order editor is available.

Finally, the weighting to be used during the estimation (curve fitting) of the model can be specified in the frequency weighting-popup menu. Default, the weighting is chosen to be the inverse of the data, so as to minimize a relative error instead of an absolute error. Additional choices include none (unit weighting to minimize an unweighted, absolute error) or advanced. The advanced weighting option enables the user to load and/or modify frequency domain weightings relatively easily. One is referred to section 3.2 for a more detailed discussion on the usage of advanced weightings.

A simple click on the estimate button will start the minimization. Progress on the iteration to fit the frequency response is displayed in the MATLAB command window. Some options associated with the Sanathanan-Koerner iteration are available under a control-button, see also Figure 4. If the minimization has been completed successfully, the model can be imported on the Model Board. Before importing the model, options associated to the frequency range for evaluation purposes, the name of the model and the additional information on the model being estimated can be modified.

3.2 Advanced weightings

The weighting used in the least squares estimation can be any frequency dependent weighting, having the same size as the frequency response used for curve fitting. It is applied *element wise* in case of multivariable frequency response. The GUI of FREQID allows the import and/or modification of a weighting relatively easily by opening a weighting window, in which each element of a (multivariable) weighting can be edited.

The weighting window can be opened by selecting the advanced weighting option in the frequency weighting-popup menu depicted in Figure 4. The weighting window will start up with the *default* weighting: inverse of data, so as to minimize relative errors during curve fitting. However, any weighting can be imported and edited elementwise in the weighting window. A snap shot of the weighting window for editing an element is depicted in Figure 5. The vertical dashed-dotted lines in

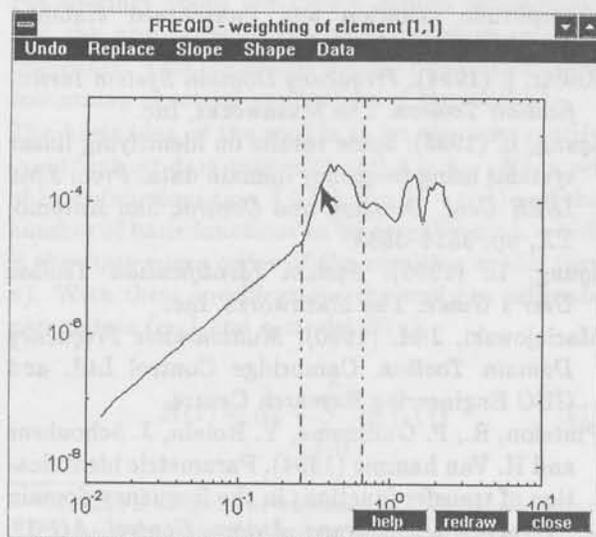


Fig. 5: Editable weighting per element

Figure 5 are used to select the frequency range to

be edited for the specific element. These lines can be moved by a simple drag & drop mouse action. The different options on the menu bar of the window depicted in Figure 5 can be used to modify the weight. In this way, the frequency weighting can for example be smoothed, integrated and differentiated. Additionally, the weighting between the two vertical dash-dotted lines can also be modified by a drag & drop mouse action, as indicated in Figure 5. The shape of the weighting caused by this drag & drop action can be influenced by the different options available under the Shape option on the menu bar. In this way, FREQID offers the possibility to tune the frequency dependent weighting in a very flexible way.

4 Evaluation of models

As a final step in estimating models, the possibility to evaluate a model on the basis of its frequency response or pole/zero plot is available within FREQID. This can be done by using either the frequency resp. or the poles/zeros check-boxes available in the main window of FREQID, see Figure 1. Turning the frequency resp. check-box on, will open the frequency response window, as depicted in Figure 6. The fre-

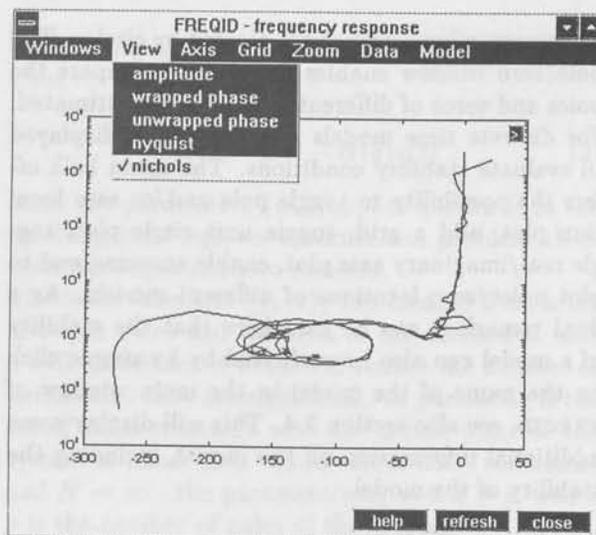


Fig. 6: Frequency response window

quency response window enables the user to view and compare both the measured frequency response and the frequency response of different models in various ways, as indicated by Figure 6. The menu balk offers the possibility to change the axis, add a grid, enable zooming and to plot various measured frequency responses (data) and/or model frequency responses in the same plot.

