Structured controller synthesis for mechanical servo-systems
algorithms, relaxations and optimality certificates

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 21 november 2006
om 12.30 uur

door

Camile Wilbert José HOL

werktuigkundig ingenieur
geboren te Enschede

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. C.W. Scherer
Prof. ir. O.H. Bosgra

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. C.W. Scherer	Technische Universiteit Delft, promotor
Prof. ir. O.H. Bosgra	Technische Universiteit Delft, promotor
Prof. dr. ir. C. Roos	Technische Universiteit Delft
Prof. dr. ir. M. Steinbuch	Technische Universiteit Eindhoven
Prof. dr. L. Vandenberghe	University of California at Los Angeles
Prof. dr. A. Packard	University of California at Berkely
Dr. ir. M. van de Wal	Philips Applied Technologies Mechatronics Department

Reservelid:

Prof. dr. ir. A. van Keulen	Technische Universiteit Delft

to Heleen

Contents

x

List of symbols

A	Open-loop A-matrix
A_{cl}	Closed-loop A-matrix
A_K	Controller A-matrix
B_1	Open-loop B-matrix of disturbance
B_2	Open-loop B-matrix of control
B_{cl}	Closed-loop B-matrix
B_K	Controller B-matrix
$\mathcal{B}(\gamma, X, K)$	Bounded Real Lemma bilinear matrix
C_1	Open-loop C-matrix of performance
C_2	Open-loop C-matrix of measurement
C^k	k-times continuously differentiable
C_{cl}	Closed-loop C-matrix
C_K	Controller C-matrix
$\det(\cdot)$	Determinant
$\mathrm{diag}(\cdot)$	diagonal augmentation
d_{opt}	Optimal value of dual problem
D_{11}	Open-loop D-matrix of disturbance-performance channel
D_{12}	Open-loop D-matrix of control-performance channel
D_{21}	Open-loop D-matrix of disturbance-measurement channel
D_{22}	Open-loop D-matrix of control-measurement channel
D_{cl}	Closed-loop D-matrix
D_K	Controller D-matrix
dB	decibel
f	Scalar-valued polynomial objective function
F	Matrix-valued polynomial objective function
$\mathcal{F}_{\mathrm{l}}$	Lower LFT
$\mathcal{F}_{\mathrm{u}}$	Upper LFT
g	Scalar-valued polynomial constraint function
G	Matrix-valued polynomial constraint function
$\mathcal{G}$	Feasible set
$\mathcal{H}_2$	Hardy space of complex functions that are analytic in $\mathbb{C}^+$ and have bounded squared integral
$\mathcal{H}_\infty$	Hardy space of complex functions that are analytic in $\mathbb{C}^+$, see page 14 for precise definiton.
Hz	Hertz
I	identity matrix
$\mathrm{Im}(\cdot)$	Image (or range) space
$\Im(\cdot)$	Imaginary part
j	Complex number such that $j^2 = -1$

K	Matrix or transfer function of the controller
$\mathbf{K}$	Controller
$\mathcal{K}$	Set of (allowed) controllers
$\mathrm{Ker}(\cdot)$	Kernel (or null) space
kg	kilogram
$\mathcal{L}_2$	Space of energy-bounded Lebesgue integrable signals
$\log(\cdot)$	Natural logarithm
m	Size of variable x
m_1	Number of disturbance variables
m_2	Number of manipulated variables
min	minutes
mm	millimeter
ms	milliseconds
n	Number of states, size of variable y
N	Newton
n_{c}	Number of controller states
N_{poly}	Number of generators in a polytope
nm	nanometer
r	Size of SOS matrix $S(x)$, size of variable z
P	Transfer function of the plant
P_{11}	Transfer function of disturbance-performance channel of the plant
P_{12}	Transfer function of control-performance channel of the plant
P_{21}	Transfer function of disturbance-measurement channel of the plant
P_{22}	Transfer function of control-measurement channel of the plant
$\mathbf{P}$	Plant
p_1	Number of performance variables
p_2	Number of measured variables
p_{opt}	Optimal value of primal problem
q	Dimension (i.e. number of columns/rows) of $F(x)$
r	Dimension (i.e. number of columns/rows) of $G(x)$
$\Re(\cdot)$	Real part
$\mathcal{RH}_\infty$	Subspace of $\mathcal{H}_\infty$ of real-rational functions
$\mathcal{RL}_\infty$	Space of real-rational and proper functions that are bounded on the imaginary axis
s	Laplace variable
sec	seconds
$s(x)$	Scalar-valued polynomial Sum-Of-Squares
$\mathcal{S}$	Set of symmetric real-valued matrices
$S(x)$	Matrix-valued polynomial Sum-Of-Squares
$s_0(x)$	Scalar-valued polynomial Sum-Of-Squares for relaxation
$\mathrm{svec}(\cdot)$	Symmetric vectorization operator
$S_0(x)$	Matrix-valued polynomial Sum-Of-Squares for relaxation
t	time
$T_{\mathcal{G}}(\cdot)$	Tangent cone
$\mathrm{Trace}(\cdot)$	Trace
$\mathrm{Trace}_r(\cdot)$	Matrix-valued generalization of the trace
u	Manipulated variable
$u(x)$	Monomial vector of SOS matrix

$v(x)$	Monomial vector of constraint function
$\mathrm{vec}(\cdot)$	vectorization operator
x	State variable, optimization variable
X	Regulator storage function matrix
$\mathcal{G}$	Domain of optimization variable x
$\mathcal{X}$	Vector space of variable x
X_{cl}	Closed-loop storage function matrix
X_{ARI}	Regulator storage function matrix of $\mathcal{H}_\infty$-solution with Riccati inequalities
y	Measured variable, optimization variable
Y	Filter storage function matrix, matrix variable
$\mathcal{Y}$	Vector space of variable y
Y_{ARI}	Filter storage function matrix of $\mathcal{H}_\infty$-solution with Riccati inequalities
z	Discrete-time Laplace variable, optimization variable
Z	Dual matrix of X
$\mathcal{Z}$	Vector space of variable z
α	Length of monomial vector $u(x)$
α_0	Length of monomial vector $u_0(x)$
β	Length of monomial vector $v(x)$
γ	$\mathcal{H}_\infty$ upper bound
δ	Uncertain variable
Δ	LFT uncertainty function
η	Length of monomial vector $w(x)$
μ	Barrier parameter
$\mu\mathrm{m}$	micrometer
$\rho_{\mathrm{sp}}(\cdot)$	Spectral radius
$\bar{\sigma}(\cdot)$	Largest singular value
τ	Time constant
$\phi(\cdot)$	Barrier function
$\psi(x)$	Scalar-valued polynomial Sum-Of-Squares for constraint qualification
$\Psi(x)$	Matrix-valued polynomial Sum-Of-Squares for constraint qualification
Ψ	Domain of optimization variable z
ω	Frequency
Ω	Domain of optimization variable y, also set of frequency points
$\mathbb{C}$	Field of complex numbers
$\mathbb{C}^0$	Imaginary axis
$\mathbb{C}^+$	Open right-half plane
$\mathbb{C}^-$	Open left-half plane
$\mathbb{R}$	Field of real numbers
$\mathbb{R}[x]$	commutative ring of polynomials in variable x
$\mathbb{N}$	Set of natural numbers
$\mathbb{Z}$	Set of integers
$\|\cdot\|_{\mathrm{F}}$	Frobenius norm
$\|\cdot\|$	norm, spectral norm
$\|\cdot\|_\infty$	$\mathcal{H}_\infty$-norm
$\otimes$	Kronecker product
$(\cdot)^T$	Transpose
$(\cdot)^*$	Complex conjugate transpose
$\langle\cdot,\cdot\rangle$	Inner product

$\langle \cdot, \cdot \rangle_r$	Matrix-valued generalization of inner product
$\partial(\cdot)$	Derivative
$\partial_+(\cdot)$	Right-derivative
$\nabla(\cdot)$	Gradient
$\nabla^2(\cdot)$	Hessian
$\succeq$	Positive semi-definite
$\succ$	Positive definite
$\preceq$	Negative semi-definite
$\prec$	Negative definite

List of acronyms

Balreal	Balanced reduction
BMI	Bilinear Matrix Inequality
CLIP	Curved Line-search Interior Point method
CC	Cone Complementarity method
flop	floating point operation
FRF	Frequency Response Function
HSV	Hankel Singular Value
IC	Integrated Circuit
IP	Interior Point
LFT	Linear Fractional Transformation
LMI	Linear Matrix Inequality
LP	Linear Program(ming)
MFCQ	Mangasarian-Fromowitz Constraint Qualification
MIMO	Multiple Input Multiple Output
NP	Nondeterministic Polynomial time
PID	Proportional Integral Differential
QQP	Quadratically constrained Quadratic Program
SDP	Semi-Definite Program(ming)
SISO	Single Input Single Output
SOF	Static Output Feedback
SOS	Sum-Of-Squares

Preface

It has been a beautiful time, the last 4.5 years. Really, such a very nice Ph.D-position that I have had, must be rare. It was great to do the research and I enjoyed working together with the people who were involved. I am happy with the obtained results. I also enjoyed very much the short internship at Philips Applied Technologies, supervising graduate students, giving lectures, being a teacher assistant for the mathematics course and Modelvorming and attending the conferences.

Therefore, I would like to thank a few people. First of all, I thank my Promotor Carsten Scherer, for the very good supervision. To learn from his insights is enjoyable in itself. Furthermore, he combines within one person the knowledge on control and systems theory with a nice personality and a lot of dedication for his Ph.D students. Thanks also to my Promotor Okko Bosgra, for giving me the opportunity to start with this project, for the advise he gave during my research and the reading of the draft version of my thesis.

Thanks to Philips Applied Technologies for supporting this research. My special thanks go out to Marc van de Wal from Philips, for his help with the experiments and for reading parts of my thesis. Thanks also to Rob Tousain and Gregor van Baars for their contributions in the discussions during the quarterly meetings.

I am also grateful to Frank Sperling and Maarten Steinbuch for their advises given at the quarterly meetings. Thanks to Ph.D-students Branko, Daniela, David, Dennis, Eduard, Gideon, Ion, Jogchem, Leon, Maarten, Maria, Monique, Martijn L., Martijn S., Matthijs, Redouane, Rob, Sjoerd and Postdoc Hakan for the good times at the University. I owe my thanks to Peter, Adrie, Agnes, Debby, Ellen, Kitty and all other colleagues at the Delft Center for Systems and Control. Thanks also to my graduate students Vital, Edgar, Maarten and Takumi.

I wish to thank the readers of parts of my thesis, Sjoerd, Shane, Maarten and Maria, for their valuable corrections and suggestions.

To my parents Berry & Huby and parents in law Harrie & Janny I am grateful for many things, especially for babysitting Manouk during the final year. Berry, thanks for the beautiful cover. Thanks to my brother Sven for the conversations about mechatronics. Thanks also to my sisters in law Ernee and Marjolein and the friends from Oldenzaal (i.e. Arno & Leonie, Elbert & Jeanette, Eelco, Hendrik-Jan, Jeroen & Melanie, Marc-Jan & Judith, Peter & Nok, Reinold, Walter & Marjan), Edwin, Ilse & Bart, Mark & Nienke, Peter H., my old hockey team (in particular Benno, Floris, Giel, Job and Juri) and all my other friends and family for bringing me joy in our spare time. Finally, many thanks to my wife Heleen and my daughters Manouk and Elisa for all the love in these years.

Chapter 1

Introduction

1.1 Motivation

1.1.1 Accurate positioning of servo-systems

Due to the ever increasing demands on throughput and the miniaturization of electronic devices, the performance requirements of several industrial servo-systems have become tighter and tighter. Examples include the positioning devices used in lithography systems for the production of Integrated Circuits (ICs). In 1965 Moore [134] observed that the number of transistors per square inch doubled every year. He conjectured that the exponential growth would continue in the future. Although recently Moore's prediction turns out to be too optimistic, the complexity of ICs is fast approaching a (presumed to be) fundamental boundary induced by physical limits [130].

This rapid development implies the need for ongoing research in IC production technology. Within a few years the patterns on the ICs will reach a size that cannot be etched any more with visible light (400-700 nm) and one must resort to ultraviolet light. A prototype of a wafer scanner using ultraviolet light is currently developed at Philips Applied Technologies in Eindhoven, the Netherlands. One of the most important servo-systems within this system is the wafer stage. This device positions a silicon disc, the wafer, under a system of lenses or mirrors and the reticle, as illustrated in Figures 1.1. The reticle contains an enlarged image of the pattern to be etched.

This image is projected onto the wafer by a refractive optical system. On top of the wafer a polymer coating is present that reacts to the light. The exposed pattern is removed with a solvent in a different machine. This process is repeated several times, since the IC consists of multiple (typically 20) layers on top of each other. To guarantee that the right vertical interconnections are made on the chip, it is crucial that the layers are positioned accurately. Furthermore, the blurring of the projected image has to be prevented. These objectives can only be achieved if very tight requirements on the motion of the wafer stage are satisfied during the exposure to the light; the required positioning accuracy is in the order of magnitude of nanometers. (A nanometer is 10^{-9} meters.)

1.1.2 Control system architecture

Combined feedforward and feedback controllers are used to achieve these performance requirements as shown in Figure 1.2. In this figure **FF** and **K** denote the feedforward and feedback controllers, respectively. **G** is the wafer stage or any other servo-system.
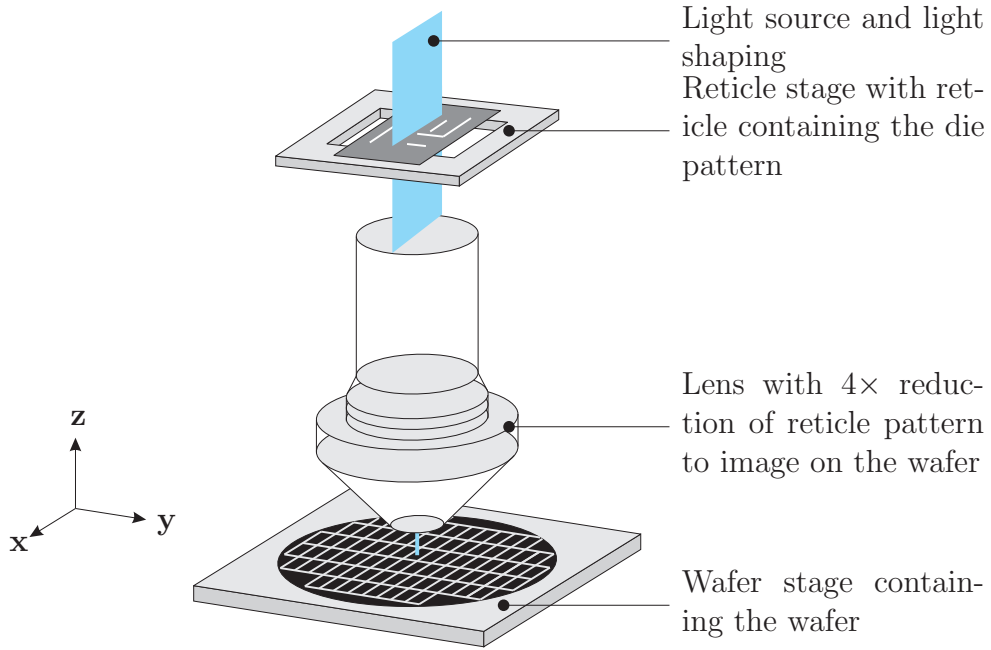
Figure 1.1: Representation of the basic layout of a wafer scanner.
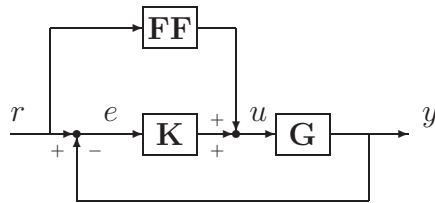
Figure 1.2: Feedback and feedforward controller configuration

The signals r, e, u and y are the reference signal, error signal, manipulated variable and measured variable, respectively. The feedforward is mainly used for trajectory following. It can also compensate for vibrations in the system induced by acceleration and deceleration of the wafer, or for repetitive disturbances. The feedback controller typically stabilizes the closed-loop system and rejects non-deterministic disturbances.

1.1.3 Controller design by tuning

The focus of this thesis is on the design of the *feedback controller*. Nowadays, for each of the six physical degrees of freedom (i.e. the three translations in $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ directions in Figure 1.1 and three rotations), a Single-Input-Single Output (SISO) controller is applied. It usually consists of a Proportional Integral Derivative (PID) Controller combined with a few notches. These controllers are computationally cheap, such that they can be implemented with high sampling rates on a real-time processor. Furthermore they are relatively easy to tune by engineers with basic control knowledge. The six SISO controllers can be considered as a single 6×6 Multiple-Input-Multiple-Ouput (MIMO) controller with only nonzero elements on the diagonal. The requirement on the controller to be a diagonal augmentation of PID controllers with a few notches is an example of a *constraint on the controller structure*.

Interaction between the physical degrees of freedom becomes important around the

desired bandwidth of the wafer stages that are currently under development. So, it is expected that the servo performance of these devices can be improved if this interaction is explicitly taken into account in the design of the controller. A systematic approach to take interaction into account in PID design is, however, lacking and compensating for it by tuning is very difficult even for the experienced control engineer. The direct tuning of a multivariable controller with dynamics on the off-diagonal elements is even more complex.

1.1.4 Controller design by $\mathcal{H}_2$- and $\mathcal{H}_\infty$-synthesis

Model-based controller design by $\mathcal{H}_2$- or $\mathcal{H}_\infty$-optimization provides the control engineer with a systematic procedure for the design of multivariable controllers. In this approach a dynamic model of the system is connected with frequency domain filters to enforce a desired shape of the closed-loop frequency responses. These frequency domain filters or frequency weights can be specified by the control designer. The interconnection of the open-loop system with these shaping filters is, under certain assumptions, called the generalized plant. After selecting the weights, a controller is computed that minimizes the $\mathcal{H}_2$- or $\mathcal{H}_\infty$-norm of the weighted closed-loop transfer functions. Although designing frequency weights that result in a good controller is not easy, it is substantially easier than direct tuning of a multivariable controller.

It has been known since the 1960's that the $\mathcal{H}_2$-optimal controllers can be designed through the solution of two quadratic matrix equations, i.e. Riccati equations. In 1989 [50] it was derived how the $\mathcal{H}_\infty$-optimal (or more precisely suboptimal) controller can be computed by solving two Riccati equations together with a coupling condition. The resulting $\mathcal{H}_2$- or $\mathcal{H}_\infty$-optimal controllers are multivariable transfer functions with McMillan degree equal to that of the model of the generalized plant, where the *McMillan degree* is the minimal length of the state vector of all state-space representations of the model.

1.1.5 Controller structure

The output of $\mathcal{H}_2$ and $\mathcal{H}_\infty$-synthesis is an unstructured controller, i.e. a controller with McMillan degree equal to that of the generalized plant and dynamics in all multivariable transfer function entries. These unstructured controllers have practical drawbacks. First of all, they are difficult to re-tune on-site and often require modification of the controller software for their implementation. This provides motivation to enforce, for instance, a PID structure on the controller with a few notches, possibly restricted to the diagonal elements of the controller.

A second disadvantage of unstructured controllers is their large computation time in real-time implementation. If the model of the system has many modes, the high McMillan degree of the resulting unstructured controller implies that computation of the controller output is expensive, which possibly results in time delays. At very high sampling rates unstructured controllers are not even implementable due to the limited on-line computation time. To solve these problems, the control engineer would like the controller to have a certain a priori fixed maximal McMillan degree. Finally, to avoid drifting of controller states when the control loop is (temporarily) open, it is often desired that the controller is asymptotically stable.

Summarizing, the control engineer often wants to enforce one or more of the following structural controller constraints:

- PID structure with a few notches

- bounds on the McMillan degree of the controller

- diagonal structure and

- stability of the controller itself.

This list is not claimed to be complete and is just an illustration for the need of controller structure in practice.

1.2   Problem formulation

The inability to enforce structure on the controller is one of the main disadvantages of $\mathcal{H}_2$- and $\mathcal{H}_\infty$-optimal control design and one of the main obstacles for a wide-spread application of model-based control in industry. The focus in this thesis is on $\mathcal{H}_\infty$-optimal synthesis, although many of the results are also applicable to $\mathcal{H}_2$-optimal control. This brings us to the topic of this thesis:

> The aim of this thesis is to provide tools for the design of model-based $\mathcal{H}_\infty$-optimal controllers with structural constraints

A more precise problem statement is given at the end of Chapter 2, which contains the necessary theoretical background for this. The structural constraint that we consider are any, or a combination, of the ones listed (with bullets) above. Particular emphasis will be put on the bound on the McMillan degree of the controller, due to its practical importance.

1.3   Outline

The outline of this thesis is as follows. Chapter 2 is a tutorial on the theoretical background that is the fundament for the remainder of this thesis. First of all, the structured controller synthesis problem will be precisely formulated in that chapter. Secondly, we will describe how the unstructured $\mathcal{H}_\infty$-suboptimal controller can be computed by convex optimization. Finally, the chapter provides the necessary background on optimization with Linear Matrix Inequalities (LMIs) and Bilinear Matrix Inequalities (BMIs) and contains a discussion of known complexity-results for the structured controller synthesis problem. On the basis of the material discussed in this chapter we present a more precise formulation of the structured $\mathcal{H}_\infty$-optimal synthesis problem considered in this thesis.

The structural controller design has been an active research area since the 1970s. In Chapter 3 it will become clear that the structured synthesis problem can be approached by techniques of a very different nature. Since we cannot discuss all methods in detail, we will only describe some recent techniques that are relevant for our purposes. References to other methods are provided. This chapter is concluded with a description of the contributions of this research work to the recent literature.

Global optimality certificates for $\mathcal{H}_\infty$-optimal controllers can be computed using the results of Chapter 4. These certificates result from a sequence of LMI relaxations whose optimal values converge from below to the closed-loop $\mathcal{H}_\infty$-norm of the optimal structured controller. The relaxations are based on a Sum-Of-Squares (SOS) decompositions

of matrix-valued polynomials. As one of the key results we show that the optimal value of any polynomial semi-definite program is equal to the optimal value of a certain dual problem that involves SOS polynomials. It is explained how LMI relaxations of these latter problems can be obtained in a straightforward fashion. The method is illustrated on static output feedback controller design examples of McMillan degree 4.

In Chapter 5 we show that lower bounds on the optimal controller performance can also be computed by solving a robust analysis problem. This result allows us to apply any robust analysis technique to the structured controller design problem. We choose relaxations based on the full block S-procedure and on SOS decompositions. Both techniques generate sequences of LMI relaxations, whose number of variables and constraints grow only bi-quadratically in the McMillan degree of the generalized plant. Furthermore both sequences are guaranteed to converge to the closed-loop $\mathcal{H}_\infty$-norm of the optimal structured controller. Based on these results, we are able to compute useful lower bounds for practically relevant control problems, as illustrated on a control problem for an active suspension system with a generalized plant of McMillan degree 27.

In Chapter 6 SOS relaxations are employed to solve robust analysis problems. We extend the relaxation results for polynomial semi-definite programs presented in Chapter 4, by showing that the significantly larger class of robust semi-definite programs are also equivalent to certain dual problems that involve SOS polynomials, in the sense that their optimal values are the same. As a consequence, guarantees for robust performance can be obtained by solving LMI relaxations of these dual problems. The conservatism in these tests is guaranteed to reduce to zero, if the number of variables in the LMIs is allowed to grow.

In Chapter 7 we present the Interior Point algorithm to solve an optimization problem with BMIs. The second-order local optimality conditions for the BMI optimization problem are also presented in this chapter. The algorithm is applied to design a controller for an active suspension system. The generalized plant has McMillan degree 27. The closed-loop $\mathcal{H}_\infty$-norm of the resulting controller is much smaller than that of two alternative methods from the literature. The experimental results on the suspension system illustrate that the method can be applied to controller design in practice.

The *sufficient* 2nd-order local optimality conditions as derived in Chapter 7 will never be satisfied, due to the inherent over-parametrization in a state-space controller. We analyze this effect in Chapter 8 for fixed-order $\mathcal{H}_\infty$-optimal synthesis using BMIs. We present a novel parametrization for which the sufficient conditions *can* be satisfied and, at the same time, covers all controllers of at most the given McMillan degree.

Finally, in Chapter 9 we present the design of SISO and 3×3 MIMO fixed-order controllers for a wafer stage prototype using three synthesis algorithms: the Simplex method, the XK-iteration as discussed in Chapter 3, and the Interior Point algorithm presented in Chapter 7. The steps towards this design will be discussed, i.e. the identification, the model reduction and the four-block $\mathcal{H}_\infty$ design. The results illustrate that the algorithms presented in this thesis can compute well-performing controllers for a very recently developed high-precision servo-system, with very tight (i.e. 5nm) requirements on the positioning accuracy. The experimental results show that the optimized fixed-order controller has much better time-domain performance than a controller obtained by frequency weighted balanced reduction of a full-order controller.

Concluding remarks and suggestions for further research are given in Chapter 10.

Chapter 2

Background and problem definition

In this chapter we present some theoretical background that is required for the remainder of this thesis. We do this by means of a step-by-step explanation of the problem considered in this thesis: the synthesis of optimal model-based controllers with a priori structure. We start in Section 2.1 by describing the feedback-loop with the generalized-plant framework. This is a well-known flexible way to model the interconnection of the physical system and the controller. The objective of the designer is to close the loop with a controller for which the closed-loop plant has a certain desirable behaviour.

These desired properties of the closed-loop system are for instance well-posedness, stability and a bound on the so-called $\mathcal{H}_\infty$-norm of the closed-loop transfer function, as formulated in Section 2.2. We confine ourselves to $\mathcal{H}_\infty$-optimal controller design and only briefly address optimal controller synthesis based on other criteria, such as $\mathcal{H}_2$-optimal and multi-objective control. The synthesis of an unstructured controller can be performed by optimization with Linear Matrix Inequalities (LMIs), as will be explained in Section 2.3. Under certain hypothesis on the system zeros, these LMIs can be converted into Riccati equations and a coupling condition.

For on-line implementation of the controller some specific structure is often required. Such structural constraints on the controller are motivated by practical issues as discussed in Section 2.4. These additional constraints make the synthesis problem significantly harder. The reason for this is that problems with Bilinear Matrix Inequalities (BMIs) are in general much harder to solve than problems with LMIs. An efficient numerical solution of LMI problems is briefly addressed in Section 2.5, together with some additional background on optimization. Finally, complexity results on structural controller synthesis are briefly discussed in Section 2.6. Although most of the material is standard, this chapter contains a derivation of the $\mathcal{H}_\infty$ Riccati equations from the LMI conditions, which to the best of our knowledge, has not been presented in the literature.

Notation The notation is fairly standard. To avoid unnecessary sub- or superscripts we use throughout the thesis the symbol x for the state variable as well as for an optimization variable. For similar reasons y denotes either the measured signal or another optimization variable. Which of the two objects is meant can always be extracted from the context. We use I_n to denote the identity matrix in $\mathbb{R}^{n \times n}$ and often simply write I, if the number of columns/rows is clear from the context.
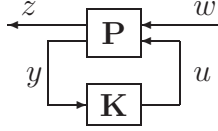
Figure 2.1: Generalized plant interconnection structure

2.1 Systems and their interconnection

2.1.1 Generalized plant

It is common practice to express a control problem with the so-called generalized plant framework as shown in Figure 2.1. This is a very flexible framework to model interconnection of the controller, the system and possibly additional weighting functions, see e.g. [168, 214]. In the set-up of Figure 2.1 w, u z and y are continuous time-signals, i.e. Lebesgue measurable functions mapping $[0, \infty) \subset \mathbb{R}$ to $\mathbb{R}^{m_1}$, $\mathbb{R}^{m_2}$, $\mathbb{R}^{p_1}$ and $\mathbb{R}^{p_2}$ respectively. We use the energy of the signals to measure their 'magnitude'. Let $\mathcal{L}_2^n[0, \infty)$ denote the vector space of Lebesgue integrable functions x mapping $[0, \infty)$ to $\mathbb{R}^n$ such that

$$\int_0^\infty x(t)^T x(t) dt$$

is finite. Since irrelevant for our purposes, we avoid technicalities on sets of measure zero and refer the reader to e.g. [36] or [161] for more details on this issue. We equip $\mathcal{L}_2^n[0, \infty)$ with the norm

$$\|x\|_2 := \sqrt{\int_0^\infty x(t)^T x(t) dt}. \tag{2.1}$$

This norm represents the 'energy' of the signal. The superscript n of $\mathcal{L}_2^n[0, \infty)$ is called the spatial dimension. We will also need the function space $\mathcal{L}_2^n(-\infty, 0]$ which is defined analogously [74, 214]. If the domain is $[0, \infty)$ we will usually write $\mathcal{L}_2$ instead of $\mathcal{L}_2^n[0, \infty)$ to simplify notation, where we also omit the spatial dimension n.

The space $\mathcal{L}_2$ is too restrictive to describe all signals of interest, since we also consider unstable systems. We therefore introduce a class of signals that have finite energy over any finite interval. For this purpose we define the truncation operator P_τ for $\tau \geq 0$ on vector valued functions by

$$(P_\tau x)(t) := \begin{cases} x(t), & 0 \leq t \leq \tau \\ 0, & t > \tau \end{cases},$$

where $x : [0, \infty) \to \mathbb{R}^m$ is an arbitrary m-dimensional signal. The extended $\mathcal{L}_2$-space is defined by

$$\mathcal{L}_{2e}[0, \infty) := \{ x : [0, \infty) \to \mathbb{R}^m | \ x \text{ is Lebesgue measurable and } P_T x \in \mathcal{L}_2[0, \infty) \text{ for all } T \geq 0 \}$$

The signals in Figure 2.1 have the following interpretation. The *generalized disturbance* signal w represents all signals that affect the system from the environment. These are for instance disturbance and reference signals or weighted versions thereof. The *generalized performance* z is the collection of signals that are considered to be important for the closed-loop performance. z usually includes error signals and the controller outputs. The *input signal* u, also referred to as *control signal* or *manipulated variable*, can be adjusted by the
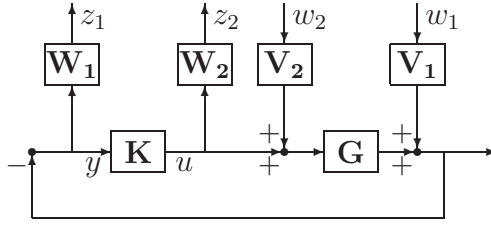
Figure 2.2: Four-block controller set-up

controller. The input to the controller is the vector y of measured signals.

The *plant* $\mathbf{P}$ (as well as all other dynamical systems considered in this thesis) is assumed to be a Linear Time-Invariant (LTI) dynamical system that admits a state space representation i.e. a mapping from an initial state $x_0 \in \mathbb{R}^n$ and input signals $w \in \mathcal{L}_{2e}^{m_1}$ and $u \in \mathcal{L}_{2e}^{m_2}$ into output signals $z \in \mathcal{L}_{2e}^{p_1}$ and $y \in \mathcal{L}_{2e}^{p_2}$, uniquely described by the equations

$$\begin{pmatrix} \dot{x} \\ z \\ y \end{pmatrix} = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array} \right) \begin{pmatrix} x \\ w \\ u \end{pmatrix}, \qquad x(0) = x_0 \in \mathbb{R}^n, \tag{2.2}$$

for some matrices $A \in \mathbb{R}^{n \times n}, B_i \in \mathbb{R}^{n \times m_i}, C_j \in \mathbb{R}^{p_i \times n}$ and $D_{ij} \in \mathbb{R}^{p_i \times m_i}$ $i, j = 1, 2$. Except for Section 2.2.3 where we define internal stability, we assume in this thesis that all dynamical systems have zero initial state $x_0 = 0$. Under this assumption the output vector $\begin{pmatrix} z^T & y^T \end{pmatrix}$ is uniquely determined by the input vector $\begin{pmatrix} w^T & u^T \end{pmatrix}$.

We only consider LTI controllers in this thesis. Let the controller $\mathbf{K}$ admit the state-space representation

$$\begin{pmatrix} \dot{x}_K \\ u \end{pmatrix} = \left(\begin{array}{c|c} A_K & B_K \\ \hline C_K & D_K \end{array} \right) \begin{pmatrix} x_K \\ y \end{pmatrix}, \qquad x_K(0) = x_{K,0} \in \mathbb{R}^{n_c}, \tag{2.3}$$

for some matrices $A_K \in \mathbb{R}^{n_c \times n_c}$, $B_K \in \mathbb{R}^{n_c \times p_2}$, $C_K \in \mathbb{R}^{m_2 \times n_c}$ and $D_K \in \mathbb{R}^{m_2 \times p_2}$. $x_{K,0}$ is the initial state of the controller, which is assumed to be zero as well.

Example 2.1 *An example of an interconnection structure that is often used for controller synthesis is shown in Figure 2.2, where $\mathbf{K}$ is the controller, $\mathbf{G}$ is the controlled system and $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{W}_1$ and $\mathbf{W}_2$ are weighting filters. This so-called* four-block *controller design problem will be used in Chapters 7 and 9 for controller design.*

2.1.2   Transfer functions

Consider a system with a state-space description

$$\begin{pmatrix} \dot{x} \\ y \end{pmatrix} = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) \begin{pmatrix} x \\ u \end{pmatrix},$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$ and $D \in \mathbb{R}^{p \times m}$. The *transfer function* $G : \mathbb{C} \to \mathbb{C}^{p \times m}$ of this system is given by

$$G(s) := D + C(sI - A)^{-1}B. \tag{2.4}$$

9

If the number of inputs m or outputs p is larger than one, we also use the terminology *transfer matrix*. $G(s)$ is well-defined for all $s \in \mathbb{C}$ for which the matrix $(sI - A)$ is invertible. For some transfer function G, suppose that $n \in \mathbb{N}$ is the minimal dimension of A to describe it by (2.4). Then G has *McMillan degree n* [183]. The McMillan degree is often called the *order* in literature, and we will also use this terminology. The corresponding quadruple (A, B, C, D) is called a *minimal realization* of G. The transfer function of a state space system is always *proper* [74], which means that $\lim_{\omega \to \infty} G(j\omega)$ exists. We write this limit as $G(\infty)$.

The transfer functions of the systems **P** and **K** are denoted by P and K, respectively. If the closed-loop system is *well-posed* (as will be defined in Section 2.2.1), the closed-loop transfer function is given by the lower Linear Fractional Transformation (LFT):

$$\mathcal{F}_l(P, K) := P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}. \tag{2.5}$$

Example 2.2 *(Example 2.1 continued.) Let V_i and W_i $i = 1, 2$ be the transfer functions of the shaping filters $\mathbf{V}_i$ and $\mathbf{W}_i$ $i = 1, 2$ respectively in Figure 2.2. The transfer function of the plant of this interconnection is*

$$P = \left(\begin{array}{cc|c} -W_1 V_1 & -W_1 G V_2 & -W_1 G \\ 0 & 0 & W_2 \\ \hline -V_1 & -G V_2 & -G \end{array} \right).$$

If $I + G(\infty)K(\infty)$ is invertible, the closed-loop transfer function is:

$$\mathcal{F}_l(P, K) = -\left(\begin{array}{cc} W_1 S V_1 & W_1 S G V_2 \\ W_2 K S V_1 & W_2 K S G V_2 \end{array} \right), \tag{2.6}$$

where $S = (I + GK)^{-1}$. Four closed-loop transfer function matrices appear in (2.6): the sensitivity S, process sensitivity SG, the control sensitivity KS and KSG.

The tuning of weighting filters in the generalized plant is certainly not trivial, although simpler than direct tuning of multi-variable controllers as mentioned in Chapter 1. In Chapters 7 and 9 the designs of these weights are presented for an active suspension system and a wafer stage respectively. For now let us assume them to be given, such that **P** is a fixed system with a state-space representation.

2.2 Closed-loop requirements

The control design problem examined in this thesis is to find a state-space controller **K** such that

- the interconnection is well-posed,

- the closed loop is stable,

- the $\mathcal{H}_\infty$- or $\mathcal{H}_2$-norm of the closed-loop is minimized and

- the controller has a certain structure.

The first two requirements are natural, in the sense that a controller that is to be implemented in practice almost always satisfies them. The third condition is used to enforce a certain desired behavior on the closed-loop. These three requirements will be explained and motivated in the following sections. The discussion of structural controller constraints is postponed to Section 2.4.
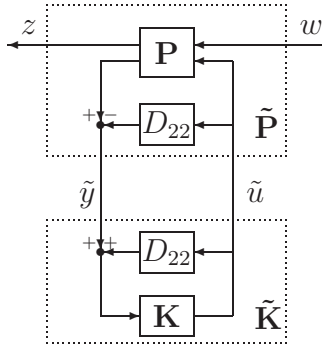
Figure 2.3: Loop transformation to eliminate D_{22}

2.2.1 Well-posedness

As a first requirement, the closed-loop system must be well-posed, which means that for all initial conditions $x(0) \in \mathbb{R}^n$ and $x_k(0) \in \mathbb{R}^{n_c}$ and all disturbance signals $w \in \mathcal{L}_2$ the equations (2.2) and (2.3) have a unique solution. The closed-loop system is well-posed if $I - D_{22}D_K$ is nonsingular [52]. The interconnection in 2.1 of the transfer functions P and K is well-posed if

$$
\begin{pmatrix} I & -K(\infty) \\ -P_{22}(\infty) & I \end{pmatrix}
$$

is nonsingular [168].

Note that the system is obviously well-posed if $D_{22} = 0$, which is often true for controller synthesis for mechanical servo-systems, as will be argued in the next section.

2.2.2 Direct Feedtrough term D_{22}

The right lower part of the plant P_{22} usually corresponds to the model of the physical system. In the four-block settings for the active suspension system in Section 7.4 and the wafer stage in Chapter 9 this is indeed the case. A mechanical servo-system is often controlled with a position measurement, which is fed back to the drives that apply a force on the system. Due to the rigid-body dynamics these systems usually have roll-off at high frequencies, which implies that the direct feed-through of the model of the system is zero. Hence for the proposed four-block set-up (and practically for all other $\mathcal{H}_\infty$ or $\mathcal{H}_2$ designs for positioning systems), the direct feed-through matrix satisfies $D_{22} = 0$. To simplify the presentation throughout this thesis, we therefore assume $D_{22} = 0$, and give at relevant places the necessary remarks about the consequences of this assumption.

Assumption 2.3 *The direct feed-through term in P from u to y is zero, i.e. $D_{22} = 0$.*

In case $D_{22} \neq 0$ we can modify the generalized plant set-up such that this assumption is satisfied. To this purpose consider the interconnection in Figure 2.1. Observe that the closed loop interconnection in terms of **P** and **K** is the same as in Figure 2.3. The plant $\tilde{\mathbf{P}}$ has state-space description

$$
\begin{pmatrix} \dot{x} \\ \hline z \\ \hline \tilde{y} \end{pmatrix} = \left(\begin{array}{c|c|c} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ \hline C_2 & D_{21} & 0 \end{array} \right) \begin{pmatrix} x \\ \hline w \\ \hline \tilde{u} \end{pmatrix} \quad x(0) = x_0
$$

11

which has the desired zero direct-feedthrough from $\tilde{u}$ to $\tilde{y}$. Suppose now we have designed a controller $A_{\tilde{K}}, B_{\tilde{K}}, C_{\tilde{K}}, D_{\tilde{K}}$ for plant $\hat{\mathbf{P}}$ with state-space representation (2.7). It is derived in Appendix A.1 that under the hypothesis that $Q := I + D_{\tilde{K}}D_{22}$ is nonsingular, the controller $\mathbf{K}$ in Figure 2.3 is given by

$$\left(\begin{array}{c|c} A_K & B_K \\ \hline C_K & D_K \end{array} \right) = \left(\begin{array}{c|c} A_{\tilde{K}} - B_{\tilde{K}}D_{22}Q^{-1}C_{\tilde{K}} & B_{\tilde{K}} - B_{\tilde{K}}D_{22}Q^{-1}D_{\tilde{K}} \\ \hline Q^{-1}C_{\tilde{K}} & Q^{-1}D_{\tilde{K}} \end{array} \right). \qquad (2.7)$$

If Q is singular, we can always perturb the controller $\tilde{K}$ such that $I + D_{\tilde{K}}D_{22}$ becomes non-singular. For small enough perturbations internal stability and $\mathcal{H}_\infty$-performance bounds on the closed-loop system are preserved. This is true since the set of internally stabilizing controllers parameterized with state-space quadruple $(A_K, B_K, C_K D_K)$ and satisfying a strict upper bound on the $\mathcal{H}_\infty$ performance, is open. This implies preservation of stability and $\mathcal{H}_\infty$-performance bounds, since the closed-loop eigenvalues and the closed-loop $\mathcal{H}_\infty$-norm depend continuously on $(A_K, B_K, C_K D_K)$. The precise definitions of internal stability and $\mathcal{H}_\infty$-performance bounds on the closed-loop are given in the next two sections, respectively.

2.2.3 Stability

We require the closed-loop system to be stable.

Closed-loop stability in terms of a state-space description

The state space realization of the closed-loop in Figure 2.1 follows from (2.2) and (2.3). Using the definitions

$$\left(\begin{array}{c|c|c} \hat{A} & \hat{B}_1 & \hat{B}_2 \\ \hline \hat{C}_1 & \hat{D}_{11} & \hat{D}_{12} \\ \hline \hat{C}_2 & \hat{D}_{21} & 0 \end{array} \right) := \left(\begin{array}{cc|cc|cc} A & 0 & B_1 & 0 & B_2 \\ 0 & 0 & 0 & I & 0 \\ \hline C_1 & 0 & D_{11} & 0 & D_{12} \\ 0 & I & 0 & 0 & 0 \\ \hline C_2 & 0 & D_{21} & 0 & 0 \end{array} \right)$$

and

$$K := \left(\begin{array}{c|c} A_K & B_K \\ \hline C_K & D_K \end{array} \right),$$

we can compactly describe the closed-loop matrices as:

$$\left(\begin{array}{c|c} A_{\mathrm{cl}}(K) & B_{\mathrm{cl}}(K) \\ \hline C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) \end{array} \right) := \left(\begin{array}{c|c} \hat{A} & \hat{B}_1 \\ \hline \hat{C}_1 & \hat{D}_{11} \end{array} \right) + \left(\begin{array}{c} \hat{B}_2 \\ \hline \hat{D}_{12} \end{array} \right) K \left(\begin{array}{c|c} \hat{C}_2 & \hat{D}_{21} \end{array} \right). \qquad (2.8)$$

By augmenting the controller state to the plant state as in $x_{\mathrm{cl}} = \left(\begin{array}{cc} x^T & x_K^T \end{array} \right)^T$, a state-space representation of the closed-loop system is then given by

$$\left(\begin{array}{c} \dot{x}_{\mathrm{cl}} \\ \hline z \end{array} \right) = \left(\begin{array}{c|c} A_{\mathrm{cl}}(K) & B_{\mathrm{cl}}(K) \\ \hline C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) \end{array} \right) \left(\begin{array}{c} x_{\mathrm{cl}} \\ \hline w \end{array} \right).$$

Remark. Observe that K in (2.8) is a *matrix* in $\mathbb{R}^{(n_c + m_1) \times (n_c \times p_2)}$. So, by augmenting the controller state to the plant state the dynamic controller synthesis problem is converted into a static controller synthesis or Static Output Feedback (SOF) problem.

The closed loop system is *internally stable* if for any pair of initial conditions $x_0 \in \mathbb{R}^n$, $x_K \in \mathbb{R}^{n_c}$ and zero disturbance $w(t) = 0$ for all $t \geq 0$, the closed-loop state $x_{\text{cl}} := \begin{pmatrix} x^T & x_K^T \end{pmatrix}^T$ converges to zero, i.e. $\lim_{t \to \infty} x_{\text{cl}}(t) = 0$. Closed-loop internal stability is equivalent to A_{cl} being *Hurwitz*, i.e. all the eigenvalues of A_{cl} are in the left-half plane $\mathbb{C}^-$ [183]. Lyapunov's stability theorem for LTI systems provides a necessary and sufficient condition in terms of linear matrix inequalities for $A_{\text{cl}}(K)$ (for fixed K) to be Hurwitz: there exists $X \in \mathcal{S}^{n+n_c}$ such that

$$X \succ 0 \text{ and } A_{\text{cl}}^T(K)X + XA_{\text{cl}}(K) \prec 0 \tag{2.9}$$

where $\mathcal{S}^{n+n_c}$ denotes the subspace of symmetric matrices in $\mathbb{R}^{(n+n_c) \times (n+n_c)}$. If F is symmetric then $F \succ 0$ ($F \succeq 0$) denotes that F is positive definite (positive semi-definite), i.e. $x^T F x > 0$ ($x^T F x \geq 0$) for all $x \neq 0$. $F \succ G$ should be read as

$$F \in \mathcal{S}^n, \ G \in \mathcal{S}^n \ \text{ and } \ F - G \succ 0.$$

The condition (2.9) is indeed a linear matrix inequality (LMI), which will be described in more detail in Section 2.5.1.

We assume in the rest of this thesis that the pairs A, B_2 and A, C_2 are stabilizable and detectable respectively, i.e.

$$\begin{pmatrix} A - sI & B_2 \end{pmatrix} \text{ and } \begin{pmatrix} A - sI \\ C_2 \end{pmatrix}$$

have full row rank and full column rank respectively for all $s \in \mathbb{C}^0 \cup \mathbb{C}^+$.

Assumption 2.4 (A, B_2) *and* (A, C_2) *are stabilizable and detectable respectively.*

There exists at least one controller K that internally stabilizes the closed-loop if and only if this assumption holds true. If $\mathbf{P}$ satisfies this property, we call it a *generalized plant* [168].

Stability in terms of transfer functions

A transfer matrix G is said to be stable if its poles are in the open left-half-plane. If the A-matrix of a state-space realization is Hurwitz, then the corresponding transfer matrix is stable. On the other hand, every minimal realization of a stable transfer function has a Hurwitz A-matrix. There exists a simple condition [168] for internal stability of the closed loop if $\mathbf{P}$ with transfer function P is a generalized plant, and K is a transfer function for $\mathbf{K}$. Under these conditions internal stability is equivalent to the condition:

$$\begin{pmatrix} I & -K \\ -P_{22} & I \end{pmatrix} \text{ has a proper and stable inverse.}$$

2.2.4 Performance

$\mathcal{H}_\infty$ **performance**

A commonly used performance criterion for optimal controller design is the $\mathcal{H}_\infty$-norm of the closed-loop transfer function from w to z in Figure 2.1. The $\mathcal{H}_\infty$-norm of a transfer function is the worst-case energy gain (as more precisely formulated in the sequel). The importance of $\mathcal{H}_\infty$ controller optimization stems from the possibility to compute robust controllers with it, i.e. controllers that are stable and satisfy the performance specification

for a whole set of system models. These robust stability and performance guarantees are based on the small-gain theorem, Theorem 2.5 in this section. A more detailed examination on robust performance and stability analysis is presented in Chapter 6.

The closed-loop $\mathcal{H}_\infty$-norm is also a convenient performance requirement for loop-shaping, since it enables to suppress undesirable large spikes in the transfer functions of the closed-loop system. This is not possible with the same ease if, for instance, the $\mathcal{H}_2$-norm is used.

Consider a state-space system (A, B, C, D), where $A \in \mathbb{R}^{n \times n}$ is Hurwitz and $D \in \mathbb{R}^{p \times m}$. Its transfer function $G(s) = D + C(sI - A)^{-1}B$ belongs to the function space $\mathcal{H}_\infty^{p \times m}$, defined as the vector space of functions $F : \mathbb{C}^+ \to \mathbb{C}^{p \times m}$, that

- are analytic in $\mathbb{C}^+$ and

- are norm-bounded on $\mathbb{C}^+$, i.e. $\sup_{s \in \mathbb{C}^+} \bar{\sigma}(G(s)) < \infty$,

where $\mathbb{C}^+$ denotes the open right-half plane and $\bar{\sigma}(A)$ denotes the maximum singular value of the matrix A. In this definition and the remainder of this thesis we omit technicalities with sets of measure zero, as has also been mentioned on page 8. Furthermore, we will omit the dimensions m and p and write $\mathcal{H}_\infty$ instead of $\mathcal{H}_\infty^{p \times m}$ for notational brevity (assuming no confusion arises as a result). The $\mathcal{H}_\infty$-norm of an element $F \in \mathcal{H}_\infty$ is

$$\|F\|_\infty := \sup_{s \in \mathbb{C}^+} \bar{\sigma}(F(s))$$

By the Maximum Modulus Principle (as for instance described in [183]) we know that this norm is determined by the values of F on the imaginary axis:

$$\|F\|_\infty = \sup_{\omega \in \mathbb{R}} \bar{\sigma}(F(j\omega)). \tag{2.10}$$

A stable transfer function G given by (2.4) belongs to the subspace of $\mathcal{H}_\infty$ of real-rational, stable and proper transfer functions, denoted by $\mathcal{RH}_\infty$. In this thesis we will also use the space $\mathcal{RL}_\infty$ of real-rational, proper transfer functions that are bounded on the imaginary axis. Note that $\mathcal{RH}_\infty \subset \mathcal{RL}_\infty$. The $\mathcal{H}_\infty$-norm of $G \in \mathcal{RH}_\infty$ with $G(s) = D + C(sI - A)^{-1}B$ is the worst-case energy gain [176]:

$$\|G\|_\infty = \sup_{w \in \mathcal{L}_2^m, \ w \neq 0, \ x(0)=0, \ \dot{x}=Ax+Bw, \ z=Cx+Dw} \frac{\|z\|_2}{\|w\|_2},$$

where $\|\cdot\|_2$ denotes the $\mathcal{L}_2$-norm of a signal, as defined in (2.1). Minimizing the $\mathcal{H}_\infty$-norm of the closed-loop system thus implies minimizing the amplification of the worst-case generalized disturbance signal w. A more important technical reason for using the $\mathcal{H}_\infty$-norm is the small-gain theorem. This theorem provides a sufficient condition for $I + G$ to have a stable and proper inverse, which is one of the fundamental theorems for robust control theory:

Theorem 2.5 *(Small Gain Theorem.) If $G \in \mathcal{RH}_\infty$ and $\|G\|_\infty < 1$ then $(I + G)$ has an inverse $(I + G)^{-1} \in \mathcal{RH}_\infty$.*

Proof. See e.g. [52, 214].

■

Bounded-Real Lemma

The $\mathcal{H}_\infty$-norm of a transfer function in $\mathcal{RH}_\infty$ can be computed by LMI optimization using the Bounded Real Lemma (BRL) which is also known as the Kalman-Yakubovich-Popov (KYP) Lemma.

Lemma 2.6 *(Bounded Real Lemma) Consider the state-space quadruple (A, B, C, D). Then A is Hurwitz and $\|D + C(sI - A)^{-1}B\|_\infty < \gamma$ if and only if there exists $X \in \mathcal{S}^n$ with $X \succ 0$ and*

$$\begin{pmatrix} A^T X + XA & XB & C^T \\ B^T X & -\gamma I & D^T \\ C & D & -\gamma I \end{pmatrix} \prec 0 \tag{2.11}$$

Proof. See for instance [52]. ∎

This is an important result in systems theory, that links a condition in the frequency domain to a finite dimensional LMI condition. If A is Hurwitz, the frequency domain condition is equivalent to the dissipation condition:

there exists $\epsilon > 0$ such that $\quad \int_0^\infty z(t)^T z(t) - (\gamma^2 - \epsilon) w(t)^T w(t) dt \leq 0 \quad$ for all $w \in \mathcal{L}_2$.

where the responses $x(t)$ and $z(t)$ for $t \in [0, \infty)$ are governed by $\dot{x} = Ax + Bw$ and $z = Cx + Dw$ respectively with $x(0) = 0$ as initial condition. The equivalence of the frequency domain inequality and dissipativity follows from Parseval's relation [116, 214] between energy in the frequency and time domain. The Bounded-Realness property of a system is a special case of 'dissipativity of a linear system with quadratic supply functions' [176], which has an LMI characterization that generalizes the one in Lemma 2.6.

Riccati Inequality

Inequality (2.11) is equivalent to a Riccati inequality together with a bound on the norm of the matrix D. To show this, we need the following Schur complement lemma, which will also be of use in the remainder of the thesis.

Lemma 2.7 *(Schur lemma)*

$$\begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \succ 0 \tag{2.12}$$

if and only if

$$Q \succ 0 \ and \ R - S^T Q^{-1} S \succ 0. \tag{2.13}$$

Proof. (2.12) obviously implies $Q \succ 0$. Hence we can perform the congruence transformation

$$\begin{pmatrix} I & 0 \\ -S^T Q^{-1} & I \end{pmatrix} \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \begin{pmatrix} I & -Q^{-1}S \\ 0 & I \end{pmatrix} = \begin{pmatrix} Q & 0 \\ 0 & R - S^T Q^{-1} S \end{pmatrix} \tag{2.14}$$

Since the inertia of a symmetric matrix is invariant under congruence transformation [94], (2.12) implies

$$\begin{pmatrix} Q & 0 \\ 0 & R - S^T Q^{-1} S \end{pmatrix} \succ 0, \tag{2.15}$$

such that (2.13) follows. On the other hand, (2.13) implies (2.15), and using the inverse congruence transformation we conclude that (2.12) holds true.

∎

Let us define R as

$$R := \gamma^2 I - D^T D. \tag{2.16}$$

By taking a Schur complement of (2.11) twice we infer that

there exists an $X \succ 0$ satisfying (2.11)

if and only if $R \succ 0$ and there exists an $X \succ 0$ with

$$(A + BR^{-1}D^T C)^T X + X(A + BR^{-1}D^T C) + XBR^{-1}B^T X + C^T (I + DR^{-1}D^T)C \prec 0. \tag{2.17}$$

Furthermore, if $R \succ 0$ and A is Hurwitz, the Riccati Inequality (2.17) has a solution $X \succ 0$ if and only if [168, 214, 203] the Hamiltonian matrix

$$H := \begin{pmatrix} A + BR^{-1}D^T C & BR^{-1}B^T \\ -C^T(I + DR^{-1}D^T)C & -(A + BR^{-1}D^T C)^T \end{pmatrix} \tag{2.18}$$

has no eigenvalues on the imaginary axis. The preceding arguments are summarized in the following lemma.

Lemma 2.8 *Let A be Hurwitz and R be defined by (2.16), then the following conditions are equivalent:*

- $\|D + C(sI - A)^{-1}B\|_\infty < \gamma$,

- $R \succ 0$ and H in (2.18) has no eigenvalues on the imaginary axis,

- $R \succ 0$ and (2.17) has a solution $X \succ 0$.

Proof. Direct consequence of results in e.g. [168] or [214].

∎

The lemma implies that we can verify if $\|D + C(sI - A)^{-1}B\|_\infty < \gamma$ by computation of $\|D\|$ and an eigenvalue computation of the Hamiltonian matrix H. This allows us to compute the $\mathcal{H}_\infty$-norm by bisection on γ and Schur decompositions of H, instead of LMI optimization. The bisection can be replaced by another one-dimensional search method presented in [31], that is often more efficient. This latter method is used for several numerical examples presented in this thesis.

Since computing the $\mathcal{H}_\infty$-norm with Riccati equations is in general computationally more efficient and reliable than LMI optimization, it is used for $\mathcal{H}_\infty$-norm computations in software packages like the Matlab Control Toolbox [7] and Slicot [14].

2.3 LMI solution for full order unstructured $\mathcal{H}_\infty$ synthesis

2.3.1 Transformation to convex problem

This section presents the LMI approach to full order $\mathcal{H}_\infty$-optimal controller synthesis problem [99, 64, 176]. For a generalized plant with transfer function P the goal is to compute a controller transfer function K, such that the closed loop is internally stable and the $\mathcal{H}_\infty$ norm of the closed-loop $\|\mathcal{F}_l(P, K)\|_\infty$ given by (2.5) is minimized:

$$
\gamma_{\text{opt}} := \quad \text{infimum} \qquad \|\mathcal{F}_l(P, K)\|_\infty
$$
$$
\text{subject to} \qquad \begin{pmatrix} I & -P_{22} \\ -K & I \end{pmatrix}^{-1} \in \mathcal{RH}_\infty \quad .
$$

Verifying existence and (if it exists) constructing the optimal controller is often complicated [165, 164, 188]. In practice it suffices to compute almost-optimal controllers.

The full order $\mathcal{H}_\infty$-optimal controller synthesis problem is equivalent to an LMI problem. We will show this equivalence using some algebraic manipulations and two lemma's: the Bounded Real-Lemma and the Projection Lemma. Furthermore we will show that, under additional assumptions on **P**, the LMI conditions reduce to the well-known Riccati equations [49].

The closed loop transfer function from w to z is given by

$$
\mathcal{F}_l(P, K)(s) = D_{\text{cl}}(K) + C_{\text{cl}}(K)(sI - A_{\text{cl}}(K))^{-1} B_{\text{cl}}(K).
$$

where $A_{\text{cl}}(K)$, $B_{\text{cl}}(K)$, $C_{\text{cl}}(K)$ and $D_{\text{cl}}(K)$ are as in (2.8). If we apply the Bounded Real Lemma, we observe that the closed-loop system is stable and has $\mathcal{H}_\infty$-norm smaller than γ if and only if there exists an $X_{\text{cl}} \in \mathcal{S}^{n+n_c}$ with $X_{\text{cl}} \succ 0$ and

$$
Q(X_{\text{cl}}) + \hat{\Phi}(X_{\text{cl}})^T K \Psi + \Psi^T K^T \hat{\Phi}(X_{\text{cl}}) \prec 0 \tag{2.19}
$$

where

$$
Q(X_{\text{cl}}) := \begin{pmatrix} \hat{A}^T X_{\text{cl}} + X_{\text{cl}} \hat{A} & X_{\text{cl}} \hat{B}_1 & \hat{C}_1^T \\ \hat{B}_1^T X_{\text{cl}} & -\gamma I & \hat{D}_{11}^T \\ \hat{C}_1 & \hat{D}_{11} & -\gamma I \end{pmatrix},
$$

$\hat{\Phi}(X_{\text{cl}}) := \begin{pmatrix} \hat{B}_2^T X_{\text{cl}} & 0 & \hat{D}_{12}^T \end{pmatrix}$ and $\Psi := \begin{pmatrix} \hat{C}_2 & \hat{D}_{21} & 0 \end{pmatrix}$. Inequality (2.19) is a so-called Bilinear Matrix Inequality (BMI), since there is a bilinear coupling between the unknown variables X_{cl} and K. Optimization problems with BMI constraints are in general difficult to solve, see Section 2.5.8. Fortunately we can eliminate this bilinear coupling by a projection, which is based on the following lemma.

Lemma 2.9 *(Projection Lemma) Suppose $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $Q \in \mathcal{S}^n$ are given and let $(B^T)_\perp$ and $C_\perp$ denote arbitrary matrices whose columns are bases for $Ker(B^T) := \{x \in \mathbb{R}^n : B^T x = 0\}$ and $Ker(C)$ respectively. Then there exists a $K \in \mathbb{R}^{m \times p}$ such that $Q + BKC + C^T K^T B^T \prec 0$ if and only if*

$$
\left((B^T)_\perp\right)^T Q (B^T)_\perp \prec 0 \ and \ (C_\perp)^T Q C_\perp \prec 0.
$$

Proof. See for instance [176]. ∎

Let the columns of the matrices $\hat{\Phi}(X_{\mathrm{cl}})_\perp$ and $\Psi_\perp$ be arbitrary bases for $\mathrm{Ker}(\hat{\Phi}(X_{\mathrm{cl}})_\perp)$ and $\mathrm{Ker}(\Psi)$ respectively. Then the lemma implies that (2.19) is equivalent to

$$\left(\hat{\Phi}(X_{\mathrm{cl}})_\perp\right)^T Q(X_{\mathrm{cl}})\hat{\Phi}(X_{\mathrm{cl}})_\perp \prec 0 \tag{2.20}$$

and

$$\Psi_\perp^T Q(X_{\mathrm{cl}})\Psi_\perp \prec 0. \tag{2.21}$$

We stress that the matrices $\hat{\Phi}(X_{\mathrm{cl}})_\perp$ and $\Psi_\perp$ can be arbitrary in the equations above, as long as their columns are bases for $\mathrm{Ker}(\hat{\Phi}(X_{\mathrm{cl}})_\perp)$ and $\mathrm{Ker}(\Psi)$ respectively.

Inequality (2.20) is a complicated condition to use in optimization, since $\mathrm{Ker}(\hat{\Phi}(X_{\mathrm{cl}})_\perp)$ depends on X_{cl}. To get rid of this dependence observe that, since X_{cl} is nonsingular, the columns of V are a basis for the kernel of $\hat{\Phi}(X_{\mathrm{cl}})$ if and only if the columns of

$$\begin{pmatrix} X_{\mathrm{cl}} & 0 & 0 \\ 0 & I_{m_1} & 0 \\ 0 & 0 & I_{p_1} \end{pmatrix} V$$

are a basis for the kernel of $\Phi := \begin{pmatrix} \hat{B}_2^T & 0 & \hat{D}_{12}^T \end{pmatrix}$. Hence, if the columns of $\Phi_\perp$ are an arbitrary basis matrix for $\mathrm{Ker}(\Phi)$ we can choose $\hat{\Phi}(X_{\mathrm{cl}})_\perp$ in (2.20) to be equal to

$$\begin{pmatrix} X_{\mathrm{cl}}^{-1} & 0 & 0 \\ 0 & I_{m_1} & 0 \\ 0 & 0 & I_{p_1} \end{pmatrix} \Phi_\perp.$$

This results in the following condition that is equivalent to (2.20), again irrespective of the choice of the basis $\Phi_\perp$:

$$(\Phi_\perp)^T \begin{pmatrix} X_{\mathrm{cl}}^{-1} & 0 & 0 \\ 0 & I_{m_1} & 0 \\ 0 & 0 & I_{p_1} \end{pmatrix} Q \begin{pmatrix} X_{\mathrm{cl}}^{-1} & 0 & 0 \\ 0 & I_{m_1} & 0 \\ 0 & 0 & I_{p_1} \end{pmatrix} \Phi_\perp =$$

$$= (\Phi_\perp)^T \begin{pmatrix} X_{\mathrm{cl}}^{-1}\hat{A}^T + \hat{A}X_{\mathrm{cl}}^{-1} & \hat{B}_1 & X_{\mathrm{cl}}^{-1}\hat{C}_1^T \\ \hat{B}_1^T & -\gamma I_{m_1} & \hat{D}_{11}^T \\ \hat{C}_1 X_{\mathrm{cl}}^{-1} & \hat{D}_{11} & -\gamma I_{p_1} \end{pmatrix} \Phi_\perp \prec 0. \tag{2.22}$$

Let us partition $X_{\mathrm{cl}} \in \mathcal{S}^{n+n_c}$ into

$$X_{\mathrm{cl}} = \begin{pmatrix} X & U \\ U^T & \hat{X} \end{pmatrix}, \tag{2.23}$$

where $X \in \mathcal{S}^n$ and define

$$Y_{\mathrm{cl}} := \begin{pmatrix} Y & V \\ V^T & \hat{Y} \end{pmatrix} := X_{\mathrm{cl}}^{-1}, \tag{2.24}$$

with $Y \in \mathcal{S}^n$.

If the columns of $N = \begin{pmatrix} N_1^T & N_2^T \end{pmatrix}^T$ are an arbitrary basis of $\mathrm{Ker}\left(\begin{pmatrix} C_2 & D_{21} \end{pmatrix} \right)$, then (2.21) is equivalent to

$$\begin{pmatrix} N & 0 \\ 0 & I_{p_1} \end{pmatrix}^T \begin{pmatrix} A^T X + XA & XB_1 & C_1^T \\ B_1^T X & -\gamma I_{m_1} & D_{11}^T \\ C_1 & D_{11} & -\gamma I_{p_1} \end{pmatrix} \begin{pmatrix} N & 0 \\ 0 & I_{p_1} \end{pmatrix} \prec 0. \tag{2.25}$$

This follows from a congruence transformation of (2.25) and the fact that the kernel of Ψ is the image of

$$\begin{pmatrix} N_1 & 0 \\ 0 & 0 \\ N_2 & 0 \\ 0 & I_{m_1} \end{pmatrix}.$$

Analogously, for arbitrary matrices $\Phi_\perp$ and $M = \begin{pmatrix} M_1^T & M_2^T \end{pmatrix}^T$ whose columns are bases of the kernels of Φ and $\begin{pmatrix} B_2^T & D_{12}^T \end{pmatrix}$ respectively, (2.22) is equivalent to

$$\begin{pmatrix} M & 0 \\ 0 & I_{m_1} \end{pmatrix}^T \begin{pmatrix} YA^T + AY & YC_1^T & B_1 \\ C_1 Y & -\gamma I & D_{11} \\ B_1^T & D_{11}^T & -\gamma I \end{pmatrix} \begin{pmatrix} M & 0 \\ 0 & I_{m_1} \end{pmatrix} \prec 0. \tag{2.26}$$

The last step towards the LMI formulation of the full-order controller synthesis problem is to express $X_{\mathrm{cl}} \succ 0$ in terms of the following constraints on X and Y:

$$\begin{pmatrix} X & I_n \\ I_n & Y \end{pmatrix} \succeq 0 \tag{2.27}$$

and

$$\mathrm{Rank} \begin{pmatrix} X & I_n \\ I_n & Y \end{pmatrix} \leq n + n_{\mathrm{c}}. \tag{2.28}$$

This is subject of the following lemma.

Lemma 2.10 *For $X \in \mathcal{S}^n$ and $Y \in \mathcal{S}^n$, there exist $U \in \mathbb{R}^{n \times n_c}$, $V \in \mathbb{R}^{n \times n_c}$ and $\hat{X} \in \mathcal{S}^{n_c}$, $\hat{Y} \in \mathcal{S}^{n_c}$ such that $X_{\mathrm{cl}} \in \mathcal{S}^{n+n_c}$ defined by (2.23) satisfies*

$$X_{\mathrm{cl}} \succ 0 \qquad and \qquad (2.24)$$

if and only if

$$(2.27) \qquad and \qquad (2.28)$$

hold true.

Proof. The proof is inspired by the book [52], which contains this lemma. See for instance also [166]. Let (2.27) and (2.28) hold true. (2.27) implies $X \succ 0$ and $Y \succ 0$, as follows from the following argument. Since obviously $X \succeq 0$ and $Y \succeq 0$, we only need to show that X and Y are nonsingular. To arrive at a contradiction suppose $Yy = 0$ for some $y \neq 0$. Then for $\alpha > \frac{1}{2}\|X\|$ we infer

$$\begin{pmatrix} -y \\ \alpha y \end{pmatrix}^T \begin{pmatrix} X & I_n \\ I_n & Y \end{pmatrix} \begin{pmatrix} -y \\ \alpha y \end{pmatrix} = y^T Xy - 2\alpha y^T y = y^T Xy - 2\alpha y^T y \leq \|y\|(\|X\| - 2\alpha) < 0,$$

19

which contradicts (2.27). Hence Y is nonsingular and thus $Y \succ 0$. A similar argument reveals that $X \succ 0$. Since $Y \succ 0$ we can use the following Schur-complement identity

$$\begin{pmatrix} X & I_n \\ I_n & Y \end{pmatrix} = \begin{pmatrix} I_n & 0 \\ Y^{-1} & I_n \end{pmatrix}^T \begin{pmatrix} X - Y^{-1} & 0 \\ 0 & Y \end{pmatrix} \begin{pmatrix} I_n & 0 \\ Y^{-1} & I_n \end{pmatrix}, \qquad (2.29)$$

which implies

$$\mathrm{Rank}\begin{pmatrix} X & I_n \\ I_n & Y \end{pmatrix} = \mathrm{Rank}(Y) + \mathrm{Rank}(X - Y^{-1}) = n + \mathrm{Rank}(X - Y^{-1}).$$

Together with (2.28), this implies $\mathrm{Rank}(X - Y^{-1}) \le n_{\mathrm{c}}$. Hence, there exists a $U \in \mathbb{R}^{n \times n_{\mathrm{c}}}$ with $X - Y^{-1} = UU^T$. For any particular choice of U and $\hat{X} = I_{n_{\mathrm{c}}}$ the following choice of X_{cl} will do the job:

$$X_{\mathrm{cl}} = \begin{pmatrix} X & U \\ U^T & I_{n_{\mathrm{c}}} \end{pmatrix} \in \mathcal{S}^{n+n_{\mathrm{c}}}. \qquad (2.30)$$

Indeed, since $X \succ 0$ and $X - UU^T = Y^{-1} \succ 0$, we infer from the Schur lemma that $X_{\mathrm{cl}} \succ 0$. Furthermore, (2.24) holds true for $V = -YU$ and $\hat{Y} = U^T YU + I_{n_{\mathrm{c}}}$.

To prove the converse, assume that $X_{\mathrm{cl}} \in \mathcal{S}^{n+n_{\mathrm{c}}}$ satisfies $X_{\mathrm{cl}} \succ 0$. Then X, Y and $\hat{Y}$ defined by (2.23) and (2.24) satisfy $X \succ 0$ and $Y \succ 0$ and

$$\begin{pmatrix} X & I_n \\ I_n & Y \end{pmatrix} = \begin{pmatrix} I_n & Y \\ 0 & V^T \end{pmatrix}^T \begin{pmatrix} X & U \\ U^T & \hat{X} \end{pmatrix} \begin{pmatrix} I_n & Y \\ 0 & V^T \end{pmatrix} =$$

$$= \begin{pmatrix} I_n & Y \\ 0 & V^T \end{pmatrix}^T X_{\mathrm{cl}} \begin{pmatrix} I_n & Y \\ 0 & V^T \end{pmatrix} \succeq 0,$$

where we exploited the identity

$$\begin{pmatrix} X & U \\ U^T & \hat{X} \end{pmatrix} \begin{pmatrix} Y & V \\ V^T & \hat{Y} \end{pmatrix} = I_{n+n_{\mathrm{c}}} \qquad (2.31)$$

in the first equality. Since $Y \succ 0$ the Schur complement relationship (2.29) holds true such that

$$\mathrm{Rank}\begin{pmatrix} X & I_n \\ I_n & Y \end{pmatrix} = \mathrm{Rank}(Y) + \mathrm{Rank}(X - Y^{-1}) = \mathrm{Rank}(Y) + \mathrm{Rank}(UV^T) \le n + n_{\mathrm{c}},$$

where the most right equality follows since $X - Y^{-1} = -UV^T Y^{-1}$, which is a consequence of the left upper part of (2.31). Since $U \in \mathbb{R}^{n \times n_{\mathrm{c}}}$ we infer $\mathrm{Rank}\left(UV^T\right) \le n_{\mathrm{c}}$ and finally conclude

$$\mathrm{Rank}\begin{pmatrix} X & I_n \\ I_n & Y \end{pmatrix} \le n + n_{\mathrm{c}}.$$

This completes the proof.

∎

For full-order synthesis $n_{\mathrm{c}} = n$ and (2.28) is trivially satisfied. Concluding this section, conditions (2.25), (2.26) and (2.27) are LMI conditions that are equivalent to the existence of a stabilizing full order controller with closed-loop $\mathcal{H}_\infty$-norm smaller than γ. Minimizing γ over these constraints (i.e. over (2.25), (2.26) and (2.27)) is an LMI problem.

2.3.2 Controller reconstruction

After having solved for X and Y, a corresponding closed-loop matrix X_{cl} can be constructed using (2.30) in Lemma 2.10 for some U satisfying

$$UU^T = X - Y^{-1} \succeq 0.$$

Since any such X_{cl} renders (2.20) and (2.21) satisfied, Lemma 2.9 implies that there exists a K such that (2.19) holds true. Computing such a controller for fixed X_{cl} is an LMI feasibility problem. The controller can also be computed directly in terms of X_{cl} and Y_{cl} and the state-space matrices of the plant, see e.g. [168]. Since the number of rows/columns of $\hat{X} = I_{n_c}$ equals that of A_K by construction, the procedure yields a controller of order at most n_c. We have therefore shown the "if" part of the following theorem.

Theorem 2.11 *There exists a stabilizing controller of McMillan degree n_c with closed-loop $\mathcal{H}_\infty$ performance smaller than γ, if and only if there exists X and Y satisfying (2.25), (2.26), (2.27) and $Rank(X - Y^{-1}) \leq n_c$.*

Proof. To prove the "only if" part, assume that there exists a stabilizing controller with transfer function K of McMillan degree of at most n_c with $\mathcal{F}_l(P, K) < \gamma$. Then there exists a realization with $A_K \in \mathbb{R}^{n_c \times n_c}$, such that $A_{\text{cl}} \in \mathbb{R}^{(n+n_c) \times (n+n_c)}$. The BRL hence implies the existence of an $X_{\text{cl}} \in \mathcal{S}^{n+n_c}$ $X_{\text{cl}} \succ 0$ satisfying (2.19). Lemma 2.10 and the projection lemma imply the existence of X and Y that satisfy (2.25), (2.26), (2.27) and $\text{Rank}(X - Y^{-1}) \leq n_c$.
∎

The above construction of a full-order controller is insightful, but not the best approach for numerical computations. Firstly, the choice $\hat{X} = I_r$ is usually not the best way to construct the closed-loop matrix X_{cl}, see [176] for some practically useful suggestions to improve the accuracy of the numerical computation. Secondly, for systems of order > 50 the current state-of-the-art LMI solvers are either too slow or too inaccurate to compute feasible X and Y of (2.25), (2.26), (2.27) with high enough accuracy[1] for the construction of a controller that satisfies the performance and stability requirements. Therefore, we derive the solutions in terms of Riccati inequalities in the next section.

2.3.3 Solution in terms of Riccati inequalities

In our experience full-order controllers for high-order systems can be better computed through the solution of two Riccati equations and a coupling condition. To derive these equations, we first transform the LMIs into Algebraic Riccati Inequalities (ARIs) under the following simplifying assumptions

$$D_{11} = 0, \quad D_{12}^T \begin{pmatrix} C_1 & D_{12} \end{pmatrix} = \begin{pmatrix} 0 & I_{m_2} \end{pmatrix} \tag{2.32}$$

and

$$D_{21} \begin{pmatrix} B_1^T & D_{21}^T \end{pmatrix} = \begin{pmatrix} 0 & I_{p_2} \end{pmatrix}. \tag{2.33}$$

---

[1]It is beyond the scope of the thesis to quantify when the accuracy is high enough. The reader is referred to [72] for the effects of round-off errors on elementary matrix manipulations.

Remark $D_{11} = 0$ makes the open-loop transfer function P_{11} strictly proper, which implies that there is no direct feed-through from w to z if the loop in Figure 2.1 is open. The most right condition in (2.32) can be interpreted as saying that the plant output $C_1 x$ and the weight D_{21} on the control effort are orthogonal, and all the control channels affect the system performances directly. Condition (2.32) implies that the system disturbances that drive the dynamics are orthogonal to the measurement noise, and furthermore that the noise channels w affect all the system measurements y [52].

The assumptions (2.32) and (2.33) imply that N and M in (2.25) and (2.26) can be chosen as

$$N = \begin{pmatrix} C_2 & D_{21} \end{pmatrix}_\perp = \begin{pmatrix} I_n & 0 \\ -D_{21}^T C_2 & (D_{21})_\perp \end{pmatrix}$$

and

$$M = \begin{pmatrix} B_2^T & D_{12}^T \end{pmatrix}_\perp = \begin{pmatrix} I_n & 0 \\ -D_{12} B_2^T & (D_{12}^T)_\perp \end{pmatrix},$$

where we can assume that the columns of $(D_{21})_\perp$ and $(D_{12}^T)_\perp$ are orthonormal bases for the kernels of D_{21} and D_{12}^T respectively, which implies that $((D_{21})_\perp)^T (D_{21})_\perp = I_k$ and $((D_{12}^T)_\perp)^T (D_{12}^T)_\perp = I_l$, where $k = m_1 - p_2 \geq 0$ and $l = p_1 - m_2 \geq 0$ are the dimensions of the kernels of $(D_{21})_\perp$ and $(D_{12}^T)_\perp$ respectively. With this N (2.25) simplifies to

$$\begin{pmatrix} A^T X + XA - XB_1 D_{21}^T C_2 - C_2^T D_{21} B_1^T X - \gamma C_2^T D_{21} D_{21}^T C_2 & (*) & C_1^T \\ ((D_{21})_\perp)^T B_1^T X + \gamma ((D_{21})_\perp)^T D_{21}^T C_2 & -\gamma ((D_{21})_\perp)^T (D_{21})_\perp & 0 \\ C_1 & 0 & -\gamma I_{m_1} \end{pmatrix} =$$

$$= \begin{pmatrix} A^T X + XA - \gamma C_2^T C_2 & XB_1 (D_{21})_\perp & C_1^T \\ ((D_{21})_\perp)^T B_1^T X & -\gamma ((D_{21})_\perp)^T (D_{21})_\perp & 0 \\ C_1 & 0 & -\gamma I_{m_1} \end{pmatrix} =$$

$$= \begin{pmatrix} A^T X + XA - \gamma C_2^T C_2 & XB_1 (D_{21})_\perp & C_1^T \\ ((D_{21})_\perp)^T B_1^T X & -\gamma I_k & 0 \\ C_1 & 0 & -\gamma I_{m_1} \end{pmatrix} \prec 0,$$

where $(*)$ denotes an element which is implied by symmetry. Taking a Schur complement twice implies that this inequality is equivalent to $\gamma > 0$ and

$$A^T X + XA - \gamma C_2^T C_2 + \frac{1}{\gamma} C_1^T C_1 + \gamma X B_1 (D_{21})_\perp ((D_{21})_\perp)^T B_1^T X \prec 0. \qquad (2.34)$$

Since $D_{21} B_1^T = 0$ and

$$\begin{pmatrix} D_{21} \\ ((D_{21})_\perp)^T \end{pmatrix} \begin{pmatrix} D_{21}^T & (D_{21})_\perp \end{pmatrix} = \begin{pmatrix} I_{p_2} & 0 \\ 0 & I_k \end{pmatrix} = I_{m_1},$$

we conclude

$$B_1 (D_{21})_\perp ((D_{21})_\perp)^T B_1^T = B_1 (D_{21})_\perp ((D_{21})_\perp)^T B_1^T + B_1 D_{21}^T D_{21} B_1^T =$$

$$= B_1 \begin{pmatrix} D_{21} \\ ((D_{21})_\perp)^T \end{pmatrix} \begin{pmatrix} D_{21}^T & (D_{21})_\perp \end{pmatrix} B_1^T = B_1 B_1^T,$$

such that (2.34) is equivalent to

$$A^T X + XA - \gamma C_2^T C_2 + \frac{1}{\gamma} C_1^T C_1 + \gamma X B_1 B_1^T X \prec 0.$$

Finally substituting $\tilde{X} := \frac{1}{\gamma} X$ and dividing by γ yields

$$A^T \tilde{X} + \tilde{X} A - C_2^T C_2 + \frac{1}{\gamma^2} C_1^T C_1 + \tilde{X} B_1 B_1^T \tilde{X} \prec 0. \tag{2.35}$$

Similarly it can be derived that for $\tilde{Y} := \frac{1}{\gamma} Y$, (2.26) is equivalent to $\gamma > 0$ and

$$A\tilde{Y} + \tilde{Y} A - B_2 B_2^T + \frac{1}{\gamma^2} B_1 B_1^T + \tilde{Y} C_1^T C_1 \tilde{Y} \prec 0. \tag{2.36}$$

If we choose $P := \tilde{X}^{-1}$ and $Q := \tilde{Y}^{-1}$ and left and right multiply (2.35) and (2.36) by $\tilde{X}^{-1}$ and $\tilde{Y}^{-1}$ respectively, we obtain the standard indefinite Riccati inequalities for $\mathcal{H}_\infty$ synthesis:

$$AP + PA^T + P(\frac{1}{\gamma^2} C_1^T C_1 - C_2^T C_2)P + B_1 B_1^T \quad \prec \quad 0, \tag{2.37}$$

$$A^T Q + QA + Q(\frac{1}{\gamma^2} B_1 B_1^T - B_2 B_2^T)Q + C_1^T C_1 \quad \prec \quad 0. \tag{2.38}$$

Furthermore we observe that the coupling condition (2.27) can be equivalently written in terms of P and Q as follows

Lemma 2.12 *Let $\gamma > 0$ be arbitrary. Then*

$$\begin{pmatrix} X & I \\ I & Y \end{pmatrix} \succeq 0 \tag{2.39}$$

is equivalent to

$$P \succ 0, \qquad Q \succ 0, \qquad \rho_{sp}(PQ) \leq \gamma^2, \tag{2.40}$$

where $P := \gamma X^{-1}$ and $Q := \gamma Y^{-1}$ and $\rho_{sp}(A)$ denotes the spectral radius of the matrix A.

Proof. Adopted from [168]. The inequality (2.39) is equivalent to $X \succ 0$, $Y \succ 0$, $Y - X^{-1} \succeq 0$ (Schur) which is nothing but

$$X \succ 0, \qquad Y \succ 0, \qquad I - R^{-T} X^{-1} R^{-1} \succeq 0, \tag{2.41}$$

where R is is any factor of Y, i.e. $Y = R^T R$. Since the nonzero spectrum of AB and BA coincide for arbitrary matrices A and B of appropriate dimensions [94], (2.41) can be rewritten as (2.40), where $P := \gamma X^{-1}$ and $Q := \gamma Y^{-1}$. ∎

The inequalities (2.37), (2.38), $P \succ 0$ and $Q \succ 0$ are strict. Hence, if $\rho_{sp}(PQ) = \gamma^2$, we can always slightly perturb P and Q such that $\rho_{sp}(PQ) < \gamma^2$. We have summarized the preceding results in the following lemma.

Lemma 2.13 *Suppose the assumptions (2.32) and (2.33) hold true and let $\gamma > 0$. Then there exist X and Y satisfying (2.25), (2.26) and (2.27) if and only if there exist P, Q satisfying (2.37) and (2.38) with*

$$P \succ 0, \qquad Q \succ 0, \qquad \rho_{sp}(PQ) < \gamma^2, \qquad (2.42)$$

Proof. The equivalence follows directly from the preceding arguments.

∎

2.3.4 Solution in terms of Riccati equations

The following result relates the solutions of the Riccati inequality and equation.

Theorem 2.14 *Consider the Riccati inequality*

$$XA + A^T X + Q + XRX \prec 0 \qquad (2.43)$$

for symmetric Q and R, $R \succeq 0$ and (A, R) is controllable. Then (2.43) has a symmetric solution X_{ARI} if and only if

$$XA + A^T X + Q + XRX = 0$$

has an anti-stabilizing solution X_{as}, i.e. $-A - RX_{as}$ is Hurwitz. If one and hence both these conditions are satisfied, then $X_{ARI} \preceq X_{as}$

Proof. See [168]

∎

Using this theorem we can replace the Riccati inequalities in a straightforward fashion by algebraic Riccati Equations, under the following additional hypothesis on the state-space representation of **P**:

$$(A, B_1) \text{ is stabilizable and } (A, C_1) \text{ is detectable.} \qquad (2.44)$$

Lemma 2.15 *Assume (2.44) holds true. The inequality (2.37) has a positive definite solution $P_{ARI} \succ 0$ if and only if the algebraic Riccati equation*

$$AP + PA^T + P(\frac{1}{\gamma^2}C_1^T C_1 - C_2^T C_2)P + B_1 B_1^T = 0 \qquad (2.45)$$

has a stabilizing solution $P_{st} \succeq 0$ (i.e. with $A + P_{st}(\frac{1}{\gamma^2}C_1^T C_1 - C_2^T C_2)$ Hurwitz). If one and hence both these conditions are satisfied, then $P_{ARI} \succeq P_{st}$.

Proof. Without loss of generality, we can assume that the state-space realization of the plant (possibly after a similarity transformation) admits the following partitioning

$$\left(\begin{array}{c|c} A & B_1 \\ \hline C_1 & 0 \\ C_2 & D_{21} \end{array}\right) = \left(\begin{array}{cc|c} A_{11} & A_{12} & B_{11} \\ 0 & A_{22} & 0 \\ \hline C_{11} & C_{12} & 0 \\ C_{21} & C_{22} & D_{21} \end{array}\right)$$

24

for some matrices A_{ij}, B_{11}, C_{ij}, $i, j = 1, 2$ of appropriate size such that A_{11} is square and (A_{11}, B_{11}) is controllable. Note that A_{22} is Hurwitz since (A, B_1) is stabilisable by (2.44).

Let $P_{\mathrm{ARI}} \succ 0$ be a solution of (2.37). Observe that $X_* := P_{\mathrm{ARI}}^{-1} \succ 0$ satisfies

$$A^T X_* + X_* A - C_2^T C_2 + \frac{1}{\gamma^2} C_1^T C_1 + X_* B_1 B_1^T X_* \prec 0. \tag{2.46}$$

Let us partition X_* according to A as

$$X_* = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}$$

and observe that $X_{11} \succ 0$. Consider the left upper block of (2.46), i.e.:

$$\begin{pmatrix} I & 0 \end{pmatrix} \left(X_* A + A^T X_* + (\frac{1}{\gamma^2} C_1^T C_1 - C_2^T C_2) + X_* B_1 B_1^T X_* \right) \begin{pmatrix} I \\ 0 \end{pmatrix} =$$

$$X_{11} A_{11} + A_{11}^T X_{11} + (\frac{1}{\gamma^2} C_{11}^T C_{11} - C_{21}^T C_{21}) + X_{11} B_{11} B_{11}^T X_{11} \prec 0 \tag{2.47}$$

Since (A_{11}, B_{11}) and hence $(A_{11}, B_{11} B_{11}^T)$ are controllable, we conclude from (2.47) and Theorem 2.14 that the ARE

$$X A_{11} + A_{11}^T X + (\frac{1}{\gamma^2} C_{11}^T C_{11} - C_{21}^T C_{21}) + X B_{11} B_{11}^T X = 0 \tag{2.48}$$

has an anti-stabilizing solution $X_{\mathrm{as}} \succeq X_{11} \succ 0$, i.e. with $\Lambda := -A_{11} - B_{11} B_{11}^T X_{\mathrm{as}}$ Hurwitz. This implies that $P_{11} := X_{\mathrm{as}}^{-1}$ satisfies

$$A_{11} P_{11} + P_{11} A_{11}^T + P_{11}(\frac{1}{\gamma^2} C_{11}^T C_{11} - C_{21}^T C_{21}) P_{11} + B_{11} B_{11}^T = 0. \tag{2.49}$$

and furthermore $A_{11} + P_{11}(\frac{1}{\gamma^2} C_{11}^T C_{11} - C_{21}^T C_{21})$ is Hurwitz since

$$\left(A_{11} + P_{11}(\frac{1}{\gamma^2} C_{11}^T C_{11} - C_{21}^T C_{21}) \right)^T = A_{11}^T + (\frac{1}{\gamma^2} C_{11}^T C_{11} - C_{21}^T C_{21}) P_{11} =$$

$$= \left(A_{11}^T X_{\mathrm{as}} + (\frac{1}{\gamma^2} C_{11}^T C_{11} - C_{21}^T C_{21}) \right) P_{11} = \left(-X_{\mathrm{as}} A_{11} - X_{\mathrm{as}} B_{11} B_{11}^T X_{\mathrm{as}} \right) P_{11} =$$

$$= X_{\mathrm{as}} \left(-A_{11} - B_{11} B_{11}^T X_{\mathrm{as}} \right) P_{11} = P_{11}^{-1} \Lambda P_{11}.$$

If we choose

$$P_{\mathrm{st}} := \begin{pmatrix} P_{11} & 0 \\ 0 & 0 \end{pmatrix} \succeq 0$$

we observe that

$$A + P_{\mathrm{st}}(\frac{1}{\gamma^2} C_1^T C_1 - C_2^T C_2) =$$

$$= \begin{pmatrix} A_{11} + P_{11}(\frac{1}{\gamma^2} C_{11}^T C_{11} - C_{21}^T C_{21}) & A_{12} + P_{11}(\frac{1}{\gamma^2} C_{11}^T C_{12} - C_{21}^T C_{22}) \\ 0 & A_{22} \end{pmatrix}$$

is Hurwitz, since $A_{11} + P_{11}(\frac{1}{\gamma^2}C_{11}^T C_{11} - C_{21}^T C_{21})$ and A_{22} are both Hurwitz. Furthermore

$$AP_{\text{st}} + P_{\text{st}}A^T + P_{\text{st}}(\frac{1}{\gamma^2}C_1^T C_1 - C_2^T C_2)P_{\text{st}} + B_1 B_1^T =$$
$$= \begin{pmatrix} A_{11}P_{11} + P_{11}A_{11}^T + P_{11}(\frac{1}{\gamma^2}C_1^T C_1 - C_2^T C_2)P_{11} + B_{11}B_{11}^T & 0 \\ 0 & 0 \end{pmatrix} = 0$$

such that indeed P_{st} is a stabilizing solution to (2.45). Finally we infer

$$P_{\text{st}} = \begin{pmatrix} X_{\text{as}}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} X_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \preceq P_{\text{ARI}},$$

where the utmost right inequality follows from a Schur complement, $P_{\text{ARI}} \succ 0$ and

$$X_{11} = \begin{pmatrix} I & 0 \end{pmatrix} P_{\text{ARI}}^{-1} \begin{pmatrix} I \\ 0 \end{pmatrix}.$$

This finishes one direction of the proof.

Now let us assume that $P_{\text{st}} \succeq 0$ is a stabilizing solution to (2.45). To simplify notation we define $V := (\frac{1}{\gamma^2}C_1^T C_1 - C_2^T C_2)$. Since $A + P_{\text{st}}V$ is Hurwitz, the Lyapunov equation

$$(A + P_{\text{st}}V)Y + Y(A + P_{\text{st}}V)^T = -I \tag{2.50}$$

has a positive definite solution $Y \succ 0$. We infer that $P_\epsilon := P_{\text{st}} + \epsilon Y$ satisfies

$$AP_\epsilon + P_\epsilon A^T + P_\epsilon V P_\epsilon + B_1 B_1^T =$$
$$= \underbrace{AP_{\text{st}} + P_{\text{st}}A^T + P_{\text{st}}V P_{\text{st}} + B_1 B_1^T}_{=0 \text{ by } (2.45)} + \epsilon(A + P_{\text{st}}V)Y + \epsilon Y(A + P_{\text{st}}V)^T + \epsilon^2 Y V Y =$$
$$= -\epsilon I + \epsilon^2 Y V Y.$$

We can always choose $\epsilon_* > 0$ small enough such that $-\epsilon_* I + \epsilon_*^2 Y V Y \prec 0$. For such an $\epsilon_* > 0$, P_{ϵ_*} is a positive definite solution to (2.37) that satisfies $P_{\epsilon_*} \succeq P_{\text{st}}$. $\blacksquare$

This result implies that we can replace the ARIs by AREs, which brings us to the solution of the full-order $\mathcal{H}_\infty$-optimal synthesis problem in terms of two Riccati equations and a coupling condition, also called the DGKF solution named by the authors of [50]: Doyle, Glover, Khargonekar and Francis.