Next to the frequency domain evaluation of the model, the poles and zeros of the models being estimated can be computed. Turning the poles/zero check-box on in the main window of FREQID, will open the poles/zero window, as depicted in Figure 7. In Figure 7, poles of a model are indicated

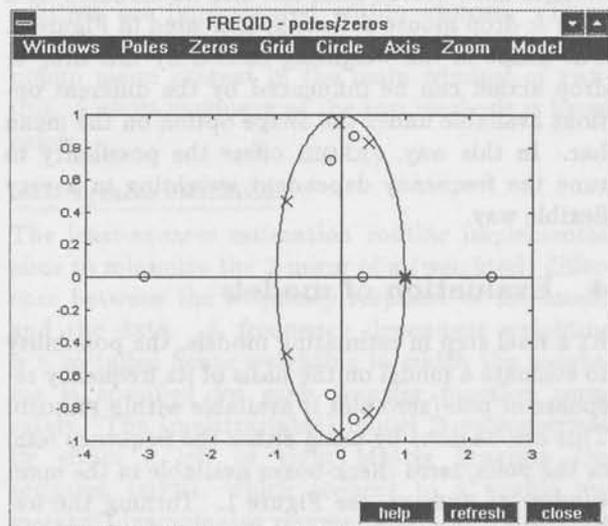


Fig. 7: poles and zeros window

by crosses, whereas zeros are plotted by circles. The pole/zero window enables the user to compare the poles and zeros of different models being estimated. For discrete time models a unit circle is displayed to evaluate stability conditions. The menu balk offers the possibility to toggle pole and/or zero location plot, add a grid, toggle unit circle plot, toggle real/imaginary axis plot, enable zooming and to plot poles/zero locations of different models. As a final remark it can be said here that the stability of a model can also be evaluated by simple click on the name of the model in the main window of FREQID, see also section 2.4. This will display some additional information on the model, including the stability of the model.

5 Summary

A MATLAB toolbox called FREQID has been presented for estimating discrete or continuous time linear (multivariable) models on the basis of (measured) frequency responses. Within FREQID, a model is being estimated by performing a curve fit routine on the available frequency domain measurement. In the current version of FREQID, this curve fit routine can be either a least-squares or a maximum amplitude criterion.

To simplify the operations involved with choosing

frequency dependent weightings, model order selection and model evaluation, FREQID is equipped with a user friendly Graphical User Interface (GUI). Additionally, the GUI serves as bookkeeper of the available frequency domain measurements and the different models being estimated.

The software of FREQID is written for MATLAB version 4.2c and the standard signal and control MATLAB toolboxes are required only. In addition, the optimization toolbox is needed only if the maximum amplitude routine is being used too.

6 Acknowledgements

Any feedback by remarks, suggestions or bug reports are a prerequisite to develop reliable and user friendly software. Therefore, the authors would like to thank David Molenaar for his contribution.

References

- Balas, G.J., J.C. Doyle, K. Glover, A. Packard and R. Smith (1995). *μ -Analysis and Synthesis Toolbox*. MUSYN Inc. and The Mathworks, Inc.
- de Callafon, R.A., D. Roover and P.M.J. Van den Hof (1996). Multivariable least squares frequency domain identification using polynomial matrix fraction descriptions. *Proc. 35th IEEE Conf. Decision and Control*, Kobe, Japan. (See also this issue, pp. 31-37).
- Hakvoort, R.G. (1994). *Matlab-toolbox CURVEFIT*. Internal Report, Delft Univ. of Technology, Mech. Eng. Systems and Control Group.
- Hakvoort, R.G. and P.M.J. Van den Hof (1994). Frequency domain curve fitting with maximum amplitude criterion and guaranteed stability. *Int. Journal of Control*, 60, 809-825.
- Kollár, I. (1994). *Frequency Domain System Identification Toolbox*. The Mathworks, Inc.
- Ljung, L. (1993). Some results on identifying linear systems using frequency domain data. *Proc. 32nd IEEE Conf. Decision and Control*, San Antonio, TX, pp. 3534-3538.
- Ljung, L. (1995). *System Identification Toolbox User's Guide*. The Mathworks, Inc.
- Maciejowski, J.M. (1990). *Multivariable Frequency Domain Toolbox*. Cambridge Control Ltd. and GEC Engineering Research Centre.
- Pintelon, R., P. Guillaume, Y. Rolain, J. Schoukens and H. Van hamme (1994). Parametric identification of transfer functions in the frequency domain - a survey. *IEEE Trans. Autom. Control*, AC-39, 2245-2260.
- Sanathanan, C.K. and J. Koerner (1963). Transfer function synthesis as a ratio of two complex polynomials. *IEEE Trans. Aut. Contr.*, AC-8, 56-58.