Theorem 2.16 *Suppose (2.32), (2.33) and (2.44) are satisfied and $\gamma > 0$. Then the following hold true:*

(i.) there exists a stabilizing controller K such that $\|\mathcal{F}_l(P,K)\|_\infty < \gamma$ if and only if the Riccati equations

$$AX + XA^T + X(\frac{1}{\gamma^2}C_1^T C_1 - C_2^T C_2)X + B_1 B_1^T = 0 \qquad (2.51)$$

and

$$A^T Y + YA + Y(\frac{1}{\gamma^2}B_1 B_1^T - B_2 B_2^T)Y + C_1^T C_1 = 0 \qquad (2.52)$$

have solutions $X_{ARE} \in \mathcal{S}^n$, $Y_{ARE} \in \mathcal{S}^n$ satisfying $X_{ARE} \succeq 0$, $Y_{ARE} \succeq 0$ and $\rho_{sp}(X_{ARE} Y_{ARE}) < \gamma^2$, and such that

$$A + X_{ARE}\left(\frac{1}{\gamma^2}C_1^T C_1 - C_2^T C_2\right) \ and \ A + \left(\frac{1}{\gamma^2}B_1^T B_1 - B_2^T B_2\right)Y_{ARE} \ are \ Hurwitz.$$
$$(2.53)$$

(ii.) If X_{ARE} and Y_{ARE} satisfy all conditions above, then the set of all stabilizing controllers $K \in \mathcal{RL}_\infty$ satisfying $\|\mathcal{F}_l(P,K)\|_\infty < \gamma$ is given by

$$\mathcal{K} := \{\mathcal{F}_l(J, Q), \ Q \in \mathcal{RH}_\infty, \ \|Q\|_\infty < \gamma\}$$

where $J(s) = D_J + C_J(sI - A_J)^{-1}B_J$ with

$$\left(\begin{array}{c|c} A_J & B_J \\ \hline C_J & D_J \end{array}\right) := \left(\begin{array}{c|cc} A_{ARE} & -ZL & ZB_2 \\ \hline F & 0 & I \\ -C_2 & I & 0 \end{array}\right),$$

$$F := -B_2^T Y_{ARE}, \ L := -X_{ARE}C_2^T, \ Z := (I - \tfrac{1}{\gamma^2}X_{ARE}Y_{ARE})^{-1} \ and$$

$$A_{ARE} := A + \frac{1}{\gamma^2}B_1 B_1^T Y_{ARE} + B_2 F + ZLC_2$$

Proof. We only give the proof of (i.) and refer the reader to [49] for the proof of (ii.). Suppose there exists an admissable controller, i.e. a controller that stabilizes the closed loop and such that $\|\mathcal{F}_l(P,K)\|_\infty < \gamma$. Theorem 2.11 for $n_c = n$ together with Lemma 2.13 imply that there exist P and Q satisfying (2.37), (2.38) and (2.42). By Lemma 2.15 we conclude that there exist solutions X_{ARE} and Y_{ARE} to (2.51) and (2.52), respectively, with $P \succeq X_{\mathrm{ARE}} \succeq 0$ and $Q \succeq Y_{\mathrm{ARE}} \succeq 0$ and such that (2.53) holds true. Since $P \succeq X_{\mathrm{ARE}}$ and $Q \succeq Y_{\mathrm{ARE}}$ we conclude $\rho_{\mathrm{sp}}(X_{\mathrm{ARE}}Y_{\mathrm{ARE}}) \leq \rho_{\mathrm{sp}}(PQ) < \gamma^2$.

On the other hand, suppose that X_{ARE} and Y_{ARE} satisfy the conditions in (i.). If we define $V := \frac{1}{\gamma^2}C_1^T C_1 - C_2^T C_2$ and $W = \frac{1}{\gamma^2}B_1 B_1^T - B_2 B_2^T$, we observe that $A + X_{\mathrm{ARE}}V$ and $A + WY_{\mathrm{ARE}}$ are stable. Hence the Lyapunov equations

$$(A + X_{\mathrm{ARE}}V)\,S + S\,(A + X_{\mathrm{ARE}}V)^T = -I \qquad (2.54)$$

and

$$T\,(A + WY_{\mathrm{ARE}}) + (A + WY_{\mathrm{ARE}})^T\,T = -I \qquad (2.55)$$

27

have positive-definite solutions $S \succ 0$ and $T \succ 0$, respectively. We infer that $P_\epsilon :=$ $X_{\mathrm{ARE}} + \epsilon S$ satisfies

$$
\begin{aligned}
AP_\epsilon + P_\epsilon A^T + P_\epsilon V P_\epsilon + B_1 B_1^T &= \\
= \underbrace{AX_{\mathrm{ARE}} + X_{\mathrm{ARE}} A^T + X_{\mathrm{ARE}} V X_{\mathrm{ARE}} + B_1 B_1^T}_{=0 \text{ by } (2.51)} &+ \epsilon(A + X_{\mathrm{ARE}} V)S + \epsilon S(A + X_{\mathrm{ARE}} V)^T + \epsilon^2 SVS = \\
&= -\epsilon I + \epsilon^2 SVS.
\end{aligned}
$$

Similarly, with $Q_\epsilon := Y_{\mathrm{ARE}} + \epsilon T$ we conclude that

$$
\begin{aligned}
A^T Q_\epsilon + Q_\epsilon A + Q_\epsilon W Q_\epsilon + C_1^T C_1 &= \\
= \underbrace{A^T Y_{\mathrm{ARE}} + Y_{\mathrm{ARE}} A + Y_{\mathrm{ARE}} W Y_{\mathrm{ARE}} + C_1^T C_1}_{=0 \text{ by } (2.52)} &+ \epsilon T(A + WY_{\mathrm{ARE}}) + \epsilon(A + WY_{\mathrm{ARE}})^T T + \epsilon^2 TWT = \\
&= -\epsilon I + \epsilon^2 TWT.
\end{aligned}
$$

Note that this construction is almost identical to that in the proof of Lemma 2.15. Furthermore we infer

$$
\rho_{\mathrm{sp}}(P_\epsilon Q_\epsilon) = \rho_{\mathrm{sp}}\left((X_{\mathrm{ARE}} + \epsilon S)(Y_{\mathrm{ARE}} + \epsilon T)\right)
$$

Now, we can always choose $\epsilon_* > 0$ small enough such that

$$
-\epsilon_* I + \epsilon_*^2 SVS \prec 0, \qquad -\epsilon_* I + \epsilon_*^2 TWT \prec 0 \qquad \text{and} \qquad \rho_{\mathrm{sp}}(P_\epsilon Q_\epsilon) < \gamma^2.
$$

For such an $\epsilon_* > 0$, P_{ϵ_*} and Q_{ϵ_*} are solutions to (2.37) and (2.38) that satisfy (2.42). Finally, Lemma 2.13 and Theorem 2.11 now directly imply that an admissable controller exists.

$\blacksquare$

The theorem parameterizes all stabilizing controllers that satisfy a bound on the closed-loop $\mathcal{H}_\infty$-performance. If any of the conditions (2.32), (2.33) or (2.44) is violated, the solution of the $\mathcal{H}_\infty$-optimal control problem in terms of algebraic Riccati equations is more involved, see [164, 188, 214].

Remark. The controller design methods in this section can not be applied if we want to enforce structural constraints on the controller, as discussed in Section 2.4.

2.3.5 $\mathcal{H}_2$ optimal control

We briefly address $\mathcal{H}_2$ synthesis, where we restrict ourselves to static output feedback. The problem is to compute a stabilizing controller such that the closed-loop transfer function is proper and minimizes its $\mathcal{H}_2$-norm. The $\mathcal{H}_2$-norm of a real-rational, strictly proper and stable transfer function G with minimal realization $(A, B, C, 0)$ is defined as

$$
\|G\|_2 := \sqrt{\frac{1}{2\pi} \mathrm{Trace} \int_{-\infty}^{\infty} G(j\omega)^* G(j\omega) d\omega}
$$

The norm can be computed in terms of the state-space realization by

$$
\|G\|_2 = \sqrt{\mathrm{Trace}(CPC^T)}
$$

where P is the unique solution to the Lyapunov equation

$$AP + PA^T + BB^T = 0.$$

The $\mathcal{H}_2$-norm has an interpretation in terms of the energy of the output signal for a sequence of unit impulses at the input. The response to a unit impulse of the ith input is

$$z_i(t) = Ce^{At}Be_i,$$

where e_i is the ith standard unit vector of $\mathbb{R}^{m_1}$. The $\mathcal{H}_2$-norm equals the sum of the energy of all impulse responses [176]:

$$\|G\|_2 = \sum_{i=1}^{m} \|z_i\|_2^2 = \text{Trace} \int_0^\infty (Ce^{At}B)^T Ce^{At}B\,dt$$

The static $\mathcal{H}_2$-optimal control problem can be formulated as follows:

$$
\begin{array}{ll}
\text{infimize} & \text{Trace}(C_{\text{cl}}(K)PC_{\text{cl}}(K)^T) \\
\text{subject to} & A_{\text{cl}}(K)P + PA_{\text{cl}}(K)^T + B_{\text{cl}}(K)B_{\text{cl}}(K)^T \prec 0 \\
& D_{\text{cl}}(K) = 0 \text{ and } P \succ 0
\end{array}
$$

where the closed loop state-space matrices $(A_{\text{cl}}(K), B_{\text{cl}}(K), C_{\text{cl}}(K), D_{\text{cl}}(K))$ are given by (2.8).

2.4 Controller structure

For several practical reasons the control designer may be interested in a controller with a certain structure. In this section we discuss three practical reasons for structural constraints on controllers:

- simplifying on-site tuning

- reducing computational delays

- strong stabilization.

2.4.1 Simplifying on-site tuning

Re-tuning of controllers is often necessary due to changes in the dynamic behavior due to wear, shipment of the servo-system, small modifications to the system or changing environmental conditions such as temperature changes and changing floor vibration spectra. To make the controller tuning process accessible for people without a background in advanced control, it is required that the there is a clear interpretation of the controller's building blocks in terms of gains, frequencies and damping ratio's of poles and zeros. The controller can then be tuned by direct modification of its parameters without redoing the $\mathcal{H}_\infty$-optimal optimization. An example of a controller with transparent structure is a diagonal augmentation of PID loops with notches.

A structured controller can often be described using a state-space description

$$(A_K(p), B_K(p), C_K(p), D_K(p)) \tag{2.56}$$

with affine[2] dependence on p. As an example we consider the diagonal augmentation of a SISO PID controller with a notch filter, which is a typical control structure for a servo system with one degree of freedom. The transfer function of such a controller is for instance given by [61]:

$$K = K_{\mathrm{PID}}K_{\mathrm{notch}},$$

where

$$K_{\mathrm{PID}}(s) := \left(p_1 + p_2\frac{1}{s} + p_3\frac{\tau s}{\tau s + 1} \right) \tag{2.57}$$

is the transfer function for a PID controller, where $p_1 > 0$, $p_2 > 0$ $p_3 > 0$. Furthermore let the transfer function for the notch filter $K_{\mathrm{notch}}(s)$ be given by

$$K_{\mathrm{notch}}(s) := \frac{s^2 + 2\omega_{\mathrm{z}}\zeta_{\mathrm{z}}s + \omega_{\mathrm{z}}^2}{s^2 + 2\omega_{\mathrm{p}}\zeta_{\mathrm{p}}s + \omega_{\mathrm{p}}^2},$$

with $\omega_{\mathrm{z}} > 0$, $\omega_{\mathrm{p}} > 0$, $\zeta_{\mathrm{z}} > 0$ and $\zeta_{\mathrm{p}} > 0$. By defining $a = 2\omega_{\mathrm{z}}\zeta_{\mathrm{z}}$, $b = \omega_{\mathrm{z}}^2$, $c = 2\omega_{\mathrm{p}}\zeta_{\mathrm{p}}$ and $d = \omega_{\mathrm{p}}^2$, we construct a single vector p of controller parameters defined by

$$p := \left(\begin{array}{cccccccc} p_1 & p_2 & p_3 & \frac{1}{\tau} & a & b & c & d \end{array} \right).$$

The notch filter K_{notch} admits the state-space representation $(A_N(p), B_N(p), C_N(p), D_N(p))$ with

$$\left(\begin{array}{c|c} A_N(p) & B_N(p) \\ \hline C_N(p) & D_N(p) \end{array} \right) := \left(\begin{array}{cc|c} 0 & 1 & 0 \\ -d & -c & 1 \\ \hline b-d & a-c & 1 \end{array} \right).$$

The PID controller K_{PID} admits the following realization

$$\left(\begin{array}{c|c} A_K(p) & B_K(p) \\ \hline C_K(p) & D_K(p) \end{array} \right) := \left(\begin{array}{cc|c} 0 & 0 & 1 \\ 0 & -\frac{1}{\tau} & \frac{1}{\tau} \\ \hline p_2 & -p_3 & p_1 + p_3 \end{array} \right).$$

The series connection of these systems is

$$\left(\begin{array}{cc} A_K(p) & B_K(p) \\ C_K(p) & D_K(p) \end{array} \right) := \left(\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 1 \\ 0 & -1/\tau & 0 & 0 & 1/\tau \\ 0 & 0 & 0 & 1 & 0 \\ p_2 & -p_3 & -d & -c & p_1 + p_3 \\ \hline p_2 & -p_3 & b-d & a-c & p_1 + p_3 \end{array} \right)$$

Observe that $A_K(p)$, $B_K(p)$, $C_K(p)$, and $D_K(p)$ are indeed affine in p.

2.4.2 Reducing computational delays

Real-time implementation of a discrete-time state-space controller represented by

$$(A_{K,\mathrm{d}}, B_{K,\mathrm{d}}, C_{K,\mathrm{d}}, D_{K,\mathrm{d}}),$$

implies that matrix-vector products must be computed by the processor. At every sample time the current state $x_{K,\mathrm{d}}$ must be multiplied by $A_{K,\mathrm{d}}$ and $C_{K,\mathrm{d}}$. Furthermore, measurement y must be multiplied by $B_{K,\mathrm{d}}$ and $D_{K,\mathrm{d}}$. These operations require computation time,

---

[2]This affine dependence is exploited for several structured controller synthesis techniques, see e.g. Section 3.2.2 and Chapter 5.

which results in a time delay. Large time delays usually limit the closed-loop bandwidth that can be achieved. To achieve high control performance it is therefore important to keep the real-time computational effort small. This motivates to compute controllers with small McMillan degree.

Remark. To reduce the number of scalar multiplications, one might use a controller state-space realization with sparse system matrices, such as the controllable canonical forms for SISO systems, or the ('numerically less sensitive') parametrization with a tri-diagonal A_K-matrix [127]. Still, the number of scalar multiplication and addition operations may be large.

2.4.3   Strong stabilization

Most controllers for mechanical systems are not asymptotically stable. Indeed often integral action is required to eliminate steady-state off-set and this gives a pole at $s = 0$. To prevent that the controller states start 'drifting' when the controller is (temporarily) disconnected from the plant it is often desirable that the remaining controller poles are in the open left-half plane $\mathbb{C}^-$. The design of a stabilizing controller that is itself stable is called *strong stabilization* in the literature, see e.g. [48] and for some recent development in this field [42]. Stability of the controller can be enforced by adding the following additional requirement to (2.19)

$$A_K^T X_K + X_K A_K \prec 0 \text{ and } X_K \succ 0 \tag{2.58}$$

which is itself a bilinear matrix inequality in X_K and A_K. Although strong stabilization is not the main focus of this thesis, strongly stabilizing controllers can be computed with the methods developed in Chapters 4 and 7. Since (2.58) is a BMI of similar type as (2.19), it can easily be added to the problem, which can be solved by the algorithms in those chapters. This also holds true for multi-objective controller design problems [212, 167, 173].

2.4.4   Consequences of constraints on controller structure

If any of the above structural constraints is added to the full-order controller synthesis problem, the procedure to transform the problem into an LMI problem as described in Section 2.3 breaks down. In particular it has been pointed out in Theorem 2.11 that a controller with a certain order r can only be computed if $\text{Rank}(X - Y^{-1}) \leq r$. This rank condition is non-convex.

Similarly, if the PID-structure or diagonal structure as described in Section 2.4.1 is required, the projection in Section 2.3 breaks down, since in the Projection Lemma, Lemma 2.9, we can not guarantee that the solution K has an a priori given structure. Finally the approach is not feasible for strong stabilization, since (2.58) together with (2.19) can not simultaneously be projected, since (to the best of our knowledge) no such projections in analogue of Lemma 2.10 exist.

As will be explained in Section 2.6, the existence of generally applicable LMI solutions to controller synthesis problems with these kind of constraints seems unlikely. However we would like to stress that no proof of non-existence is known. Only in specific cases a transformation into LMIs is known, for which some important examples will be given in Section 2.7.

2.5 Optimization with matrix inequalities

For a generalized plant as in (2.2) and a structured controller $K(p)$ parameterized by $p \in \mathbb{R}^{l_c}$, the closed-loop matrices are

$$\begin{pmatrix} A_{cl}(p) & B_{cl}(p) \\ C_{cl}(p) & D_{cl}(p) \end{pmatrix} := \begin{pmatrix} \hat{A} & \hat{B}_1 \\ \hat{C}_1 & \hat{D}_{11} \end{pmatrix} + \begin{pmatrix} \hat{B}_2 \\ \hat{D}_{21} \end{pmatrix} K(p) \begin{pmatrix} \hat{C}_2 & \hat{D}_{12} \end{pmatrix}. \tag{2.59}$$

As follows from Section 2.3, the $\mathcal{H}_\infty$-optimal structured controller synthesis problem can be formulated as the BMI optimization problem

$$\begin{array}{ll} \text{infimize} & \gamma \\ \text{subject to} & X \succ 0 \\ & \begin{pmatrix} A_{cl}(p)^T X + X A_{cl}(p) & X B_{cl}(p) & C_{cl}(p)^T \\ B_{cl}(p)^T X_{cl} & -\gamma I & D_{cl}(p)^T \\ C_{cl}(p) & D_{cl}(p) & -\gamma I \end{pmatrix} \prec 0 \end{array}$$

We have seen in Section 2.3 that the full-order control problem can be formulated as an LMI problem. It is therefore not surprising that optimization with LMIs and BMIs plays an important role in this thesis. In this section we will discuss these problems in more detail.

2.5.1 Linear Matrix Inequalities

An LMI is an inequality

$$F(x) := F_0 + F_1 x_1 + \ldots + F_m x_m \prec 0 \tag{2.60}$$

where $F_i \in \mathcal{S}^N$, $i = 0, 1, \ldots, m$ are data matrices and $x \in \mathbb{R}^m$ is the unknown optimization variable (which should not be confused with the state vector in the previous sections). Inequality (2.60) is a convex constraint, by which we mean that the feasible set $\mathcal{G} := \{x \in \mathbb{R}^m : F(x) \prec 0\}$ is convex. (See [155] for a definition of convex sets.) This follows from affinity of F and convexity of the set of symmetric positive semi-definite $N \times N$ matrices. This set is referred to as the positive semi-definite cone. A cone in a vector space $\mathcal{V}$ [155] is a set $\mathcal{C} \subset \mathcal{V}$ that is closed under multiplication by a positive scalar, i.e.

$$x \in \mathcal{C} \Rightarrow \lambda x \in \mathcal{C} \text{ for all } \lambda > 0.$$

Optimization over constraints that restrict matrices to the semi-definite cone is referred to as Semi-Definite Programming (SDP)

As a consequence, the LMI optimization problem

$$\inf_{F(x) \prec 0} c^T x \tag{2.61}$$

is convex. Let p_{opt} be the optimal value of (2.61), where we use the convention that $p_{opt} = \infty$ if the problem is infeasible, i.e. if $\mathcal{G} = \{x \in \mathbb{R}^m : F(x) \prec 0\}$ is empty. Furthermore we write $p_{opt} = -\infty$ if (2.61) is unbounded, i.e. if $c^T x$ can be made arbitrarily small by choosing a suitable $x \in \mathcal{G}$.

Convex optimization problems have the nice property that a local optimum (if it exists) is also a global optimum.

Remark. The reader is referred to the book [29] for more details on convex optimization. Furthermore, in [26] convex optimization is applied to a very wide range of controller synthesis problems. For application of convex optimization to other application areas like trust topology design or combinatorial optimization the book [9] is recommended. The application of LMI optimization to control problems in particular is described in [27] and [176].

2.5.2 LMI optimization

The ellipsoid algorithm is an optimization method for LMI problems that provides an almost optimal solution or an infeasibility certificate in *polynomial time*, if we know that the feasible region $\mathcal{G} = \{x \in \mathbb{R}^m : F(x) \prec 0\}$ is contained in a ball of radius ρ:

$$\mathcal{G} \subset \{x \in \mathbb{R}^m | \, \|x\| < \rho\}.$$

Finding a solution in polynomial time roughly means that the number of basic arithmetic operations that is required to compute the solution is a polynomial function of the size of the data matrices of the problem, i.e. in m and N for (2.61). The reader is referred to [9] for the precise statements and for a proof of the polynomial complexity of the ellipsoid algorithm.

In spite of this guaranteed bound on the efficiency, the performance of the ellipsoid algorithm is not very fast in practice [9, 27].

The so-called Interior Point methods, first developed for linear programming, turn out to be much faster in practice, while at the same time have guaranteed polynomial-time bounds on the numerical complexity (see e.g. [157]). They solve a sequence of 'simple' problems to get a solution to the original more 'difficult' problem. The difficult problem is the minimization of the unconstrained non-smooth function

$$\psi(x) := \begin{cases} c^T x & \text{for all } x \in \mathbb{R}^m \text{ with } F(x) \prec 0 \\ \infty & \text{otherwise} \end{cases} \tag{2.62}$$

which is obviously equivalent to (2.61). The sequence of subproblems that are solved are the unconstrained minimization of a so-called *barrier function* $\phi(x, \mu)$. For the LMI problems as in (2.61) often the following barrier function is used:[3]

$$\phi(x, \mu) := \begin{cases} \frac{c^T x}{\mu} - \log(\det(F(x))) & \text{for all } x \in \mathbb{R}^m \text{ with } F(x) \prec 0 \\ \infty & \text{otherwise} \end{cases},$$

where μ is the so-called *barrier parameter*. It is not hard to show that this barrier function satisfies the following properties for all $\mu > 0$:

- $\phi(x, \mu) = \infty$ if $F(x)$ is not negative definite,

- $\lim_{k \to \infty} \phi(x_k, \mu) = \infty$ for every sequence $x_k \in \mathcal{G}$, $k = 1, 2, \ldots$ with a limit on the boundary of $\mathcal{G}$,

- $\phi(x, \mu)$ is self-concordant, i.e. for all $x \in \mathcal{G}$, $\mu > 0$ and $h \in \mathbb{R}^n$,

$$y(t) := \phi(x + th, \mu) \text{ satisfies } [y'''(0)]^2 \le 4y''(0)^3,$$

---

[3]see Figure 7.1 on page 116 for an example of a barrier function.

where $\mathcal{G} = \{x \in \mathbb{R}^m : F(x) \prec 0\}$. These properties are crucial in the proof for polynomial complexity of the IP algorithm for general convex problems and in particular for LMI problems, as described in the book by Nesterov and Nemirovskii [137], see also [29], [9] and [103]. A conceptional algorithm of an Interior Point method is

1. given $x_0 \in \mathcal{G}$, $\mu_0 > 0$ large enough, $k = 0$ and $\alpha > 1$

2. find an approximate minimizer x_* of $\phi(x, \mu_k)$, with Newton's method and line search, starting at $x = x_k$

3. set $\mu_{k+1} = \frac{\mu_k}{\alpha}$, $x_{k+1} = x_*$, $k = k + 1$ and go to step 2

Note that for small μ_k the minimum in Step 2 can be efficiently computed if the initial point x_k is not too far away (in some specific norm [103]) from a sub-optimal solution. This is the fundamental reason for the need to carefully decrease μ. For $\mu \downarrow 0$, $\phi(\cdot, \mu)$ converges (point-wise) to ψ in (2.62), which illustrates that the subproblems to be solved are simpler (i.e. smooth) estimates of the original 'difficult problem'. The reader is referred to [209] and [156] for Interior-Point optimization for Linear Programs and to [205] for a recent collection of texts on Semi-Definite Programming.

2.5.3 Lagrange duality of LMI problems

The LMI minimization problem (2.61) is related to a certain maximization problem, which is referred to by the *Lagrange dual* problem:

$$\max_{S \succeq 0 \text{ and } c_i + \langle S, F_i \rangle = 0 \text{ for all } i=1,\ldots,n} \langle S, F_0 \rangle, \tag{2.63}$$

with optimal value d_{opt}, where $\langle \cdot, \cdot \rangle$ is the standard inner product on the set of symmetric matrices defined as $\langle A, B \rangle := \text{Trace}(AB)$. Recall that the optimal value of the *primal* problem (2.61) is p_{opt}. *Weak duality* implies $p_{\text{opt}} \geq d_{\text{opt}}$ and always holds true. If these optimal values are actually equal (i.e. $p_{\text{opt}} = d_{\text{opt}}$), then we say that *strong duality* holds.

Slater's constraint qualification is a condition that implies strong duality, see e.g. [29]. For the LMI problem in (2.61), Slater's conditions reads as follows:

there exists an x with $F(x) \prec 0$.

Since the constraints in (2.61) are strict, Slater's condition reduces in this case to feasibility of the primal problem (2.61). The concept of duality is very important in optimization and in particular in optimization for control, as will become clear in the later chapters of this thesis. In [196] and [6] some existing results in control theory such as the Bounded-Real Lemma are (re-)proven using duality theory.

Remark. For the problem of infimizing γ subject to (2.25), (2.26) and (2.27), Slater's condition is satisfied if Assumption 2.4 holds true, i.e. if $\mathbf{P}$ is a generalised plant. Indeed if $\mathbf{P}$ is a generalised plant, there exists a full-order stabilizing controller $K \in \mathbb{R}^{(n+m_2) \times (n+p_2)}$. By Lyapunov's theorem there hence exists an $X_{\text{cl}} \in \mathcal{S}^{2n}$ with $X_{\text{cl}} \succ 0$ such that

$$A_{\text{cl}}^T(K) X_{\text{cl}} + X_{\text{cl}} A_{\text{cl}}(K) \prec 0. \tag{2.64}$$

Let X and Y be defined through (2.23) and (2.24) and let the columns of $\begin{pmatrix} N_1^T & N_2^T \end{pmatrix}^T$ and $\begin{pmatrix} M_1^T & M_2^T \end{pmatrix}^T$ be arbitrary bases of the kernels of $\begin{pmatrix} C_2 & D_{21} \end{pmatrix}$ and $\begin{pmatrix} B_2^T & D_{12}^T \end{pmatrix}$, respectively. (2.64) implies that

$$N^T \left(X(A + B_2 D_K C_2) + U B_K C_2 + (A + B_2 D_K C_2)^T X + (U B_K C_2)^T \right) N =$$
$$= N^T (XA + A^T X) N \prec 0.$$

A similar argument reveals that $M^T(AY + YA^T)M \prec 0$. Hence by choosing γ large enough, (2.25) and (2.26) are satisfied. Since (2.27) holds true, we can slightly perturb X and Y such that (2.25) and (2.26) are still satisfied and

$$\begin{pmatrix} X & I \\ I & Y \end{pmatrix} \succ 0.$$

Hence, Slater's condition is indeed satisfied.

2.5.4 Lagrange duality for general SDPs

LMI dualization is actually a special case of Lagrange dualization for general SDPs of the form $\inf_{x \in \mathbb{R}^m,\ G(x) \preceq 0} f(x)$ where $f : \mathbb{R}^m \to \mathbb{R}$ and $G : \mathbb{R}^m \to \mathcal{S}^N$ are arbitrary scalar and matrix-valued functions respectively. Using weak Lagrange duality we can construct lower bounds on the optimal value p_{opt} of this problem. Indeed, for an arbitrary $S \succ 0$ the optimal value $p_{\text{opt}} := \inf_{x \in \mathbb{R}^m,\ G(x) \preceq 0} f(x)$ is bounded below by the infimum over $x \in \mathbb{R}^m$ of the Lagrangian defined by

$$L(x, S) := f(x) + \text{Trace}(S^T G(x)).$$

This follows since for all $S \succeq 0$

$$p_{\text{opt}} := \inf_{x \in \mathbb{R}^m,\ G(x) \preceq 0} f(x) \geq \inf_{x \in \mathbb{R}^m,\ G(x) \preceq 0} f(x) + \text{Trace}(S^T G(x)) \geq$$
$$\geq \inf_{x \in \mathbb{R}^m} f(x) + \text{Trace}(S^T G(x)) = \inf_{x \in \mathbb{R}^m} L(x, S).$$

If G and f are convex and $\{x \in \mathbb{R}^m | G(x) \prec 0\}$ is nonempty, then strong duality holds true [29]:

$$p_{\text{opt}} = \max_{S \succeq 0} \inf_{x \in \mathbb{R}^m} L(x, S). \tag{2.65}$$

2.5.5 Optimality

In this section we consider the nonstrict version of (2.61), i.e. $\inf_{G(x) \preceq 0} f(x)$. For completeness we give the definition of an *optimal value* of this problem. p_{opt} is the optimal value if

- $f(x) \geq p_{\text{opt}}$ for all $\{x \mid G(x) \preceq 0\}$ and

- for every $\epsilon > 0$ there exists an x_* such that $G(x_*) \preceq 0$ and $f(x_*) < p_{\text{opt}} + \epsilon$.

An $x_{\text{opt}} \in \mathbb{R}^m$ is *globally optimal* for the SDP $\inf_{G(x) \preceq 0} f(x)$ if it is feasible (i.e. $G(x_{\text{opt}}) \preceq 0$) and $f(x_{\text{opt}}) = p_{\text{opt}}$. An $x_{\text{opt}} \in \mathbb{R}^m$ is *locally optimal* if there exists a $\rho > 0$ and a ball around x_{opt} with radius ρ, i.e. $\mathcal{B}_\rho := \{x \in \mathbb{R}^n | \ \|x_{\text{opt}} - x\| \leq \rho\}$, such that x_{opt} is globally optimal for $\inf_{G(x) \preceq 0, x \in \mathcal{B}_\rho} f(x)$. In Section 4.5.6 we will derive conditions for global optimality for polynomial semi-definite programs in general and the structured $\mathcal{H}_\infty$-optimal controller synthesis problem in particular. Additionally, we derive computationally less demanding conditions for local optimality of these problems in Section 7.3.

2.5.6 S-procedure

An important application of Lagrange duality is the S-procedure [211], which is an important tool for robustness analysis. See [109] for some historical remarks. The S-procedure gives a sufficient condition for a quadratic polynomial $f_0(x)$ to be nonnegative on a set $\mathcal{G} := \{x \in \mathbb{R}^m : f_i(x) \leq 0, \; i = 1, \ldots, p\}$, where $f_i, \; i = 1, \ldots, p$ are also quadratic. Any finite set of quadratic polynomials $f_i, \; i = 0, 1, \ldots, p$ can always be written as

$$f_i(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}^T \begin{pmatrix} r_i & s_i^T \\ s_i & Q_i \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} \tag{2.66}$$

for some $r_i \in \mathbb{R}$, $s_i \in \mathbb{R}^n$ and $Q_i \in \mathcal{S}^m$, $i = 0, 1, \ldots, p$.

Lemma 2.17 *(S-procedure) Let quadratic polynomials f_i be given as in (2.66) for some $r_i \in \mathbb{R}$, $s_i \in \mathbb{R}^n$ and $Q_i \in \mathcal{S}^m$, $i = 1, \ldots, p$. Let $\mathcal{G} := \{x \in \mathbb{R}^m : f_i(x) \leq 0, \; i = 1, \ldots, p\}$. If*

$$\exists \lambda_i \geq 0, \;\; i = 1, \ldots, p, \;\; such \; that \;\; \begin{pmatrix} r_0 & s_0^T \\ s_0 & Q_0 \end{pmatrix} + \sum_{i=1}^{p} \lambda_i \begin{pmatrix} r_i & s_i^T \\ s_i & Q_i \end{pmatrix} \succeq 0 \tag{2.67}$$

then

$$f_0(x) \geq 0 \; for \; all \; x \in \mathcal{G}. \tag{2.68}$$

The converse (i.e. ((2.68)⇒(2.67)) holds true if $p = 1$ and the following constraint qualification is satisfied:

$$\{x \in \mathbb{R}^m : f_1(x) < 0\} \; is \; nonempty.$$

Furthermore, the converse ((2.68)⇒(2.67)) holds true for $p \geq 1$, if

$$\{x \in \mathbb{R}^m : f_i(x) < 0, \; i = 1, \ldots, p\} \; is \; nonempty \tag{2.69}$$

and $Q_i \succeq 0$, $i = 0, 1, \ldots, p$.

Proof. See e.g. [172].

$\blacksquare$

Remark. If $Q_i \succeq 0$, $i = 0, 1, \ldots, p$ then all f_i, $i = 1, \ldots, p$ are convex quadratic functions. Lemma 2.17 is just the standard strong Lagrange duality result for convex quadratic optimization problems. The surprising fact is that in the non-convex case the sufficient condition is also necessary for $p = 1$.

2.5.7 Full block S-procedure

A nontrivial extension of the S-procedure is its full-block version, independently proved by [98] and [170]. This result has far-reaching consequences for robust controller analysis and is the basis for many relaxation schemes, see e.g. [171, 172] and the references therein.

Typically in robust analysis the objective is to verify that a certain system property such as stability or $\mathcal{L}_2$ performance is satisfied, if the system varies in a certain set of
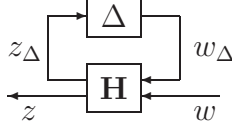
Figure 2.4: Generalized plant interconnection structure for robust analysis

systems parameterized by a so-called "uncertain" parameter. The generalized-plant set-up for robustness analysis as depicted in Figure 2.4 is often used for this purpose, where δ is the uncertain parameter, $\Delta(\delta)$ is an affine function of the uncertainty δ and the performance channel is from w to z. Note that Figures 2.4 and 2.1 can be combined for robust controller synthesis. Suppose

$$H(p) = \left(\begin{array}{cc} A & B \\ C(p) & D(p) \end{array} \right),$$

for some matrices A, B, affine functions $C(p)$ and $D(p)$ and auxiliary variable $p \in \mathbb{R}^{\mathbb{N}}$. Many robust analysis problems can be reformulated as the search for a $p \in \mathbb{R}^{\mathbb{N}}$ such that the map from w to z given by the upper LFT $\mathcal{F}_u(\Delta(\delta), H(p)) = D(p) + C(p)\Delta(\delta)(I - A\Delta(\delta))^{-1}B$ satisfies the following quadratic performance criterion

$$\left(\begin{array}{c} I \\ \mathcal{F}_u(\Delta(\delta), H(p)) \end{array} \right)^T Q \left(\begin{array}{c} I \\ \mathcal{F}_u(\Delta(\delta), H(p)) \end{array} \right) \prec 0 \tag{2.70}$$

for some symmetric matrix Q and for all δ varying in some compact set $\boldsymbol{\delta} \subset \mathbb{R}^d$.

Theorem 2.18 *Let the uncertain parameter δ be varying in the set $\boldsymbol{\delta} \subset \mathbb{R}^d$ and let $p \in \mathbb{R}^{\mathbb{N}}$ be fixed. If $\boldsymbol{\delta}$ is compact and $I - A\Delta(\delta)$ is nonsingular for all $\delta \in \boldsymbol{\delta}$ then (2.70) holds true for all $\delta \in \boldsymbol{\delta}$ if and only if there exists a multiplier P satisfying*

$$\left(\begin{array}{c} \Delta(\delta) \\ I \end{array} \right)^T P \left(\begin{array}{c} \Delta(\delta) \\ I \end{array} \right) \succ 0 \text{ for all } \delta \in \boldsymbol{\delta} \tag{2.71}$$

and

$$\left(\begin{array}{cc} I & 0 \\ A & B \end{array} \right)^T P \left(\begin{array}{cc} I & 0 \\ A & B \end{array} \right) + \left(\begin{array}{cc} 0 & I \\ C(p) & D(p) \end{array} \right)^T Q \left(\begin{array}{cc} 0 & I \\ C(p) & D(p) \end{array} \right) \prec 0. \tag{2.72}$$

Proof. See [170] or [98].

∎

Remark. If $\boldsymbol{\delta}$ is not compact, the S-procedure is not necessarily lossless. For certain simple uncertainty block structures, losslessness can still be shown [172].

2.5.8 Bilinear Matrix Inequalities

A bilinear matrix inequality is an inequality of the form

$$F(x, y) := F_0 + F_1 x_1 + \ldots + F_m x_m + G_1 y_1 + \ldots + G_n y_n + \sum_{i=1}^{m} \sum_{j=1}^{n} H_{ij} x_i y_j \preceq 0, \tag{2.73}$$

where F_i, G_j and H_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, m$ are symmetric matrices of the same size. The seemingly innocent difference with the LMI in (2.60) is the reason for a huge difference between LMI and BMI optimization problems: the first are convex problems and the second are not necessarily convex. It has been shown in [191] that some instances of BMI problems are *NP-hard* (Nondeterministic Polynomial) in the sense of [9]. This will be further addressed in the next section.

We will present an Interior Point method and sum-of-squares relaxations for BMI optimization in Chapters 4 and 7. Other approaches to solve BMI problems are for instance sequential LMI optimization as discussed in Section 3.2 and the Branch and Bound method discussed in Section 3.3.

In Section 2.3 we have seen that the fixed-order control problem can either be written as a BMI problem or an LMI problem with a rank constraint. It is interesting to mention that these two latter problems are intimately related, in the sense that any BMI problem can be equivalently formulated as an LMI problem with a rank constraint and vice versa [132].

2.6   Complexity of the structured synthesis problem

We will use in this section the terms *polynomial-time* and *NP-hard*, and refer the reader to [9] for an explanation of these concepts.

For simplicity, we consider the static output-feedback stabilization problem: given matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{p \times n}$ determine whether there exists a matrix $K \in \mathbb{R}^{m \times p}$ such that $A + BKC$ is Hurwitz. The complexity of this problem is still open. More precisely, it is still unknown whether there exists an algorithm that solves any instance of this problem in polynomial-time.

Blondel and Tsitsiklis [20] have shown that this control problem is NP-hard if the coefficients of the controller K are constrained to lie in pre-specified intervals. See e.g. [21] for a definition of NP-hardness and a survey on computational complexity results in control. Blondel and Tsitsiklis conjecture that the SOF problem is also NP-hard if the extra constraints are removed. To date, however, no such result is known.

Another result on the negative side, that supports to some extent the conjecture that the SOF problem is NP-hard, is due to Fu and Luo [63]. They show that for all $n \in \mathbb{N}$ there exist affine mappings $F : \mathcal{S}^n \to \mathcal{S}^n$ and $G : \mathcal{S}^n \to \mathcal{S}^n$, such that verifying the feasibility of

$$F(X) \prec 0, \quad G(Y) = 0 \text{ and } XY = I_n \tag{2.74}$$

is also NP-hard. The existence of a static output feedback stabilization problem is equivalent to (2.74) for some affine F and G. Finally, as mentioned in the previous section, Toker and Özbay [191] proved that some instances of BMI problems are NP-hard.

A result on the positive side is the yet unpublished work of Mesbahi [131], who claims to have found an SDP algorithm for the static output feedback problem.

2.7   Structured synthesis under additional hypotheses

Under additional hypothesis the fixed-order $\mathcal{H}_\infty$-optimal controller synthesis problem simplifies significantly and can be solved by convex optimization. We discuss in the sequel two such situations.

State Feedback

For the state-feedback problem, the structured controller synthesis problem is easily solved. State feedback corresponds to $C_2 = I$ and $D_{21} = 0$. We can then choose $N = \begin{pmatrix} 0 \\ I_{m_2} \end{pmatrix}$ such that (2.25) reduces to

$$\begin{pmatrix} -\gamma I & D_{11}^T \\ D_{11} & -\gamma I \end{pmatrix} \prec 0.$$

Since this equation is implied by (2.25), (2.26) becomes redundant. Hence for any feasible X we can always choose $Y = X^{-1}$, which implies from Lemma 2.10 that the sub-optimal state feedback controller can always be chosen as a static matrix. More precisely, if there exists a dynamic state-feedback controller with closed-loop $\mathcal{H}_\infty$-norm smaller than γ, we can always find a static feedback with the same performance bound.

Block-diagonal control

Let us consider the important problem of synthesis of a block-diagonal controller

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} F & 0 \\ 0 & G \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

with dynamic $G \in \mathcal{RL}_\infty^{m_2 \times p_2}$ and static $F \in \mathbb{R}^{m_3 \times p_3}$ for the plant P with suitable partitioning as in

$$P = \left(\begin{array}{c|cc} P_{11} & P_{12} & P_{13} \\ \hline P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{array} \right),$$

where $P_{ij} \in \mathcal{RL}_\infty^{p_i \times m_j}$ $i, j = 1, 2, 3$. Under the additional hypothesis that P_{23}, P_{32} and P_{33} are identically zero, the structured controller synthesis problem can be solved by convex optimization. The reader is referred to [169] for the details.

2.8   Problem statement of this thesis work

We are now in the position to formulate the problem of this thesis more precisely. As has been clarified in the previous sections, the structured $\mathcal{H}_\infty$-optimal control problem is a very hard optimization problem. For this reason we need to extend the existing theory on optimization for controller synthesis to be able to develop suitable algorithms to solve this problem. We embed these new theoretical results into novel algorithms for controller optimization, as well as for computation of optimality certificates of controllers.

Starting point will be the static output-feedback $\mathcal{H}_\infty$-optimal control problem in terms of BMIs:

infimize $\qquad \gamma$

subject to $\qquad \mathcal{B}(\gamma, X, K) := - \begin{pmatrix} A_{\mathrm{cl}}(K)^T X + X A_{\mathrm{cl}}(K) & X B_{\mathrm{cl}}(K) & C_{\mathrm{cl}}(K)^T \\ B_{\mathrm{cl}}(K)^T X_{\mathrm{cl}} & -\gamma I & D_{\mathrm{cl}}(K)^T \\ C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) & -\gamma I \end{pmatrix} \succ 0$

$\qquad X \succ 0$

(2.75)

where $(A_{\mathrm{cl}}(K), B_{\mathrm{cl}}(K), C_{\mathrm{cl}}(K), D_{\mathrm{cl}}(K))$ are defined in (2.8). We choose this formulation for the following reasons

- it is very flexible, in the sense that almost all of the relevant structured $\mathcal{H}_\infty$-optimal controller synthesis problems can be molded into this format

- the constraints are smooth in the decision variables X, K and γ, which is a desirable property for the development of optimization algorithms for this problem

- the constraints are quadratic in the decision variables, such that relaxation schemes for matrix-valued polynomial optimization, as are developed in this thesis, can be applied.

The ultimate goal is to find a globally optimal solution controller, or stated otherwise, a controller for which a global optimality certificate can be found. The performance of such a controller can definitely not be improved by any other controller of the same structure. Global optimality allows, therefore, to decide that some of the structural constraints must be relaxed to reach the desired performance. With local search methods, without global optimality certificates, such a decision can never be made with certainty. One of the topics of this thesis is therefore the development of algorithms to compute global optimality certificates.

For systems with large McMillan degree, e.g. 30 and larger, global controller optimization seems not feasible with the current computer power. For these kind of systems local controller optimization techniques have therefore been developed in this thesis work. The aim is then to compute within reasonable time a locally optimal controller. We have made an effort to reduce the computation time of these algorithms by *exploiting the control-theoretic characteristics of the underlying problem*, i.e. the structure that is present in the objective function or constraints that is a consequence of it being a control problem. An example is for instance the Lyapunov-like structure in the BMIs of (2.75).

Remark. Exploiting the control-theoretic characteristics of the underlying problem is sometimes denoted by 'structure exploitation' in the literature. Since the term 'structure' might be confused with the (conceptually different) structure of a controller (i.e. diagonal, PID etc.), we use the terminology 'exploiting the control-theoretic characteristics' instead.

Summarizing, the research question posed in Chapter 1 can be made more specific as follows: how can we develop

① techniques for solving (2.75) with the focus on (global or local) convergence guarantees

② schemes to assess local and global optimality of the resulting controllers

③ algorithms for (2.75) with the focus on reducing the complexity of the computations, such that they can be employed to design controllers for industrial servo-systems?

Chapter 3

Literature survey

The synthesis of fixed-order (or static-output feedback) controllers satisfying stability requirements and / or closed loop norm bounds have been an active area of research since as early as the beginning of the 70's. A whole variety of techniques has been developed, some of them of a surprisingly different nature. The existence of $\mathcal{H}_\infty$-optimal fixed-order controller synthesis algorithms based on (seemingly) totally different concepts can be (partially) explained by the different ways that a bound on the closed-loop $\mathcal{H}_\infty$ performance can be expressed, e.g., in the frequency domain, with Riccati equations and with matrix inequalities. In view of the large number of different techniques, only a few of them can be discussed in detail in this chapter. Relevant references to other algorithms are provided.

Probably the computationally cheapest and most straightforward way to compute reduced-order or static controllers is by controller reduction as explained in Section 3.1. The reduction is based on weighted balancing followed by truncation or residualization. This section also addresses an approach to quantify the loss of closed-loop performance in terms of the reduction error. In Section 3.2 we discuss algorithms based on sequential LMI optimization. A branch and bound algorithm will be described in Section 3.3. Nonsmooth optimization of structured controllers is discussed in Section 3.4 and we briefly address several other algorithms in Section 3.5. We conclude this chapter with a presentation of the contributions of this thesis with respect to the existing literature in Section 3.6.

3.1 Controller reduction

A computationally very cheap method to compute fixed-order controllers is reduction of some full order $\mathcal{H}_\infty$-optimal controller. In this method a balanced realization of a (possibly weighted) full order $\mathcal{H}_\infty$-optimal controller is computed to evaluate, in a heuristic way, the contribution of each controller state to the closed-loop performance. The controller states that have a small contribution are then eliminated by either residualization or truncation.