ORTTOOL - A Matlab[†] toolbox for system identification with generalized orthonormal basis functions[§]

Peter S.C. Heuberger[‡] and Paul M.J. Van den Hof

*Mechanical Engineering Systems and Control Group
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands.
E-mail: p.s.c.heuberger@wbmt.tudelft.nl*

Abstract. A system identification toolbox for MATLAB is presented, for estimating linear time invariant models using generalized orthonormal functions. The toolbox is supported by a user-friendly graphical user interface and communicates with MathWorks' System Identification toolbox (SITB), version 4.0. With ORTTOOL it is possible to identify linear parametric SISO models on the basis of experimental data (open loop identification), using basis functions that can be flexibly chosen by the user. The tool has options for input of data and basis functions, settings of order, initial conditions, number of iterations, model reduction algorithms, a number of simple validation tests and various graphics.

1 Introduction

ORTTOOL is a graphical user interface written in the MATLAB interpreter language. It is intended to accommodate identification of single input-single output systems, using orthonormal basis functions. For specifics about the theory behind this method and the properties of the basis functions, see the references. This document merely discusses aspects concerning practical application of the method.

The basic idea of the tool is to let the user specify input/output data (say $u(t), y(t), t = 1 \dots N$), a set of basis functions (say $\{f_k(z), k = 1 \dots \infty\}$) and the number of basis functions to be actually used, which is the state space order of the resulting model (say n). With these specifications the tool can estimate parameters $\{\alpha_k\}$ and a model $H(z)$,

$$H(z) = \alpha_0 + \sum_{k=1}^n \alpha_k f_k(z) \quad (1)$$

[†]MATLAB is a registered trademark of the Mathworks, Inc.

[§]The software described in this paper is available through anonymous ftp at: <ftp-mesc.wbmt.tudelft.nl>, directory <pub/matlab/orttool>.

[‡]The work of Peter Heuberger is financially supported by the Dutch Technology Foundation (STW) under contract DWT55.3618. He is on partial leave from the Dutch National Institute of Public Health and the Environment (RIVM).

by minimizing in least square sense the prediction error :

$$e(t) := y(t) - H(q)u(t) \quad (2)$$

Since the parameters $\{\alpha_k\}$ appear quadratic in the resulting least squares optimization problem there exists a unique analytic solution.

The basis functions $\{f_k(z)\}$ are created from a set of stable poles and the crux of the method is that a well-chosen set of poles will reduce the number of parameters that are significant. In particular if the set of poles coincides with the actual poles of the system at hand, then -under white noise conditions and $N \rightarrow \infty$ - the parameters $\alpha_k = 0, k > p$, where p is the number of poles of the system.

It is helpful to keep in mind that the most straight forward example of the basis functions is the set of pulse functions $\{z^{-k}\}$, in which case the method is equivalent to the standard estimation of FIR (finite impulse response) models. See section 2 for the construction of the set of basis functions. The tool is built up in one main window, in which at all times 4 so called frames are visible. In total there are 7 frames, one of which is the so-called Message frame. This frame is used for intermediate messages, warnings and error messages. It is always visible. The other 6 frames are divided in 3 groups:

1. Basis frame and Data frame.
2. Order frame and Reduction frame.
3. Initial frame and Iteration frame.

Of each group only one frame is visible at a time. Each frame has a button to toggle the visibility. This set-up was taken to reduce the amount of screen space, while keeping as much information at hand as possible. In Figure 1 the primary frames are displayed.

Next to these frames the tool has a number of pull-down menu buttons at the top, as shown in Figure 2. In the next sections all the frames and menu buttons will be discussed in detail.

The tool keeps track of all the results created and bases used. These are stored in memory and can be saved in files or copied to SITB. The undertaken actions can be logged in a log-file.

2 Background

In this section a short explanation is given on the construction of the basis functions for the SISO case. This concept has a straightforward multivariable extension. See the references for details.

The sets of functions used in this approach are based on the fact that each inner (i.e. stable all-pass) transfer function gives rise to a basis for \mathcal{H}_2 in the following way.

Let $[A, B, C, D]$ be a *balanced realization* of an inner transfer function $G_b(z)$, with McMillan degree m . Define $V_1(z) := [zI - A]^{-1}B$. Then the set of functions defined by the scalar elements of

$$V_k(z) := V_1(z)G_b^{k-1}(z) \quad V_k(z) \in \mathbb{R}^m \quad (3)$$

forms a basis for the strictly proper transfer functions in \mathcal{H}_2 .

Hence for every transfer function $H(z) \in \mathcal{H}_2$ there exist unique parameters α_0 and $\{L_k \in \mathbb{R}^m, k = 1, \dots, \infty\}$ such that

$$H(z) = \alpha_0 + \sum_{k=1}^{\infty} L_k^T V_k(z) \quad (4)$$

Well known examples of this concept are the cases $G_b = z^{-1}$ and $G_b(z) = \frac{1-az}{z-a}$ which result in the basis of pulse functions $\{z^{-k}\}$ respectively the basis of Laguerre functions $\{\sqrt{1-a^2} \frac{(1-az)^k}{(z-a)^{k+1}}\}$.

When applying this expansion formulation to system identification it is assumed that $H(z)$ can be adequately approximated with a finite number of basis functions

$$\hat{H}(z) = \alpha_0 + \sum_{k=1}^r L_k^T V_k(z) \quad (5)$$

Rewriting the vector functions $\{V_k(z)\}$ and vectors $\{L_k\}$ in scalar functions $\{f_i(z)\}$ and scalars $\{\alpha_i\}$ gives the expression (1).

In the SISO case an inner function is completely parametrized by its poles $\{a_i\}$, since every inner function $G_b(z)$ can be written as the Blaschke product

$$G_b(z) = \pm \prod_{i=1}^p \frac{1 - a_i z}{z - a_i} \quad (6)$$

The approach is motivated by the fact that -if the inner function c.q. the set of poles is well chosen- the number of significant parameters will be small, where well chosen indicates that the poles of the inner function are close to the poles of the actual system at hand. This implies that a-priori knowledge about major time-constants can be directly used in this approach. In practical situations such knowledge is often available. See the references for an analysis of the bias and variance of the resulting estimates.

3 Utilities

In order to apply the method the following steps have to be performed:

- specification of input/output data
- choice of basis, in terms of the poles of the inner function
- the number of basis functions to incorporate

There are however a number of practical issues to be considered. These are reflected upon in the next paragraphs.

3.1 Data

The frame titled DATA is meant for importing input/output data. The variable(s) can be located in 3 different locations denoted by:

Workspace	MATLAB's main level storage space.
File	A file in MATLAB's .mat format.
Ident	Copy the ONLY data set selected in the SITB board(s).

The name of the input and output variable have to be entered in text fields labeled *Input* respectively *Output*. Furthermore one can specify a range to be used. In order to preprocess the data there is an option to remove means from the data on the chosen range.

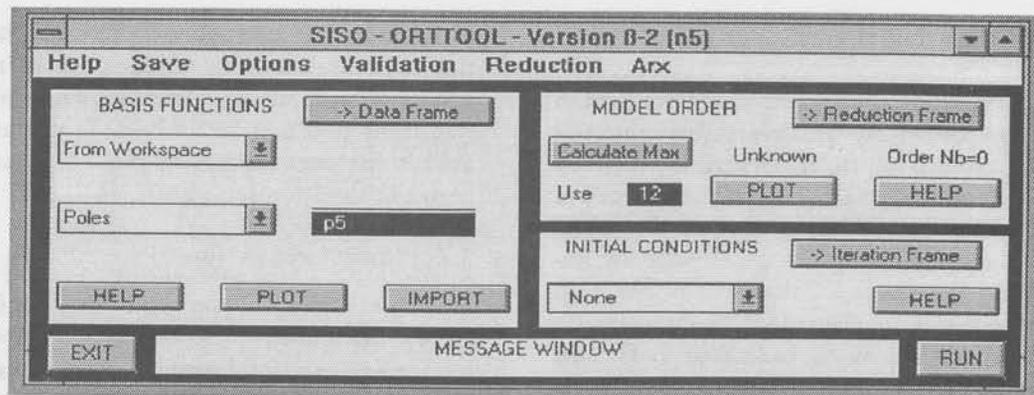


Fig. 1: ORTTOOL main window: primary frames.