3.1.1 Controller reduction by direct balanced truncation

Model reduction

A well-known approach to model reduction is based on the so-called *Gramian matrices*. Let (A, B, C, D) be a minimal realization of a stable transfer function $G(s) = D + C(sI - A)^{-1}B$ with McMillan degree n. Then the controllability and observability Gramians of the realization are symmetric positive definite matrices $P \in \mathcal{S}^n$ and $Q \in \mathcal{S}^n$ respectively,

defined by

$$P = \int_0^\infty e^{At} B B^T e^{A^T t} dt$$

and

$$Q = \int_0^\infty e^{A^T t} C^T C e^{At} dt.$$

P and Q are related to the minimum input energy to reach a specific state and output energy respectively. To be more specific, $x_0^T P^{-1} x_0$ is the minimum required energy of $u \in \mathcal{L}_2(-\infty, 0]$ to reach the state x_0 at $t = 0$ from the zero state at $t = -\infty$, i.e. $x_0^T P^{-1} x_0$ is equal to the optimal value of the minimization problem [69]

$$
\begin{array}{ll}
\text{infimize} & \int_{-\infty}^0 u(t)^T u(t) dt \\
\text{subject to} & u \in \mathcal{L}_2(-\infty, 0], \ x(0) = x_0, \ \lim_{t \to -\infty} x(t) = 0 \\
& \dot{x}(t) = Ax(t) + Bu(t), \ \text{for all } t \leq 0
\end{array}
$$

Furthermore, $x_0^T Q x_0$ is the energy of the free response of the output $y(t)$ from initial condition $x(0) = x_0$ [69]:

$$x_0^T Q x_0 = \int_0^\infty y(t)^T y(t) dt, \text{ where } x(0) = x_0 \text{ and } \dot{x}(t) = Ax(t), \ y(t) = Cx(t) \text{ for all } t \geq 0.$$

It is easily verified that P and Q satisfy

$$AP + PA^T + BB^T = 0$$

and

$$A^T Q + QA + C^T C = 0.$$

The Gramians are not invariant under similarity transformations. In fact, there always exists a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ such that the transformed system

$$\left(\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline \tilde{C} & \tilde{D} \end{array} \right) = \left(\begin{array}{c|c} T^{-1}AT & T^{-1}B \\ \hline CT & D \end{array} \right)$$

has diagonal Gramians $\tilde{P} = T^{-1} P T^{-T}$ and $\tilde{Q} = T^T Q T$ that satisfy [69]

$$\tilde{P} = \tilde{Q} = \Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & & 0 \\ 0 & 0 & \sigma_3 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \sigma_n \end{pmatrix},$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$, are the *Hankel singular values* (HSVs) of G. The HSVs are invariant under similarity transformations.

The realization $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ with state vector $\tilde{x}$ is called the *balanced realization* of G. The balanced realization is not unique. If the HSVs of the system are distinct, then it is unique up to similarity transformation with a sign matrix [69], i.e. a diagonal matrix with diagonal elements equal to either 1 or -1.

Balanced reduction of the system is done by elimination of the states $x_e = \left(\begin{array}{cc} 0 & I_{n-r} \end{array} \right) \hat{x}$ and hence retaining the states $x_r = \left(\begin{array}{cc} I_r & 0 \end{array} \right) \hat{x}$ corresponding to the r largest Hankel

singular values. Let us introduce to this purpose the partitioning of the state-space matrices implied by $x = \begin{pmatrix} x_{\mathrm{r}}^T & x_{\mathrm{e}}^T \end{pmatrix}^T$:

$$\left(\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline \tilde{C} & \tilde{D} \end{array} \right) = \left(\begin{array}{cc|c} \tilde{A}_{\mathrm{rr}} & \tilde{A}_{\mathrm{re}} & \tilde{B}_{\mathrm{r}} \\ \tilde{A}_{\mathrm{er}} & \tilde{A}_{\mathrm{ee}} & \tilde{B}_{\mathrm{e}} \\ \hline \tilde{C}_{\mathrm{r}} & \tilde{C}_{\mathrm{e}} & \tilde{D} \end{array} \right) \tag{3.1}$$

Model reduction by *Truncation* eliminates the states x_{e}, which results in the following reduced system:

$$\left(\begin{array}{c|c} A_{\mathrm{T}} & B_{\mathrm{T}} \\ \hline C_{\mathrm{T}} & D_{\mathrm{T}} \end{array} \right) := \left(\begin{array}{c|c} \tilde{A}_{\mathrm{rr}} & \tilde{B}_{\mathrm{r}} \\ \hline \tilde{C}_{\mathrm{r}} & \tilde{D} \end{array} \right) = \left(\begin{array}{c|c} \begin{pmatrix} I_r & 0 \end{pmatrix} T^{-1} & 0 \\ \hline 0 & I \end{array} \right) \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) \left(\begin{array}{c|c} T \begin{pmatrix} I_r & 0 \end{pmatrix}^T & 0 \\ \hline 0 & I \end{array} \right) \tag{3.2}$$

In contrast to truncation, *Residualization* approximates the dynamics in x_{e}, i.e. $\dot{x}_{\mathrm{e}} = \tilde{A}_{\mathrm{er}} x_{\mathrm{r}} + \tilde{A}_{\mathrm{ee}} x_{\mathrm{e}} + \tilde{B}_{\mathrm{e}} u$, by the algebraic relation

$$0 = \tilde{A}_{\mathrm{er}} x_{\mathrm{r}} + \tilde{A}_{\mathrm{ee}} x_{\mathrm{e}} + \tilde{B}_{\mathrm{e}} u,$$

which yields the reduced system

$$\left(\begin{array}{c|c} A_{\mathrm{R}} & B_{\mathrm{R}} \\ \hline C_{\mathrm{R}} & D_{\mathrm{R}} \end{array} \right) := \left(\begin{array}{c|c} \tilde{A}_{\mathrm{rr}} - \tilde{A}_{\mathrm{re}} \tilde{A}_{\mathrm{ee}}^{-1} \tilde{A}_{\mathrm{er}} & \tilde{B}_{\mathrm{r}} - \tilde{A}_{\mathrm{re}} \tilde{A}_{\mathrm{ee}}^{-1} \tilde{B}_{\mathrm{e}} \\ \hline \tilde{C}_{\mathrm{r}} - \tilde{C}_{\mathrm{e}} \tilde{A}_{\mathrm{ee}}^{-1} \tilde{A}_{\mathrm{er}} & \tilde{D} - \tilde{C}_{\mathrm{e}} \tilde{A}_{\mathrm{ee}}^{-1} \tilde{B}_{\mathrm{e}} \end{array} \right). \tag{3.3}$$

One of the main differences between reduction based on truncation and residualization is that truncation preserves the direct-feedthrough of the original system whereas the residualization preserves the steady-state gain, i.e. $G(\infty) = G_{\mathrm{T}}(\infty)$ and $G(0) = G_{\mathrm{R}}(0)$, where

$$G_{\mathrm{T}}(s) \quad := \quad D_{\mathrm{T}} + C_{\mathrm{T}}(sI - A_{\mathrm{T}})^{-1} B_{\mathrm{T}} \tag{3.4}$$
$$G_{\mathrm{R}}(s) \quad := \quad D_{\mathrm{R}} + C_{\mathrm{R}}(sI - A_{\mathrm{R}})^{-1} B_{\mathrm{R}}. \tag{3.5}$$

This is easily verified from (3.2) and (3.3). It was mentioned above that the balanced realization is unique up to a similarity transformation with sign matrix if the HSVs $\sigma_1 > \sigma_2 > \ldots > \sigma_n \geq 0$ are distinct. Under this assumption the transfer function of the rth-order reduced model does not depend on the specific balanced realization [206]. This immediately follows from (3.2)-(3.5).

A nice property of both balanced truncation and residualization is their preservation of stability, i.e. if the original system G is stable and the HSVs are distinct, then the reduced system is also stable. Furthermore, the $\mathcal{H}_\infty$ approximation error satisfies an upper bound, equal to 'twice the sum of the tail' of the Hankel singular values. These facts are the content of the following lemma.

Lemma 3.1 *Suppose the stable transfer function $G(s)$ admits a balanced state-space representation $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ with $\tilde{A}$ Hurwitz and with Gramians $\tilde{P} = \tilde{Q} = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$, where the Hankel singular values satisfy $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$ and $\sigma_r > \sigma_{r+1}$ for some $1 \leq r < n$. Consider the balanced truncated and residualized state space systems (A_T, B_T, C_T, D_T) and (A_R, B_R, C_R, D_R) given by (3.2) and (3.3) respectively, using the partitioning in (3.1) with $\tilde{A}_{rr} = \begin{pmatrix} I_r & 0 \end{pmatrix} \tilde{A} \begin{pmatrix} I_r & 0 \end{pmatrix}^T$. Then A_T and A_R are stable. Furthermore $G_T(s)$ and $G_R(s)$ given by (3.4) and (3.5) respectively satisfy*

$$\|G - G_T\|_\infty \leq 2 \sum_{i=r+1}^{n} \sigma_i \tag{3.6}$$

and

$$\|G - G_R\|_\infty \leq 2 \sum_{i=r+1}^{n} \sigma_i \tag{3.7}$$

Proof. see [69] and [74] for truncation and residualization, respectively.

∎

Direct controller reduction

One of the most straightforward controller reduction methods is direct truncation or residualizion of the balanced controller. This is computationally cheap, if compared to other methods for fixed-order controller synthesis. P and Q can easily be computed for systems with state dimension 100, using for instance the Bartels-Stewart algorithm as described in Section 7.2.2. For systems with larger state dimensions, dedicated Lyapunov solvers [13] that exploit sparsity might be more suitable.

A disadvantage of the method is that the resulting controller is not guaranteed to be stabilizing, even though the original controller is. Furthermore the closed-loop performance with the reduced controller is often much worse than the full-order optimal one. To improve this, the so-called *frequency weighted reduction* has been developed to take into account that the controller is part of a closed-loop interconnection, as explained in the next two sections.

3.1.2 Frequency weighted controller reduction

This section describes frequency weighted controller reduction [57, 206, 207, 139, 213, 1], see [139] for more references on this subject. Consider the series interconnection VKW where $K \in \mathcal{RH}_\infty$ is the transfer function of the full order controller and $V \in \mathcal{RH}_\infty$ and $W \in \mathcal{RH}_\infty$ are frequency weighting filters, all assumed to be real rational, proper and stable. Let (A_K, B_K, C_K, D_K), (A_V, B_V, C_V, D_V) and (A_W, B_W, C_W, D_W) be minimal realizations of K, V and W respectively, with $A_K \in \mathbb{R}^{n_c \times n_c}$, $A_V \in \mathbb{R}^{\mu \times \mu}$ and $A_W \in \mathbb{R}^{\nu \times \nu}$. A realization of VKW is then given by

$$\left(\begin{array}{ccc|c} A_V & B_V C_K & B_V D_K C_W & B_V D_K D_W \\ 0 & A_K & B_K C_W & B_K D_W \\ 0 & 0 & A_W & B_W \\ \hline C_V & D_V C_K & D_V D_K C_W & D_V D_K D_W \end{array} \right).$$

Let the Gramians of this state-space system be partitioned corresponding to the states of V, W and K as

$$P = \left(\begin{array}{ccc} P_{VV} & P_{VK} & P_{VW} \\ P_{VK}^T & P_{KK} & P_{KW} \\ P_{VW}^T & P_{KW}^T & P_{WW} \end{array} \right) \qquad \text{and} \qquad Q = \left(\begin{array}{ccc} Q_{VV} & Q_{VK} & Q_{VW} \\ Q_{VK}^T & Q_{KK} & Q_{KW} \\ Q_{VW}^T & Q_{KW}^T & Q_W \end{array} \right).$$

Then there exists a similarity transformation

$$T = \left(\begin{array}{ccc} I_\mu & 0 & 0 \\ 0 & T_K & 0 \\ 0 & 0 & I_\nu \end{array} \right)$$

44

with $T_K \in \mathbb{R}^{n_c \times n_c}$ such that the Gramians transform into

$$
\hat{P} = T^{-1}PT^{-T} = \begin{pmatrix} \hat{P}_{VV} & \hat{P}_{VK} & \hat{P}_{VW} \\ \hat{P}_{VK}^T & \Sigma_k & \hat{P}_{KW} \\ \hat{P}_{VW}^T & \hat{P}_{KW}^T & \hat{P}_{WW} \end{pmatrix}
$$

$$
\hat{Q} = T^T Q T = \begin{pmatrix} \hat{Q}_{VV} & \hat{Q}_{VK} & \hat{Q}_{VW} \\ \hat{Q}_{VK}^T & \Sigma_k & \hat{Q}_{KW} \\ \hat{Q}_{VW}^T & \hat{Q}_{KW}^T & \hat{Q}_{WW} \end{pmatrix}
$$

where $\Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_{n_c})$, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{n_c} \geq 0$. The reason for not fully diagonalizing P and Q is to prevent mixing of the states of K with those of V and W.

A *Frequency Weighted reduced* controller of order r is obtained using truncation or residualization of the last $n_c - r$ controller states after a similarity transformation with T_K. In case of truncation this leads to:

$$
\left(\begin{array}{c|c} A_{\mathrm{FW}} & B_{\mathrm{FW}} \\ \hline C_{\mathrm{FW}} & D_{\mathrm{FW}} \end{array} \right) := \left(\begin{array}{c|c} \left(\begin{array}{cc} I_r & 0 \end{array} \right) T_K^{-1} & 0 \\ \hline 0 & I \end{array} \right) \left(\begin{array}{c|c} A_K & B_K \\ \hline C_K & D_K \end{array} \right) \left(\begin{array}{c|c} T_K \left(\begin{array}{cc} I_r & 0 \end{array} \right)^T & 0 \\ \hline 0 & I \end{array} \right). \tag{3.8}
$$

Let us denote the corresponding transfer function by K_{FW}. For single sided weighting, i.e., if either $V = I$ or $W = I$, the transfer function K_{FW} is stable if K is stable [57]. See [139] and the references therein for $\mathcal{H}_\infty$-norm error bounds on some special cases of one-sided weighted reduction. For double-sided dynamic weighting, the reduced controller is not necessarily stable even if K is, let alone that a $\mathcal{H}_\infty$-bound on the reduction error is known in general. See [184] for an example of an unstable controller obtained by double-sided frequency weighted truncation of a stable controller.

3.1.3 Closed-loop Controller reduction

This section describes the so-called *closed-loop balanced reduction* as presented in [206,208]. Consider a generalized plant

$$
P = \left(\begin{array}{c|c} P_{11} & P_{12} \\ \hline P_{21} & P_{22} \end{array} \right) \tag{3.9}
$$

and a stabilizing controller K with McMillan degree n_c, obtained by e.g. full order $\mathcal{H}_\infty$-optimal controller synthesis. Assume the plant and controller admit state space realizations (2.2) and (2.3) respectively. With $x_{\mathrm{cl}} = \left(\begin{array}{cc} x^T & x_K^T \end{array} \right)^T$, the closed-loop system admits the realization

$$
\left(\begin{array}{c} \dot{x}_{\mathrm{cl}} \\ z \end{array} \right) = \left(\begin{array}{c|c} A_{\mathrm{cl}}(K) & B_{\mathrm{cl}}(K) \\ \hline C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) \end{array} \right) \left(\begin{array}{c} x_{\mathrm{cl}} \\ w \end{array} \right),
$$

where

$$
\left(\begin{array}{c|c} A_{\mathrm{cl}}(K) & B_{\mathrm{cl}}(K) \\ \hline C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) \end{array} \right) := \left(\begin{array}{cc|c} A + B_2 D_K C_2 & B_2 C_K & B_1 + B_2 D_K D_{21} \\ B_K C_2 & A_K & B_K D_{21} \\ \hline C_1 + D_{12} D_K C_2 & D_{12} C_K & D_{11} + D_{12} D_K D_{21} \end{array} \right). \tag{3.10}
$$

The controllability Gramian P_{cl} and observability Gramian Q_{cl} are the solutions of the following Lyapunov equations:

$$
\begin{aligned}
A_{\mathrm{cl}}(K)P_{\mathrm{cl}} + P_{\mathrm{cl}}A_{\mathrm{cl}}(K)^T + B_{\mathrm{cl}}(K)B_{\mathrm{cl}}(K)^T &= 0, \\
A_{\mathrm{cl}}^T(K)Q_{\mathrm{cl}} + Q_{\mathrm{cl}}A_{\mathrm{cl}}(K) + C_{\mathrm{cl}}(K)^T C_{\mathrm{cl}}(K) &= 0.
\end{aligned}
$$

Let us partition these Gramians according to the closed-loop states $x_{\mathrm{cl}} = \begin{pmatrix} x^T & x_K^T \end{pmatrix}^T$ as:

$$P_{\mathrm{cl}} = \begin{bmatrix} P_{\mathrm{PP}} & P_{\mathrm{PK}} \\ P_{\mathrm{PK}}^T & P_{\mathrm{KK}} \end{bmatrix} \text{ and } Q_{\mathrm{cl}} = \begin{bmatrix} Q_{\mathrm{PP}} & Q_{\mathrm{PK}} \\ P_{\mathrm{PK}}^T & Q_{\mathrm{KK}} \end{bmatrix}.$$

There exists a balancing similarity transformation $T_{\mathrm{K}} \in \mathbb{R}^{n_{\mathrm{c}} \times n_{\mathrm{c}}}$ of the controller states $z_k = T_{\mathrm{K}}^{-1} x_{\mathrm{K}}$ such that [206]:

$$T_{\mathrm{K}}^{-1} P_{\mathrm{KK}} T_{\mathrm{K}}^{-T} = T_{\mathrm{K}}^T Q_{\mathrm{KK}} T_{\mathrm{K}} = \Sigma_{\mathrm{KK}} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{n_{\mathrm{c}}}).$$

After a similarity transformation of the closed-loop system with:

$$T_{\mathrm{cl}} := \begin{bmatrix} I & 0 \\ 0 & T_{\mathrm{K}} \end{bmatrix},$$

the closed loop state-space quadruple becomes $(T_{\mathrm{cl}}^{-1} A_{\mathrm{cl}} T_{\mathrm{cl}}, T_{\mathrm{cl}}^{-1} B_{\mathrm{cl}}, C_{\mathrm{cl}} T_{\mathrm{cl}}, D_{\mathrm{cl}})$ and the Gramians P_{cl} and Q_{cl} transform into:

$$\hat{P}_{\mathrm{cl}} = \begin{bmatrix} \hat{P}_{\mathrm{PP}} & \hat{P}_{\mathrm{PK}} \\ \hat{P}_{\mathrm{PK}}^T & \Sigma_{KK} \end{bmatrix}, \qquad \hat{Q}_{\mathrm{cl}} = \begin{bmatrix} \hat{Q}_{\mathrm{PP}} & \hat{Q}_{\mathrm{PK}} \\ \hat{Q}_{\mathrm{PK}}^T & \Sigma_{KK} \end{bmatrix},$$

where $\Sigma_{\mathrm{KK}} = \mathrm{diag}(\sigma_1, \ldots, \sigma_{n_{\mathrm{c}}})$, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{n_{\mathrm{c}}} \geq 0$. The element σ_i is in a heuristic sense representative for the contribution of the ith (transformed) controller state to the closed-loop input-output behavior.

The reduced controller of maximal McMillan degree r is obtained using truncation or residualization of the last $n_{\mathrm{c}} - r$ controller states of $T_K^{-1} x_K$. The resulting Closed Loop Truncated controller is given by

$$\left(\begin{array}{c|c} A_{\mathrm{CLT}} & B_{\mathrm{CLT}} \\ \hline C_{\mathrm{CLT}} & D_{\mathrm{CLT}} \end{array} \right) := \left(\begin{array}{c|c} \begin{pmatrix} I_r & 0 \end{pmatrix} T^{-1} & 0 \\ \hline 0 & I \end{array} \right) \left(\begin{array}{c|c} A_K & B_K \\ \hline C_K & D_K \end{array} \right) \left(\begin{array}{c|c} T \begin{pmatrix} I_r & 0 \end{pmatrix}^T & 0 \\ \hline 0 & I \end{array} \right). \quad (3.11)$$

In a similar fashion a Closed Loop Residualized controller can be obtained.

Evaluation

To the best of our knowledge, there are no generally valid a priori bounds known on the closed-loop $\mathcal{H}_\infty$-performance of a closed-loop reduced controller in terms of σ_i, $i = 1, \ldots, n_{\mathrm{c}}$. The Closed Loop Truncated (or Residualized) controller may not be stabilizing, even if the original controller is. Despite this lack of theoretical guarantees for performance, the resulting reduced controller is often very acceptable, i.e. it stabilizes the plant and results in a reasonable closed-loop $\mathcal{H}_\infty$-norm. Closed-loop reduction is, therefore, in our opinion well-suited to generate initial guesses, for instance for the optimization algorithms that will be presented in Section 3.4 and Chapter 7 of this thesis. Since the reduction involves the solution of two Lyapunov Equations, it is very efficient even for systems with large McMillan degree, if compared with the other algorithms for fixed order controller synthesis.

See [163] for some results on the connection between the closed-loop reduction of this section and frequency weighted reduction of the previous section.

3.1.4 Bounds on closed-loop $\mathcal{H}_\infty$-norm after controller reduction

For certain bounds on the suitably weighted controller reduction error, the closed-loop system with a reduced controller is guaranteed to be stable and is bounded in $\mathcal{H}_\infty$-norm. Several conditions of this kind have been presented in the literature [23, 22, 70, 126, 77, 139, 201, 197]. We will describe the method of Goddard and Glover [70] based on coprime factorization, which has also been discussed in [197, 139, 214]. Recall from Theorem 2.16 that the set of all stabilizing controllers satisfying $\|\mathcal{F}_l(P, K)\|_\infty < \gamma$ is given by

$$\mathcal{K} := \{\mathcal{F}_l(J, Q), \ \ Q \in \mathcal{RH}_\infty, \ \|Q\|_\infty < \gamma\}$$

where J is as in Theorem 2.16. Since J_{21} has a stable and proper inverse [214], $\mathcal{F}_l(J, Q)$ can be written as

$$\mathcal{F}_l(J, Q) = (M_{12} + M_{11}Q)(M_{22} + M_{21}Q)^{-1}$$

where

$$\left(\begin{array}{cc} M_{11} & M_{12} \\ M_{21} & M_{22} \end{array} \right) = \left(\begin{array}{cc} J_{12} - J_{11}J_{21}^{-1}J_{22} & J_{11}J_{21}^{-1} \\ -J_{21}^{-1}J_{22} & J_{21}^{-1} \end{array} \right). \tag{3.12}$$

Now $K_c = M_{12}M_{22}^{-1}$ is a coprime factorization of the so-called *central controller*. A reduced order controller K_r is an admissible controller (i.e. K_r stabilizes P and satisfies $\|\mathcal{F}_l(P, K_r)\|_\infty < \gamma$) if it admits a factorization $K_r = UV^{-1}$ with $U, V \in \mathcal{RH}_\infty$ satisfying [70, 214]

$$\left\| \left(\begin{array}{cc} \frac{1}{\gamma}I_{m_2} & 0 \\ 0 & I_{p_2} \end{array} \right) M^{-1} \left(\begin{array}{c} M_{12} - U \\ M_{22} - V \end{array} \right) \right\|_\infty < \frac{1}{\sqrt{2}} \tag{3.13}$$

Computing the reduced controller which minimizes the left-hand side of (3.13) and using it in a controller synthesis scheme is in general not straightforward.

3.2 Sequential LMI optimization

Motivated by the successful application of Interior Point optimization for LMI problems, several researchers have investigated fixed-order $\mathcal{H}_\infty$-optimal controller synthesis by sequential LMI optimization. In our experience and as illustrated in Sections 7.4 and 9.4, these methods work well for systems up to about McMillan degree 30, but for systems with larger orders the solution of the LMI sub-problems is either computationally too expensive or results in inaccurate solutions, depending on which LMI solver is used.

The dual iteration method [96], the min / max algorithm [68], the XY-centering algorithm [100], the path-following method [79, 80], the log-det heuristic [59], the augmented Lagrangian method [4], the alternating projections method [75, 15], the Cone Complementarity algorithm [56], and the VK-iteration [54] are methods using sequential LMI iterations. Since optimization results with the latter two algorithms will be presented in this thesis, we discuss them in more detail in the following sections.

3.2.1 Cone-Complementarity algorithm

In this section we consider the static controller synthesis. As was shown in Section 2.3.1, the fixed order synthesis problem can easily be transformed into the static synthesis problem. For a static controller the number of controller states satisfies $n_c = 0$, which together with (2.28) implies $X - Y^{-1} = 0$. The cone-complementarity algorithm is motivated by the following simple fact.

Lemma 3.2 *If (2.27) is satisfied for $X \in \mathcal{S}^n, Y \in \mathcal{S}^n$ then:*

(i.) $\mathrm{Trace}(XY) \geq n$

(ii.) $X - Y^{-1} = 0 \Leftrightarrow \mathrm{Trace}(XY) = n$

Proof. (2.27) implies $Y \succ 0$ and $X - Y^{-1} \succeq 0$. Hence for any R with $Y = R^T R$ we conclude $RXR^T \succeq I_n$. Hence

$$\mathrm{Trace}(XY) = \mathrm{Trace}(XR^T R) = \mathrm{Trace}(RXR^T) \succeq \mathrm{Trace}(I_n) = n$$

which shows (i.). To prove (ii.), observe that $X - Y^{-1} = 0$ implies that $XY = I_n$, which immediately shows $\mathrm{Trace}(XY) = n$. On the other hand, if $\mathrm{Trace}(XY) = n$ and $Y = R^T R$, we infer $\mathrm{Trace}(RXR^T) = \mathrm{Trace}(XY) = n$. Combining this with $RXR^T \succeq I_n$ (from (2.27)) implies $RXR^T = I_n$ such that indeed $X - Y^{-1} = 0$.

$\blacksquare$

The lemma implies the following result. If a static controller with closed-loop $\mathcal{H}_\infty$-optimal performance of γ^* exists, then the optimal value of

$$\begin{array}{ll} \text{infimize} & \mathrm{Trace}(XY) \\ \text{subject to} & (2.25), (2.26) \text{ and } (2.27) \text{ for fixed } \gamma = \gamma^* \end{array} \qquad (3.14)$$

is n. Furthermore if this optimal value is attained at (X_*, Y_*), then there exist an $X_{\mathrm{cl}} \in \mathcal{S}^n$ and a corresponding static controller K that render (2.19) satisfied. Because the constraints in (3.14) are all LMIs, the feasible region of this optimization problem is convex. The objective functional $\mathrm{Trace}(XY)$ is, however, non-convex. The cone complementarity method uses sequential linearization of the objective functional $\mathrm{Trace}(XY)$ at the current iterate (X_k, Y_k):

$$\mathrm{Trace}(XY) = \mathrm{Trace}(X_k, Y_k) + \mathrm{Trace}\left[(X - X_k)Y_k + X_k(Y - Y_k)\right] + \text{ higher order terms.}$$

A conceptual description of the algorithm is as follows:

Algorithm 3.3 *(Cone complementarity algorithm for static controllers)*

1. *Initialization: fix γ at a desired value larger than the $\mathcal{H}_\infty$-optimal full-order performance. Set $k = 0$ and find (X_0, Y_0) satisfying (2.25), (2.26) and (2.27),*

2. *find any suboptimal solution (X_{k+1}, Y_{k+1}) to the minimization problem*

$$\begin{array}{ll} \text{infimize} & \mathrm{Trace}(XY_k + X_kY) \\ \text{subject to} & (2.25), (2.26) \ (2.27) \end{array}$$

3. *set $t_k = \frac{1}{2}\mathrm{Trace}\left(X_{k+1}Y_k + X_kY_{k+1}\right)$,*

4. *if a stopping criterion is satisfied exit, otherwise set $k = k + 1$ and return to step 2.*

In Steps 1 and 2, an LMI feasibility and minimization problem have to be solved respectively. The other steps are straightforward. The sequence t_k is non-increasing and bounded below by n. It converges therefore to some value $t_{\mathrm{opt}} \geq n$ for $k \to \infty$. If $t_{\mathrm{opt}} = n$ and X_k and Y_k converge to some X_{opt} and Y_{opt} for $k \to \infty$, then $X_{\mathrm{opt}}Y_{\mathrm{opt}} = I_n$. For the static output feedback stabilization problem it is shown in [56] that at every step k

$$\mathrm{Rank}\begin{pmatrix} X_k & I \\ I & Y_k \end{pmatrix} \leq 2n - \max(p_2, m_2).$$

By a bisection over γ the closed-loop $\mathcal{H}_\infty$-performance can be minimized.

Evaluation

Convergence of t_k in Algorithm 3.3 to the global optimal value of (3.14) is not guaranteed. Hence t_k might not converge to n, even if a static controller with performance γ exists.

If, for a certain value of γ, the sequence t_k does not converge to n, it is not possible, in general, to construct a stabilizing controller on the basis of the final iterates $X_{k_{\max}}$ and $Y_{k_{\max}}$, where $k_{\max}$ is the maximum number of iterations. This is in contrast with the next method, which minimizes γ directly and enables to construct a stabilizing controller of order n_c at each iteration.

The algorithm is also quite sensitive with respect to round-off errors in the LMI optimization. Due to these errors, t_k does never converge exactly to n in practice. Even if the final optimal value is very close to n, numerical errors in the solution (X_k, Y_k) make it often impossible to construct an X_{cl} that is (together with an appropriate K) feasible for (2.19). Hence, for a successful controller construction it is often required to use an accurate LMI solver such as LMILAB [65]. The LMI sub-problems in Algorithm 3.3 are, however, too large to be solved by LMILAB in reasonable time if the McMillan degree of the plant exceeds about 40. For a generalized plant with McMillan degree 27 the method is still applicable, as will be illustrated in Section 7.4.

3.2.2 VK-iteration approach

VK-iteration [54] is similar to the well-known controller-scaling iteration (DK-iteration) for robust control synthesis. In the VK-iterations γ is minimized subject to $X_{\mathrm{cl}} \succ 0$ and the BMI inequality (2.19), by iteratively solving LMI subproblems in X_{cl} (which is also denoted by V) and K. VK iteration is, therefore, a *coordinate search* over the 'coordinates' X_{cl} and K. The algorithm consists of the following steps:

Algorithm 3.4 *(VK-iteration):*

1. *find a stabilizing initial controller K_0, set $k = 0$,*

2. *search for a $(X_{\mathrm{cl}})_k \succ 0$ that minimizes γ and satisfies (2.19) for fixed $K = K_k$,*

3. *search for controller K_{k+1} that minimizes γ and satisfies (2.19) for fixed $X_{\mathrm{cl}} = (X_{\mathrm{cl}})_k$,*

4. *if $k = k_{\max}$ then stop, otherwise set $k = k + 1$ return to step 2.*

$k_{\max}$ is a tuning parameter of the algorithm.

Evaluation

To the best of our knowledge there is no convergence result known for this algorithm. The algorithm has been successfully applied to systems up to McMillan degree four [54, 8]. However, in our experiments with plants of larger McMillan degree the algorithm often got stuck at controllers that are far from optimal and that can easily be improved by other algorithms. This is illustrated in Chapter 9 by means of a controller design problem for a wafer stage.

The accuracy of the LMI-solver is important, since the computed controller may not be stabilizing if the errors in the LMI variables are large.

An advantage of VK-iteration if compared to the cone complementarity algorithm is the possibility to explicitly add structural constraints other than the McMillan degree of the controller. Examples of such constraints are diagonal controllers and strong stabilization.

3.3 Branch and bound method

The *Branch and Bound* (BB) method is a global optimization algorithm. It has been applied by Tuan and Apkarian [193] to a general class of BMIs given by:

$$\text{infimize} \quad c^T x + d^T y \tag{3.15}$$
$$\text{subject to} \quad F(x,y) \preceq 0 \tag{3.16}$$
$$x_{\min} \leq x \leq x_{\max} \tag{3.17}$$

where

$$F(x,y) := F_0 + F_1 x_1 + \ldots + F_m x_m + G_1 y_1 + \ldots + G_n y_n + \sum_{i=1}^{m}\sum_{j=1}^{n} H_{ij} x_i y_j$$

and $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ are unknown variables. The vectors $c \in \mathbb{R}^m$, $d \in \mathbb{R}^m$ and symmetric matrices F_i, G_j and H_{ij}, $i = 1, \ldots, m$, $i = 1, \ldots, n$ and $x_{\min}, x_{\max}, \in \mathbb{R}^m$ are given. Assume without loss of generality that $m \leq n$. This BMI problem formulation includes almost all practical static-output feedback controller synthesis problems, if the controller variables are a priori restricted to lie within a box $\mathcal{K}_{\text{box}}$. Indeed with $x = \text{vec}(K)$, $K \in \mathcal{K}_{\text{box}}$ is equivalent to $x_{\min} \leq x \leq x_{\max}$ for some $x_{\min}, x_{\max}, \in \mathbb{R}^m$, where the inequality is element-wise, i.e. $a \leq b$ for $a, b \in \mathbb{R}^m$ is defined as

$$a_i \leq b_i \text{ for all } i \in 1, \ldots, m.$$

Furthermore let us choose the vector y as $y := \begin{pmatrix} \gamma & \text{svec}(X)^T \end{pmatrix}^T$, where svec is the vectorization of a symmetric matrix, i.e. if $A \in \mathcal{S}^r$ is defined by

$$A := \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,r} \\ A_{1,2} & A_{2,2} & & A_{2,r} \\ \vdots & & \ddots & \vdots \\ A_{1,n} & A_{2,r} & \cdots & A_{r,r} \end{pmatrix}$$

then

$$\text{svec}(A) := [A_{1,1}, A_{1,2}, \ldots, A_{1,r}, A_{2,2}, A_{2,3}, \ldots, A_{2,r}, \ldots, A_{r,r}]^T. \tag{3.18}$$

With these choices, $X_{\text{cl}} \succ 0$ and (2.19) are constraints that can be combined by diagonal augmentation into a constraint of the form (3.16).

We will give a conceptual description of the algorithm. The reader is referred to [193] for more details. The method consists of the following steps:

1. branching,

2. bounding,

3. discarding.

These steps are illustrated in Figure 3.1, where $x \in \mathbb{R}^2$ are the branching variables. The branching consists of sub-partitioning the box $\{x \in \mathbb{R}^m | x_{\min} \leq x \leq x_{\max}\}$ into small boxes. On these boxes lower and upper bounds on the achievable performance within these subsets are computed by solving two LMI problems. Those boxes that lead to a larger lower bound than the best upper bound can be discarded. By further refinement of the non-discarded boxes the best upper and worst lower bound get closer and closer to each other.

It is shown in [193] that for fixed $\epsilon > 0$ the difference between the best upper and worst lower bound is smaller than ϵ in a finite number of iterations. Any controller in a non-discarded set is at this stage a global ϵ-suboptimal controller. Although global convergence is a nice property of the method, it comes at a large price: the computation is so expensive that optimization of controllers for real-life systems seems out of reach. The complexity of the algorithm is strongly increasing with the number of branching variables, i.e. the controller variables. For each bounding step LMI problems have to be solved. As mentioned in Section 3.2, the state-of-the-art LMI software can not solve these problems with high accuracy for systems with large McMillan degree.

3.4 Nonsmooth optimization

Let the closed-loop matrices $(A_{\mathrm{cl}}(K), B_{\mathrm{cl}}(K), C_{\mathrm{cl}}(K), D_{\mathrm{cl}}(K))$ be as in (2.8). Observe that these matrices depend affinely on the matrix K. The $\mathcal{H}_\infty$-optimal structured controller problem is

$$\begin{aligned}
&\text{infimize} && \gamma \\
&\text{subject to} && \begin{pmatrix} A_{\mathrm{cl}}(K)^T X + X A_{\mathrm{cl}}(K) & X B_{\mathrm{cl}}(K) & C_{\mathrm{cl}}(K)^T \\ B_{\mathrm{cl}}(K)^T X_{\mathrm{cl}} & -\gamma I & D_{\mathrm{cl}}(K)^T \\ C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) & -\gamma I \end{pmatrix} \prec 0, \ X \succ 0.
\end{aligned} \tag{3.19}$$

A disadvantage of optimizing over γ, X and K is the large number of decision variables in X. The number of decision variables in X for a 60th order plant with a 15th order controller is $\frac{1}{2}(n + n_{\mathrm{c}})(n + n_{\mathrm{c}} + 1) = 2850$.

$\mathcal{H}_\infty$-optimal controller design with non-smooth objective and constraints

One way to avoid the introduction of the Lyapunov variable X is to directly use the $\mathcal{H}_\infty$-norm as optimization objective and to consider

$$\begin{aligned}
&\text{infimize} && \max_{\omega \in \mathbb{R}}(\bar{\sigma}\,(\mathcal{F}_l(P(j\omega), K(j\omega)))) \\
&\text{subject to} && \max(\Re(\lambda(A_{\mathrm{cl}}(K)))) < 0
\end{aligned} \tag{3.20}$$

where $\bar{\sigma}(\cdot)$ denotes the maximum singular value and $\max(\Re(\lambda(A)))$ denotes the maximum of the real parts of the eigenvalues of the matrix A. A disadvantage of the formulation is that both the objective and the constraints contain non-smooth functions, as is illustrated with the following example.

Example 3.5 *Let us consider the static output feedback stabilisation problem with $K \in \mathbb{R}$, $A_{\mathrm{cl}}(K) := A + B_2 K C_2$ and*

$$\left(\begin{array}{c|c} A & B_2 \\ \hline C_2 & 0 \end{array} \right) = \left(\begin{array}{ccc|c} -4.8 & -1.6875 & -0.2875 & 1 \\ 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ \hline 1 & 0.25 & 0.15625 & 0 \end{array} \right).$$

Figure 3.2 displays the root-locus plot, i.e. the eigenvalues $\lambda(A_{\mathrm{cl}}(K))$ in the complex plane for $K \in [0, \infty)$. For $K_* = 0.5321$ the eigenvalues are

$$\lambda(A_{\mathrm{cl}}(0.5321)) = \left(\begin{array}{ccc} -3.4881 & -0.92199 & -0.92199 \end{array} \right).$$

At K_* the left-derivative of $g(K) := \max(\Re(\lambda(A_{cl}(K))))$ does not exist, such that $g(K)$ is non-differentiable at K_*. In Figure 3.3 the eigenvalues are plotted for $K \in [0,1]$, which illustrates that $g(K)$ indeed is non-differentiable at $K_* = 0.5321$.

Non-smooth dependence of $\max_{\omega \in \mathbb{R}}(\bar{\sigma}(\mathcal{F}_l(P(j\omega), K(j\omega))))$ on K often occurs if the maximum is attained at two or more different frequencies. In our experience this is often encountered during controller optimization. It is therefore important to take this into consideration.

Implications of non-smoothness for optimization

The non-smoothness implies that algorithms are not applicable that compute their search direction by differentiation of the objective function and constraints. Of course one could just ignore the non-differentiability and compute approximate derivatives by numerical perturbation, but in our experience these kind of algorithms often get stuck at non-smooth points that are not locally optimal.

This motivates to use dedicated optimization methods for nonsmooth optimization. These kind of algorithms have been applied to the fixed-order controller synthesis problem by for instance Rotunno and De Callafon [158,159], by Burke, Lewis, Overton and Henrion [32–34] using gradient sampling techniques and by Apkarian and Noll [3] using the *Multi-Directional Search* method. The latter is originally due to Torczon [192] and will be discussed in the sequel.

Multi-Directional Search method

The Multi-Directional Search (MDS) method [192] is very similar to the Simplex method [53] and is an algorithm for unconstrained optimization, e.g. for solving $\inf_{x \in \mathbb{R}^m} f_p(x)$, where $f_p : \mathbb{R}^m \to \mathbb{R}$. The MDS algorithm has tuning parameters $\alpha \in (1, \infty)$ and $\beta \in (0, 1)$ and can be summarized as follows

Algorithm 3.6 *(MDS algorithm)* Set $x_0^0 = x_{init}$, $x_j^0 = x_0^0 + e_j$ where e_j, $j = 1, \ldots, m$ are basis vectors in $\mathbb{R}^m$. Choose a large enough $N \in \mathbb{N}$ and set $k = 0$. Perform the following steps:

1. *Compute $p_{x,opt}^k := \min_{j \in \{0,1,\ldots,m\}} f_p(x_j^k)$ and let j_{opt}^k be any index such that $p_{x,opt}^k = f_p(x_{j_{opt}}^k)$.*

2. *compute $r_j = x_{j_{opt}}^k - (x_j^k - x_{j_{opt}}^k)$ and $p_{r,opt}^k := \min_{j \in \{0,1,\ldots,m\}} f_p(r_j^k)$. If $p_{r,opt}^k < p_{x,opt}^k$ then go to Step 3, otherwise go to Step 4.*

3. *compute $q_j^k := x_{j_{opt}}^k - \alpha(x_j^k - x_{j_{opt}}^k)$, $j = 0, 1, \ldots, m$ and $p_{q,opt}^k := \min_{j \in \{0,1,\ldots,m\}} f_p(q_j^k)$. If $p_{q,opt}^k < p_{r,opt}^k$ set $x_j^{k+1} := q_j^k$, $j = 0, 1, \ldots, m$, otherwise set $x_j^{k+1} := r_j^k$, $j = 0, 1, \ldots, m$. Goto step 5*

4. *compute $c_j^k := x_{j_{opt}}^k - \beta(x_j^k - x_{j_{opt}}^k)$ and set $x_j^{k+1} := c_j^k$, $j = 0, 1, \ldots, m$.*

5. *set $k = k + 1$. If $k > N$ stop, otherwise return to Step 1.*

Steps 2, 3 and 4 can be geometrically interpreted as reflection, expansion and contraction steps as illustrated in Figure 3.4.

The problem in (3.20) is a constrained optimization problem. We present in the sequel a penalty technique [136, 125] to replace the constrained problem by an unconstrained one, which can be solved with the MDS algorithm. To simplify notation we define $x := \text{vec}(K)$, $f(x) : \max_{\omega \in \mathbb{R}}(\bar{\sigma}(\mathcal{F}_l(P(j\omega), K(j\omega))))$ and $g(x) := \max(\Re(\lambda(A)))$, such that (3.20) can be written as $\inf_{g(x) \leq 0} f(x)$. Let $x_{\text{init}} = \text{vec}(K_{\text{init}})$ be an initial, not necessarily stabilizing, controller. Then we solve the unconstrained problem $\inf_{x \in \mathbb{R}^m} f_p(x)$ for

$$f_p(x) := f(x) + \rho \max(0, g(x)), \tag{3.21}$$

where $\rho > 0$ is (fixed) tuning parameter of the algorithm. For guidelines for the choice of ρ the reader is for instance referred to [125].

3.5 Other fixed-order methods

In this section we briefly address some important alternative techniques for structured $\mathcal{H}_\infty$-synthesis. The reader is referred to the references for more details. The first two methods are classical and use a sequence of Lyapunov equations to solve the static $\mathcal{H}_2$-optimal control problem. The other algorithms have been developed more recently. They optimize the closed-loop $\mathcal{H}_\infty$-norm.

Anderson-Moore algorithm

This method solves the $\mathcal{H}_2$-optimal static output feedback controller synthesis problem by iteratively solving Lyapunov equations. Using the solutions to these Lyapunov equations, a descent direction is computed for the controller variables, based on the implicit function theorem. The method has been described in the book "Linear Optimal Control" of Anderson and Moore [2]. See [150] for a further analysis of the search steps that are computed in this method. Pensar and Toivonen [145] have made an extension of the algorithm to fixed-order $\mathcal{H}_\infty$-synthesis.

Homotopy method on the Projection Equations

This method [95, 204, 43, 44, 202] has been developed for the static $\mathcal{H}_2$-optimal controller synthesis problem. The so-called projection equations are first-order necessary conditions for local optimality of this problem. These are solved by a homotopy [154] scheme that consists of a sequence of Lyapunov-like equations. Conditions for $\mathcal{H}_\infty$-optimal control have been derived in [16].

Frequency gridding

In [73] a PID controller is tuned in a model matching problem. In this approach a family of transfer functions $\{K_p(s)| \ p \in \mathbb{R}^M\}$ of PID controllers is parameterized by $p \in \mathbb{R}^M$ (for some $M \in \mathbb{N}$). Within this family a $p \in \mathbb{R}^M$ is searched, such that the loop gain $G(s)K_{p_{\text{opt}}}(s)$ matches the desired loop shape transfer function $L(s)$ at certain frequencies. More precisely, for a finite frequency grid $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ for some $N \in \mathbb{N}$ the largest maximum singular value of the frequency response of $V(GK(p) - L)$ is minimized:

$$\inf_{p \in \mathbb{R}^M} \sup_{\omega \in \Omega} \bar{\sigma}\left(V(j\omega)(G(j\omega)K_p(j\omega) - L(j\omega))\right), \tag{3.22}$$

where V is a weighting function. If $K_p(j\omega)$ is affine in p for fixed $\omega \in \mathbb{R}$, then (3.22) can be converted into an LMI problem. As addressed in Section 2.4.1, this is often the case. Note that an optimal K_p is not necessarily stabilizing.

Interpolation approach

To design fixed-order controllers, interpolation techniques for $\mathcal{H}_\infty$-controller design have been modified to construct interpolants with a bound on the McMillan degree [135]. These methods cannot straightforwardly be extended to MIMO control for the generalized plants as described in Section 2.1.1.

Evolutionary, genetic and randomized algorithms

These algorithms [111, 35, 115] optimize a controller using a random number generator. These methods seem not well suited to optimize controllers with a moderate or large number of variables.

Interior point methods

Interior point methods [93, 121] will be extensively discussed in Chapter 7.

3.6 Contribution of this thesis

It is surprising that, despite the large number of research publications on the topic, the fixed order $\mathcal{H}_\infty$-optimal control problem is (in our opinion) still not fully solved. We will precisely formulate these unsolved issues and point to the (partial) answers that are presented in this thesis.

Firstly, most of the algorithms that have been published in the literature are local optimization methods. Convergence of these algorithms to the global optimal controller is not guaranteed. It is therefore not possible to judge the quality of the resulting optimized controller, in the sense that it is unknown if there exists a controller with the same structure that yields a much smaller closed-loop $\mathcal{H}_\infty$-norm. The knowledge of a global sub-optimality certificate may prevent a futile and time-consuming re-optimization with different guesses, since they imply that no better controller with the specified structure exists. We will present in Chapters 4 and 5 a methodology for the computation of such global optimality certificates. It consists of a sequence of relaxations problems with Linear Matrix Inequalities (LMIs), whose optimal values are lower bounds of the achievable closed-loop $\mathcal{H}_\infty$-norm and converge in the limit to this value. The proposed scheme is based on the decomposition of polynomial matrices as a sum of squares. Our efforts to reduce the complexity of the relaxations resulted in a scheme that is not only applicable to examples with small McMillan degree, but also to (moderately sized) control problems encountered in engineering practice.

This latter scheme is based on a reformulation of the fixed-order $\mathcal{H}_\infty$-controller synthesis problem as a robust analysis problem. This is of independent interest, since it enables to apply the wide range of the robust analysis techniques to obtain optimal $\mathcal{H}_\infty$-controller values. If these relaxations are exact, it is under certain conditions possible to construct a globally optimal controller.