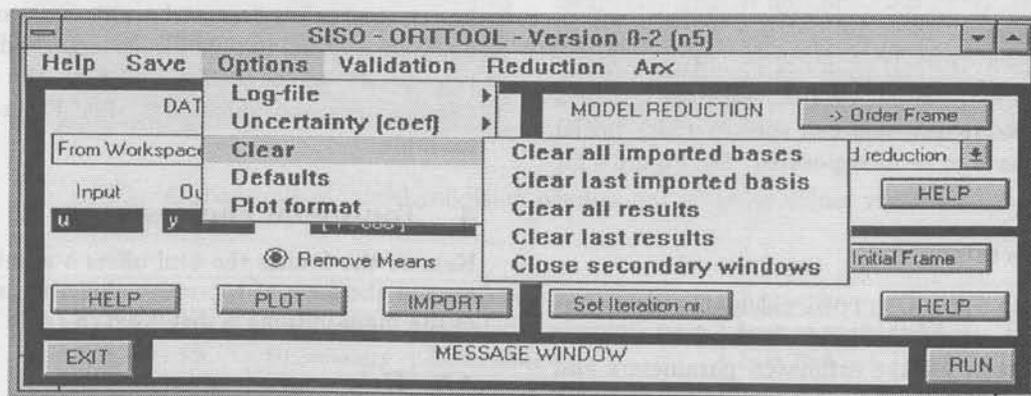


Fig. 2: ORTTOOL main window: pulldown menu's.

3.2 Basis

The frame titled BASIS FUNCTIONS is meant for importing the characteristics of the system based orthonormal basis. As stated these basis functions are completely determined by a set of stable poles (SISO). The program offers 4 different ways to enter the characteristics, but these are all converted to poles before processing. The variable(s) can be located in 4 different locations denoted by:

Workspace	MATLAB's main level storage space.
File	A file in MATLAB's .mat format.
Iterate	Convert the last created results into basis functions.
Ident	Copy the ONLY model selected in the SITB board(s).

3.3 Model order

As stated, the resulting model order is equal to the number of basis functions used. However this number is limited in 2 ways. The first limitation comes from the number of data points N . This number should be larger than the number of functions n ,

to ensure the existence of a unique solution, i.e. to keep an overdetermined set of equations. The second limitation comes from the dynamic character of the basis functions. The basis functions should have damped out sufficiently in the interval $[t_1, t_N]$. If this is not the case the optimization problem may become badly conditioned, but also the resulting model can inhibit undesired behavior. Consider for example the basis functions depicted in Figure 3, when $N = 100$. Such situations should be avoided. The Model Order frame, used to specify the number of basis functions, includes an option to calculate the maximum number of functions (given a basis).

3.4 Initial conditions

Especially if the number of measurements is relatively small, the effect of initial conditions on the output data may be substantial. This effect strongly depends on the underlying system, since slow poles may cause slow damping of the effect. The total number of initial conditions is equal to the number n of orthogonal functions used. Since this would almost double the number of parameters, the tool has an option in the Initial Condition frame to check

wether and which initial conditions are of importance.

The check is based on the regression matrix used in the estimation procedure. This regression matrix R consists of 2 parts $R = [R_u | R_x]$ where R_u depends solely on the input data and R_x solely on the initial conditions. The mutual importance of the initial conditions is determined by a singular value decomposition of the matrix R_x . Initial conditions with small singular values are discarded. The importance of the initial conditions on the estimated coefficients is determined by evaluating the norm of the projection R_p of the matrix R_x on R_u . This norm should neither be large nor small, since a large norm (≈ 1) implies that the initial condition is independent of R_u , while a small norm (≈ 0) shows it is dependent of R_u . In both cases the initial condition will not influence the estimation result. With the aid of a graphics window the user can specify which initial conditions have to be incorporated. See Figure 4 for an example.

3.5 Uncertainty

Analogous to SITB, ORTTOOL calculates the uncertainty in the estimation result in the form of a covariance matrix for the estimated parameters and the variance of the error signal. To this end a third order prewhitening filter is applied to the prediction error signal. The resulting uncertainty bounds can be visualized in SITB.

3.6 Iterations

The theory behind this tool encourages an iterative procedure, improving the a-priori knowledge (i.e. the poles of the basis) by each estimation step, by adaptation of the basis. Although there is yet no convergence result to support such a set-up, it would heuristically be the 'way to go' (i.e. try to force as much knowledge as possible into the basis in order

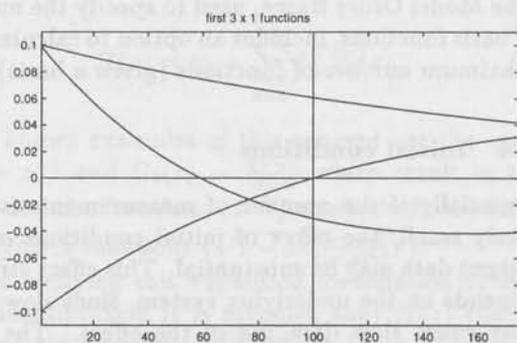


Fig. 3: Example of 3 basis functions.

to facilitate the estimation). The tool has the option to iterate a number of times, where in each step a large order model is estimated, which after reduction leads to a new set of basis functions, that are used in the next estimation step. This number can be set in the Iteration frame.

3.7 Model reduction

In general the estimation result will be a (relatively) large order model, which has to be reduced in order for further use, for instance when using an iterative set-up. To this end the tool features a model reduction option, where a choice from 4 model reduction methods can be made. Three of these are based on so called balanced realizations and the last one is an approximate realization method. The user can specify the reduction order on the basis of a (Hankel) singular value plot. See Figure 4 for an example.

4 Top-menu choices

Next to the frames the tool offers a number of buttons in the form of top-menu choices. An overview of the main options is displayed in table 1.

4.1 Help

Pressing the Help button invokes MATLAB's HTHELP, a hypertext utility for MATLAB help and HTML viewing, with links to various subjects. This utility is part of the MATLAB toolbox UITOOLS.

4.2 Save

Currently the toolbox has only options to save the data and the created results (estimation and reduction). Results can either be saved to files in the standard MATLAB binary format or copied directly to SITB. The data can only be copied to SITB. When saving results, the user can choose between saving only the last created result or all results. Furthermore the format can be specified, either the state space format or the Theta format as used by SITB.

4.3 Options

- Log-file
With the options of this menu a file can be opened in which the most important actions are saved. It can be closed, re-opened, viewed and deleted.
- Uncertainty
In the current version of ORTTOOL, only uncertainty for the estimated parameters is calculated. The uncertainty in the initial conditions is not calculated.

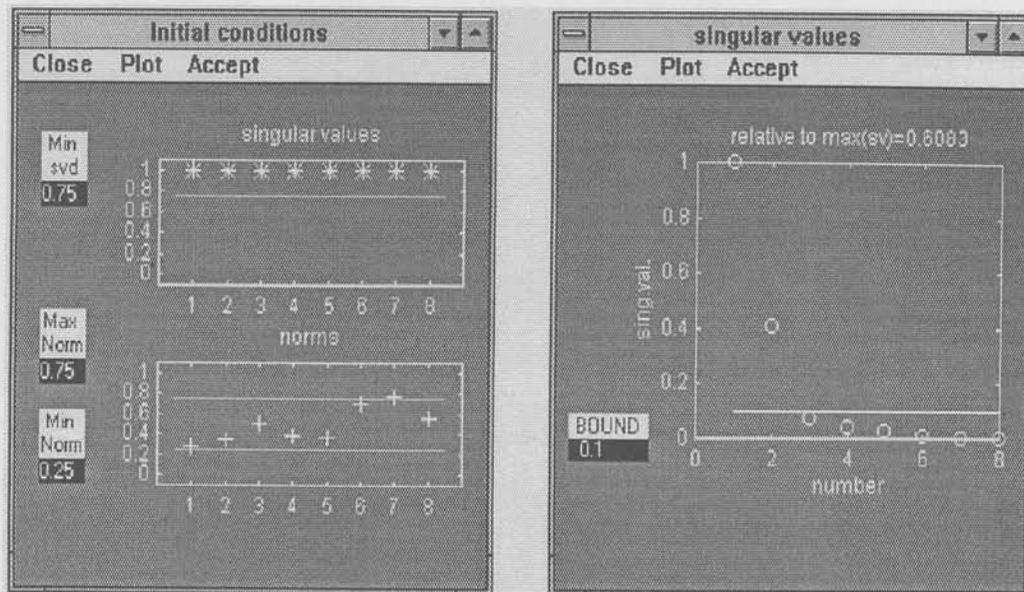


Fig. 4: Example of initial condition window and singular values window.

Help	Save	Options	Validation	Reduction	Arx
	all in A,B,C,D all in THETA last in A,B,C,D last in THETA all to IDENT last to IDENT Data to IDENT	Log-file Uncertainty Clear Defaults Plot Format	Simulation+Impulse resp. Bode plots Prediction error Coefficients Poles+zeros	Last created model Last Full order model Compare with Full model	

Table 1: Orttool Top-menu

- **Clear**
The tool keeps all bases and results in memory. With the clear option these may be removed from memory. Furthermore an option is included to delete all extra graphics ('secondary') windows.
- **Defaults**
With the Default button the user can change the default parameters and filenames used by the tool.
- **Plot Format**
This option is used for the default type of hard-copy files. The user can choose between a number of Postscript and Windows formats.

4.4 Validation

The validation option offers a number of plot choices, to display characteristics of the created model, where possible together with results ob-

tained directly from the data. To this end several SITB functions are used.

4.5 Reduction

With this button a reduced order model is created, based on the settings in the Reduction frame. The user can choose between the last created model or the last created full order model.

4.6 Arx

This button can be used to create an initial model guess, applying an Arx model. The order of the model must be specified in the Model Order frame. This option uses the SITB function ARX.

5 Remaining Options

- **Exit button**
This option quits ORTTOL. Be sure to save your results before issuing this command. A

question dialog box is used to confirm this action.