In Chapter 6 the Sum-Of-Squares decompositions are used to generate a family of relaxations for robust analysis problems. These relaxations are guaranteed to be asymptotically

exact. This result allows us to systematically reduce the conservatism for a very wide range of robust analysis problems.

Secondly, exploitation of the structure of the $\mathcal{H}_\infty$-optimal control problem is in our opinion crucial to develop a synthesis method that:

- allows the incorporation of various additional constraints such as controller degree bounds and strong stabilization and

- is applicable to systems with high McMillan degree.

To the best of our knowledge such an algorithm does not exist in the literature. In Chapter 7 we present an Interior Point method for the fixed-order $\mathcal{H}_\infty$-optimal control synthesis problem in terms of *Bilinear Matrix Inequalities* (BMIs). Exploitation of the control-theoretic characteristics of the problem yields a novel Interior Point optimization algorithm based on Sylvester equations, that has connections with both the 'classical' methods based on Lyapunov equations (which are a subclass of Sylvester equations) and the 'modern' techniques based on Interior Point optimization. The approach is applied to design a controller for an active suspension system. Experimental results as well as a comparison with two other state-of-the-art fixed-order synthesis techniques are presented.

The interior-point algorithm converges to a solution satisfying the first-order necessary optimality conditions, which has motivated us to derive (local) second-order optimality conditions for the BMI problem. These conditions are based on elementary arguments and are usually computationally cheaper than the global optimality certificates presented in Chapters 4 and 5.

Thirdly, the effects of the inherent over-parametrization of state-space controllers can hamper convergence of fixed-order controller optimization algorithms. To the best of our knowledge this topic has not been addressed within the literature on structured controller synthesis. In Chapter 8 we analyze the effects of over-parametrization and present a remedy in terms of a reduced controller parametrization.

Finally, we present in Chapter 9 an application of structured controller synthesis to the controller design for a wafer stage prototype. We describe the design of a SISO and 3×3 MIMO controllers and present the result of experiments. The results illustrate that the algorithms presented in this thesis are suited to design high-performing controllers for mechanical servo-systems used in industry.
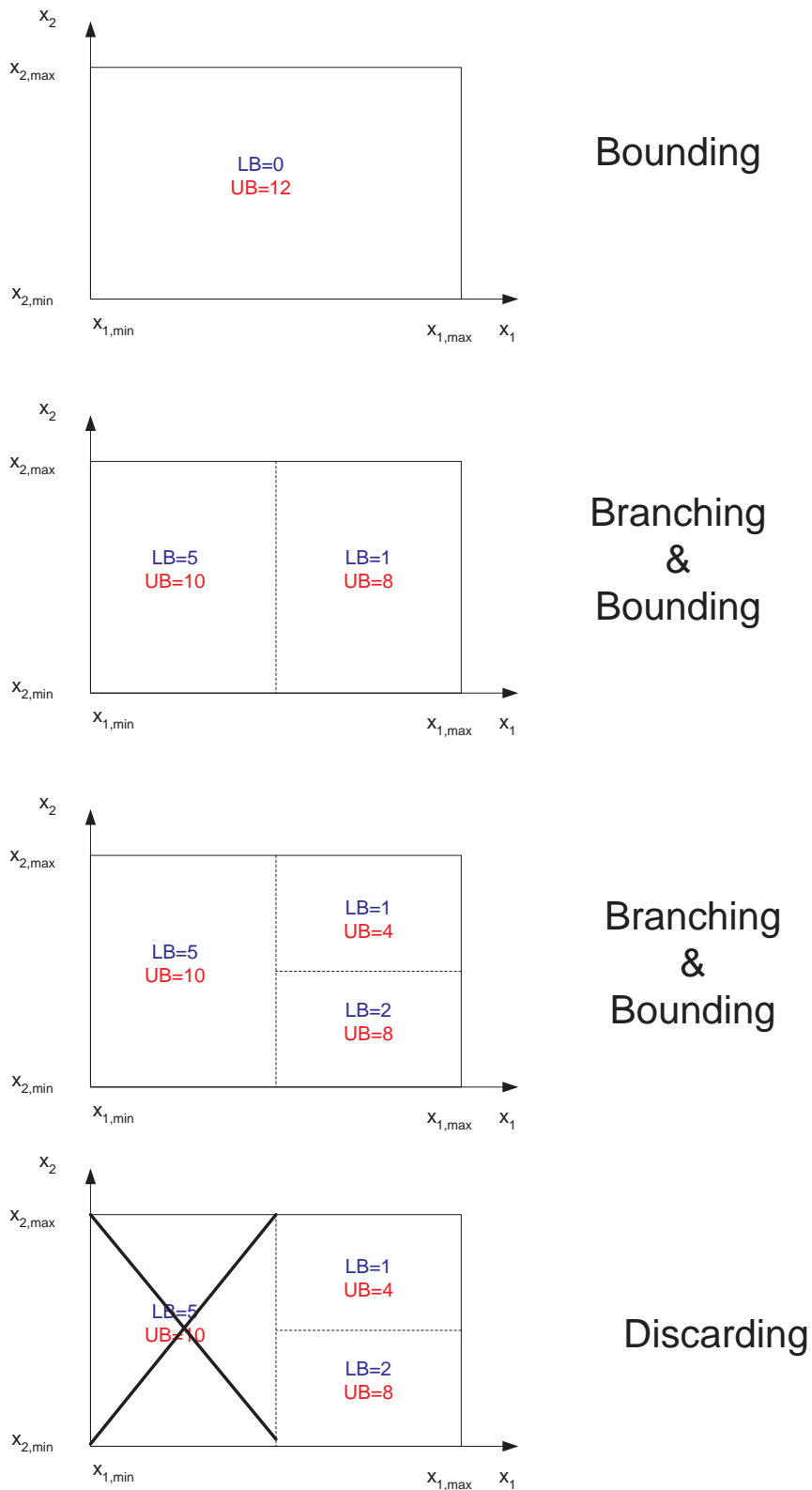
Figure 3.1: The steps in the branch and bound algorithm. The boxes are subdivided in smaller boxes (branching), upper **UB** and lower bounds **LB** are computed (bounding) and boxes with a lower bound **LB** larger than the globally smallest upper bound are discarded (discarding).
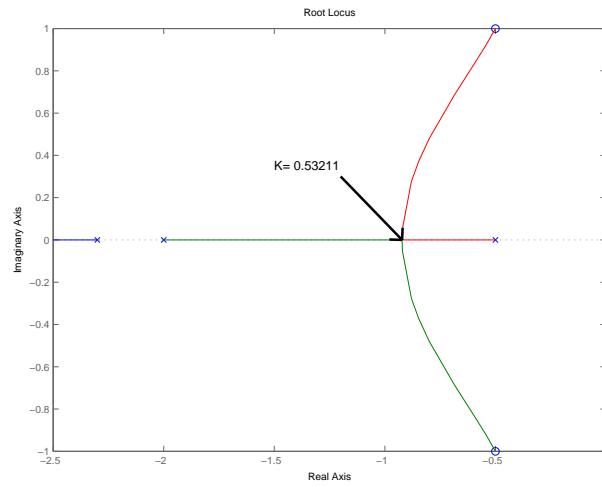
Figure 3.2: Root-locus plot of Example 3.5
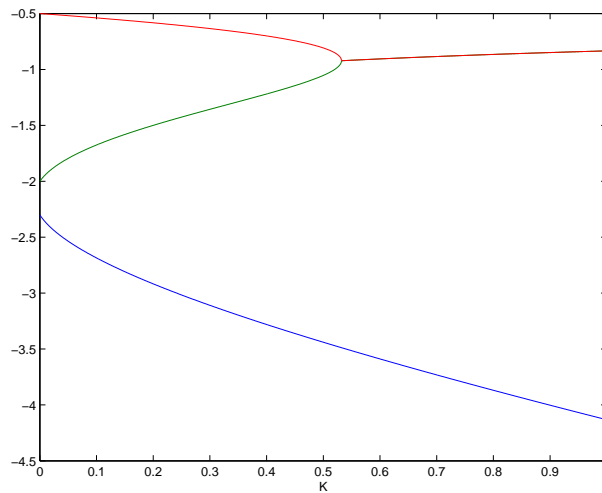
Figure 3.3: Real parts of the eigenvalues of $A_{\mathrm{cl}}(K)$ for $K \in [0, 1]$. The three eigenvalues are depicted. For $K > 0.5321$ two eigenvalues coincide.
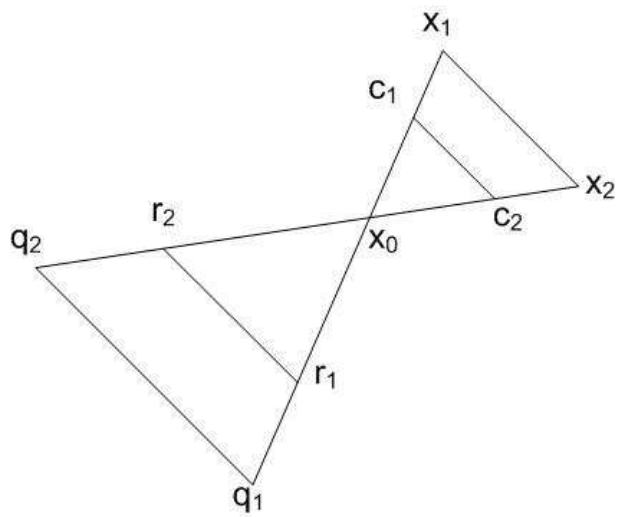
Figure 3.4: The reflection, expansion and contraction steps in the simplex algorithm.

Chapter 4

Polynomial optimization by Sum-Of-Squares

4.1 Introduction

Many analysis and synthesis problems in control can be reduced to scalar polynomially constrained polynomial programs. This class of problems is still very general and contains non-convex instances that are hard to solve. Only (rather) recently it has been suggested how to construct LMI relaxations of such optimization problems based on the Sum-Of-Squares (SOS) decomposition of multivariable polynomials [144,37,119,39,38,85,83]. These relaxation schemes have been applied to various non-convex problems in control such as Lyapunov stability of nonlinear dynamic systems [144, 141, 38]. Scalar polynomial techniques have also been applied to the SISO fixed-order controller synthesis problem by Dorato [47] for stabilization and by Henrion *et. al.* [86, 82] for $\mathcal{H}_\infty$ performance. These techniques cannot be extended to MIMO (Multiple Input Multiple Output) controller synthesis in a straightforward fashion. Furthermore, the latter method critically depends on the choice of a certain central polynomial.

In particular in control engineering many problems actually involve, apart from scalar polynomial constraints, semi-definite constraints on symmetric-valued polynomial matrices. Since multiple polynomial SDP-constraints can be easily collected into one single polynomial SDP-constraint, these problems can be described as

$$
\begin{array}{ll}
\text{infimize} & f(x) \\
\text{subject to} & G(x) \preceq 0
\end{array}
\tag{4.1}
$$

where $x \in \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^m \to \mathcal{S}^r$ are general scalar, symmetric-matrix-valued polynomials (to be precisely defined in the sequel). In this and the following chapters we make the assumption:

Assumption 4.1 *The set*

$$
\{x \in \mathbb{R}^m | \ G(x) \preceq 0\}
$$

is nonempty.

Let us denote the optimal value of (4.1) by p_{opt}. Examples of these kind of problems are the spectral factorization of multidimensional transfer functions to assess dissipativity of linear shift-invariant distributed systems [146], multi-objective control and input-output selection, where the integer constraints of type $p \in \{0, 1\}$ are replaced by quadratic constraints of the form $p(p - 1) = 0$.

The main result of this chapter is a sequence of LMI relaxations of the possibly non-convex problem (4.1) on the basis of matrix SOS decompositions, whose optimal values are converging from below to the optimal value p_{opt}. This result is an extension of SOS relaxations of problems with scalar polynomial constraints [142,37,119] to polynomial semi-definite constraints. These latter relaxations can also directly be applied to polynomial SDPs, after scalarization of the semi-definite constraints. However, a major advantage of using matrix-valued SOS decompositions is that the size of the constructed LMI relaxations grows at most quadratically in the dimension r (i.e. the number of columns / rows) of $G(x)$.

As described in Section 2.4, the structured $\mathcal{H}_\infty$-optimal controller synthesis problem can also be molded into (4.1). The LMI relaxations presented in this chapter can be applied to compute lower bounds on the optimal fixed order $\mathcal{H}_\infty$ performance. The computation of such bounds enables to equip local optimization algorithms, such as the one presented in Chapter 7, with a stopping criterion, because the bounds provide information on the difference of the performance of the computed controller and the optimal fixed order $\mathcal{H}_\infty$ performance. This is important since algorithms based on local optimization can in general not be guaranteed to converge to a globally optimal solution. A trivial lower bound on the fixed-order $\mathcal{H}_\infty$-optimal controller performance is, of course, the optimal full order performance. Boyd and Vandenberghe [28] proposed lower bounds based on convex relaxations of the fixed-order synthesis problem, though it is unknown how to systematically improve these relaxations. For the sequence of LMI relaxations presented in this chapter the relaxation gap is systematically reduced to zero. To the best of our knowledge such sequences did not exist in the literature before the publication of our results in [89,88,90], except for Branch and Bound schemes. The application of the BB algorithm to practical control problems seems currently out of reach, as mentioned in Section 3.3.

The outline of this chapter is as follows. The Sum-Of-Squares relaxations for polynomial SDPs are introduced in Section 4.2. They can be interpreted as a polynomial extension to standard Lagrange dualization. In Section 4.3 we introduce the concept of a sum of squares of matrix-valued polynomials. Section 4.4 covers the main contribution of this chapter, namely the extension of the exactness result of SOS relaxations for scalar polynomial problems to polynomial SDPs. This allows us to construct a sequence of LMI relaxations whose optimal values converge to the optimal value of the polynomial SDP, as explained in Section 4.5. In Section 4.6 we exploit these results to compute lower bounds on the globally optimal fixed-order $\mathcal{H}_\infty$-optimal control performance. In Section 4.7 we apply the method to a set of (randomly generated) fourth-order systems. We conclude with a brief evaluation of the results in Section 4.8.

4.2 Polynomial semi-definite programming

The fixed order synthesis problem and many important non-convex optimization problems can be written as polynomial semi-definite programs. A scalar multivariate polynomial $p(x)$ for $x \in \mathbb{R}^m$ can be written as

$$p(x) = \sum_{i=1}^{N} c_i x_1^{\alpha_{i1}} x_2^{\alpha_{i2}} \ldots x_m^{\alpha_{im}} \tag{4.2}$$

where $c_i \in \mathbb{R}$, $i = 1, \ldots, N$ are the 'coefficients' and $\alpha_{ij} \in \mathbb{N}_0 := (0 \cup \mathbb{N})$, $i = 1, \ldots, N$, $j = 1, \ldots, m$, are the 'exponents'. The set of all polynomials in the variable x is denoted

by $\mathbb{R}[x]$. Polynomials are actually finite linear combinations of 'monomials', which are products of powers of x_i, $i = 1, \ldots, m$ given by

$$h(x) = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_m^{\alpha_m}$$

for some $\alpha_j \in \mathbb{N}_0$, $j = 1, \ldots, m$. For $\alpha \in \mathbb{R}^m$ we will sometimes use the notation x^α to denote $x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_m^{\alpha_m}$. The total degree of a monomial $h(x) = x^\alpha$ is the sum of the exponents $\mathrm{tdeg}(h) := \sum_{j=1}^m \alpha_j$ and the total degree of a polynomial is defined as the maximum of the total degrees of its monomials. We write $f(x) = g(x)$ to denote that the two polynomials $f(x)$ and $g(x)$ are the same, i.e. if they both admit a representation as in (4.2) with the same exponent matrix $\alpha \in \mathbb{N}^{N \times m}$ and coefficient vector $c \in \mathbb{R}^N$.

A symmetric $r \times r$ matrix-valued polynomial can also be written in the same fashion, with the only difference that the coefficients $c_i \in \mathcal{S}^r$ are symmetric matrices. LMI problems (also denoted by *Linear SemiDefinite Programs*) are a subclass of polynomial SDPs. Any LMI problem can be written as (4.1) for affine f and G. As mentioned in Section 2.5.2, LMI problems can be solved efficiently. Such efficiency results are definitely out of reach for the general problem formulation (4.1).

As explained in Chapter 2, the $\mathcal{H}_\infty$-optimal control problem can be formulated as

$$\begin{array}{ll} \text{infimize} & \gamma \\ \text{subject to} & X_{\mathrm{cl}} \succ 0, \ \mathcal{B}(\gamma, X_{\mathrm{cl}}, K) \succ 0, \end{array} \tag{4.3}$$

where $\mathcal{B}(\gamma, X_{\mathrm{cl}}, K)$ represents the 'constraint function corresponding to the Bounded Real Lemma:

$$\mathcal{B}(\gamma, X, K) := - \begin{pmatrix} A_{\mathrm{cl}}(K)^T X + X A_{\mathrm{cl}}(K) & X B_{\mathrm{cl}}(K) & C_{\mathrm{cl}}(K)^T \\ B_{\mathrm{cl}}(K)^T X_{\mathrm{cl}} & -\gamma I & D_{\mathrm{cl}}(K)^T \\ C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) & -\gamma I \end{pmatrix}, \tag{4.4}$$

where $(A_{\mathrm{cl}}(K), B_{\mathrm{cl}}(K), C_{\mathrm{cl}}(K), D_{\mathrm{cl}}(K))$ are defined as in (2.8). This is a polynomial semi-definite program as in (4.1) over the variable $x = (\gamma, X, K)$, with an affine objective function $f(x) = \gamma$ and a polynomial (more specifically bilinear) semi-definite constraint $G(x) \prec 0$ where

$$G(x) := \begin{pmatrix} -\mathcal{B}(\gamma, X_{\mathrm{cl}}, K) & 0 \\ 0 & -X_{\mathrm{cl}} \end{pmatrix}. \tag{4.5}$$

Although the inequalities in (4.3) are strict, we work in the remainder of this chapter with (4.1), which has non-strict inequalities.

As shown in Section 2.5.4 weak Lagrange duality implies that for any matrix $S \succeq 0$, the value $\inf_{x \in \mathbb{R}^{n_x}} f(x) + \langle S, G(x) \rangle$ is a lower bound on the optimal value p_{opt} of (4.1), where $\langle A, B \rangle := \mathrm{Trace}(AB)$ is the standard inner product for symmetric matrices with the same number of rows/columns. However, not even the maximization of this lower bound over $S \succeq 0$ closes the duality gap, due to non-convexity of the problem. This is the reason for considering, instead, Lagrange multiplier matrices $S(x) \succeq 0$ which are polynomial functions of x. Still $\inf_{x \in \mathbb{R}^{n_x}} f(x) + \langle S(x), G(x) \rangle$ defines a lower bound of p_{opt}, and the best lower bound that is achievable in this fashion is given by the supremal d for which there exists a polynomial matrix $S(x) \succeq 0$ such that

$$f(x) + \langle S(x), G(x) \rangle - d > 0 \quad \text{for all} \quad x \in \mathbb{R}^m. \tag{4.6}$$

Remark. The extension of a constant Lagrange multiplier to a polynomial multiplier can be nicely illustrated by means of Figure 4.1. Strong duality for convex optimization
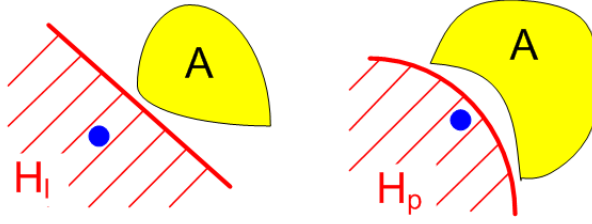
61

Figure 4.1: Separation of a point and a convex set by a halfspace (left) and of a point and a non-convex set by a semi-algebraic set (right)

problems is based on a geometric version of the Hahn-Banach theorem known as Eidelheit's separation theorem (as explained in e.g. [124]). A special case of this theorem is the following: for any point in $x^* \in \mathbb{R}^m$ and a compact convex set $\mathcal{A} \subset \mathbb{R}^m$ that are disjoint, there exists a nonzero $s \in \mathbb{R}^m$ and $b \in \mathbb{R}$ such that the halfspace $H_l := \{x \in \mathbb{R}^m | \langle s, x \rangle + b \leq 0\}$ contains x^* but does not contain points of $\mathcal{A}$, where $\langle \cdot, \cdot \rangle$ is the standard inner product in $\mathbb{R}^m$ as defined by $\langle u, v \rangle = u^T v$.

As is clear from the right-hand side of Figure 4.1, separation by an *affine functional* $\langle s, x \rangle + b$ of a point and a compact *non-convex* set that are disjoint is not possible in general. Instead of an affine functional one could try to construct a *polynomial* $p(x)$ such that the semi-algebraic set $H_p := \{x \in \mathbb{R}^m | p(x) \leq 0\}$ contains x^* but does not contain points of $\mathcal{A}$. An example of such a set is depicted on the right-hand side of Figure 4.1. This separation is illustrated in the next example for a convex and a nonconvex polynomial optimization problem.

Example 4.2 *Consider the convex optimization problem $p_{opt} = \min_{G(x) \preceq 0} f(x)$, $x \in \mathbb{R}$, with a linear SDP constraint $G(x) := diag(x-1, -x-1)$ and objective $f(x) := x^2 + 3x$. The separation is illustrated in Figure 4.2. In this figure the feasible region $\mathcal{G} = \{x \in \mathbb{R} | G(x) \preceq 0\} = [-1, 1]$ is depicted as a bar on the x-axis. Also shown is the so-called* epigraph *of $f(x)$ [155], i.e. the set $epi(f) := \{(x, r) \in \mathbb{R}^{m+1} | r \geq f(x)\}$, which is in this case a subset of $\mathbb{R}^2$. Since $f(x)$ is convex, $epi(f)$ is a convex set [155]. The optimal value $p_{opt} = -2$ is attained at $x^* = -1$, which is depicted with a star at $(x_{opt}, p_{opt}) = (-1, -2)$. Since strong duality holds the optimal dual multiplier $S = diag(0, 1)$ satisfies $\min(f(x) + \langle S, G(x) \rangle) = p_{opt}$. This implies that $f(x) \geq p_{opt} - \langle S, G(x) \rangle = x - 1 =: h(x)$. Geometrically this means that $epi(f)$ lies above the dotted line defined by the linear functional $h(x)$. In conclusion, the halfspace $H_l := \{(x, y) \in \mathbb{R}^2 | y \leq h(x)\}$ contains (x_{opt}, p_{opt}) but does not contain any interior points of $epi(f)$. This simple example illustrates the interpretation of Lagrange duality by separation with affine functionals. Note that for nonlinear, convex $G(x)$ the set $\{(x, y) \in \mathbb{R}^2 | y \leq h(x)\}$ with $h(x) := p_{opt} + \langle S(x), G(x) \rangle$ is not necessarily a halfspace. In this case a geometric interpretation with a halfspace is still possible using a different graph, see [125].*

The situation for a non-convex polynomial optimization problem $\inf_{G(x) \preceq 0} f(x)$ is shown in Figure 4.3 for the same $G(x)$ as above and $f(x) := -x^6 + 4x^4 - 4x^2 + \frac{1}{5}x$. An optimal polynomial multiplier $S(x)$ with

$$S(x) = \begin{pmatrix} S_1(x) & 0 \\ 0 & S_2(x) \end{pmatrix} \tag{4.7}$$
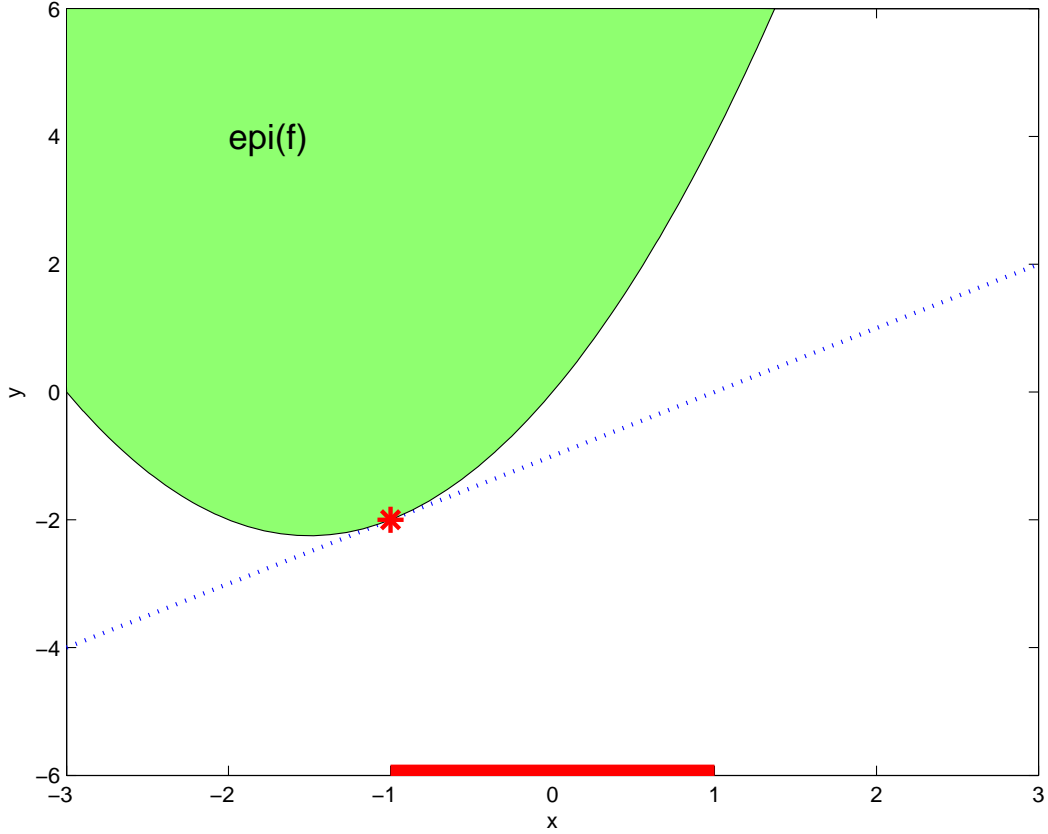
Figure 4.2: Interpretation of Lagrange duality by separation by a hyperplane. Depicted are feasible region (bar on x-axis), optimal value ($*$), epi(f) (solid), and separating curve (dotted).

where

$$S_1(x) := 0.511 + 0.563x - 1.296x^2 - 1.700x^3 + 0.854x^4 + 2.211x^5 + 1.050x^6$$
$$S_2(x) := 0.716 - 0.665x - 1.499x^2 + 1.986x^3 + 0.007x^4 - 1.406x^5 + 1.050x^6$$

has been computed with the LMI relaxation that will be presented in Section 4.5. S and $d := -1.3504$ satisfy (4.6), such that d is a global lower bound on the optimal value. Hence $x_{opt} = -0.8340$ is almost optimal since it satisfies $G(x_{opt}) \preceq 0$ and $f(x_{opt}) = -1.3503 := p_{opt}$.

In this case the interior of epi(f) is separated from the almost optimal point (x_{opt}, p_{opt}) by the curve described by $h(x) := p_{opt} - \langle S(x), G(x) \rangle$, as is nicely illustrated in Figure 4.3. This follows from $f(x) + \langle S(x), G(x) \rangle - p_{opt} \geq 0$ for all $x \in \mathbb{R}$, which indeed implies $f(x) \geq h(x)$ for all $x \in \mathbb{R}$, such that $H_p := \{(x, y) \in \mathbb{R}^2 | y \leq h(x)\}$ contains (x_{opt}, p_{opt}) but does not contain any interior points of epi(f).

In order to render the determination of the lower bound d satisfying (4.6) computational, we need a characterization of positivity of the matrix-valued polynomial $S(x)$ and the scalar polynomial (4.6). For this purpose we introduce the concept of Sum-Of-Squares matrix-valued polynomials in the next section. First we give some results on sum-of squares of scalar-valued polynomials. A scalar polynomial $s : \mathbb{R}^m \to \mathbb{R}$ is a Sum-Of-Squares if there
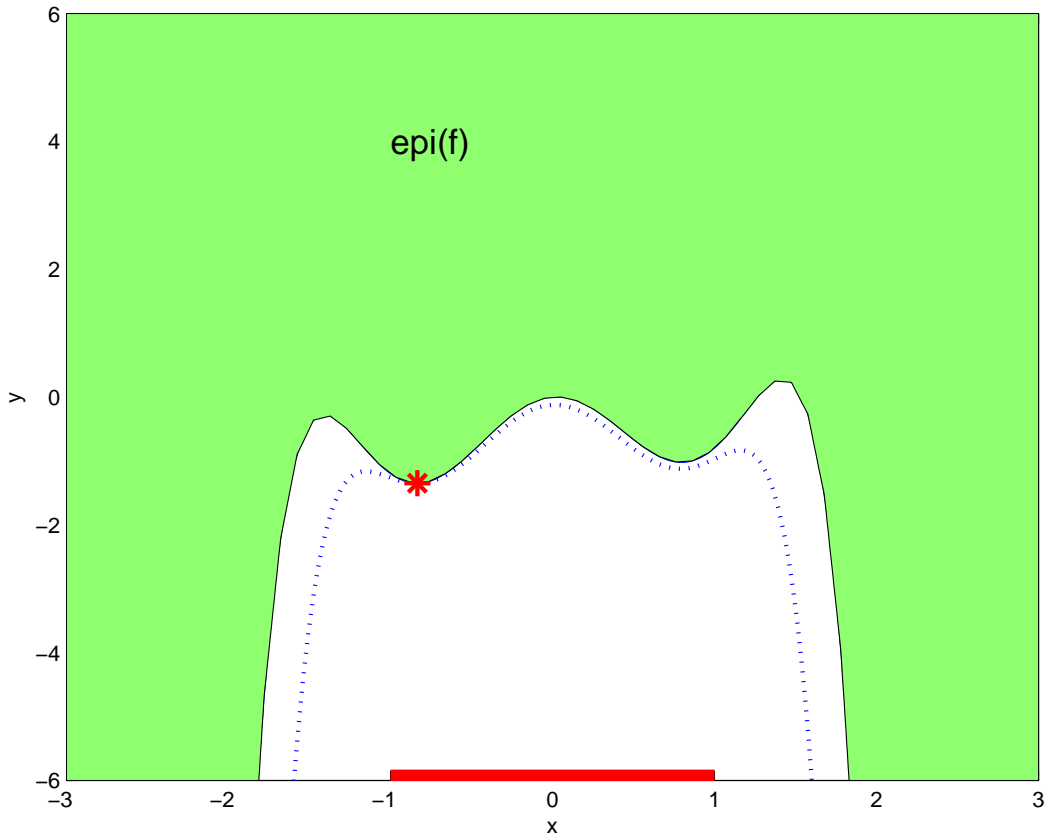
Figure 4.3: Interpretation of polynomial relaxations by separation with a semi-algebraic set. Depicted are feasible region (bar on x-axis), optimal value ($*$), epi(f) (solid), and separating curve (dotted).

exist some number $N \in \mathbb{N}$ and polynomials t_j, $j = 1, \ldots, N$ such that

$$s(x) = \sum_{j=1}^{N} t_j(x).$$

Remark. The characterization of non-negative polynomials, even for scalar polynomials, is a nontrivial problem, strongly related to Hilbert's 17th problem. In 1900 Hilbert posed his famous open questions in mathematics at the Paris Congress. Hilbert's 17th problem is related to the question whether any homogeneous nonnegative polynomial can be written as a sum of squares of rational functions. In 1927 Artin proved that this is indeed always possible. See [153] for a historical overview.

The following result due to Putinar is important in the sequel. It characterizes functions that are positive on sets described by polynomial inequalities that satisfy a constraint qualification.

Theorem 4.3 *Suppose that $g_i(x)$, $i = 0, 1, \ldots, r$ are polynomials such that $\mathcal{G} := \{x \in \mathbb{R}^m | \ g_i(x) \leq 0, \ i = 0, 1, \ldots, r\}$ is nonempty. Furthermore, suppose that the following*

constraint qualification holds true:

$$\exists \ SOS \ \psi_i(x), i = 0, 1, \ldots, r \ \textit{such that:} \ \left\{ x \in \mathbb{R}^m | \ \sum_{i=1}^{r} \psi_i(x)g_i(x) - \psi_0(x) \leq 0 \right\} \ \textit{is compact}$$

(4.8)

If a polynomial f is strictly positive on $\mathcal{G}$, i.e. satisfies $f(x) > 0$ for all $x \in \mathcal{G}$, then there exists SOS polynomials s_i, $i = 0, 1, \ldots, r$, such that

$$s_0(x) = f(x) + \sum_{i=1}^{r} s_i(x)g_i(x)$$

Proof. See [149], [101] or [180]. ∎

Remark. The constraint qualification (4.8) can be equivalently formulated as follows [180]:

$$\exists M > 0, \ \text{SOS polynomials} \ \psi_i(x), i = 0, 1, \ldots, r \ \text{such that:} \ \psi_0(x) = M - \|x\|^2 + \sum_{i=1}^{r} \psi_i(x)g_i(x).$$

(4.9)

4.3 Sum of squares of polynomial matrices

A symmetric matrix-valued polynomial matrix $S(x) : \mathbb{R}^m \to \mathcal{S}^r$ is said to be a SOS if there exists a (not necessarily square and typically tall) polynomial matrix $T(x)$ such that

$$S(x) = T(x)^T T(x).$$

If $r = 1$ and if $T_j(x)$, $j = 1, \ldots, N$, denote the components of the column vector $T(x)$ of length r we infer $S(x) = \sum_{j=1}^{N} T_j(x)^2$, which motivates our terminology since we are indeed dealing with a generalization of classical scalar SOS representations. Clearly, any SOS polynomial matrix is globally positive semi-definite, but as for $r = 1$ the converse is in general not true, see the remark on Hilbert's 17$^{\text{th}}$ problem in the previous section.

Example 4.4 *The matrix valued polynomial*

$$S(x) = \begin{pmatrix} 2 + x^2 & 3x \\ 3x & 3x^2(2 + 2x + 3x^2) \end{pmatrix}$$

(4.10)

is a SOS since $S(x) = T(x)^T T(x)$ where

$$T(x) = \begin{pmatrix} 1 & 2x \\ 1 & x \\ x & 0 \\ 0 & 3x^2 + x \end{pmatrix}$$

Let us briefly describe how one easily verifies whether a given polynomial matrix is SOS by solving a linear SDP. We present a construction that allows flexibility in the selection of the underlying monomial bases. Let us hence choose a polynomial vector $u(x) = \text{col}(u_1(x), \ldots, u_\xi(x))$ whose components $u_k(x)$, $k = 1, \ldots, \xi$ are pairwise different x-monomials. It suffices to take only those monomials whose squares have lower total degree than the total degree of $S(x)$. Usually even fewer monomials are needed, as will be explained in more detail in Section 4.5.4.

Then $S(x)$ of dimension $r \times r$ is said to be SOS with respect to the monomial basis $u(x)$ if there exist real matrices T_k, $k = 1, \ldots, \xi$, such that

$$S(x) = T(x)^T T(x) \;\; \text{with} \;\; T(x) = \sum_{k=1}^{\xi} T_k u_k(x) = \sum_{k=1}^{\xi} T_k (u_k(x) \otimes I_r).$$

If $U = (T_1 \;\; \cdots \;\; T_\xi)$ and if P denotes the permutation matrix that guarantees $u(x) \otimes I_r = P[I_r \otimes u(x)]$ we infer with $W = (UP)^T(UP) \succeq 0$ that

$$S(x) = [I_r \otimes u(x)]^T W [I_r \otimes u(x)]. \tag{4.11}$$

For the numerical verification of whether $S(x)$ is SOS it is convenient to introduce the following bilinear mapping

$$\langle ., . \rangle_r : \mathbb{R}^{p \times qr} \times \mathbb{R}^{p \times qr} \to \mathbb{R}^{r \times r}, \;\; \langle A, B \rangle_r = \text{Trace}_r(A^T B)$$

where

$$\text{Trace}_r(C) := \begin{pmatrix} \text{Trace}(C_{11}) & \cdots & \text{Trace}(C_{1r}) \\ \vdots & \ddots & \vdots \\ \text{Trace}(C_{r1}) & \cdots & \text{Trace}(C_{rr}) \end{pmatrix} \tag{4.12}$$

for $C \in \mathbb{R}^{qr \times qr}$, $C_{jk} \in \mathbb{R}^{q \times q}$, $j, k = 1, \ldots, r$. For later reference and with the Kronecker product $\otimes$ it is easy to see that

$$\text{Trace}_r(A(I_r \otimes B)) = \text{Trace}_r((I_r \otimes B)A) \;\; \text{for} \;\; A \in \mathbb{R}^{rp \times rq}, \;\; B \in \mathbb{R}^{q \times p}. \tag{4.13}$$

A proof is given in Appendix A.2.1. For symmetric $C \in \mathbb{R}^{qr \times qr}$ and $B \in \mathbb{R}^{q \times p}$, we can choose $A = (I_r \otimes B)^T C$ to infer

$$\text{Trace}_r((I_r \otimes B)^T C (I_r \otimes B)) = \langle C, I_r \otimes BB^T \rangle_r. \tag{4.14}$$

Note that the left-hand side just equals $(I_r \otimes B)^T C (I_r \otimes B)$ if B is a column. Finally,

$$\text{if } C \in \mathcal{S}^{qr} \text{ and } D \in \mathcal{S}^q \text{ are positive semi-definite then } \langle C, I_r \otimes D \rangle_r \succeq 0. \tag{4.15}$$

Indeed, if we decompose $D = BB^T$ we infer $(I_r \otimes B)^T C (I_r \otimes B) \succeq 0$. Since the trace operator from $\mathbb{R}^{n \times n}$ into $\mathbb{R}$ is completely positive [41], we conclude that $\text{Trace}_r((I_r \otimes B)^T C (I_r \otimes B)) \succeq 0$ and then (4.14) indeed implies (4.15). See Appendix A.2.1 for some additional elementary properties of Trace_r and the Kronecker product that will be frequently used in the remainder of this thesis. See Appendix A.2.2 for some remarks on complete positivity.

Let us now choose pairwise different monomials $u_j(x)$, $j = 1, \ldots, \xi$, and collect them into the vector $u(x) = (u_1(x)^T, \ldots, u_\xi(x)^T)^T$. Then $S(x)$ is said to be SOS with respect to

the monomial basis $u(x)$ if there exist real matrices T_j, $j = 1, \ldots, \xi$, with r columns such that

$$S(x) = T(x)^T T(x) \quad \text{with} \quad T(x) = \sum_{j=1}^{\xi} T_j u_j(x)$$

If $w_j(x)$, $j = 1, \ldots, \eta$, denote the pairwise different monomials that appear in $u(x)u(x)^T$ one can determine the unique symmetric matrices Z_j with

$$u(x)u(x)^T = \sum_{j=1}^{\eta} Z_j w_j(x).$$

Now we are ready to prove the following result which reduces the question of whether $S(x)$ is SOS with respect to $u(x)$ to an LMI feasibility problem.

Lemma 4.5 *The polynomial matrix $S(x)$ of dimension r is SOS with respect to the monomial basis $u(x)$ iff there exist symmetric S_j with $S(x) = \sum_{j=1}^{\eta} S_j w_j(x)$ and the following linear system has a solution $W \succeq 0$:*

$$\langle W, I_r \otimes Z_j \rangle_r = S_j, \quad j = 1, \ldots, \eta. \tag{4.16}$$

If W solves (4.16) then $S(x) = \langle W, I_r \otimes u(x)u(x)^T \rangle_r = (I_r \otimes u(x))^T W (I_r \otimes u(x))$.

    Proof. If $S(x)$ is SOS with respect to $u(x)$ and if we collect the coefficients of $T(x)$ as $U = (T_1 \ \cdots \ T_\xi)$ then $T(x) = U(u(x) \otimes I_r)$. If P denotes the permutation matrix that guarantees $u(x) \otimes I_r = P(I_r \otimes u(x))$, we infer with $W := (UP)^T(UP) \succeq 0$ that $S(x) = (I_r \otimes u(x))^T W (I_r \otimes u(x))$. With (4.14) we obtain

$$S(x) = (I_r \otimes u(x))^T W (I_r \otimes u(x)) = \langle W, I_r \otimes (u(x)u(x)^T) \rangle_r = \sum_{j=1}^{\eta} \langle W, I_p \otimes Z_j \rangle w_j(x),$$

which proves necessity. Conversely, if W satisfies (4.16) we can reverse the arguments to obtain the claimed representation. If $W \succeq 0$ the factorization $U^T U = PWP^T$ then leads to $S(x) = T(x)^T T(x)$ with $T(x) = U(u(x) \otimes I_r)$. ∎

Remark. The construction in Lemma 4.5 is a natural extension of the representation of scalar SOS polynomials with the Gram Matrix, see e.g. [147].

Example 4.6 *(Example 4.4 continued) For the Sum-Of-Squares polynomial $S(x)$ of Example 4.4 and polynomial basis $u(x) = \begin{pmatrix} 1 & x & x^2 \end{pmatrix}^T$ we obtain*

$$W = \begin{pmatrix} 2 & 0 & 0 & 0 & 3 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 6 & 3 \\ 0 & 0 & 0 & 0 & 3 & 9 \end{pmatrix}.$$

4.4 Polynomial Lagrange duality with SOS

Let us now just replace the inequality in (4.6) and $S(x) \succeq 0$ for all $x \in \mathbb{R}^m$ by the requirement that the corresponding polynomials or polynomial matrices are SOS. This leads to the following optimization problem:

$$\begin{array}{ll} \text{supremize} & d \\ \text{subject to} & S(x) \text{ and } s_0(x) := f(x) + \langle S(x), G(x) \rangle - d \text{ are SOS} \end{array} \tag{4.17}$$

with optimal value d_{opt}. We have seen in Section 4.2 that d_{opt} defines an upper bound on the optimal p_{opt} of (4.1). The main result of this chapter is to prove equality under a rather mild constraint qualification on $G(x)$, similar to the qualification in Theorem 4.3. On the basis of this result we will discuss in Section 4.5 how to construct families of relaxations in order to determine a whole sequence of lower bounds for problem (4.1) which converges to p_{opt}.

Theorem 4.7 *Let Assumption 4.1 hold true and p_{opt} be the optimal solution of (4.1). Furthermore, suppose that the following constraint qualification holds true:*

$$\exists M > 0, \text{ matrix SOS } \Psi(x) \text{ and SOS } \psi(x) \text{ such that: } \psi(x) = M - \|x\|^2 + \langle \Psi(x), G(x) \rangle. \tag{4.18}$$

Then the optimal value d_{opt} of (4.17) equals the optimal value p_{opt} of (4.1).

Proof. The value of (4.17) is not larger than p_{opt}. Since trivial for $p_{\mathrm{opt}} = \infty$, we assume that $G(x) \preceq 0$ is feasible. Choose any $\epsilon > 0$ and some $\hat{x}$ with $G(\hat{x}) \preceq 0$ and $f(\hat{x}) \le p_{\mathrm{opt}} + \epsilon$. Let us now suppose that $S(x)$ and $f(x) + \langle S(x), G(x) \rangle - d_{\mathrm{opt}}$ are SOS. Then

$$p_{\mathrm{opt}} + \epsilon - d_{\mathrm{opt}} \ge f(\hat{x}) - d_{\mathrm{opt}} \ge f(\hat{x}) + \langle S(\hat{x}), G(\hat{x}) \rangle - d_{\mathrm{opt}} \ge 0$$

and thus $p_{\mathrm{opt}} + \epsilon \ge d_{\mathrm{opt}}$. Since ϵ was arbitrary we infer $p_{\mathrm{opt}} \ge d_{\mathrm{opt}}$.

To prove the converse we first reveal that as a consequence of the constraint qualification, if

$$G(x) \preceq 0 \text{ is replaced by } \tilde{G}(x) := \mathrm{diag}(G(x), \|x\|^2 - M^2) \preceq 0$$

then (4.1) is not modified. In a first step of the proof let us show that the same is true for the SOS reformulation (4.17). Indeed suppose $S(x)$ and $s_0(x) = f(x) - d + \langle S(x), G(x) \rangle$ are SOS. Then $\tilde{S}(x) := \mathrm{diag}(S(x), 0)$ satisfies $s_0(x) = f(x) - d + \langle \tilde{S}(x), \hat{G}(x) \rangle$.

Conversely suppose $f(x) - d + \langle \tilde{S}(x), \tilde{G}(x) \rangle = \tilde{s}_0(x)$ with SOS polynomials $\tilde{S}(x)$ and $\tilde{s}_0(x)$. Now we make explicit use of $M^2 - \|x\|^2 = \psi(x) - \langle \Psi(x), G(x) \rangle$ with SOS matrices $\psi(x), \Psi(x)$. Let us partition

$$\tilde{S}(x) = \begin{pmatrix} S_{\mathrm{part}}(x) & * \\ * & s_{\mathrm{part}}(x) \end{pmatrix}$$

with $s_{\mathrm{part}}(x)$ a scalar polynomial. Define $S(x) := (S_{\mathrm{part}}(x) + s_{\mathrm{part}}(x)\Psi(x))$, of dimension r. It is easy to verify that both $s_{\mathrm{part}}(x)$ and $S(x)$ are both SOS and satisfy $\langle \tilde{S}(x), \tilde{G}(x) \rangle = \langle S(x), G(x) \rangle - s_{\mathrm{part}}(x)\psi(x)$. This implies $f(x) - d + \langle S(x), G(x) \rangle = \tilde{s}_0(x) + s_{\mathrm{part}}(x)\psi(x)$ and it remains to observe that $\tilde{s}_0(x) + s_{\mathrm{part}}(x)\psi(x)$ is SOS.