- Run button
The Run button starts the estimation, using the settings in the various frames, i.e.
 - Imported basis
 - Imported data
 - Model order
 - Iteration number
 - Reduction method (if iterating)
 - Reduction order (if iterating)
 - Initial condition settings
- Graphics
ORTTOOL can create secondary graphics windows at various instances, e.g. to display validation results, basis functions, data, singular values etc. Each window has a button to create a hard-copy.
- Message frame
The Message frame is used to display intermediate steps, warnings and errors. The errors are also included in the log-file.
- Help buttons
Each of the frames has a HELP button to display specific information.

6 Future extensions

In the near future the choice of the poles and the number of functions will become more flexible in order to avoid problems as explained in section 2.1. Furthermore the tool will be expanded to include the estimation of MIMO ($p \times m$) systems. This will offer two possible approaches to estimate a model

$$H(z) = \sum \alpha_k f_k(z)$$

1. MIMO basis functions:

$$\alpha_k^T \in \mathbb{R}^p \text{ and } f_k^T(z) \in \mathbb{R}^m$$

2. Scalar basis functions:

$$\alpha_k \in \mathbb{R}^{p \times m} \text{ and } f_k(z) \in \mathbb{R}$$

7 Summary

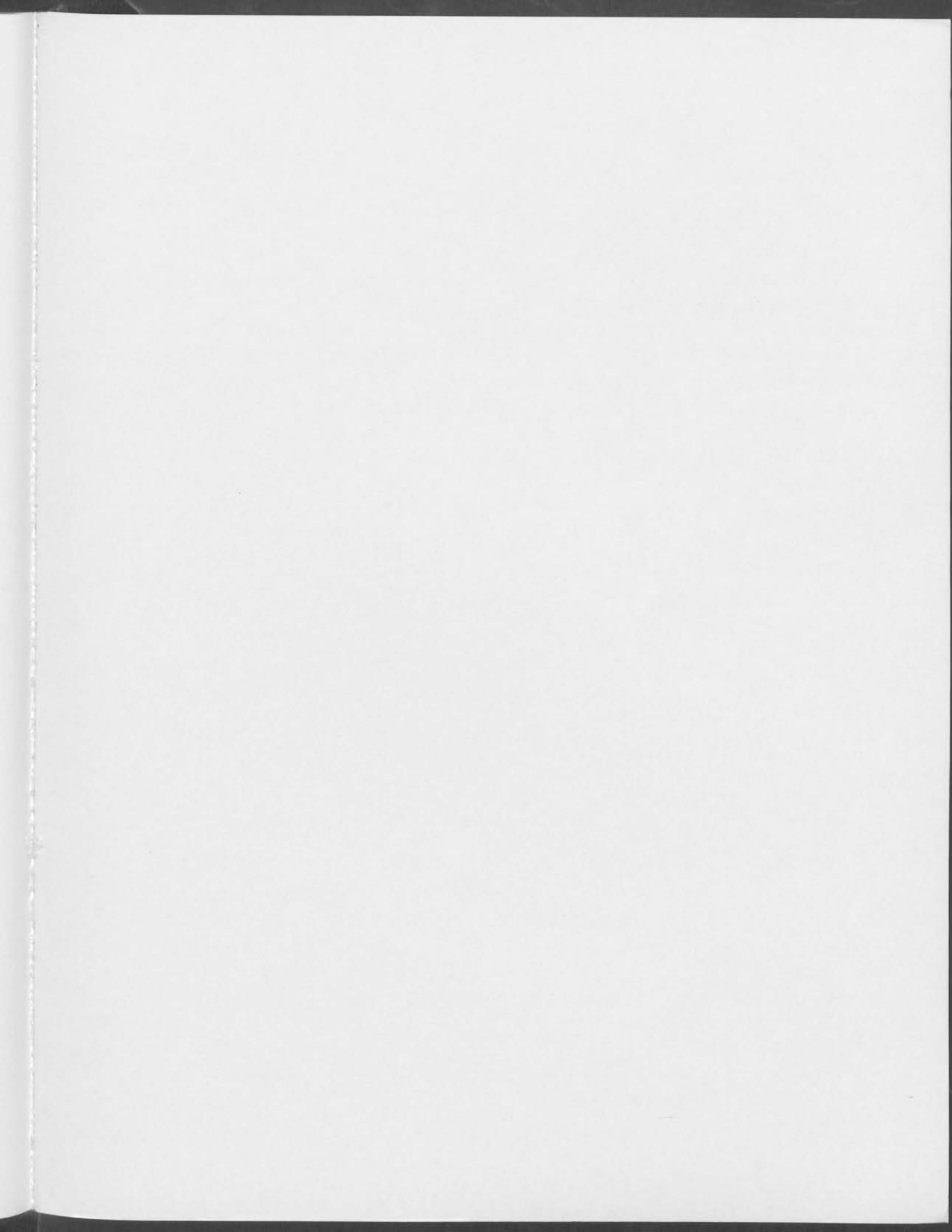
A MATLAB toolbox has been presented for open-loop system identification on the basis of orthonormal basis functions. It is operated through a graphical user interface, that facilitates exchange of models and data with SITB. The current version only supports the estimation of SISO models. The tool includes options for the estimation of initial conditions, model reduction and an iterative approach.