Therefore, from now on we can assume without loss of generality that

$$v_1^T G(x) v_1 = \|x\|^2 - M^2 \tag{4.19}$$

where v_1 is the last standard unit vector. To show $p_{\text{opt}} \leq d_{\text{opt}}$, choose a sequence of unit vectors $v_2, v_3, \ldots$ such that v_i, $i = 1, 2, \ldots$ is dense in the Euclidian unit sphere $\{v \in \mathbb{R}^r : \|v\| = 1\}$. Define $\mathcal{G}_N := \{x \in \mathbb{R}^m : v_i^T G(x) v_i \leq 0, \quad i = 1, \ldots, N\}$ to infer that $\mathcal{G}_N$ is compact (by (4.19)) and that $\mathcal{G}_N \supset \mathcal{G}_{N+1} \supset \mathcal{G}$ for $N = 1, 2, \ldots$. Therefore $p_N := \min\{f(x) : x \in \mathcal{G}_N\}$ is attained by some x_N and

$$p_N \leq p_{N+1} \leq p_{\text{opt}} \quad \text{for all} \quad N = 1, 2, \ldots.$$

Choose a subsequence N_ν with $x_{N_\nu} \to x_*$ to infer $p_{\text{opt}} \geq \lim_{\nu \to \infty} f(x_{N_\nu}) = f(x_*) =: p_*$. Then $p_* = p_{\text{opt}}$ follows if we can show that $G(x_*) \preceq 0$. In fact, otherwise there exists a unit vector v with $\delta := v^T G(x_*) v > 0$. By convergence there exists some K with $\|G(x_{N_\nu})\| \leq K$ for all ν. By density, there exists a sufficiently large ν such that $K\|v_i - v\|^2 + 2K\|v_i - v\| < \delta/2$ for some $i \in \{1, \ldots, N_\nu\}$. Since $v^T G(x_{N_\nu}) v \to v^T G(x_*) v$ we can increase ν to even guarantee $v^T G(x_{N_\nu}) v \geq \delta/2$ and arrive at the following contradiction:

$$
\begin{aligned}
0 \geq v_i^T G(x_{N_\nu}) v_i &= \\
&= (v_i - v)^T G(x_{N_\nu})(v_i - v) + 2v^T G(x_{N_\nu})(v_i - v) + v^T G(x_{N_\nu}) v \geq \\
&\geq -K\|v_i - v\|^2 - 2K\|v_i - v\| + \delta/2 > 0.
\end{aligned}
$$

When replacing the original SDP constraint $G(x) \preceq 0$ by the scalar constraints $v^T G(x) v \leq 0$, $i = 1, \ldots, N_0$ for some N_0, we are now in the position to apply Putinar's scalar representation result given in Theorem 4.3 since, due to (4.19), the constraint qualifications (4.9) and (equivalently) (4.8) are trivially satisfied. Hence for arbitrary $\epsilon > 0$, there exist a large enough $N \in \mathbb{N}$ and polynomials $t_i(x)$, $i = 1, \ldots, N$, such that

$$w(x) := f(x) - d_{\text{opt}} + \epsilon + \sum_{i=1}^{N} [t_i(x)^T t_i(x)](v_i^T G(x) v_i) \quad \text{is SOS in} \quad x. \qquad (4.20)$$

With the SOS matrix

$$S_N(x) := \sum_{i=1}^{N} v_i t_i(x)^T t_i(x) v_i^T = \begin{pmatrix} t_1(x)v_1^T \\ \vdots \\ t_N(x)v_N^T \end{pmatrix}^T \begin{pmatrix} t_1(x)v_1^T \\ \vdots \\ t_N(x)v_N^T \end{pmatrix}$$

we conclude that $f(x) - d_{\text{opt}} + \epsilon + \langle S_N(x), G(x) \rangle$ equals $w(x)$ in (4.20) and is thus SOS. This implies that the optimal value d_{opt} of (4.17) is at least $p_{\text{opt}} - \epsilon$, and since $\epsilon > 0$ was arbitrary the proof is finished.

$\blacksquare$

Theorem 4.7 is a natural extension of Theorem 4.3 to polynomial SDPs. Lasserre's approach [119] for minimizing $f(x)$ over scalar polynomial constraints $g_i(x) \leq 0$, $i = 1, \ldots, m$, is recovered with $G(x) = \text{diag}(g_1(x), \ldots, g_m(x))$. To the best of our knowledge, this generalization has, except for our publications [90, 88] and the recent independent work of Kojima [114], not been presented anywhere else in the literature. The constraint qualification in Theorem 4.7 is a natural generalization of that used by Schweighofer [180].

Remark. As in the scalar case, the constraint qualification (4.18) can be equivalently

formulated as follows: there exist an SOS matrix $\Psi(x)$ and an SOS polynomial $\psi(x)$ such that

$$\{x \in \mathbb{R}^{n_x} | \; \langle \Psi(x), G(x) \rangle - \psi(x) \leq 0\} \; \text{ is compact.} \qquad (4.21)$$

This is shown in Appendix A.2.3

Example 4.8 *We reconsider the non-convex optimization problem in Example 4.2, i.e.*

$$\inf_{G(x) \preceq 0} f(x),$$

with $f(x) = -x^6 + 4x^4 - 4x^2 + \frac{1}{5}x$ and $G(x) = diag(x-1, -x-1)$. Figure 4.4 displays $f(x)$ and $s_0(x) := f(x) - d_{opt} + \langle S(x), G(x) \rangle$, where $S(x)$ is as defined in (4.7). It can numerically
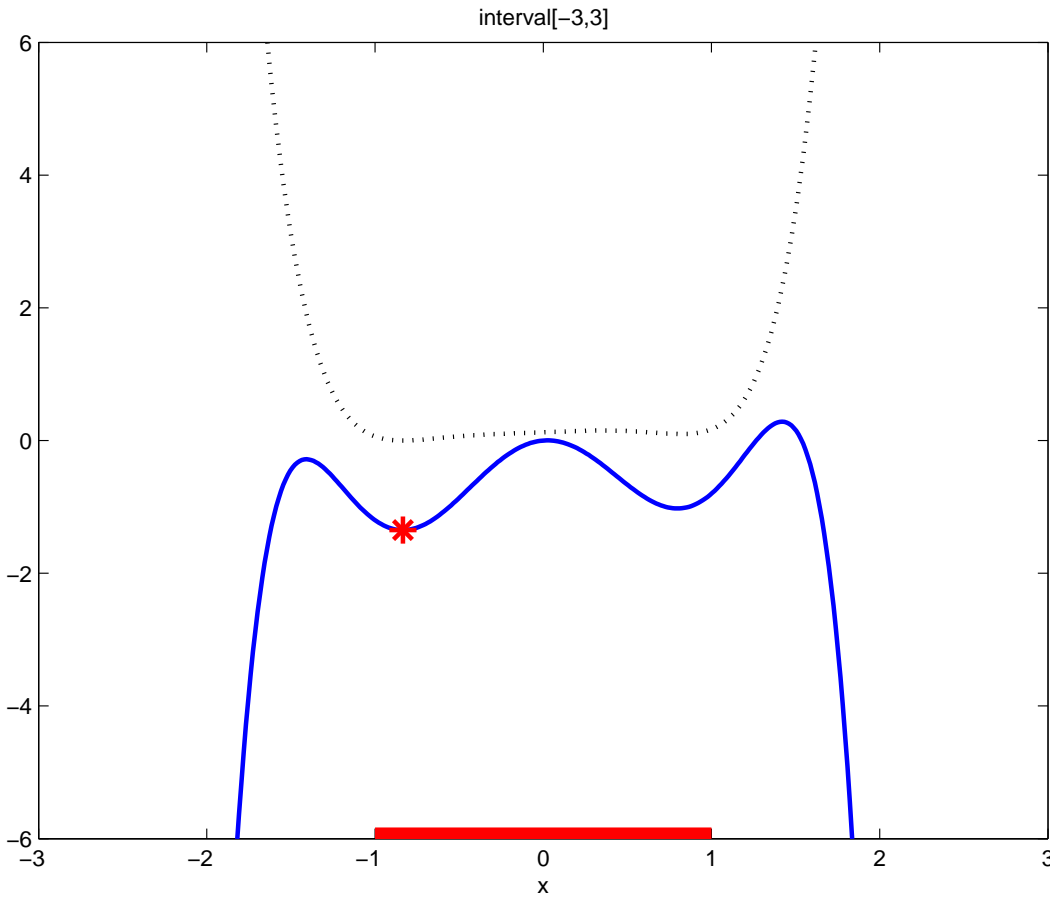
Figure 4.4: SOS relaxation example with the feasible region (bar on x-axis), $f(x)$ (solid line), $s_0(x) := f(x) - d_{opt} + \langle S(x), G(x) \rangle$ dotted line and optimal value p_{opt} ($*$).

be verified that $s_0(x)$ is a Sum-Of-Squares and as a consequence it is nonnegative. The function $f(x) - d_{opt}$ is negative for some values outside the feasible domain $\mathcal{G}$, e.g. for all $x > 1.6$. Hence to render $s_0(x) = f(x) - d_{opt} + \langle S(x), G(x) \rangle$ nonnegative the term $\langle S(x), G(x) \rangle$ must be positive for those values. This is possible since $G(x)$ is indefinite outside $\mathcal{G}$. This illustrates that, to render $s_0(x)$ nonnegative, $S(x)$ 'pulls $f(x)$ up' outside the feasible domain $\mathcal{G}$ with the term $\langle S(x), G(x) \rangle$.

Remark. The Sum-Of-Squares relaxations can be interpreted as adding redundant constraints to the problem and then applying standard Lagrange dualization. Consider

$$\inf_{G(x) \preceq 0} f(x)$$

with optimal value p_{opt} and an arbitrary $\epsilon > 0$. If $G(x)$ satisfies the constraint qualification (4.18), then there exist SOS $S(x)$ and $s_0(x)$ with

$$f(x) + \langle S(x), G(x) \rangle - p_{\text{opt}} + \epsilon = s_0(x).$$

Let us introduce the new optimization problem

$$\inf_{G(x) \preceq 0, \ g(x) \leq 0} f(x) \tag{4.22}$$

with $g(x) := \langle S(x), G(x) \rangle$ and optimal value p_*. Since $g(x) \leq 0$ is implied by $G(x) \preceq 0$, the constraint $g(x) \leq 0$ is redundant. Hence the optimal values of the two problems coincide: $p_{\text{opt}} = p_*$. The Lagrange relaxation of (4.4) with *constant* Lagrange multipliers $S_1 = 0$ and $s_2 = 1$ results in

$$f(x) + \langle S_1, G(x) \rangle + \langle s_2, g(x) \rangle = f(x) + \langle S(x), G(x) \rangle = s_0(x) + p_{\text{opt}} - \epsilon.$$

This illustrates that if we add suitable redundant constraints to the optimization problem, then we can use a constant multiplier $S \succeq 0$ in (4.17) instead of a polynomial multiplier.

4.5 LMI relaxations based on SOS

4.5.1 Construction of LMI relaxation families

With monomial a vector $v(x) \in \mathbb{R}^\zeta$ and some real coefficient matrix $B \in \mathbb{R}^{r \times r\zeta}$ let us represent the constraint functions as $G(x) = B(I_r \otimes v(x))$. The constraint in (4.17) can be reformulated as

$$f(x) + \langle S(x), G(x) \rangle - d = s_0(x), \qquad \text{with } s_0(x) \text{ and } S(x) \text{ being SOS.}$$

Moreover, let us choose monomial vectors $u(x)$ and $u_0(x)$ of length ξ and ξ_0 to parameterize $S(x)$ and $s_0(x)$ with respect to $u(x)$ and $u_0(x)$ with $W \succeq 0$ and $W_0 \succeq 0$ respectively as in Section 4.3. We infer

$$s_0(x) = \langle W_0, u_0(x)u_0(x)^T \rangle$$

and

$$
\begin{aligned}
\langle S(x), G(x) \rangle &= \langle \langle W, I_r \otimes u(x)u(x)^T \rangle_r, B(I_r \otimes v(x)) \rangle \\
&= \text{Trace}\left(W \left(B \left(I_r \otimes v(x) \right) \right) \otimes u(x)u(x)^T \right) \\
&= \text{Trace}\left(W \left(B \otimes I_\xi \right) \left(I_r \otimes v(x) \otimes u(x)u(x)^T \right) \right),
\end{aligned}
$$

where we used the properties of the Kronecker product and Trace_r given in Appendix A.2.1. Let us now choose the pairwise different monomials

$$w_0(x) = 1, \qquad w_1(x), \ldots, w_\eta(x)$$

71

to allow for the representations

$$u_0(x)u_0(x)^T = \sum_{j=0}^{\eta} P_j^0 w_j(x), \qquad v(x) \otimes u(x)u(x)^T = \sum_{j=0}^{\eta} P_j w_j(x) \qquad (4.23)$$

and

$$f(x) = \sum_{j=0}^{\eta} a_j w_j(x). \qquad (4.24)$$

Then there exist SOS polynomials $S(x)$ and $s_0(x)$ with respect to $u(x)$ and $u_0(x)$, such that

$$f(x) + \langle S(x), I_q \otimes G(x) \rangle - d = s_0(x) \qquad (4.25)$$

if and only if there exists a solution to the following LMI system:

$$W_0 \succeq 0, \qquad W \succeq 0, \qquad (4.26)$$
$$a_j + \langle W_0, P_j^0 \rangle + \langle W, (B \otimes I_\xi)(I_r \otimes P_j) \rangle - \delta_j d = 0, \ j = 0, 1, \ldots, \eta, \qquad (4.27)$$

where $\delta_0 = 1$ and $\delta_j = 0$ for $j = 1, \ldots, \eta$. We can hence easily maximize d over these LMI constraints to determine a lower bound on the optimal value of (4.1). The next theorem shows that these lower bounds are guaranteed to converge to the optimal value p_{opt} of (4.1), if we choose $u_0(x)$ and $u(x)$ to comprise all monomials up to a certain degree, and if we let the degree bound grow to infinity.

Theorem 4.9 *Let* $\{u_1(x), u_2(x), \ldots\}$ *be a sequence of monomial vectors such that*

$$span\left(\{u_1, u_2, \ldots\}\right) = \mathbb{R}[x].$$

Consider the sequence

$$d_N := \quad supremum \quad d$$
$$subject \ to \quad s_0(x) = f(x) + \langle S(x), G(x) \rangle - d$$
$$with \ s_0(x) \ and \ S(x) \ SOS \ w.r.t. \ \Big(\ u_1(x), \ u_1(x), \ \ldots, \ u_N(x) \ \Big)^T$$
$$(4.28)$$

Then $\lim_{N \to \infty} d_N = p_{opt}$, *where* p_{opt} *is the optimal value of* (4.1).

Proof. Obviously $d_N \leq d_{N+1} \leq p_{\mathrm{opt}}$ for all $N \in \mathbb{N}$. Furthermore Theorem 4.7 implies that for arbitrary $\epsilon > 0$ there exist SOS matrices $S(x)$ and $s_0(x)$ such that $s_0(x) = f(x) + p_{\mathrm{opt}} - \epsilon + \langle S(x), G(x) \rangle$. Since

$$span\left(\{u_1, u_2, \ldots\}\right) = \mathbb{R}[x]$$

we infer that there exists a large enough $M \in \mathbb{N}$ such that $s_0 \in span\left(\{u_1, u_2, \ldots, u_M\}\right)$ and $S_{ij} \in span\left(\{u_1, u_2, \ldots, u_M\}\right)$, $i = 1, \ldots, r$, $j = 1, \ldots, r$, where S_{ij} denotes the entry of the ith row and j^{th} column of the matrix-valued polynomial S. This implies that S and s_0 are SOS with respect to $\left(\ u_1(x), \ u_1(x), \ \ldots, \ u_M(x) \ \right)^T$. We conclude that d_M in (4.28) satisfies $d_M \geq p_{\mathrm{opt}} - \epsilon$. Since ϵ was arbitrary, the result follows.

∎

4.5.2 Size of the LMI problem

The size of the LMI relaxation for (4.17) is easily determined as follows. The constraints are (4.26) and (4.27). The conditions on the matrices W_0 and W to be nonnegative definite in (4.26) comprise one inequality in $\mathcal{S}^{\xi_0}$ and one in $\mathcal{S}^{r\xi}$, where (as mentioned earlier in the text) r, ξ_0 and ξ denote the number of rows in $G(x)$ and the number of monomials for the SOS matrices $s_0(x)$ and $S(x)$ respectively. Equation (4.27) adds another $\eta + 1$ scalar equation constraints to the LMI problem.

The decision variables in the LMI relaxation are $d \in \mathbb{R}$ and the matrices for the SOS representation $W_0 \in \mathcal{S}^{\xi_0}$ and $W \in \mathcal{S}^{r\xi}$. Since a symmetric matrix in $\mathcal{S}^{r\xi}$ can be parameterized by a vector in $\mathbb{R}^{\frac{1}{2}r\xi(r\xi+1)}$, we end up in total with

$$1 + \frac{1}{2}\left(\xi_0(\xi_0 + 1) + r\xi(r\xi + 1)\right) \tag{4.29}$$

scalar variables in our LMI problem.

4.5.3 Comparison with scalarization

In this section we shed some light on the benefits of exploiting the matrix structure in the SOS relaxations compared to straightforward scalarization. In particular we explain why scalarization fails to lead to the desired properties (of quadratic growth in the matrix sizes) of the corresponding LMI relaxations. Observe that $G(x) \preceq 0$ is equivalent to $M_i(G(x)) \leq 0, i = 1, \ldots, N$ where $M_i(A), i = 1, \ldots, N$ are all the principal (nonempty) minors of a matrix $A \in \mathbb{R}^{r \times r}$ [94]. The maximum of the total degrees of the minors $M_i(G(x)), i = 1, \ldots, N$ is often much higher than the total degree of $G(x)$. A larger polynomial degree often requires to use a larger monomial basis and hence more variables and constraints in the LMI relaxation to obtain good approximations of p_{opt}. This is illustrated by the following example (inspired by a personal communication with Didier Henrion).

Example 4.10 *We compute lower bounds on* $p_{opt} = \inf_{G(x) \preceq 0} f(x)$ *where* $x = \begin{pmatrix} x_1 & x_2 \end{pmatrix}^T \in \mathbb{R}^2$, $f(x) = x_1 + x_2$ *and*

$$G(x) = \begin{pmatrix} -1 & -x_1^2 & 0 \\ -x_1^2 & -9 + x_2^2 & 0 \\ 0 & 0 & -1 + \frac{x_2^2 + x_1^2}{100} \end{pmatrix}.$$

Table 4.1 shows lower bounds on the optimal value and the sizes of the LMI problems for SOS relaxations based on (4.17) and based on two ways of scalarization, where we used for "Scalar 1"

$$g_1(x) = \det\left(\begin{pmatrix} I_2 & 0 \end{pmatrix} G \begin{pmatrix} I_2 \\ 0 \end{pmatrix}\right), \quad g_2(x) = \text{Trace}\left(\begin{pmatrix} I_2 & 0 \end{pmatrix} G \begin{pmatrix} I_2 \\ 0 \end{pmatrix}\right) \quad \text{and } g_3(x) = G_{3,3}$$

and for "Scalar 2" the minors $g_i(x) := M_i(G(x)) \leq 0, i = 1, \ldots, 7$. *We choose monomial bases* $u_0(x)$ *as shown in the table and* $u(x) = 1$ *to represent* $s_0(x)$ *and* $S(x)$ *respectively as in Lemma 4.5. An upper bound on the optimal value* p_{opt} *is* $F(-1.148, -2.695) = -3.843$, *which was obtained by gridding. As is clear from the table, the matrix-valued relaxation is (almost) exact for* $u_0 = (1, x_1, x_2)^T$, *obtained by LMI optimization with 18 constraints and 13 variables. The scalarised relaxations are exact if* $u_0 = (1, x_1, x_2, x_1^2, x_1 x_2, x_2^2)^T$, *which required the solution of an LMI problem with 27 constraints and 16 variables. The table shows that "Scalar 2" requires even more LMI variables to make the relaxation gap small.*

Table 4.1: Optimal values and LMI size for matrix and scalar relaxations

Relax- ation	optim. value d_{opt}	monomial in $u_0(x)^T$	LMI constr	LMI vars
Matrix	-3.85	$(1, x_1, x_2)$	18	13
Scalar 1	-12.65	$(1, x_1, x_2)$	16	10
Scalar 1	-3.85	$(1, x_1, x_2, x_1^2, x_1 x_2, x_2^2)$	27	16
Scalar 2	-2.6e4	$(1, x_1, x_2)$	29	14
Scalar 2	-3.85	$(1, x_1, x_2, x_1^2, x_1 x_2, x_2^2)$	48	29

4.5.4 Strict feasibility and variable reduction

Strict feasibility

We solve the LMI relaxations, i.e. supremizing d with constraints (4.26) and (4.27), by Interior Point solvers as described in Section 2.5.2. For these solvers absence of strictly feasible points, i.e. the absence of $d \in \mathbb{R}$, $W_0 \in \mathcal{S}^{\xi_0}$ and $W \in \mathcal{S}^{r\xi}$ that satisfy (4.26) strictly and satisfy (4.27), often results in numerical problems. Hence guaranteeing the existence of strictly feasible points is important to obtain numerically accurate results.

Variable elimination

One of the reasons for the absence of strictly feasible points is the occurrence of monomials in the basis $u_0(x)$ that are not needed in an SOS decomposition of $f(x) + \langle S(x), G(x) \rangle$. To eliminate unnecessary monomials in the basis $u_0(x)$ the *Newton polytope* is a useful concept (as defined in e.g. [189]). The Newton polytope of a polynomial $f : \mathbb{R}^m \to \mathbb{R}$ given by $f(x) := \sum_{i=1}^{N} c_i x^{\alpha_i}$, is the convex hull of its exponents $\alpha_i \in \mathbb{R}^m$, $i = 1, \ldots, N$:

$$\mathbf{New}(f(x)) := \mathrm{co}\{\alpha_1, \alpha_2, \ldots, \alpha_N\} \subset \mathbb{R}^m.$$

If $f(x)$ is a scalar SOS polynomial, then it is SOS with respect to any monomial basis $u_0(x)$ that contains all monomials $m_j(x) = x_j^\alpha$ whose exponents $\alpha_j \in (\mathbb{N}_0)^m$ satisfy $2\alpha_j \in \mathbf{New}(f(x))$ [152], see also [148]. This motivates to define $\mathcal{M}_{f(x)}$ as the set of monomials that contains all monomials that can appear in a SOS representation of f:

$$\mathcal{M}_{f(x)} := \left\{ x^\alpha \mid \alpha \in (\mathbb{N}_0)^m, \ 2\alpha \in \mathbf{New}(f(x)) \right\}$$

This implies that the only monomials that are needed in $u_0(x)$ are in $\mathcal{M}_{f(x)+\langle S(x), G(x)\rangle - d}$. Note that $\mathcal{M}_{f(x)+\langle S(x), G(x)\rangle - d}$ depends on d and the coefficients W of the SOS polynomial $S(x) = \langle W, I_r \otimes (u(x)u(x)^T)\rangle_r$, which are not known a priori. In those cases we let

$$\mathcal{M}_{f(x)+\langle S(x), G(x)\rangle - d}$$

be the union of all monomials that can occur for all instances of the unknown parameters.

We therefore choose the vector u_0 to contain precisely those monomials in the set $\mathcal{M}_{f(x)+\langle S(x), G(x)\rangle - d}$. However, even when we use the monomials in $\mathcal{M}_{f(x)+\langle S(x), G(x)\rangle - d}$ as basis for $u_0(x)$ for some problem instances a strictly feasible point, i.e. that satisfies (4.26) strictly and satisfies (4.27), does still not exist. Consider for instance the optimization

problem $\inf_{G(x) \preceq 0} f(x)$, where $x \in \mathbb{R}$, $f(x) := x$ and

$$G(x) = \begin{pmatrix} -x^3 & 0 & 0 \\ 0 & x^2 - 1 & 0 \\ 0 & 0 & -x \end{pmatrix}$$

and consider SOS basis $u(x) = 1$. We choose $u_0(x)$ according to the prescription above, e.g. $u_0(x) = \begin{pmatrix} 1 & x \end{pmatrix}^T$. Then the problem in (4.17) is to find the supremal d, $W_0 \succeq 0$ and nonnegative numbers $\lambda_i \geq 0$, $i = 1, 2, 3$ such that

$$x - d - \lambda_1 x^3 + \lambda_2(x^2 - 1) - \lambda_3 x = u_0^T(x) W_0 u_0(x).$$

Since this equation implies that $\lambda_1 = 0$, this problem is not strictly feasible. We therefore propose to use the following variable elimination procedure.

Algorithm 4.11 *Initialization: set $i = 0$, $\mathcal{W}_0^i = \mathcal{S}^{\xi_0}$ and $\mathcal{W}^i = \mathcal{S}^{r\xi}$.*

1. compute (by e.g. vectorization followed by Gauss elimination [72]) linearly independent $W^k \in \mathcal{W}^i$ and $W_0^k \in \mathcal{W}_0^i$, $k = 0, \ldots, M$ for some $M \in \mathbb{N}_0$ such that

$$\mathcal{V}^i := \left\{ \left(d, W^0 + \sum_{k=1}^{M} v_k W^k, W_0^0 + \sum_{k=1}^{M} v_k W_0^k \right) \in \mathbb{R} \times \mathcal{S}^{r\xi} \times \mathcal{S}^{\xi} \middle| d \in \mathbb{R}, \ v \in \mathbb{R}^M \right\}$$

parameterizes all solutions of (4.27) restricted to $\mathcal{W}^i \times \mathcal{W}_0^i$, i.e.

$$\mathcal{V}^i = \left\{ (d, W, W_0) \in \mathbb{R} \times \mathcal{W}^i \times \mathcal{W}_0^i \middle| (d, W, W_0) \ satisfy \ (4.27) \right\}$$

2. Set

$$\mathcal{I}_0^i := \left\{ l \in \{1, \ldots \xi_0\} \middle| e_l^T W_0^k e_l = 0 \ \ for \ all \ \ k \in \{0, \ldots M\} \right\}$$

and

$$\mathcal{I}^i := \left\{ l \in \{1, \ldots r\xi\} \middle| e_l^T W^k e_l = 0 \ \ for \ all \ \ k \in \{0, \ldots M\} \right\}.$$

Furthermore set

$$\mathcal{W}_0^{i+1} = \left\{ W_0 \in \mathcal{W}_0^i \middle| W_0 e_l = 0 \ for \ all \ l \in \mathcal{I}_0^i \right\}$$

and

$$\mathcal{W}^{i+1} = \left\{ W \in \mathcal{W}^i \middle| W e_l = 0 \ for \ all \ l \in \mathcal{I}^i \right\}$$

3. if $\mathcal{W}^i \neq \mathcal{W}^{i-1}$ or $\mathcal{W}_0^i \neq \mathcal{W}_0^{i-1}$, set $i = i + 1$ and go to step 1, otherwise stop.

The steps in the algorithm can be interpreted as follows. In step 1 a parametrization of the solution set $\mathcal{V}^i$ of the linear equations (4.27) is constructed, where the additional restriction $W \in \mathcal{W}^i$ and $W_0 \in \mathcal{W}_0^i$ is made. Since $\mathcal{W}_0^i = \mathcal{S}^{\xi_0}$ and $\mathcal{W}^i = \mathcal{S}^{r\xi}$ for $i = 0$, this restricts W and W_0 to be symmetric matrices of appropriate size in the first iteration of the algorithm. In step 2 the indices $\mathcal{I}^i$ of the diagonal elements of W that are identically zero in $\mathcal{V}^i$ are identified. All variables in the corresponding rows and columns of W must necessarily be zero, since W is positive semi-definite. These variables are therefore eliminated, by enforcing that the columns and rows of the matrices in $\mathcal{W}^{i+1}$ are zero for those indices. In a similar fashion rows and columns of W_0 are eliminated. The loop is repeated if at least one row or column in either W or W_0 has been eliminated in the last iteration.

As the original conditions $W \succeq 0$ and $W_0 \succeq 0$ are often not strictly feasible together with (4.27), it should be clear that we actually implement the modified version

$$V_0 \succeq 0, \qquad V \succeq 0,$$

$$a_j + \langle E_{I_0} V_0 E_{I_0}^T, P_j^0 \rangle + \langle E_I V E_I^T, (B \otimes I_\xi)(I_r \otimes P_j) \rangle - \delta_j d = 0, \; j = 0, 1, \ldots, \eta,$$

where the optimization variables V_0 and V are symmetric matrices of appropriate size, $\mathcal{I}$ and $\mathcal{I}_0$ are the final index sets in Algorithm 4.11 and E_{I_0} is a full column rank matrix whose columns consist of the basis vectors $\{e_i \mid i \in \mathcal{I}_0\}$. Similarly, E_I is a full column rank matrix whose columns consist of the basis vectors $\{e_i \mid i \in \mathcal{I}\}$.

4.5.5   Choice of monomial basis

A straightforward way to choose the monomial basis vector $u(x)$ is to collect all monomials up to a certain total degree p, and then apply the elimination technique discussed in the previous section. If x consists of a large number of variables, this will lead to large LMI problems even for small p. Indeed the number of all monomials in $x \in \mathbb{R}^m$ up to a given total degree p is given by [147]

$$\binom{m + p}{m}, \tag{4.30}$$

where $\binom{\cdot}{\cdot}$ denotes the binomial coefficient defined as

$$\binom{a}{b} := \frac{a!}{b!(a-b)!}$$

for $a > b$. Observe that

$$\binom{m+p}{m} = \frac{(m+p)(m+p-1)\ldots 2 \cdot 1}{m(m-1)\ldots 2 \cdot 1} = (m+p)(m+p-1)\ldots(m+2)(m+1).$$

Although this is for fixed p polynomial growth in m, we observe that the growth rate is already large for moderate total degree p. If the number of variables m is large, the length of the monomial vector is therefore huge even for small p, e.g if $m = 20$, $p = 3$ the number of monomials up to order p is

$$\binom{m+p}{m} = \binom{23}{20} = 1771.$$

In [144] a comparison is given of the number of variables and constraints in two different LMI formulations of the SOS relaxation. It is clear from Section 4.5.2 that this implies that the number of variables and constraints in the LMI relaxation also grows fast in m. To keep the LMI relaxations computationally tractable, it is therefore important to use a more refined selection of the monomial basis than bluntly collecting all monomials up to a certain total degree. It is however not known how to find the monomial basis of a certain length that gives the best lower bound in (4.17). This is a practically important and interesting question that is not yet fully answered. The elimination procedure described in the previous section is a first step in this direction, since it eliminates monomials in $u_0(x)$ and $u(x)$ that cannot occur in a sum-of squares representation. A more sophisticated selection

procedure that selects monomials based on their expected impact on the quality of the lower bounds is however beyond the scope of this thesis. Unless specified differently, we collect all monomials of a certain degree in $u(x)$ and $u_0(x)$, followed by the elimination procedure of the previous section.

Remark. If $G(x)$ is diagonal, a priori bounds on the degree of the polynomial $S(x)$ to achieve an exact lower bound with (4.17) have been presented in [138, 181] and [174]. However, these bounds get very large if we approach the optimal value of the SDP.

4.5.6 Constructing optimal solutions

In this section we present a method to construct a globally optimal solution, if the relaxation gap is zero. If such solutions exist, they can be computed on the basis of the kernel of the optimal W_0 in (4.26) and (4.27). The procedure, to be explained below, is based on the following theorem. See [37] for a similar result.

Theorem 4.12 *Suppose that (4.1) attains its optimal value p_{opt} at x_{opt}. Let d_{opt}, $W_{(0,opt)}$ and W_{opt} be an optimal solution to the SOS relaxation for some monomial bases $u_0(x)$ and $u(x)$. If $p_{opt} = d_{opt}$, then $u_0(x_{opt}) \in Ker(W_{(0,opt)})$.*

Proof. With $s_{(0,\text{opt})}(x) = u_0(x)^T W_{(0,\text{opt})} u_0(x)$ and $S_{\text{opt}}(x) = [I_r \otimes u(x)]^T W_{\text{opt}} [I_r \otimes u(x)]$ we conclude that

$$s_{(0,\text{opt})}(x) = f(x) + \langle S_{\text{opt}}(x), G(x) \rangle - d_{\text{opt}}$$

If we plug in x_{opt}, we infer

$$s_{(0,\text{opt})}(x_{\text{opt}}) = \langle S_{\text{opt}}(x_{\text{opt}}), G(x_{\text{opt}}) \rangle \tag{4.31}$$

Since $s_{(0,\text{opt})}$ is SOS, $s_{(0,\text{opt})}(x_{\text{opt}}) \geq 0$. On the other hand, $\langle S_{\text{opt}}(x_{\text{opt}}), G(x_{\text{opt}}) \rangle \leq 0$, because S_{opt} is SOS and $G(x_{\text{opt}}) \preceq 0$ by feasibility of x_{opt} for (4.1). Hence (4.31) implies

$$u_0(x_{\text{opt}})^T W_{(0,\text{opt})} u_0(x_{\text{opt}}) = s_{(0,\text{opt})}(x_{\text{opt}}) = 0$$

This shows that $u_0(x_{\text{opt}}) \in \text{Ker}(W_{(0,\text{opt})})$.
$\blacksquare$

Remark. Note that the theorem includes the possibility that the optimal value is attained at more than one points.

With this result we are in the position to present the exactness test. We assume that the first $m + 1$ elements of the bases $u_0(x)$ are

$$\begin{pmatrix} I_{m+1} & 0 \end{pmatrix} u_0(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}$$

and suppose that these are not eliminated by the procedure described Section 4.5.4. If the kernel of $W_{(0,\text{opt})}$ has dimension 1, the construction of an optimal x_{opt} is very simple. If y is the kernel vector of $W_{(0,\text{opt})}$ with its first element equal to 1, an optimal solution is x_{opt} given by:

$$\begin{pmatrix} 1 \\ x_{\text{opt}} \end{pmatrix} = \begin{pmatrix} I_{m+1} & 0 \end{pmatrix} y. \tag{4.32}$$

The construction is more involved if the dimension of the kernel l is larger than one. The solution can be found by solving a polynomial semi-definite feasibility problem with l variables. If l is small compared to the number of variables m in x, this may be a relatively easy problem.

Example 4.13 *We will construct the optimal solution to the non-convex optimization problem of Example 4.2, i.e.*

$$\inf_{G(x) \preceq 0} f(x),$$

where $f(x) = -x^6 + 4x^4 - 4x^2 + \frac{1}{5}x$ *and* $G(x) = diag(x - 1, -x - 1)$. *The best lower bound for a monomial basis* $u_0(x) = u(x) = (1, x, x^2, x^3)$ *is* -1.3503, *and the corresponding* $W_{0,opt}$ *is*

$$W_{0,opt} = \begin{pmatrix} 0.46513 & -0.43787 & -0.67852 & 0.61419 \\ -0.43787 & 1.0641 & 0.70175 & -1.4376 \\ -0.67852 & 0.70175 & 1.0167 & -0.95408 \\ 0.61419 & -1.4376 & -0.95408 & 1.9739 \end{pmatrix}.$$

We decompose this matrix as

$$W_{0,opt} = Y^T \begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & 0 & \sigma_4 \end{pmatrix} Y,$$

with $\sigma_1 = 3.85$, $\sigma_2 = 0.646$ $\sigma_3 = 0.0188$ *and* $\sigma_4 = 8.87 \cdot 10^{-10}$. *The dimension of the (approximate) kernel is therefore 1, since there is a large gap between* σ_3 *and* σ_4. *The corresponding scaled kernel vector is*

$$y = \begin{pmatrix} 1 \\ -0.83505 \\ 0.69731 \\ -0.58229 \end{pmatrix}$$

The optimal solution is therefore $x_{opt} = -0.835$. *Indeed* $f(x_{opt}) = -1.35$.

Remark. A different approach to extract global solutions are presented in [84, 120]. In these papers the dual variables of the LMI relaxation are used to extract the solution.

4.6 SOS relaxtaions for fixed-order H_∞ controller synthesis

An SOS relaxation of the $\mathcal{H}_\infty$-fixed order synthesis problem is:

$$\begin{aligned} \text{supremize} \quad & d \\ \text{subject to} \quad & \gamma + \langle S(\gamma, X, K), G(\gamma, X, K) \rangle - d = s_0(\gamma, X, K) \\ & S(\gamma, X, K) \text{ and } s_0(\gamma, X, K) \text{ are SOS} \end{aligned}$$

where

$$G(\gamma, X, K) := diag\left(-X, -\mathcal{B}(\gamma, X, K), \gamma^2 + \|X\|_F^2 + \|K\|_F^2 - \rho\right),$$

$\|A\|_{\mathrm{F}}$ denotes the Frobenius norm of A and $\mathcal{B}$ is as in (4.4). The additional constraint $\gamma^2 + \|X\|_{\mathrm{F}}^2 + \|K\|_{\mathrm{F}}^2 - \rho \leq 0$ is added to bound the size of the variables. It guarantees that the constraint qualification in (4.18) is satisfied. If we choose ρ large enough, this extra constraint does hardly affect the optimal value.

Remark. The symmetric matrix-valued polynomials with matrix-valued arguments such as $S(\gamma, X, K)$ must be read as SOS polynomials $S(x)$ with $x = \begin{pmatrix} \gamma & \mathrm{svec}(X)^T & \mathrm{vec}(K)^T \end{pmatrix}^T$, where svec is as defined in (3.18) on page 50.

This definition of matrix-valued polynomials with a matrix-valued argument is a natural extension of a matrix-valued polynomial. In a similar fashion we define rational functions with a matrix-valued argument. Since $\mathrm{vec}(K) \mapsto K$ is a linear bijection, standard results on polynomials and rational functions such as Weierstrass' Approximation Theorem and the S-procedure hold obviously true for polynomials and rational functions with a matrix-valued argument, as will be used in the next chapter. Noncommutative polynomials in the sense of [81] are conceptually different. The interested reader is referred to that paper for Sum-Of-Squares results for these kind of polynomials.

Remark. One of the strengths of the proposed relaxation scheme is the possibility to construct globally optimal structured controllers, if there is no relaxation gap. The controller can be constructed with the procedure presented in Section 4.5.6, by extracting the corresponding controller variables from x_{opt} in (4.32). We will see in the next section that a globally optimal controller can often be constructed with (4.32), even if the total degree of the sum-of squares polynomials is only 2.

4.7 Application

We computed lower bounds for randomly generated 4th order plants with $n = 4, p_1 = 3, m_1 = 2, p_2 = 1, m_2 = 1$ and computed upper bounds by gridding, as shown in Table 4.2. The table shows lower bounds on the optimal static-output feedback $\mathcal{H}_\infty$, that were computed by SOS relaxations. The bounds for the $\mathcal{H}_2$-norm have been computed with a similar approach, see also [90]. To keep the size of the LMI problems small, we used the very simple SOS bases $u(x) = \begin{pmatrix} 1 & K \end{pmatrix}$ and $u_0(x) = \begin{pmatrix} 1 & \gamma & K & \mathrm{svec}(X) & K\mathrm{svec}(X) \end{pmatrix}^T$, where $K \in \mathbb{R}$ is a static controller and $X \in \mathcal{S}^4$. The number of decision variables in the LMI is 469. Table 4.2 reveals that the lower bound on the globally optimal $\mathcal{H}_2$ performance is for all systems larger than the trivial lower bound of full order performance and in 6 out of 11 cases it is equal to the upper bound, such that the relaxation gap is zero. It is not yet clear whether there is a theoretical explanation for the lower bounds being no worse than the full order performance for this choice of bases. The lower bounds on the globally optimal $\mathcal{H}_\infty$ performance is only 7 out of 13 times better than the full order performance.

Extracting controllers

We construct globally optimal controllers with the method presented in Section 4.5.6. We choose a monomial basis vector $u_0(x)$ such that its first $m + 1$ columns are given by

$$\begin{pmatrix} I_{m+1} & 0 \end{pmatrix} u_0(x) = \begin{pmatrix} 1 \\ x \end{pmatrix},$$

Table 4.2: Upper bounds based on gridding and lower bounds for static output feedback together with Full Order $\mathcal{H}_2$ and $\mathcal{H}_\infty$ performance for randomly generated 4th order systems

$\mathcal{H}_2$			$\mathcal{H}_\infty$		
Upper bound	Lower bound	Full Order performance	Upper bound	Lower bound	Full Order performance
6.74	6.67	4.6	9.9	7.71	8.14
11.1	11.1	9.71	15	14.6	12.5
14.4	14.4	11.5	15.9	15.5	13.3
8.78	8.49	7.36	30.7	15.1	30.7
7.95	7.94	2.69	5.61	2.83	4.6
8.88	8.88	7.68	21.2	19.4	21.2
16.1	16.1	15.4	55.9	42.5	29.5
10.6	10.6	8.1	22.6	12	22.6
5.02	5.02	2.14	7.35	6.37	3.98
44.1	44.1	42.7	199	166	155
12	12	9.6	36.1	35.4	35.4
3.07	3.07	1.25	2.65	2.5	1.94
5.8	5.8	4.35	8.84	7.58	8.13

where
$$x = \begin{pmatrix} 1 & \gamma & K & \text{svec}(X) \end{pmatrix}^T.$$

The kernel of W_0 can not be computed exactly due to numerical errors. Instead we decompose it as $W_0 = Y \Sigma Y^T$, with orthogonal Y and diagonal $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_{\xi_0}) \succeq 0$ and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\xi_0}$. Let $\{e_1, \ldots, e_{\xi_0}\}$ be a basis for $\mathbb{R}^{\xi_0}$. Then we determine a candidate global optimal solution by (4.32), where

$$\begin{pmatrix} 1 \\ y \end{pmatrix} = \frac{1}{e_1 Y e_{\xi_0}} Y e_{\xi_0}.$$

The corresponding candidate controller is extracted as the second element in the vector x_{opt}. Tables 4.3 and 4.4 show the results for the $\mathcal{H}_2$- and $\mathcal{H}_\infty$-norm optimal control problems respectively. The first column shows the lower bound computed by the SOS relaxation, as was also shown in Table 4.2. The second and third columns show the best controllers and the corresponding closed-loop norms found by gridding, respectively. The last two columns of the tables show the extracted controller and the corresponding closed-loop norms, respectively. The entries (in Table 4.4) with value ∞ correspond to cases where the controller is not stabilizing. From Table 4.3 we observe that by this procedure we were able to extract an almost-optimal controller in 11 out of 13 times. The constructed $\mathcal{H}_\infty$-controllers are much less often close to optimal, probably because the relaxation gaps are much more often large.

4.8   Conclusion

In this chapter we have made a first step towards the computation of global optimality certificates for structured controller synthesis. This gives a partial answer to research question ② in Section 2.8. In addition, we have developed an SOS-based relaxation scheme

Table 4.3: Lower bounds and upper bounds for static output feedback based on gridding and controller extraction for the $\mathcal{H}_2$ performance for randomly generated 4$^{\text{th}}$ order system

Lower bound	Upper bounds			
	Gridding		Controller extraction	
$\mathcal{H}_2$-norm	Controller	$\mathcal{H}_2$-norm	Controller	$\mathcal{H}_2$-norm
6.67	−1.83	6.74	−39.3	14.7
11.1	0.406	11.1	0.392	11.1
14.4	−0.276	14.4	−0.271	14.4
8.49	−4.21	8.78	−17.4	8.89
7.94	−2.6	7.95	−10.6	11.7
8.88	−0.307	8.88	−0.315	8.88
16.1	0.096	16.1	0.0997	16.1
10.6	−0.834	10.6	−0.825	10.6
5.02	−3.84	5.02	−3.84	5.02
44.1	0.902	44.1	0.888	44.1
12	−0.276	12	−0.264	12
3.07	−1.73	3.07	−1.75	3.07
5.8	−3.56	5.8	−3.58	5.8

Table 4.4: Lower bounds and upper bounds for static output feedback based on gridding and controller extraction for the $\mathcal{H}_\infty$ performance for randomly generated 4$^{\text{th}}$ order system

Lower bound	Upper bounds			
	Gridding		Controller extraction	
$\mathcal{H}_\infty$-norm	Controller	$\mathcal{H}_\infty$-norm	Controller	$\mathcal{H}_\infty$-norm
7.71	−1.67	9.9	−3.06	15
14.6	0.995	15	1.1	15.2
15.5	0.344	15.9	1.3	18.6
15.1	0.902	30.7	0.992	31.4
2.83	−2.26	5.61	−2.23	5.62
19.4	0.282	21.2	−0.629	∞
42.5	0.127	55.9	0.52	∞
12	0.003	22.6	−2.16	∞
6.37	−2.48	7.35	−3.24	7.75
166	1.12	199	0.923	232
35.4	−0.307	36.1	−0.618	42
2.5	−1.61	2.65	−2.06	5.93
7.58	0.003	8.84	0.931	22.6

for polynomial SDPs, which are applicable to a wide range of very hard problems in control engineering.

For small-order systems the SOS relaxations are surprisingly good for very simple monomial bases, in the sense that they are often close to the optimal fixed-order performance and significantly larger than the full order performance. Especially for the $\mathcal{H}_2$-optimal control problem, verifiable globally optimal controllers have been computed with only small monomial basis vectors.

In our numerical experience the size of the LMI problems of the SOS relaxations grows fast, such that good lower bounds can be only computed for systems with state dimension up to about 4. This is to be expected, since the relaxation scheme is applicable to the whole class of polynomial SDPs, which contain a wide range of very hard problems. To reduce the computational complexity it is crucial to *exploit the control-theoretic characteristics* of the problem. We reveal in the next chapter that this can be done by avoiding the need of constructing SOS polynomials in the Lyapunov variable X.

Chapter 5

$\mathcal{H}_\infty$-synthesis by robust analysis

The LMI relaxations described in the previous chapter have a number of variables that grows fast with the number of plant states. As a consequence, good lower bounds on the structured controller performance can only be computed for systems with small McMillan degree up to about 4. In this chapter we present a two-stage dualization technique which converts the fixed order problem into a robust analysis problem, which can then be relaxed with Sum-Of-Squares decompositions. The resulting SOS relaxations grow only *quadratically* in the number of states, which is the same growth rate as for the LMI solution of the full-order problem.

In Section 5.1 we convert the structured synthesis problem into a robust analysis problem, where the controller variables are the 'uncertain' parameters. Relaxing the robust analysis problem by Sum-Of-Squares as described in Section 5.2 leads to a family of LMI problems. The optimal values of these problems are global lower bounds on the optimal closed-loop $\mathcal{H}_\infty$ performance which converge to this performance value for increasing monomial bases.

The conversion to a robust analysis problem is of independent interest: it allows us to apply the wide spectrum of robust analysis techniques to the fixed order controller design problem. In particular using a family of LMI relaxations based on the S-procedure we can compute global lower bounds on the optimal closed loop $\mathcal{H}_\infty$ performance. This will be explained in Section 5.3. Again it is guaranteed that optimal values of this family converge from below to the globally optimal fixed order $\mathcal{H}_\infty$-norm. We apply the method in Section 5.5 to a fourth-order example and to the fixed-order $\mathcal{H}_\infty$-optimal controller design problem of an active suspension system with a 27th-order plant. The results presented in this chapter have been published in [89], [88] and [91].

5.1 Conversion to robustness analysis

For the closed-loop matrices described by (2.8) our starting point is the same BMI formulation of the $\mathcal{H}_\infty$ structured controller synthesis problem as in the previous chapter. Consider a generalized plant with state-space realization

$$\left(\frac{\dot{x}}{\begin{array}{c} z \\ y \end{array}} \right) = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array} \right) \left(\frac{x}{\begin{array}{c} w \\ u \end{array}} \right), \qquad x(0) = 0, \tag{5.1}$$

with $A \in \mathbb{R}^{n+n_c \times n+n_c}$, $B_i \in \mathbb{R}^{n+n_c \times m_i}$, $C \in \mathbb{R}^{p_i \times n+n_c}$ and $D_{ij} \in \mathbb{R}^{p_i \times m_i}$, $i, j \in \{1, 2\}$. With static feedback matrix $K \in \mathbb{R}^{(n_c+m_2) \times (n_c+p_2)}$, the closed loop matrices are given by

$$\left(\begin{array}{c|c} A_{\mathrm{cl}}(K) & B_{\mathrm{cl}}(K) \\ \hline C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) \end{array} \right) = \left(\begin{array}{c|c} A & B_1 \\ \hline C_1 & D_{11} \end{array} \right) + \left(\begin{array}{c} B_2 \\ \hline D_{21} \end{array} \right) K \left(\begin{array}{c|c} C_2 & D_{12} \end{array} \right). \tag{5.2}$$

The structured $\mathcal{H}_\infty$-optimal control problem can then be formulated as

$$\inf_{A_{\mathrm{cl}}(K)\text{is Hurwitz}} \left\| D_{\mathrm{cl}}(K) + C_{\mathrm{cl}}(K)(sI - A_{\mathrm{cl}}(K))^{-1} B_{\mathrm{cl}}(K) \right\|_\infty \tag{5.3}$$

As explained in Chapter 2, this problem can be formulated in terms of BMIs using the 'Bounded Real Lemma inequality'

$$\mathcal{B}(\gamma, X, K) := - \left(\begin{array}{ccc} A_{\mathrm{cl}}(K)^T X + X A_{\mathrm{cl}}(K) & X B_{\mathrm{cl}}(K) & C_{\mathrm{cl}}(K)^T \\ B_{\mathrm{cl}}(K)^T X & -\gamma I & D_{\mathrm{cl}}(K)^T \\ C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) & -\gamma \end{array} \right) \succ 0. \tag{5.4}$$

As explained later in this chapter, we can reduce the complexity of our relaxations by taking first a Schur complement on the Bounded-Real Lemma inequality $\mathcal{B}(\gamma, X, K) \succ 0$. Indeed if using Lemma 2.7 we formulate (5.4) equivalently as

$$\mathcal{B}_{\mathrm{S}}(t, X, K) \succ 0, \tag{5.5}$$

where

$$\mathcal{B}_{\mathrm{S}}(t, X, K) := \left(\begin{array}{cc} -A_{\mathrm{cl}}(K)^T X - X A_{\mathrm{cl}}(K) & -X B_{\mathrm{cl}}(K) \\ -B_{\mathrm{cl}}(K)^T X & tI \end{array} \right) - \left(\begin{array}{c} C_{\mathrm{cl}}(K)^T \\ D_{\mathrm{cl}}(K)^T \end{array} \right) \left(\begin{array}{cc} C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) \end{array} \right)$$

where $t = \gamma^2$ and the closed loop matrices $(A_{\mathrm{cl}}(K), B_{\mathrm{cl}}(K), C_{\mathrm{cl}}(K), D_{\mathrm{cl}}(K))$ are defined by (5.2). Hence (5.3) is equivalent to

$$\begin{array}{ll} \text{infimize} & \gamma \\ \text{subject to} & X \succ 0 \text{ and } \mathcal{B}_{\mathrm{S}}(t, X, K) \succ 0. \end{array} \tag{5.6}$$

For technical reasons that will be clarified in the sequel we need the additional assumption that $(A_{\mathrm{cl}}(K), B_{\mathrm{cl}}(K))$ is controllable for every $K \in \mathcal{K}$. Furthermore we restrict the controller variables, as was also assumed in the previous chapter, to a compact set $\mathcal{K}$. Compactness can, for instance, be realized by restricting the controller variables to a Euclidean ball of radius $\rho > 0$

$$\mathcal{K} := \left\{ K \in \mathbb{R}^{(n_c+m_2) \times (n_c+p_2)} \mid \|K\| \leq \rho \right\}.$$

Some additional remarks on these assumptions will be given later on in this chapter. The BMI problem to be solved can then be formulated as

$$\begin{array}{ll} \text{infimize} & t \\ \text{subject to} & X \succ 0, \ \mathcal{B}_{\mathrm{S}}(t, X, K) \succ 0, \ K \in \mathcal{K} \text{ and } t > 0 \end{array} \tag{5.7}$$

with optimal value p_{opt}. We added the obviously redundant constraint $t > 0$ for reasons to be explained in the next section.

For translating the nonlinear synthesis problem in (5.7) into an equivalent robustness analysis problem, the key idea is to apply *partial* Lagrange dualization: fix the controller variables K and dualize with respect to the Lyapunov variable X. We will show that one

is required to determine parameter-dependent dual variables, in full analogy to computing parameter-dependent Lyapunov functions for LPV systems. As the main advantage, this reformulation allows us to suggest novel SOS relaxations that grow only *quadratically* in the number of the system states, in stark contrast to the relaxations of the previous chapter, which were based on Lagrange dualization in all decision variables of (5.7). The number of monomials and hence the number of LMI variables of the relaxations of the previous chapter grow according to the binomial formula (4.30).

5.1.1   Partial dualization

For fixed $K = K_0 \in \mathcal{K}$, (5.7) is an LMI problem in X and t:

$$
\begin{aligned}
\text{infimize} \quad & t \\
\text{subject to} \quad & X \succ 0, \qquad \mathcal{B}_S(t, X, K_0) \succ 0, \qquad t > 0,
\end{aligned} \tag{5.8}
$$

Let us partition the dual variable Z for the constraint $\mathcal{B}_S(t, X, K) \succ 0$ in (5.7) as

$$
Z = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{12}^T & Z_{22} \end{pmatrix} \in \mathcal{S}^{n+n_c+m_1}, \text{ with } Z_{11} \in \mathcal{S}^{n+n_c}. \tag{5.9}
$$

Then the Langrange dual reads as follows:

$$
\begin{aligned}
\text{supremize} \quad & \text{Trace}\left(\begin{bmatrix} C_{cl}(K_0) & D_{cl}(K_0) \end{bmatrix} Z \begin{bmatrix} C_{cl}(K_0) & D_{cl}(K_0) \end{bmatrix}^T \right) \\
\text{subject to} \quad & A_{cl}(K_0)Z_{11} + Z_{11}A_{cl}(K_0)^T + B_{cl}(K_0)Z_{12}^T + Z_{12}B_{cl}(K_0)^T \succeq 0 \\
& \text{Trace}(Z_{22}) \leq 1, \qquad Z \succeq 0,
\end{aligned} \tag{5.10}
$$

see Appendix A.3.1 for a derivation. The reason for the redundant constraint $t > 0$ in the primal problem is to arrive at a dual problem without inequality constraints.

Let $d_{opt}(K_0)$ denote the dual optimal value of (5.10). Note that (5.10) is strictly feasible for all $K_0 \in \mathcal{K}$ as is shown in Appendix A.3.2. This implies (due to strong duality) $d_{opt}(K_0) = p_{opt}(K_0)$ and, as a consequence, we draw the following conclusion. Given any $d \in \mathbb{R}$ suppose that the matrix-valued function $Z(K)$ satisfies

$$
\text{Trace}\left(\begin{bmatrix} C_{cl}(K) & D_{cl}(K) \end{bmatrix} Z(K) \begin{bmatrix} C_{cl}(K) & D_{cl}(K) \end{bmatrix}^T \right) \; > \; d, \quad (5.11)
$$

$$
A_{cl}(K)Z_{11}(K) + Z_{11}(K)A_{cl}(K)^T + B_{cl}(K)Z_{12}(K)^T + Z_{12}(K)B_{cl}^T(K) \; \succ \; 0, \quad (5.12)
$$

$$
\text{Trace}(Z_{22}(K)) < 1, \qquad Z(K) \; \succ \; 0, \quad (5.13)
$$

for all $K \in \mathcal{K}$. Then it is clear that $d_{opt}(K) \geq d$ and hence $p_{opt}(K) \geq d$ hold for all $K \in \mathcal{K}$. Therefore d is a lower bound on the best achievable controller performance. It is thus natural to maximize d over some class of functions $Z(.)$ in order to determine tight lower bounds on the value of (5.7). Our construction allows us to show that this lower bound is actually tight if optimizing over symmetric matrix-valued polynomials $Z(K)$.

Theorem 5.1 *Let p_{opt} be the optimal solution of (5.7) and d_{opt} be the supremal d for which there exists a polynomial matrix $Z(K) \in \mathcal{S}^{n+n_c+m_1}$ satisfying (5.11)-(5.13) for all $K \in \mathcal{K}$. Then $p_{opt} = d_{opt}$.*

Proof. We have already seen that $p_{\text{opt}} \geq d_{\text{opt}}$. Now suppose $p_{\text{opt}} \geq d_{\text{opt}} + \epsilon$ for some $\epsilon > 0$. For any fixed $K_0 \in \mathcal{K}$, the optimal value of (5.8) and hence that of (5.10) is not smaller than p_{opt}. Since (5.10) is strictly feasible there exists Y^0 (partitioned as (5.9)) with

$$\text{Trace}\left(\begin{bmatrix} C_{\text{cl}}(K_0) & D_{\text{cl}}(K_0) \end{bmatrix} Y^0 \begin{bmatrix} C_{\text{cl}}(K_0) & D_{\text{cl}}(K_0) \end{bmatrix}^T \right) > p_{\text{opt}} - \frac{\epsilon}{2},$$

$$A_{\text{cl}}(K_0)Y_{11}^0 + Y_{11}^0 A_{\text{cl}}(K_0)^T + B_{\text{cl}}(K_0)Y_{12}^{0\,T} + Y_{12}^0 B^T(K_0) \succ 0,$$

$$\text{Trace}(Y_{22}^0) < 1, \qquad Y^0 \succ 0.$$

Since the inequalities are strict and $\mathcal{K}$ is compact, we can use a partition of unity argument [160] to show that there actually exists a *continuous* function $Y(K)$ such that

$$\text{Trace}\left(\begin{bmatrix} C_{\text{cl}}(K) & D_{\text{cl}}(K) \end{bmatrix} Y(K) \begin{bmatrix} C_{\text{cl}}(K) & D_{\text{cl}}(K) \end{bmatrix}^T \right) > p_{\text{opt}} - \frac{\epsilon}{2}, \quad (5.14)$$

$$A_{\text{cl}}(K)Y_{11}(K) + Y_{11}(K)A_{\text{cl}}(K)^T + B_{\text{cl}}(K)Y_{12}(K)^T + Y_{12}(K)B_{\text{cl}}^T(K) \succ 0, \quad (5.15)$$

$$\text{Trace}(Y_{22}(K)) < 1, \qquad Y(K) \succ 0, \quad (5.16)$$

for all $K \in \mathcal{K}$. Due to the *Weierstrass' Approximation Theorem* of continuous functions by polynomials [116] on compacta, we can even choose $Y(K)$ to be a matrix polynomial. This allows us to conclude $d_{\text{opt}} \geq p_{\text{opt}} - \frac{\epsilon}{2}$, a contradiction which finishes the proof. ∎

In actual computations we optimize over functions $Z(\cdot)$ belonging to an increasing sequence of finite-dimensional subspaces of matrix-valued polynomials. Then the difference of the computed lower bound with the actual optimal $\mathcal{H}_\infty$-performance is non-decreasing. If we restrict the search to a subspace of degree-bounded matrix polynomials, and if we let the bound on the degree grow to infinity, Theorem 5.1 guarantees that the corresponding lower bounds converge from below to the globally optimal $\mathcal{H}_\infty$ performance.

We have thus reduced the $\mathcal{H}_\infty$-synthesis problem to a robust analysis problem with complicating variables K and polynomial robustness certificates $Z(K)$. In Section 5.2 we will discuss how (5.11)-(5.13) can be relaxed to standard LMI constraints via suitable SOS tests and in Section 5.3 we will discuss relaxations based on the S-procedure.

Remark. The robustness certificates provide global lower bounds and are not constructive, i.e. they do not yield a globally optimal controller. However, in Section 5.3 we will show that under certain conditions the relaxations do make it possible to construct globally optimal controllers.

Remark. The proposed partial dualization technique is not at all restricted to fixed-order $\mathcal{H}_\infty$-optimal control. Straightforward variations do apply to a whole variety of other interesting problems, such as designing structured controllers for any performance criterion that admits an LMI representation of its analysis problem (as e.g. general quadratic performance, $\mathcal{H}_2$-performance, multi-objective control [212, 167, 173] or placement of closed-loop poles in LMI regions [40]).

Remark. We require the controller parameters to lie in a compact set in order to be able to apply Weierstrass' Approximation Theorem. From a practical point of view this is actually not restrictive, since the controller parameters cannot be too large for digital

implementation. Moreover one can exploit the flexibility in choosing the set $\mathcal{K}$ in order to incorporate the suggested lower bound computations in branch-and-bound techniques, see Section 3.3.

Remark. The controllability assumption is needed to prove that the dual (5.10) is strictly feasible for all $K \in \mathcal{K}$. Controllability can be verified by a Hautus test: $(A_{\mathrm{cl}}(K), B_{\mathrm{cl}}(K))$ is controllable for all $K \in \mathcal{K}$ if and only if

$$P_{\mathrm{H}}(\lambda, K) := \left(\begin{array}{cc} A_{\mathrm{cl}}(K) - \lambda I & B_{\mathrm{cl}}(K) \end{array} \right) \text{ has full row rank for all } \lambda \in \mathbb{C}, \; K \in \mathcal{K}. \quad (5.17)$$

This property can be verified by SOS decompositions as described in Chapter 4. Indeed suppose $M > 0$ is chosen with $\|A_{\mathrm{cl}}(K)\| \leq M$ for all $K \in \mathcal{K}$ (with $\|\cdot\|$ denoting the spectral norm). Then (5.17) holds true if and only if the real-valued polynomial

$$\begin{aligned} F_{\mathrm{H}}(a, b, K) : \;\; &= \;\; |\det \left(P_{\mathrm{H}}(a + bi, K) P_{\mathrm{H}}(a + bi, K)^* \right)|^2 \\ &= \det \left(P_{\mathrm{H}}(a + bi, K) P_{\mathrm{H}}(a + bi, K)^* \right)^* \det \left(P_{\mathrm{H}}(a + bi, K) P_{\mathrm{H}}(a + bi, K)^* \right) \end{aligned}$$

is strictly positive on $[-M, M] \times [-M, M] \times \mathcal{K}$. This can be tested with SOS decompositions, provided that $\mathcal{K}$ has a representation that satisfies the constraint qualification (4.18). The upper bound M on the spectral norm of A on $\mathcal{K}$ can also be computed with SOS techniques.

Remark. The utmost right constraint in (5.13) ($Z(K) \succ 0$ for all $K \in \mathcal{K}$) can be replaced by the (generally much) stronger condition that $Z(K)$ is SOS. As we will see in Section 5.2 this may reduce the complexity of our relaxation problems. Theorem 5.1 is still true after the replacement, since for any matrix-valued polynomial $Z(K)$ that satisfies (5.11)-(5.13) for all $K \in \mathcal{K}$, we can find a unique matrix valued function $R(K)$ on $\mathcal{K}$ that is the Cholesky factor of $Z(K)$ for all $K \in \mathcal{K}$. Furthermore $R(K)$ is continuous on $\mathcal{K}$ if $Z(K)$ is, because the Cholesky factor of a matrix can be computed by a sequence of continuity preserving operations on the coefficients of the matrix [72]. Again by Weierstrass' Approximation Theorem there exists an approximation of the continuous $R(K)$ on $\mathcal{K}$ by a polynomial $\tilde{R}(K)$ such $Z(K) := \tilde{R}(K)^T \tilde{R}(K)$ satisfies (5.11)-(5.13). The constructed matrix-valued polynomial $Z(K)$ is then indeed SOS.

5.1.2 Finite-dimensional approximation

Suppose that $Z_j : \mathbb{R}^{(n_{\mathrm{c}}+m_2) \times (n_{\mathrm{c}}+p_2)} \mapsto \mathcal{S}^{n+m_1}$, $j = 1, 2, \ldots, N_{\mathrm{par}}$, is a set of linearly independent symmetric-valued polynomial functions in K (such as a basis for the real vector space of all symmetric matrix polynomials of a certain maximal total degree). Let us now restrict the search of $Z(.)$ in Theorem 5.1 to the subspace

$$\mathcal{Z}_{N_{\mathrm{par}}} := \left\{ Z(., z) \mid Z(., z) := \sum_{j=1}^{N_{\mathrm{par}}} z_j Z_j(.), \; z = (z_1, \ldots, z_{N_{\mathrm{par}}}) \in \mathbb{R}^{N_{\mathrm{par}}} \right\}. \quad (5.18)$$

Then (5.11)-(5.13) are polynomial inequalities in K that are affine in the coefficients z for the indicated parametrization of the elements $Z(., z)$ in $\mathcal{Z}_{N_{\mathrm{par}}}$. With $y := \mathrm{col}(d, z)$, $c := \mathrm{col}(1, 0_{n_z})$ and

$$F(K, y) := \mathrm{diag}(F_{11}(K, z) - d, F_{22}(K, z), Z(K, z), 1 - \mathrm{Trace}(Z_{22}(K, z))), \quad (5.19)$$

where

$$F_{11}(K, z) := \text{Trace} \left(\begin{bmatrix} C_{\text{cl}}(K) & D_{\text{cl}}(K) \end{bmatrix} Z(K, z) \begin{bmatrix} C_{\text{cl}}(K) & D_{\text{cl}}(K) \end{bmatrix}^T \right) \tag{5.20}$$

and

$$F_{22}(K, z) := Z_{11}(K, z)A_{\text{cl}}(K)^T + A_{\text{cl}}(K)Z_{11}(K, z) + Z_{12}(K, z)B_{\text{cl}}(K)^T + B_{\text{cl}}(K)Z_{12}^T(K, z)^T, \tag{5.21}$$

the problem to be solved can be compactly written as follows:

$$\begin{array}{ll} \text{supremize} & c^T y \\ \text{subject to} & F(K, y) \succ 0 \text{ for all } K \in \mathcal{K}. \end{array} \tag{5.22}$$

This problem involves a semi-infinite semi-definite constraint on a matrix polynomial in K, *only*. This allows us to construct relaxations that rely on SOS-decomposition for polynomials in K. This is the key to keep the size of the resulting LMI-problem *quadratic* in the number of system states, as opposed to the large growth of the size of the LMI-problem for the direct approach discussed in Chapter 4.

Remark. Additional structure in the elements of the feedback matrix K can be taken into account as follows. Suppose that the controller variables can be parameterized with an affine function $K(p)$ as in (2.56), $K : \mathbb{R}^{l_c} \to \mathbb{R}^{(n_c+m_2)\times(n_c+p_2)}$ for some $l_c \leq (n_c+m_2)(n_c+p_2)$. Then it is easy to see that partial dualization in p transforms the structured $\mathcal{H}_\infty$-optimal control problem into a problem of type (5.22). This is explained in full detail in our publication [89].

5.2 Relaxations based on Sum-Of-Squares

In this section we present SOS-relaxations for the robust analysis problem (5.22) with compact $\mathcal{K}$ described as $\mathcal{K} := \{K \in \mathbb{R}^{(n_c+m_2)\times(n_c+p_2)} | G(K) \preceq 0\}$ for some matrix-valued polynomial $G : \mathbb{R}^{(n_c+m_2)\times(n_c+p_2)} \to \mathcal{S}^r$. Note that this description of $\mathcal{K}$ allows us to represent any set described by non-strict inequalities which include polytopes, norm-bounded sets or any combination thereof.

Since SOS-relaxations for robust analysis problems are addressed in detail in Chapter 6, we briefly sketch in this section the ideas. We replace the constraints in (5.22) by requiring the existence of $\epsilon > 0$ and SOS matrices $S_0(K)$ and $S(K)$, $S_0 : \mathbb{R}^{(n_c+m_2)\times(n_c+p_2)} \to \mathcal{S}^q$, $S : \mathbb{R}^{(n_c+m_2)\times(n_c+p_2)} \to \mathcal{S}^{qr}$ such that

$$F(K, y) + \langle S(K)G(K) \rangle_q - \epsilon I = S_0(K), \tag{5.23}$$

where q is the number of rows/columns of $F(K, y)$. Using similar weak duality arguments as in Section 4.2 it is easy to see that the optimal value of (5.22) is bounded from below by the largest achievable $c^T y$ for which there exist $\epsilon > 0$ and a SOS matrix $S(K)$ with (5.23). In Chapter 6 we will show that this relaxation is exact under the constraint qualification (4.18). This is the key step to see that one can construct a family of LMI relaxations for computing lower bounds arbitrarily close to the optimal $\mathcal{H}_\infty$-norm, of which their sizes grow quadratically in the state dimension n. The construction of the LMI relaxations is similar to the one for matrix valued polynomial SDPs and will be described in more detail in Section 6.3.

Remark. We have performed partial dualization in the high dimensional variables X in order to arrive at the formulation of (5.22). It is interesting to interpret the replacement of the semi-infinite constraint in (5.22) by (5.23) as the result of a second Lagrange relaxation step in the low dimensional variable K. In this sense the suggested relaxation can be viewed as a full SOS Lagrange dualization of the original nonlinear semi-definite program, and splitting into two steps yields the desired quadratic growth (of the size of the LMIs) in the McMillan degree n.

Remark. In Section 2.2.2 it has been discussed how to transform a generalized plant with nonzero D_{22} to a plant with zero D_{22}, using (2.7). This mapping results in a rational dependence of the left-hand sides of (5.11) and (5.13) on K, such that we cannot directly apply the SOS test. Under the well-posedness condition that $I - D_k(K)D_{22}$ is nonsingular for all $K \in \mathcal{K}$, we can multiply (5.11) and (5.13) by $\det(I - D_k(K)D_{22})^2$. In this case $\det(I - D_k(K)D_{22})^2 > 0$ for all $K \in \mathcal{K}$. Hence the resulting inequalities for all $K \in \mathcal{K}$

$$(\det(I - D_{22}D_k(K))I)^T(Z_{11}(K)A(K)^T + A_{\mathrm{cl}}(K)Z_{11}(K) + Z_{12}(K)B_{\mathrm{cl}}^T(K)+$$
$$+ B_{\mathrm{cl}}(K)Z_{12}(K)^T)(\det(I - D_{22}D_k(K))I) \succ 0,$$

$$(\det(I - D_{22}D_k(K)))^2\mathrm{Trace}(\begin{bmatrix} C(K) & D(K) \end{bmatrix} Z(K) \begin{bmatrix} C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) \end{bmatrix}^T) > d$$

and

$$Z(K) \succeq 0$$

are equivalent to (5.11) and (5.13) and polynomial in K, such that our problem is a robust LMI problem and we can apply the SOS technique. It is easy to see that testing well-posedness is a robust LMI problem as well. In a similar fashion we can introduce a rational parametrization for $Z(K)$. Let $Z_j : \mathbb{R}^{(n_{\mathrm{c}}+m_2)\times(n_{\mathrm{c}}+p_2)} \mapsto \mathcal{S}^{n+n_{\mathrm{c}}+m_1}, j = 1, 2, \ldots, N_{\mathrm{par}}$, be a set of linearly independent symmetric valued rational (instead of polynomial as in Section 5.1.2) functions in K without poles in $\mathcal{K}$. By multiplication of the inequalities (5.11)-(5.13) with the smallest common denominator of $Z(K)$ that is strictly positive for all $K \in \mathcal{K}$, their left-hand sides become polynomials in K. See [106] for similar ideas to transform a positivity test of a rational function to a positivity test on a polynomial.

5.3   Relaxations based on the S-procedure

We present in this section relaxations based on the S-procedure if $\mathcal{K}$ is the polytope

$$\mathrm{co}\{K_1, K_2, \ldots, K_{N_{\mathrm{poly}}}\}.$$

We use a parametrization of $Z(K)$ as a linear combination of fixed rational functions in K, and we apply the full block S-procedure as described in Section 2.5.7 to solve the robust analysis problem. However, we do not need the S-procedure to tackle this particular problem; we can use the SOS relaxations of the previous section or any other robust analysis test.

Let $Z_j : \mathbb{R}^{(n_{\mathrm{c}}+m_2)\times(n_{\mathrm{c}}+p_2)} \mapsto \mathcal{S}^{n+n_{\mathrm{c}}+m_1}, j \in \{1, 2, \ldots, N_{\mathrm{par}}\}$, be independent matrix-valued rational functions in K without poles in $\mathcal{K}$. Consider the vector space

$$\mathcal{Z}_{N_{\mathrm{par}}} := \left\{ \sum_{j=1}^{N_{\mathrm{par}}} z_j Z_j(K) \,\middle|\, z \in \mathbb{R}^{N_{\mathrm{par}}} \right\}, \tag{5.24}$$

We search over $Z \in \mathcal{Z}_{N_{\text{par}}}$. With a slight abuse of notation we use $Z(K, z)$ to represent the function $Z(K, z) = \sum_{j=1}^{N_{\text{par}}} z_j Z_j(K)$ parameterized by $z \in \mathbb{R}^{N_{\text{par}}}$.

Full block S-procedure

If we use the parametrization (5.24) for $Z(K, z)$, we observe that the inequalities (5.11), (5.12) and (5.13) are rational inequalities in K depending affinely on the coefficients z in the parametrization of $Z(K, z)$. We construct an LFT

$$
\begin{aligned}
F_{\text{LFT}}(z, K) &= \mathcal{F}_{\text{u}} \left(\Delta(K), \underbrace{\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C}(z) & \mathbf{D}(z) \end{pmatrix}}_{H(z)} \right) \\
&= \mathbf{D}(z) + \mathbf{C}(z)\Delta(K)(I - \mathbf{A}\Delta(K))^{-1}\mathbf{B},
\end{aligned}
$$

where $\mathcal{F}_{\text{u}}$ is the upper LFT as explained in Section 2.1.2, $\Delta(K)$ is linear in K and $I - \mathbf{A}\Delta(K)$ is nonsingular for all $K \in \mathcal{K}$, such that (5.11), (5.12) and (5.13) for all $K \in \mathcal{K}$ are equivalent to

$$
\begin{pmatrix} I \\ F_{\text{LFT}}(z, K) \end{pmatrix}^T P_d \begin{pmatrix} I \\ F_{\text{LFT}}(z, K) \end{pmatrix} \prec 0 \text{ for all } K \in \mathcal{K}, \tag{5.25}
$$

where P_d is a fixed matrix. The construction of such an LFT and a corresponding matrix P_d is presented at the end of this section. Since $\mathcal{K}$ is compact we can apply the full-block S-procedure as described in Section 2.5.7, to infer that our problem is equivalent to maximizing d over $z \in \mathbb{R}^{N_{\text{par}}}$ and multipliers P_{mult} with

$$
\begin{pmatrix} \Delta(K) \\ I \end{pmatrix}^T P_{\text{mult}} \begin{pmatrix} \Delta(K) \\ I \end{pmatrix} \succ 0 \text{ for all } K \in \mathcal{K}, \tag{5.26}
$$

and

$$
\begin{pmatrix} I & 0 \\ \mathbf{A} & \mathbf{B} \end{pmatrix}^T P_{\text{mult}} \begin{pmatrix} I & 0 \\ \mathbf{A} & \mathbf{B} \end{pmatrix} + \begin{pmatrix} 0 & I \\ \mathbf{C}(z) & \mathbf{D}(z) \end{pmatrix}^T P_d \begin{pmatrix} 0 & I \\ \mathbf{C}(z) & \mathbf{D}(z) \end{pmatrix} \prec 0. \tag{5.27}
$$

Equation (5.26) is a semi-infinite constraint on the multiplier P_{mult}. To render the lower bound computationally tractable we introduce a (standard) inner approximation of the set of multipliers. The set $\mathbf{P}_{\text{mult}}$ of multipliers P_{mult} such that

$$
\begin{pmatrix} I \\ 0 \end{pmatrix}^T P_{\text{mult}} \begin{pmatrix} I \\ 0 \end{pmatrix} \prec 0, \tag{5.28}
$$

$$
\begin{pmatrix} \Delta(K_k) \\ I \end{pmatrix}^T P_{\text{mult}} \begin{pmatrix} \Delta(K_k) \\ I \end{pmatrix} \succ 0, \quad k = 1, \ldots, N_{\text{poly}} \tag{5.29}
$$

is an inner approximation for $\{P_{\text{mult}} | P_{\text{mult}} \text{ satisfies } (5.26)\}$, as can be shown by an elementary convexity argument [176]. Hence the optimal value of the LMI problem

$$
\begin{array}{ll}
\text{supremize} & d \\
\text{subject to} & (5.27), (5.28) \text{ and } (5.29)
\end{array}
$$

is a lower bound on the square of the optimal closed-loop $\mathcal{H}_\infty$-norm. If we optimize over many controller variables, the number of generators N_{poly} of $\mathcal{K}$ may be large. To avoid explosion of the size of the relaxed problem, we can exploit the block diagonal structure in $\Delta(K)$.

Remark. If $A_{\text{cl}}(K)$ is stable for all $K \in \mathcal{K}$, we can eliminate several parameters in the problem. Inequality (5.12) then reduces to

$$Z_{11}(K)A_{\text{cl}}(K)^T + A_{\text{cl}}(K)Z_{11}(K) + Z_{12}(K)B_{\text{cl}}(K)^T + B_{\text{cl}}(K)Z_{21}^T(K) = 0,$$

and $Z_{11}(K)$ can be explicitly described as a rational function in terms of $Z_{12}(K)$. This drastically reduces the number of free variables in the polynomial Z. The function $Z_{11}(K)$ constructed in this fashion has a rational dependence on the controller variables and affine dependence on $Z_{12}(K)$.

LFT construction

In this section we present an LFT representation to obtain relaxations based on the S-procedure of (5.11), (5.12) and (5.13). For this purpose let us define, similar to F in (5.19) the matrix-valued polynomial

$$F_{\text{LFT}}(K, z) := \begin{pmatrix} F_{\text{LFT},11}(K, z) - V & 0 & 0 & 0 \\ 0 & F_{\text{LFT},22}(K, z) & 0 & 0 \\ 0 & 0 & Z(K, z) & 0 \\ 0 & 0 & 0 & W - Z_{22}(K, z) \end{pmatrix}$$

where

$$F_{\text{LFT},11}(K, z) := Z(K, z) \begin{pmatrix} C_{\text{cl}}(K) & D_{\text{cl}}(K) \end{pmatrix}^T \begin{pmatrix} C_{\text{cl}}(K) & D_{\text{cl}}(K) \end{pmatrix}$$

and

$$F_{\text{LFT},22}(K, z) := Z_{11}(K, z)A_{\text{cl}}(K)^T + Z_{12}(K, z)B_{\text{cl}}(K)^T.$$

$V \in \mathcal{S}^{n+n_c+m_1}$ and $W \in \mathcal{S}^{m_1}$ are auxiliary variables. Let us furthermore choose P_d in (5.25) as

$$P_d := \left(\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & -I_{n+n_c+m_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -I_{n+n_c} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -I_{n+n_c+m_1} & 0 \\ 0 & 0 & 0 & 0 & d & 0 & 0 & -I_{m_1} \\ \hline -I_{n+n_c+m_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -I_{n+n_c} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -I_{n+n_c+m_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -I_{m_1} & 0 & 0 & 0 & 0 \end{array} \right).$$

$$(5.30)$$

It is easy to see that existence of V, W and z with (5.25), $\text{Trace}(V) > d$ and $\text{Trace}(W) < 1$ is equivalent to (5.11), (5.12) and (5.13) for all $K \in \mathcal{K}$. To keep the LMIs small, it is important to construct the LFT $\mathcal{F}_u(\Delta(K), H(z)) = F_{\text{LFT}}(K, z)$ with a small sized uncertainty block $\Delta(K)$. For this purpose we decompose F_{LFT} as

$$F_{\text{LFT}}(K, z) = L_0 + \sum_{i=1}^{4} L_i Z(K, z) R_i(K),$$

where $L_0 \in S^{n+n_c+m_1}$ is given by

$$L_0 := \begin{pmatrix} -V & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & W \end{pmatrix}$$

and $L_i \in \mathbb{R}^{(3(n+n_c+m_1))\times(n+n_c+m_1)}$, $i = 1, \ldots, 4$ are defined as

$$L_1 := \begin{pmatrix} I_{n+n_c+m_1} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad L_2 := \begin{pmatrix} 0 & 0 \\ I_{n+n_c} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad L_3 := \begin{pmatrix} 0 \\ 0 \\ I_{n+n_c+m_1} \\ 0 \end{pmatrix} \quad L_4 := \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & I_{m_1} \end{pmatrix}.$$

Finally, $R_i \in \mathbb{R}^{(n+n_c+m_1)\times(3(n+n_c+m_1))}$, $i = 1, \ldots, 4$ are chosen as follows:

$$R_1(K) := \left(\begin{pmatrix} C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) \end{pmatrix}^T \begin{pmatrix} C_{\mathrm{cl}}(K) & D_{\mathrm{cl}}(K) \end{pmatrix} \ 0 \ 0 \ 0 \right),$$

$$R_2(K) := \begin{pmatrix} 0 & A_{\mathrm{cl}}(K)^T & 0 & 0 \\ 0 & B_{\mathrm{cl}}(K)^T & 0 & 0 \end{pmatrix}, \qquad R_3 := \begin{pmatrix} 0 & 0 & I_{n+n_c+m_1} & 0 \end{pmatrix}$$

and

$$R_4 := \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{m_1} \end{pmatrix},$$

respectively. Note that the size of the zero blocks in the matrices given above are chosen to match the partitioning of the left-upper block of P_d in (5.30). The LFTs for $R_1(K)$ and $R_2(K)$ are

$$R_1(K) = \mathcal{F}_{\mathrm{u}} \left(\begin{pmatrix} K^T & 0 \\ 0 & K \end{pmatrix}, \begin{pmatrix} 0 & D_{12}^T D_{12} & D_{12}^T C_1 & D_{12}^T D_{11} & 0 & 0 & 0 \\ 0 & 0 & C_2 & D_{21} & 0 & 0 & 0 \\ C_2^T & C_1^T D_{12} & C_1^T C_1 & C_1^T D_{11} & 0 & 0 & 0 \\ D_{21}^T & D_{11}^T D_{12} & D_{11}^T C_1 & D_{11}^T D_{11} & 0 & 0 & 0 \end{pmatrix} \right)$$

and

$$R_2(K) = \mathcal{F}_{\mathrm{u}} \left(K^T, \begin{pmatrix} 0 & 0 & B_2^T & 0 & 0 \\ C_2^T & 0 & A^T & 0 & 0 \\ D_{21}^T & 0 & B_1^T & 0 & 0 \end{pmatrix} \right),$$

respectively.

Next, we construct the LFT for $Z(K)$. The size of its uncertainty block depends on the degree of numerator and denominator of $Z(K)$. Suppose, for instance, that Z is parameterized by $Z(K, z) = Z_D(z) + Z_C(z)K(I - Z_A K)^{-1}Z_B$. Then, we can construct an LFT

$$Z(K, z) = \mathcal{F}_{\mathrm{u}} \left(\begin{pmatrix} Z_A & Z_B \\ Z_C(z) & Z_D(z) \end{pmatrix}, K \right),$$

with an uncertainty block of size $(n_c + m_2) \times (n_c + p_2)$. We compose each term $L_i Z(K) R_i(K)$ by multiplication of the LFT $Z(K)$ with the LFT $R_i(K)$, which results in a (usually larger) new LFT [214].

Remark. The LFT construction procedure in [89] is different, since the original problem is formulated in terms of (5.4) instead of (5.5). This different formulation may, depending on the size of the data matrices, result in a smaller LFT.

5.4 Complexity, conservatism and exactness

5.4.1 Complexity for approach based on SOS relaxations

The number of variables and constraints in the LMI relaxations grows quadratically in the state dimension n. Indeed $F(K, y)$ in (5.19) and hence $S_0(K)$ in (5.23) have $q :=
2 + 2n + 2n_c + m_1$ rows and columns. If the monomial bases for $S_0(K)$ and $S(K)$ are ξ_0 and ξ, respectively, then the variables in the SOS relaxation are $y \in \mathbb{R}^{1+N_{par}}$, $W_0 \in
\mathcal{S}^{\xi_0(2+2n+2n_c+m_1)}$ and $W \in \mathcal{S}^{\xi(2+2n+2n_c+m_1)r}$, as will be explained in Chapter 6. If we count a symmetric matrix in $\mathcal{S}^l$ as $\frac{1}{2}l(l + 1)$ variables, we arrive at the total number of 'scalar variables'

$$
1 + N_{par} + \frac{1}{2}(\xi_0(2 + 2n + 2n_c + m_1) + 1)(\xi_0(2 + 2n + 2n_c + m_1)) +
$$
$$
+ (\xi(2 + 2n + 2n_c + m_1)r + 1)(\xi(2 + 2n + 2n_c + m_1)r). \quad (5.31)
$$

The constraints in the LMI problem are $W_0 \succeq 0$ and $W \succeq 0$ and equation constraints similar to (4.27).

5.4.2 Complexity for approach based on S-procedure

The variables in the relaxation based on the S-procedure are $P_{mult} \in \mathcal{S}^{r_{LFT}+c_{LFT}}$, $z \in \mathbb{R}^{N_{par}}$, $V \in \mathcal{S}^{n+n_c+m_1}$ and $W \in \mathcal{S}^{m_1}$, where r_{LFT} and c_{LFT} are the number of rows and columns in the uncertainty block $\Delta(K)$ respectively. The total number of 'scalar variables' is

$$
1 + N_{par} + \frac{1}{2}(r_{LFT} + c_{LFT} + 1)(r_{LFT} + c_{LFT}) + \frac{1}{2}(n + n_c + m_1 + 1)(n + n_c + m_1) + \frac{1}{2}(m_1 + 1)m_1
$$

The constraints are (5.27) and (5.28) and (5.29). Since $F_{LFT}(z, K)$ and $\mathbf{D}(z)$ are $(2 + 2n + m_1) \times (2 + 2n + m_1)$ matrices and the number of rows in $\mathbf{C}(z)$ is r_{LFT}, the constraint (5.27) is in $\mathcal{S}^{r_{LFT}+(2+2n+m_1)}$. The constraint (5.28) is in $\mathcal{S}^{r_{LFT}}$ and the N_{poly} constraints in (5.29) are all in $\mathcal{S}^{c_{LFT}}$.

It is clear that the number of LMI variables depends on the number N_{par} of basis elements of $\mathcal{Z}_{N_{par}}$ in (5.18) and (5.24). In addition, the choice of the monomial basis $u(x)$ of $S(x)$ affects the number of LMI variables and constraints. It is not a priori known for which choice of bases $\mathcal{Z}_{N_{par}}$ and $u(x)$, the suggested relaxations generate good lower bounds. Typically one expects that for 'large' sets $\mathcal{K}$ a large dimension of $\mathcal{Z}_{N_{par}}$ is needed to obtain a good lower bound.

5.4.3 Conservatism and exactness

There are two sources of conservatism in our approach. The first is the approximation of the matrix valued function $Z(K)$ by a finite-order rational matrix function and applies to the SOS-based method and to the approach based on the S-procedure as well. To prove its asymptotic exactness we can apply Weierstrass' Approximation Theorem again. It implies that the sequence of optimal values of

$$
\begin{aligned}
&\text{supremize} &&d \\
&\text{subject to} &&Z \in \mathcal{Z}_{N_{par}} \\
& &&(5.11),\ (5.12) \text{ and } (5.13) \text{ for all } K \in \mathcal{K}
\end{aligned}
$$

converges to p_{opt} for $N_{\text{par}} \to \infty$. The size of the LMI problems of this family is increasing due to the increase of the dimension of $\mathcal{Z}_{N_{\text{par}}}$ for larger N_{par}.

By applying SOS relaxations, a second source of conservatism is introduced. It will be shown in the next chapter how this conservatism can be systematically reduced under mild conditions, by using a sequence of monomial basis of increasing dimension to parameterize the SOS polynomials.

The second source of conservatism when applying the S-procedure is the error made by the inner approximation of the multipliers. As will be shown in the next chapter as well, these errors can be reduced under similar conditions as for the SOS approach, by using an asymptotically exact family of multiplier parameterizations, again by introducing more variables in the LMI problem.

If the relaxation of the robust analysis problem is exact, a test presented in [172] enables to construct a worst-case perturbation under certain conditions. For the problem under consideration this results in a "worst-case" controller. If the closed-loop $\mathcal{H}_\infty$-norm equals the lower bound d_{opt}, then this controller is optimal and the lower bound computation exact. It is interesting to compare this method to construct a globally optimal controller with the one discussed in Section 4.5.6. This is however beyond the scope of this thesis.

5.5 Application

We present lower bound computations for the fixed order $\mathcal{H}_\infty$ problem of a fourth order system and a 27th order active suspension system. The results of the relaxations based on SOS decompositions and the S-procedure have also been published in [88] and [89] respectively.

5.5.1 Fourth order system

We consider an example with

$$
\left(\begin{array}{c|c|c} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ \hline C_2 & D_{21} & D_{22} \end{array} \right) := \left(\begin{array}{cccc|c|cc} -7 & 4 & 0 & 0.2 & 0.9 & 0.2 & 0 \\ -0.5 & -2 & 0 & 0 & 2 & 0.2 & 0 \\ 3 & 4 & -0.5 & 0 & 0.1 & 0.1 & 0 \\ 3 & 4 & 2 & -1 & -4 & 0 & -0.2 \\ \hline 0 & -10 & -3 & 0 & 0 & 3 & -4 \\ \hline 0.8 & 0.1 & 0 & 0 & 0.3 & 0 & 0 \end{array} \right)
\tag{5.32}
$$

and computed lower bounds on the closed-loop $\mathcal{H}_\infty$-performance of all stabilizing static controllers in a compact subset of $\mathbb{R}^{2 \times 1}$. The open loop $\mathcal{H}_\infty$ norm is 47.6. We first compute an initial feedback law $K_{\text{init}} = \left(\begin{array}{cc} -38 & -28 \end{array} \right)^T$, which gives a performance of 0.60. We compute bounds for the box $\rho \mathcal{K}_{\text{box}}$, where $\rho = 5$ and of the unit box $\mathcal{K}_{\text{box}} := \{ K \in \mathbb{R}^2 | \ \|K_i\| \le 1 \}$ around the initial controller, i.e. $K := K_{\text{init}} + \left(\begin{array}{cc} K_1 & K_2 \end{array} \right)^T$. Observe that $K \in \rho \mathcal{K}$ is equivalent to $G(K) \preceq 0$ where

$$
G(K) := \left(\begin{array}{cc} K_1^2 - \rho^2 & 0 \\ 0 & K_2^2 - \rho^2 \end{array} \right).
\tag{5.33}
$$

SOS relaxations

We consider

$$
F_{\text{r}}(K, y) = \text{diag}(F_{11}(K, z) - d, F_{22}(K, z), 1 - \text{Trace}(Z_{22}(K, z)))
$$

94

where F_{11} and F_{22} as in (5.20) and (5.21), together with the constraint that

$$Z(K) \text{ is SOS.} \tag{5.34}$$

As argued in Section 5.1.1, replacement of F by F_r is not restrictive, if we add (5.34) as additional constraint. This reduces the number of variables in the LMI relaxation, since F_r has 6 rows and columns opposed to 11 rows and columns of F. We used the monomial vector $u(K) = 1$. The monomial vector $u_0(K)$ is chosen such that it contains the monomials[1] in $\mathcal{M}_{F(K,Y)+\langle S(K),G(K)\rangle_q - \epsilon I}$, see (5.23).

The resulting lower bound, the number of variables and constraints in the LMI are shown in Table 5.1 for various degrees in $Z(K)$. The first column presents the monomials in $Z(K)$, where $(1, K_1, K_1^2)$ should be interpreted as the parametrization

$$Z(K,z) = Z_0 + \sum_{j=1}^{N_{\text{par}}} \left(z_j E_j + z_{j+N_{\text{par}}} K_1 E_j + z_{j+2N_{\text{par}}} K_1^2 E_j \right),$$

where $N_{\text{par}} = \dim(\mathcal{S}^{n+m_1}) = \frac{1}{2}(n + m_1 + 1)(n + m_1)$ and E_j, $j = 1, \ldots, N_{\text{par}}$, is a basis for $\mathcal{S}^{n+m_1}$. For this parametrization the number of variables in y grows quadratically with n, which illustrates the aforementioned *quadratic growth* in the number of state variables.

By a gridding technique we have found an optimal controller $K_{\text{opt}} = \begin{pmatrix} 1.33 & 0.69 \end{pmatrix}^T \in \mathcal{K}$ with performance 0.1832. The best lower bound on the optimal closed-loop $\mathcal{H}_\infty$-norm is 0.1811, which is slightly smaller than the optimal performance 0.1832. The number of variables and constraints in our implementation of the LMI relaxations are also shown in the table. These may deviate from those in Section 5.4, due to some minor modifications to the implementation. The same holds true for the number of LMI variables and constraints given in the remaining tables of this chapter. The number of constraints is very large, but apparently due to the sparseness each LMI problem can be solved within 80 seconds.