Acknowledgements

The authors acknowledge the contributions of Edwin van Donkelaar and Roel van der Stok to the research and the software implementations involved in this project.

References

- Heuberger, P.S.C. and O.H. Bosgra (1990). Approximate system identification using system based orthonormal functions. *Proc. 29th IEEE Conf. Decis. Control*, Honolulu, HI, pp. 1086–1092.
- Heuberger, P.S.C. (1991). *On Approximate System Identification with System Based Orthonormal Functions*. Dr. Dissertation, Delft University of Technology, The Netherlands, 1991.
- Heuberger, P.S.C., P.M.J. Van den Hof and O.H. Bosgra (1995). A generalized orthonormal basis for linear dynamical systems. *IEEE Trans. Autom. Control*, AC-40, 451–465.
- Ljung (1995). *System Identification Toolbox User's Guide*, Version 4.0, The Mathworks, Inc., Natick, Mass., 1995.
- Ninness, B.M. and F. Gustafsson (1994). *A Unifying Construction of Orthonormal Bases for System Identification*. Report EE9433, Dept. Electrical Engineering, Univ. Newcastle, Australia.
- Van den Hof, P.M.J., P.S.C. Heuberger and J. Bokor (1995). System identification with generalized orthonormal basis functions. *Automatica*, 31, 1821–1834.
- Wahlberg, B. (1991). System identification using Laguerre models. *IEEE Trans. Autom. Control*, AC-36, 551–562.
- Wahlberg, B. (1994a). System identification using Kautz models. *IEEE Trans. Autom. Control*, AC-39, 1276–1282.
- Wahlberg, B. (1994b). Laguerre and Kautz models. In: *Prepr. 10th IFAC Symp. System Identification*, Copenhagen, Denmark, Vol. 3, pp. 1–12.



CONTENTS Volume 9, December 1996

Multivariable closed-loop identification: from indirect identification to dual-Youla parametrization <i>P.M.J. Van den Hof and R.A. de Callafon</i>	1
Asymptotic variance expressions for closed-loop identification and their relevance in identification for control <i>M. Gevers, L. Ljung and P.M.J. Van den Hof</i>	9
Analysis of closed-loop identification with a tailor-made parametrization <i>E.T. van Donkelaar and P.M.J. Van den Hof</i>	17
Identification in view of control design of a CD player <i>H.G.M. Dötsch, P.M.J. Van den Hof, O.H. Bosgra and M. Steinbuch</i>	25
Multivariable least squares frequency domain identification using polynomial matrix fraction descriptions <i>R.A. de Callafon, D. de Roover and P.M.J. Van den Hof</i>	31
Model set determination from a batch of plant models <i>T. Zhou and P.M.J. Van den Hof</i>	39
Closed-loop balanced reduction with application to a compact disc mechanism <i>P.M.R. Wortelboer, M. Steinbuch and O.H. Bosgra</i>	47
Alternative parametrization in modelling and analysis of a Stewart platform <i>S. Koekebakker, P.C. Teerhuis and A.J.J. van der Weiden</i>	59
High-performance motion control of a flexible mechanical servomechanism <i>D. de Roover, O.H. Bosgra, F.B. Sperlring and M. Steinbuch</i>	69
Dualization of the internal model principle in compensator and observer theory <i>D. de Roover and O.H. Bosgra</i>	79
Synthesis of a robust iterative learning controller using an H_{∞} approach <i>D. de Roover</i>	89
Robust generalized H_2 control for uncertain and linear parametrically varying systems with full block scalings <i>C.W. Scherer</i>	97
Robust linear parametrically varying flight control system design with bounded rates <i>D.M.C. Willemsen, S. Bennani and C.W. Scherer</i>	105
Parametrically varying flight control system design with full block scalings <i>R.G.E. Njio, C.W. Scherer and S. Bennani</i>	113
CLOSID - A closed-loop system identification toolbox for Matlab <i>P.M.J. Van den Hof, R.A. de Callafon and E.T. van Donkelaar</i>	121
FREQID - Frequency domain identification toolbox for use with Matlab <i>R.A. de Callafon and P.M.J. Van den Hof</i>	129
ORTTOOL - A Matlab toolbox for system identification with generalized orthonormal basis functions <i>P.S.C. Heuberger and P.M.J. Van den Hof</i>	135