For comparison, we have also applied the direct SOS relaxation as described in the previous chapter to this example. A lower bound of 0.1832 is computed as the optimal value of the LMI relaxation

$$
\begin{aligned}
&\text{supremize} && d \\
&\text{subject to} && \gamma + \langle S(\gamma, X, K), G(\gamma, X, K)\rangle - d = s_0(\gamma, X, K) \\
& && S(\gamma, X, K), \quad s_0(\gamma, X, K) \text{ are SOS w.r.t. } u(\gamma, X, K), \ u_0(\gamma, X, K)
\end{aligned}
$$

where

$$G(\gamma, X, K) := \operatorname{diag}\left(-X, -\mathcal{B}(\gamma, X, K), K_1^2 - \rho^2, K_2^2 - \rho^2\right),$$

and with monomial basis

$$u(\gamma, X, K) := \begin{pmatrix} 1, & \gamma, & K_1, & K_2, & K_1^2, & K_1 K_2, & K_2^2 \end{pmatrix}$$

and $u_0(\gamma, X, K)$ contains all monomials in $\mathcal{M}_{\gamma + \langle S(\gamma,X,K),G(\gamma,X,K)\rangle - d}$. The number of LMI variables is 1380. This lower bound of 0.1832 is slightly better than the best lower bound 0.1811 in Table 5.1 and is equal to the upper bound 0.1832.

---

[1]See Section 4.5.4 for the definition of $\mathcal{M}$.

Table 5.1: SOS lower bounds $\sqrt{d_{\mathrm{opt}}}$ on optimal closed-loop $\mathcal{H}_\infty$-norm for the fourth order system (5.32), with various monomial bases of $Z(K)$.

Basis of $Z(K)$	$\sqrt{d_{\mathrm{opt}}}$	# LMI variables	# LMI constr.	comp. time
(1)	0.0077	1039	11.338	53.4s
$(1, K_1, K_1^2)$	0.1513	1109	12.028	60.7s
$(1, K_2, K_2^2)$	0.0083	1109	12.028	55.6s
$(1, K_1, K_1^2, K_1 K_2, K_1^2 K_2, K_2^2)$	0.1811	1377	14.291	79.9s

S-procedure relaxations

Using the S-procedure relaxations as discussed in Section 5.3, we were unfortunately not able to reproduce these exact lower bounds, because the size of Δ and hence the resulting LMI problem gets too large if we use a polynomial $Z(K)$ even of modest total degree. Only a relaxation with constant basis (1), i.e. $Z(K) = Z_0 \in \mathcal{S}^5$ appears to be computable with the current computers. This relaxation yields a lower bound of 0 for $\rho = 5$. A reasonable lower bound of 0.3882 could only be established for $\rho = 0.1$. Note that this bound does not contradict the optimal performance 0.1832 above, since this performance is realized by a controller in the box for $\rho = 5$ and is outside the box for $\rho = 0.1$.

5.5.2 Active suspension system

As a second example, we consider the control of the active suspension system. A detailed description of this system is postponed to Section 7.4, where we present a fixed-order controller design for this system by interior-point optimization. The dynamic model of the suspension system has 17 states and the weights of our 4-block $\mathcal{H}_\infty$ design contributes with 10 states, which adds up to 27 states of the generalized plant. The full order design has closed-loop $\mathcal{H}_\infty$-norm 2.48. We computed a 5th order controller by closed-loop balanced residualization, as described in Section 3.1.1 with performance 3.41. We computed lower bounds for changes in two diagonal elements of the state-space matrices of the controller

$$
K(p_1, p_2) = \left(\begin{array}{c|c} A_K(p_1, p_2) & B_K \\ \hline C_K & D_K \end{array} \right)
$$

$$
= \left(\begin{array}{ccccc|c} -78.2 & 1129.2 & 173.24 & -97.751 & -130.36 & 6.6086 \\ -1240.9 & -78.2 + p_1 & 111.45 & 125.12 & 76.16 & 21.445 \\ 0 & 0 & -6.0294 & 164.81 + p_2 & 159 & -11.126 \\ 0 & 0 & 0 & -204.56 & 49.031 & -12.405 \\ 0 & 0 & 0 & -458.3 & -204.56 & -9.4469 \\ \hline -0.067565 & 0.19822 & -1.0047 & -0.069722 & 0.19324 & 0.0062862 \end{array} \right),
$$

where p_1 and p_2 are free scalar controller variables.

SOS relaxations

Table 5.2 shows the computed lower bounds for various bases for $Z(K)$ and various controller sets $\rho\mathcal{K}_{\mathrm{ball}} := \{p \mid \mathbb{R}^2, \|p\| \le \rho\}$, $\rho \in \{5, 10, 50, 100\}$, together with the number of LMI variables. Again we used the monomial vector $u(K) = 1$ and $u_0(K) = \mathbf{New}(F(K, Y) + \langle S(K), G(K) \rangle_q - \epsilon I)$. We observe that the lower bounds are larger than the

Table 5.2: SOS lower bounds $\sqrt{d_{\mathrm{opt}}}$ on optimal closed-loop $\mathcal{H}_\infty$-norm for the active suspension system, with various monomial bases of $Z(K)$. The controller sets $\mathcal{K} = \rho \mathcal{K}_{\mathrm{ball}}$ are balls of various radii $\rho \in \{5, 10, 50, 100\}$.

Basis	Radius ρ				# LMI variables
of $Z(K)$	5	10	50	100	
(1)	3.2271	3.0176	2.2445	1.9790	1719
$(1, K_1)$	3.2570	3.0732	2.3975	2.1585	2313
$(1, K_2)$	3.2412	3.0468	2.3725	2.2398	2313

Table 5.3: S-procedure lower bounds $\sqrt{d_{\mathrm{opt}}}$ on optimal closed-loop $\mathcal{H}_\infty$-norm for the active suspension system, with $Z(K) = Z_0$. The controller sets $\mathcal{K} = \rho \mathcal{K}_{\mathrm{box}}$ are boxes of various sizes $\rho \in \{0.1, 2.5, 2.8\}$.

Basis	Radius ρ		size Δ	# LMI variables
of $Z(K)$	0.1	2.5		
(1)	3.41	3.19	4	613

full-order performance for radii $\rho \in \{5, 10\}$ with either $u(K) = (1, K_1)$ or $u(K) = (1, K_2)$. Hence relevant lower bounds can be computed for reasonably large parameter sets $\mathcal{K}$.

S-procedure relaxations

Lower bounds have also been obtained with the S-procedure. The results in Section 5.3 cannot be applied to compute lower bounds for $\mathcal{K}_{\mathrm{ball}}$, since this ball is not a polytope. We therefore replace it by the box

$$\mathcal{K}_{\mathrm{box}} := \left\{ p | \ \mathbb{R}^2, |p_i| \leq 1, \ i = 1, 2 \right\}.$$

The resulting lower bounds for various boxes $\rho \mathcal{K}_{\mathrm{box}}$ are shown in Table 5.3. We observe that the lower bounds are worse than the lower bounds by SOS decompositions, which follows from comparison of Tables 5.2 and 5.3 and by observing $\mathcal{K}_{\mathrm{box}} \subset \sqrt{2} \mathcal{K}_{\mathrm{ball}}$.

5.5.3 Conclusions on application

The relaxations based on SOS decompositions appear to be better than those based on the S-procedure, in the sense that they generate better lower bounds with roughly the same size of the LMI problem. The S-procedure however, allows a more straightforward implementation for rationally parameterized $Z(K)$, because it can directly be applied to inequalities that are rational in K.

We have applied the method to systems of McMillan degree 4 and 27 respectively. The first example illustrates that close to exact lower bounds can efficiently be computed, with simple monomial bases for the SOS polynomials $Z(K)$, $S(K)$ and $S_0(K)$. The second example showed the feasibility of the approach for plants with moderate McMillan degree, in the sense that we can compute nontrivial lower bounds by solving LMI problems with about 2300 variables. Hence the lower bound relaxations based on partial dualization are computable for this practically relevant control problem.

5.6 Conclusions

We suggested novel relaxation schemes for the structured $\mathcal{H}_\infty$-optimal controller synthesis problem, based on SOS matrices and the S-procedure. Both grow quadratically in the dimension of the system state and can be guaranteed to be asymptotically exact under mild conditions. With these schemes we can assess global optimality of controllers. This concludes our work on the research question on global optimality certificates, as formulated in item ② in Section 2.8. To the best of our knowledge this has been an unexplored research area before the publication of our results in [89, 88, 90].

As mentioned in Section 2.6, it is widely accepted that structured controller synthesis is a very hard problem. This partially explains the large effort we had to make to reduce the complexity of the LMI relaxations. The developed relaxation scheme is applicable to control problems encountered in practice. The LMI problems may still be quite large if for instance N_{par} is large or the controller has many variables. Due to the strength of the global results with certainty, we expect that controller synthesis tools based on the proposed principles will become more important in control engineering.

Chapter 6

Sum-of-squares relaxations for Robust SDPs

We leave the main track on fixed-order synthesis and focus in this chapter on robust control: we present a nontrivial extension of the results on Sum-Of-Squares relaxations in the previous two chapters to robust Semi-Definite Programs. These are SDPs with coefficient matrices depending on an uncertain parameter that varies, typically, in a compact set. Several robust stability and performance analysis problems with either parametric or dynamic, time-invariant or time-varying deterministic uncertainties can be re-formulated as robust SDPs [55, 9, 172].

Various schemes (such as multiplier relaxation in structured singular value theory [51, 140, 97, 98, 170]) and more recently Sum-Of-Squares decompositions [38, 39, 86] have been applied to construct efficiently computable relaxations for these programs. In general it cannot be expected that these relaxations are exact, and the only known techniques to systematically reduce the relaxation gap with guaranteed convergence is restricted to boxes [19, 18] or to finitely generated polytopes with known generators [171, 172]. As the main goal of this chapter we show how such asymptotically exact relaxation families can be constructed on the basis of matrix SOS decompositions for the much larger class of uncertainty sets, such as norm-bounded and structured uncertainty as encountered in μ-analysis and various other possibly non-convex compact uncertainty sets. In contrast to approaches based on scalarization and a subsequent application of existing relaxation techniques [119, 143, 180], we will be able to show that the size of the constructed LMI relaxations grows at most bi-quadratically in the dimension of the polynomial matrices that occur in the problem formulation. Moreover the techniques in [172] can be applied in order to verify whether a given finite relaxation does not involve any conservatism.

The outline of this chapter is as follows. After the problem description in Section 6.1, we present the main result in Section 6.2: a reformulation of the robust SDP in terms of matrix-valued SOS polynomials. In Section 6.3 we construct LMI-relaxations to this latter problem, whose optimal values converge from above to the optimal value of the robust SDP. In Section 6.4 we show the benefits of the formulation with matrix-valued SOS polynomials: it has enabled us to prove a guaranteed bi-quadratic growth of the size of the LMI relaxations in the dimension of the polynomial matrices. In Section 6.5 we reveal how the SOS relaxations can be exploited to systematically reduce the conservatism in standard robust analysis tests based on the full-block S-procedure and more specific D-scalings. Finally in Section 6.6 we apply the method to compute upper bounds on a robust SDP and to assess robust analysis of a 4th order helicopter model. Both these examples

are taken from the literature. The results in this chapter and the numerical examples have partially been published in [174], [92] and [175].

6.1 Robust polynomial SDPs

In this chapter we consider the following robust polynomial SDP with optimal value p_{opt}:

$$
\begin{array}{ll}
\text{infimize} & c^T y \\
\text{subject to} & F(x, y) \succ 0 \ \text{ for all } \ x \in \mathbb{R}^m \ \text{ with } \ G(x) \preceq 0.
\end{array}
\tag{6.1}
$$

Here $F : \mathbb{R}^m \times \mathbb{R}^n \mapsto \mathcal{S}^q$ and $G : \mathbb{R}^m \mapsto \mathcal{S}^r$ are symmetric-valued functions which depend polynomially on the uncertainty parameter $x \in \mathbb{R}^m$ while F depends affinely on the design parameter $y \in \mathbb{R}^n$. Therefore $F(x, y) \succ 0$ is a standard linear matrix inequality (LMI) in y for fixed x, while the robust counterpart requires to satisfy the LMI for all x in the uncertainty set

$$
\mathcal{G} = \{x \in \mathbb{R}^m : \ G(x) \preceq 0\}.
\tag{6.2}
$$

The set $\mathcal{G}$ admits a very general description in terms of a polynomial semi-definite constraint. We stress that in many interesting practical cases $\mathcal{G}$ turns out to admit an LMI representation (i.e. G is affine) or is even just a compact polytope (i.e. G is diagonal and affine). Polynomial SDPs as considered in Chapter 4 are recovered from (6.1) with $F(x, y) = f(x) - y$ and $c = -1$. If F depends also affinely on the uncertainty x and $\mathcal{G}$ is the convex hull of a moderate number of explicitly given finite number of points in $\mathbb{R}^m$ (also called the generators [155] of $\mathcal{G}$), it is clear that (6.1) amounts to solving a standard LMI problem. If the number of extreme points to describe $\mathcal{G}$ is large, it is often possible to construct efficiently computable relaxations with a priori guarantees on the relaxation error [10, 11]. The situation drastically differs if the uncertainties enter nonlinearly, since then such guarantees are not known.

If $F(x, y)$ is rationally dependent on x and a well-posedness condition is satisfied, we can render the Robust SDP polynomial in x: multiply $F(x, y)$ by its smallest common denominator that is positive on $\mathcal{G}$. Complex-valued uncertainty are reduced to real values in a standard fashion.

Example 6.1 *As an example of a robust analysis problem that can be molded into (6.1), consider the uncertain system $\dot{z}(t) = A(\delta(t))z(t)$, where $A(\cdot)$ is a matrix-valued polynomial and $\delta(\cdot)$ varies in the set of continuously differentiable parameter curves $\delta : [0, \infty) \mapsto \mathbb{R}^m$ with $\delta(t) \in \boldsymbol{\Delta}$ for all $t \in [0, \infty)$ and compact uncertainty set $\boldsymbol{\Delta}$. Consider for instance the following combination of polytopic and norm-bounded uncertainties*

$$
\boldsymbol{\Delta} := \{\delta = \begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix}^T : \ \delta_1 \in \mathbb{R}^{m_1}, \ \delta_2 \in \mathbb{R}^{m_2}, \ H(\delta_1) \preceq 0, \ \|\delta_2\| \leq 1\}
$$

where H is affine, $m_1 + m_2 = m$ and such that $\boldsymbol{\Delta}$ is compact. Then the system is uniformly exponentially stable if there exists a $Y \succ 0$ such that $A(\delta)^T Y + Y A(\delta) \prec 0$ for all $\delta \in \boldsymbol{\Delta}$ [46]. With $x_1 = \delta_1$, $x_2 = \delta_2$, $y = svec(Y)$, $c = 0$, $G(x) = diag(H(x_1), x_2^T x_2 - 1)$ and $F(x, y) = diag(Y, -A(x)^T Y - Y A(x))$ this problem is of type (6.1).

6.2 Construction of an exact SOS reformulation

In this section we show that under a constraint qualification on G, (6.1) is equivalent to a certain SOS problem, in the sense that their optimal values are equal. On the basis of these

results we are able to construct a sequence of finite-dimensional LMI relaxation problems, whose optimal values converge to the optimal value p_{opt} of (6.1). Hence, we can construct LMI relaxations of robust SDPs, with a guarantee of reducing the relaxation gap to zero.

We will show in the proof of Theorem 6.3, that the optimal value d_{opt} of the following optimization problem is an upper bound on the optimal value of (6.1):

$$
\begin{aligned}
&\text{infimize} && c^T y \\
&\text{subject to} && \epsilon > 0, \ S(x) \text{ and } F(x,y) + \langle S(x), I_q \otimes G(x) \rangle_q - \epsilon I_q \text{ are SOS in } x.
\end{aligned}
\tag{6.3}
$$

The main result of this chapter is to prove equality (i.e. $p_{\text{opt}} = d_{\text{opt}}$) under the constraint qualification (4.18), as discussed in Section 4.4. Before presenting this main result, let us first assume that G is diagonal, i.e.

$$
G(x) := \text{diag}(g_1(x), g_2(x), \ldots, g_r(x)).
\tag{6.4}
$$

Feasibility of y_* for (6.1) comes down to computationally verifying whether

$$
F(x, y_*) \succ 0 \ \text{ for all } \ G(x) \prec 0.
\tag{6.5}
$$

The following theorem shows that this is possible using a representation with matrix-valued SOS polynomials.

Theorem 6.2 *Suppose $G(x)$ is as in (6.4) for some $g_i(x)$, $i = 1, \ldots, r$ and satisfies (4.18). Then (6.5) implies that there exist $\epsilon > 0$ and matrix SOS $S_0(x), S_1(x), \ldots, S_r(x)$ such that*

$$
F(x, y_*) - \epsilon I_q + \sum_{i=1}^{r} S_i(x) g_i(x) = S_0(x).
\tag{6.6}
$$

Proof. The proof is a straightforward extension of Theorem 2 in our paper [174], see also our publication [175] ∎

Now let us drop the assumption on $G(x)$ being diagonal. This brings us to the central result in this chapter.

Theorem 6.3 *Suppose that G satisfies (4.18). If p_{opt} and d_{opt} are the optimal values of (6.1) and (6.3) respectively, then $p_{opt} = d_{opt}$.*

Remark. Theorem 6.3 combines the results on polynomial SDP as described in Chapter 4 and robust LMI problems with polytopic uncertainty regions (with diagonal and affine G) [172] to a very general formulation with a wide range of applications in robust controller analysis.

Proof. We use similar arguments as in the proof of Theorem 4.7. We first prove $p_{\text{opt}} \leq d_{\text{opt}}$, by showing that the constraint in (6.1) is implied by the constraint in (6.3). Consider arbitrary $\epsilon > 0$, $y_* \in \mathbb{R}^n$ and x_* with $G(x_*) \preceq 0$. Let us now suppose that $S(x)$ and $S_0(x) = F(x, y_*) - \epsilon I_q + \langle S(x), I_q \otimes G(x) \rangle_q$ are SOS. Due to (4.15) one infers

$$
F(x_*, y_*) \succeq F(x_*, y_*) - \epsilon I_q + \langle S(x_*), I_q \otimes G(x_*) \rangle_q = S_0(x) \succeq 0.
$$

Since x_* with $G(x_*) \preceq 0$ was arbitrary, the implication is shown. We conclude that $p_{\text{opt}} \leq d_{\text{opt}}$.

To prove $p_{\mathrm{opt}} \geq d_{\mathrm{opt}}$, note that, as a consequence of the constraint qualification, if $G(x) \preceq 0$ is replaced by

$$\tilde{G}(x) := \mathrm{diag}(G(x), \|x\|^2 - M) \preceq 0$$

then the value of (6.1) is not modified. In a first step of the proof of $p_{\mathrm{opt}} \geq d_{\mathrm{opt}}$, let us show that the same is true for the SOS reformulation (6.3).

Indeed suppose $F(x,y) - \epsilon I_q + \langle S(x), I_q \otimes G(x) \rangle_r = S_0(x)$ with SOS matrices $S_0(x)$ and $S(x)$. If we partition $S(x) = (S_{jk}(x))_{jk}$ into $r \times r$-blocks then $\tilde{S}(x) := (\mathrm{diag}(S_{jk}(x), 0))_{jk}$ satisfies $\langle \tilde{S}(x), I_q \otimes \tilde{G}(x) \rangle_q = \langle S(x), I_q \otimes \tilde{G}(x) \rangle_q = \langle S(x), I_q \otimes G(x) \rangle_q$ and therefore $F(x,y) - \epsilon I_q + \langle \tilde{S}(x), I_q \otimes \tilde{G}(x) \rangle_q - \epsilon I_q = S_0(x)$. Conversely suppose $F(x,y) - \epsilon I_q + \langle \tilde{S}(x), I_q \otimes \tilde{G}(x) \rangle_q = \tilde{S}_0(x)$ with SOS matrices $\tilde{S}_0(x)$, $\tilde{S}(x)$. Now we make explicit use of (4.18) with SOS matrices $\psi(x)$, $\Psi(x)$. Let us partition

$$\tilde{S}(x) = \left(\begin{pmatrix} S_{jk}(x) & * \\ * & s_{jk}(x) \end{pmatrix} \right)_{jk}$$

into blocks of size $(r+1) \times (r+1)$ and define

$$S(x) := (S_{jk}(x) + s_{jk}(x)\Psi(x))_{jk}$$

and $s(x) = (s_{jk}(x))_{jk}$ of dimension qr and q respectively. It is easy to verify that both matrices are SOS and satisfy

$$\langle \tilde{S}(x), I_q \otimes \tilde{G}(x) \rangle_q = \langle S(x), I_q \otimes G(x) \rangle_q - s(x)\psi(x).$$

This implies $F(x,y) - \epsilon I_q + \langle S(x), I_q \otimes G(x) \rangle_q = \tilde{S}_0(x) + s(x)\psi(x)$ and it remains to observe that $\tilde{S}_0(x) + s(x)\psi(x)$ is SOS.

Therefore, from now on we can assume without loss of generality that

$$v_1^T G(x) v_1 = \|x\|^2 - M \tag{6.7}$$

where $v_1 := e_r$ is the last standard unit vector. It remains to show $p_{\mathrm{opt}} \geq d_{\mathrm{opt}}$, and for this purpose it suffices to choose an arbitrary y_* which is feasible for (6.1) and to prove that y_* is as well feasible for (6.3).

Let us hence assume $F(x, y_*) \succ 0$ for all $x \in \mathcal{G}$. Choose a sequence of unit vectors $v_2, v_3, \ldots$ such that v_i, $i = 1, 2, \ldots$ is dense in the unit sphere $\{v \in \mathbb{R}^r : \|v\| = 1\}$. Define

$$\mathcal{G}_N := \{x \in \mathbb{R}^m : v_i^T G(x) v_i \leq 0, \ i = 1, \ldots, N\}$$

to infer that $\mathcal{G}_N$ is compact (by (4.19)) and that $\mathcal{G}_N \supset \mathcal{G}_{N+1} \supset \mathcal{G}$ for $N = 1, 2, \ldots$. Therefore $q_N := \min\{\lambda_{\min}(F(x, y_*)) : x \in \mathcal{G}_N\}$ is attained by some x_N and $q_N \leq q_{N+1}$ for all $N = 1, 2, \ldots$. Let us prove that there exists some N_0 for which $q_{N_0} > 0$ which implies

$$F(x, y_*) \succ 0 \ \text{ for all } \ x \in \mathcal{G}_{N_0}. \tag{6.8}$$

Indeed otherwise $q_N \leq 0$ for all $N = 1, 2, \ldots$ and hence $\lim_{N \to \infty} q_N \leq 0$. Choose a subsequence N_ν with $x_{N_\nu} \to x_0$ to infer $0 \geq \lim_{\nu \to \infty} \lambda_{\min}(F(x_{N_\nu}, y_*)) = \lambda_{\min}(F(x_0, y_*))$. This contradicts the choice of y_* if we can show that $G(x_0) \preceq 0$. In fact, otherwise there exists a v with unit norm $\|v\| = 1$ such that $\delta := v^T G(x_0) v > 0$. By convergence there exists some $L \in \mathbb{R}$ with $\|G(x_{N_\nu})\| \leq L$ for all ν. By density there exists a sufficiently large

ν such that $L\|v_i - v\|^2 + 2L\|v_i - v\| < \delta/2$ for some $i \in \{1, \ldots, N_\nu\}$. Since $v^T G(x_{N_\nu})v \to v^T G(x_0)v$ we can increase ν to even guarantee $v^T G(x_{N_\nu})v \geq \delta/2$ and arrive at the following contradiction:

$$0 \geq v_i^T G(x_{N_\nu})v_i =$$
$$= (v_i - v)^T G(x_{N_\nu})(v_i - v) + 2v^T G(x_{N_\nu})(v_i - v) + v^T G(x_{N_\nu})v \geq$$
$$\geq -L\|v_i - v\|^2 - 2L\|v_i - v\| + \delta/2 > 0.$$

We are now in the position to apply Theorem 6.3 to (6.8) since, due to (4.19), the constraint qualification is trivially satisfied. Hence there exist $\epsilon > 0$ and polynomial matrices $U_i(x)$ with q columns, $i = 1, \ldots, N_0$, such that

$$F(x, y_*) - \epsilon I + \sum_{i=1}^{N_0} [U_i(x)^T U_i(x)](v_i^T G(x)v_i) \tag{6.9}$$

is SOS in x. With some Kronecker product manipulations and (4.13) we conclude

$$[U_i(x)^T U_i(x)](v_i^T G(x)v_i) = \text{Trace}_q\left([U_i(x)^T U_i(x)] \otimes (v_i^T G(x)v_i)\right) =$$
$$= \text{Trace}_q\left(([U_i(x)^T U_i(x)] \otimes v_i^T)(I_q \otimes G(x))(I_q \otimes v_i)\right) =$$
$$= \text{Trace}_q\left((I_q \otimes v_i)([U_i(x)^T U_i(x)] \otimes v_i^T)(I_q \otimes G(x))\right) =$$
$$= \text{Trace}_q\left(([U_i(x)^T U_i(x)] \otimes v_i v_i^T)(I_q \otimes G(x))\right) =$$
$$= \langle (U_i(x) \otimes v_i^T)^T (U_i(x) \otimes v_i^T), I_q \otimes G(x) \rangle_q.$$

With the SOS polynomial matrix

$$S(x) := \sum_{i=1}^{N_0} (U_i(x) \otimes v_i^T)^T (U_i(x) \otimes v_i^T)$$

we infer that $F(x, y_*) - \epsilon I + \langle S(x), I_q \otimes G(x) \rangle$ equals the left-hand side in (6.9) and is hence SOS in x. Therefore y_* is feasible for (6.3).

∎

6.3 Construction of LMI relaxations

We briefly discuss how to modify the LMIs in Section 4.5 to compute upper bounds for (6.3), by choosing a fixed monomial bases $u(x)$ for $S(x)$. The procedure in Section 4.5.4 gives us the required monomial basis vector $u_0(x)$ to represent $F(x, y) + \langle S(x), I_q \otimes G(x) \rangle_q - \epsilon I_q$ as a sum of squares. Moreover we choose some monomial basis vector $v(x)$ and matrix B such that $G(x) = B(I_r \otimes v(x))$. Finally choose pairwise different monomials $w_1(x), \ldots, w_\eta(x)$ such that there exist matrices P_j^0, P_j and symmetrically-valued affine mappings $A_j(y, \epsilon)$

with

$$u_0(x)u_0(x)^T = \sum_{j=1}^{\eta} P_j^0 w_j(x),$$

$$v(x) \otimes u(x)u(x)^T = \sum_{j=1}^{\eta} P_j w_j(x),$$

$$F(x,y) - \epsilon I = \sum_{j=1}^{\eta} A_j(y,\epsilon) w_j(x).$$

Furthermore we construct SOS polynomials

$$S_0(x) = \langle W_0, I_q \otimes u_0(x)u_0(x)^T \rangle_q$$

and

$$S(x) = \langle W, I_{qr} \otimes u(x)u(x)^T \rangle_{qr},$$

parameterized with matrices $W_0 \in \mathcal{S}^{q\xi_0}$ and $W \in \mathcal{S}^{qr\xi}$. We then infer $S_0(x) = \sum_{j=1}^{\eta} \langle W_0, I_q \otimes P_j^0 \rangle_q w_j(x)$ as well as

$$\begin{aligned}
\langle S(x), I_q \otimes G(x) \rangle_q &= \mathrm{Trace}_q \left[(I_{qr} \otimes u(x))^T W (I_{qr} \otimes u(x))(I_q \otimes G(x)) \right] \\
&= \mathrm{Trace}_q \left[(I_q \otimes [I_r \otimes u(x)^T]) W (I_q \otimes ([I_r \otimes u(x)]G(x))) \right] \\
&= \mathrm{Trace}_q \left[W(I_q \otimes [G(x) \otimes u(x)][I_r \otimes u(x)^T]) \right] \\
&= \mathrm{Trace}_q \left[W(I_q \otimes [B(I_r \otimes v(x))] \otimes u(x)u(x)^T) \right] \\
&= \mathrm{Trace}_q \left[W(I_q \otimes B \otimes I_\xi)(I_{qr} \otimes v(x) \otimes u(x)u(x)^T) \right] \\
&= \sum_{j=1}^{\eta} \langle W, (I_q \otimes B \otimes I_\xi)(I_{qr} \otimes P_j) \rangle_q w_j(x).
\end{aligned}$$

Therefore there exist $\epsilon > 0$ and SOS matrices $S(x)$, $S_0(x)$ with respect to $u(x)$, $u_0(x)$ such that $F(x,y) - \epsilon I + \langle S(x), I_q \otimes G_i(x) \rangle_q = S_0(x)$ iff the following LMI system is feasible:

$$\epsilon > 0, \quad W_0 \succeq 0, \quad W \succeq 0, \tag{6.10}$$

$$A_j(y,\epsilon) + \langle W, (I_q \otimes B \otimes I_\xi)(I_{qr} \otimes P_j) \rangle_q = \langle W_0, (I_q \otimes P_j^0) \rangle_q, \quad j = 1, ..., \eta. \tag{6.11}$$

If we minimize $c^T y$ over these LMI constraints we determine an upper bound of the value d_{opt} of (6.3) and hence also an upper bound of the value p_{opt} of (6.1), even without constraint qualification. Let $\{u_1(x), u_2(x), ...\}$ be a sequence of monomial vectors such that

$$\mathrm{span}(\{u_1, u_2, ...\}) = \mathbb{R}[x].$$

Let us choose for some ξ the monomial basis vector $u(x) = \begin{pmatrix} u_1(x) & u_2(x) & ..., u_N(x) \end{pmatrix}^T$ and let $u_0(x)$ contain all monomials in $\mathcal{M}_{F(x,y)+\langle S(x), I_q \otimes G(x) \rangle_q - \epsilon I_q}$[1] Let d_N denote the corresponding upper bound, i.e. the optimal value of the corresponding LMI relaxation. Then the sequence d_N is non-increasing and is guaranteed to converge to p_{opt} for increasing N, if the constraint qualification (4.18) is satisfied. The proof of this result is completely analogous to Theorem 4.9 and therefore omitted.

---

[1] See Section 4.5.4 for the definition of $\mathcal{M}_{F(x,y)+\langle S(x), I_q \otimes G(x) \rangle_q - \epsilon I_q}$.

Let us finally observe that the size of the relaxation is determined by three SDP constraints (6.10) in $\mathcal{S}^1$, $\mathcal{S}^{q\xi_0}$ and $\mathcal{S}^{qr\xi}$ and the η affine equation constraints (6.11) in $\mathcal{S}^q$. Moreover it involves the unknowns y, ϵ, W_0, W of sizes n, 1, $q\xi_0$ and $qr\xi$ respectively which sum up to

$$n + 1 + 0.5\,q\xi_0(q\xi_0 + 1) + 0.5\,qr\xi(qr\xi + 1)$$

scalar decision variables. Generically this number is reduced by $0.5\,\eta q(q+1)$ through (6.11). The variable elimination technique as described in 4.5.4 can be applied to (6.10) and (6.11). As stressed in the introduction of this chapter the size of the LMI relaxation indeed only grows bi-quadratically in the dimension q of $F(x,y)$ or r of $G(x)$ respectively.

6.4 Comparison with scalarization

We compare in this section the size of the LMI relaxations (6.11) with a SOS decomposition after scalarization of both $F(x)$ and $G(x)$, similar the scalarization of polynomial SDPs as described in Section 4.5.3. Indeed, we can equivalently replace the semi-definite constraint $G(x) \preceq 0$ by its minors $M_i(G(x)) \leq 0, i = 1, \ldots, N$. Furthermore if we define $f(v,x,y) := v^T F(x,y)v$ and

$$h_i(v,x) = M_i(G(x)) \quad i = 1, \ldots, N, \quad h_{N+1}(v,x) = 1 - v^T v, \quad h_{N+2}(v,x) = v^T v - 2, \quad (6.12)$$

then (6.1) is equivalent to infimizing $c^T y$ subject to

$$f(v,x,y) > 0 \text{ for all } (x,v) \text{ with } h_i(v,x) \leq 0, \quad i = 1, \ldots, N+2. \quad (6.13)$$

If $h_i, i = 1, \ldots, N+2$ satisfy a constraint qualification, then Theorem 4.3 and (6.13) imply that there exist SOS polynomials $s_i(v,x)$, $i = 1, \ldots, N+2$, such that

$$f(v,x,y) + \sum_{i=1}^{N+2} s_i(v,x)h_i(v,x) \quad \text{is SOS.} \quad (6.14)$$

However, although $f(v,x,y)$ and $h_i(v,x)$ are quadratic in v, no available result guarantees that the SOS polynomials $s_i(v,x)$, $i = 1, \ldots, N+2$, can be chosen quadratic in v without loosing the relaxation's exactness. Theorem 6.3 implies that one can indeed confine the search to $s_{N+1}(v,x) = 0$, $s_{N+2}(v,x) = 0$ and to $s_j(v,x) = v^T S_j(x)v$, $j = 0, 1, \ldots, r$, which are homogenously quadratic in v, without violating $p_{\text{opt}} = d_{\text{opt}}$. Furthermore the theorem shows that the minors of $G(x)$ (or any other scalarization technique for $G(x) \preceq 0$) are not required for exact SOS relaxations. Hence Theorem 6.3 extents Theorem 4.7 in the sense that bi-quadratic growth of the LMI relaxations in the size of both $F(x)$ and $G(x)$ is realized.

6.5 Relaxations based on the S-procedure

6.5.1 S-procedure

In this section we intend to reveal the relation of the suggested approach to relaxations based on the so-called full-block S-procedure, as presented in Section 2.5.7. Indeed, even

if $F(x, y)$ is rational (instead of merely polynomial) in x without pole at $x = 0$, one can construct the LFT

$$F(x, y) = \mathcal{L}_{\mathrm{u}} \left(\Delta(x), \begin{pmatrix} A & B \\ C(y) & D(y) \end{pmatrix} \right) := D(y) + C(y)\Delta(x)(I - A\Delta(x))^{-1}B \quad (6.15)$$

where A, B are fixed fixed matrices and $C(y)$, $D(y)$, $\Delta(x)$ are matrix-valued affine mappings in y and x respectively. Let us stress that many robust control problems involve constraints that are naturally formulated in this fashion which hence forms an excellent starting point to construct relaxations [51, 214].

6.5.2 SOS relaxations based on S-procedure

We assume the constraint qualification of Theorem 6.3 to hold such that $\mathcal{G}$ is compact. One can then apply the full block S-procedure as described in Section 2.5.7 to infer that

$$\det(I - A\Delta(x)) \neq 0 \quad \text{and} \quad F(x, y) \succ 0 \quad \text{for all} \quad x \in \mathcal{G}$$

if and only if there exists a symmetric multiplier matrix P_{mult} such that

$$\begin{pmatrix} \Delta(x) \\ I \end{pmatrix}^T P_{\mathrm{mult}} \begin{pmatrix} \Delta(x) \\ I \end{pmatrix} \succ 0 \quad \text{for all} \quad x \in \mathcal{G}, \quad (6.16)$$

$$\begin{pmatrix} I & 0 \\ A & B \end{pmatrix}^T P_{\mathrm{mult}} \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} - \begin{pmatrix} 0 & I \\ C(y) & D(y) \end{pmatrix}^T \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \begin{pmatrix} 0 & I \\ C(y) & D(y) \end{pmatrix} \prec 0. \quad (6.17)$$

Although the uncertainties x enter the original problem in a rational fashion, we observe that (6.16) is quadratic in x and affine in P_{mult}. We can hence apply Theorem 6.1 to infer that (6.16) holds iff there exist $\epsilon > 0$ and SOS matrices $S(x)$, $S_0(x)$ with

$$\begin{pmatrix} \Delta(x) \\ I \end{pmatrix}^T P_{\mathrm{mult}} \begin{pmatrix} \Delta(x) \\ I \end{pmatrix} + \langle S(x), I_q \otimes G(x) \rangle_q - \epsilon I = S_0(x). \quad (6.18)$$

If constraining $S(x)$, $S_0(x)$ to be SOS with respect to fixed monomial basis vectors $u(x)$, $u_0(x)$ it is possible to turn (6.18) into a genuine LMI constraint (Section 6.3). By infimizing $c^T y$ over these LMIs combined with (6.17), one determines an upper bound on the optimal value p_{opt} of (6.1), which is again guaranteed to converge to p_{opt} if the length of the monomial basis vector $u(x)$ and $u_0(x)$ grow to infinity.

In practice the relaxation is often already exact for a short vectors $u(x)$ and $u_0(x)$. Since the suggested LMI relaxation falls in the general class as discussed in [172], we can directly apply all results in this reference in order to numerically verify exactness in practice. It is an interesting topic to relate this exactness test with the one in Section 4.5.6. This is however beyond the scope of this thesis.

6.5.3 Connections to standard relaxations

The presented SOS relaxations are extensions of various standard relaxations in the literature. We will discuss how the well-known D-scalings [140] can be recovered as special cases of the SOS relaxations.

Consider the uncertainty set

$$\mathcal{G} := \{ x \in \mathbb{R}^{q^2} \mid G(x) \preceq 0 \}$$

where $G(x) := \Delta(x)^T \Delta(x) - I_q$, $\Delta(x) := \sum_{i=1}^{q^2} x_i E_i$, and $\{E_1, \ldots, E_{q^2}\}$ is the standard basis of $\mathbb{R}^{q \times q}$. For arbitrary $p > 0$ and $\epsilon > 0$, (6.16) holds true if P_{mult} is given by

$$P(p, \epsilon) := \begin{pmatrix} -pI_q & 0 \\ 0 & (p + \epsilon)I_q \end{pmatrix}.$$

We will now show that the relaxations based on the scalings $P(p, \epsilon)$ for all $p > 0$ and $\epsilon > 0$ are a subset of the SOS relaxations. This is done by showing that $P_{\text{mult}} = P(p, \epsilon)$ render (6.18) feasible for all $p > 0$ and $\epsilon > 0$. For this purpose, we choose $S_0(x) := 0$ and

$$S(x) := \sum_{i=1}^{q^2} p \otimes E_i \otimes E_i.$$

These choices imply that the left-hand side of (6.18) equals

$$\begin{pmatrix} \Delta(x) \\ I_q \end{pmatrix}^T P(p, \epsilon) \begin{pmatrix} \Delta(x) \\ I_q \end{pmatrix} + \langle S(x), I_q \otimes G(x) \rangle_q - \epsilon I_q =$$

$$= \epsilon I_q + p \left(-\Delta(x)^T \Delta(x) + I_q \right) + p \sum_{i=1}^{q^2} \langle E_i \otimes E_i, I_q \otimes [\Delta(x)^T \Delta(x) - I_q] \rangle_q - \epsilon I_q =$$

$$= -p \left(\Delta(x)^T \Delta(x) - I_q \right) + p \sum_{i=1}^{q^2} E_i \text{Trace}(E_i \left(\Delta(x)^T \Delta(x) - I_q \right)) =$$

$$= -p \left(\Delta(x)^T \Delta(x) - I_q \right) + p \left(\Delta(x)^T \Delta(x) - I_q \right) = 0,$$

where we used the identity $\text{Trace}_q(A \otimes B) = A\text{Trace}(B)$ for arbitrary $A \in \mathbb{R}^{q \times q}$ and arbitrary real square matrix B, as is shown in Appendix A.2.1. This implies that (6.18) is satisfied.

6.6 Application

6.6.1 An example from the literature

We choose a variation of an example in [140, 172] and determine the infimal y with

$$\begin{pmatrix} y & f_a(x) \\ f_a(x) & y \end{pmatrix} \succ 0 \quad \text{for all} \quad x \in \mathcal{G} := \{ x \in \mathbb{R}^2 : g_j(x) \leq 0, \ j = 1, \ldots, 4 \}$$

for 20 equidistant values of $a \in [0.5, 1]$, where

$$f_a(x_1, x_2) := 1 - 2 \frac{a x_1^2 x_2}{2 - 2a x_2 + a x_1^2 x_2 - x_1^2} - \frac{x_2 \left(2a x_2 + x_1^2 + a x_1^2 x_2 - 2 \right)}{2 - 2a^2 x_2^2 - x_1^2 + x_1^2 a^2 x_2^2},$$

$$g_1(x) := -0.8 + x_1, \ g_2(x) = -0.7 - x_1, \ g_3(x) = -0.7 + x_2, \ g_4(x) = -0.65 - x_2.$$

Relax-ation	Line in Figure 6.1	Monomial bases				
		$u_0(x)^T$	$u_1(x)^T$	$u_2(x)^T$	$u_3(x)^T$	$u_4(x)^T$
a	–	$(1, x_1, x_2)$	$(1, x_1)$	$(1, x_1)$	1	1
b	$\cdots$	$(1, x_1, x_2)$	1	1	$\begin{pmatrix} 1 & x_1 \end{pmatrix}$	$\begin{pmatrix} 1 & x_2 \end{pmatrix}$
c	$--$	$(1, x_1, x_2)$	$(1, x_1, x_2)$	$(1, x_1, x_2)$	$(1, x_1, x_2)$	$(1, x_1, x_2)$
d	$--$	$(1, x_1, x_2, x_1 x_2)$	$(1, x_1, x_2)$	$(1, x_1, x_2)$	$(1, x_1, x_2)$	$(1, x_1, x_2)$

Table 6.1: SOS bases employed for relaxations.

Note that the optimal value equals $\sup_{x \in \mathcal{G}} |f_a(x)|$. Moreover, one easily determines a linear fractional representation of f_a with $\Delta(x) = \mathrm{diag}(x_1 I_2, x_2 I_2)$ and

$$
\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) = \left(\begin{array}{cccc|c} 0 & 1 & 0 & 1 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 2a & 0 & a & 0 & 0 \\ 0 & -2a & 0 & -a & 1 \\ \hline 0 & 0 & 0 & 1 & 1 \end{array} \right).
$$

Since the polytope $\mathcal{G}$ is compact, the constraint qualification is satisfied and we can either apply the relaxation in Section 6.5.2 (labeled by **a**, **b**, **c**) or we can use the direct approach as in Section 6.3 (labeled by **d**), both based on Theorem 6.3. The corresponding upper bounds have been computed for SOS matrices $S_j(x)$, $j = 0, 1, \ldots, 4$, with respect to the monomial bases $u_j(x)$, $j = 0, 1, \ldots, 4$, as given in Table 6.1.

Figure 6.1 depicts the computed upper bounds on p_{opt}, together with lower bounds that are obtained by constructing a worst-case uncertainty as described in [172]. Clearly **a**, **b** suffer from a relaxation gap, while both **c** and **d** are exact as confirmed by the exactness test of [172]. For comparison we show in Figure 6.2 lower and upper bounds computed with a relaxation as presented in [172], which is based on Pólya's Theorem. First, second and third order relaxations are shown in **a** (solid), **b** (dotted) and **c** (dashed) respectively. We observe that up to $\alpha < 0.85$ all three relaxations are exact. For $\alpha > 0.85$ the third order relaxation is exact.

6.6.2 Robust analysis for a helicopter model

As a second example, we consider the stability analysis of an LPV model of a closed-loop Vertical TakeOff and Landing (VTOL) helicopter [65,98]. The linearized longitudinal dynamic equations of the helicopter, after applying a static feedback law as in [98], are $\dot{z} = A(\delta)z$ where $\delta = \begin{pmatrix} \delta_1, \delta_2, \delta_3 \end{pmatrix}$ and

$$
A(\delta) = \begin{pmatrix} -0.0366 & -0.096 & 0.018 & -0.45 \\ 0.0482 & a_3(\delta) & 0.0024 & -4.02 \\ 0.10 & a_1(\delta) & -0.707 & a_2(\delta) \\ 0 & 0 & 1 & 0 \end{pmatrix}
$$

and $a_1(\delta) = 14.0 + 0.05\delta_1$, $a_2(\delta) = 1.42 + 0.01\delta_2$ and $a_3 = -18.2 - 0.0399\delta_3$. We analyze its stability for all uncertainties satisfying $\|\delta\| \leq \gamma$ and $|\dot{\delta}_k| \leq \rho$, $k = 1, 2, 3$ for fixed values of γ and ρ. We consider quadratic Lyapunov functions $z^T P(\delta) z$ where P depends affinely on δ:

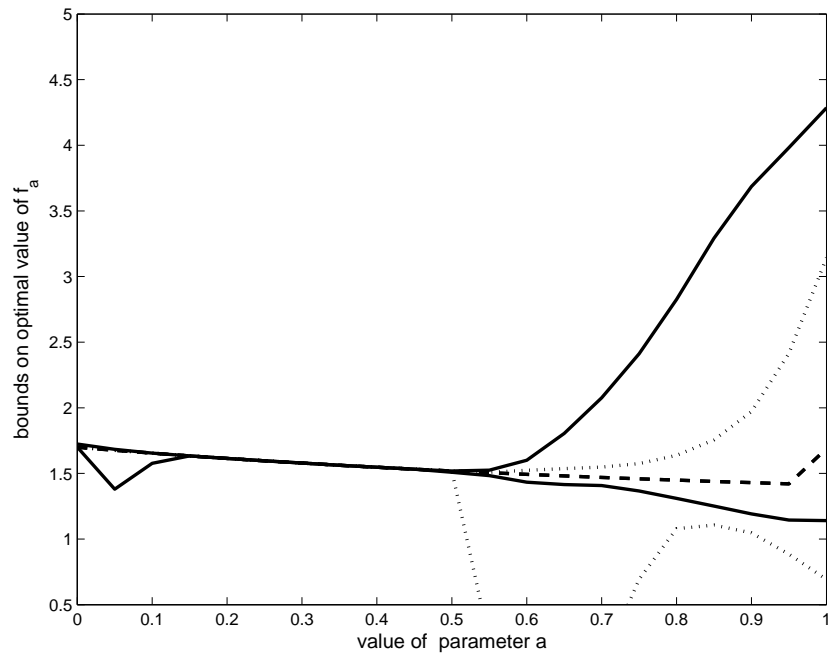$$
P(\delta) = \sum_{i=1}^{4} P_i m_i(\delta),
$$

Figure 6.1: Upper and lower bounds for SOS relaxations **a** (solid), **b** (dotted), **c** and **d** (dashed).
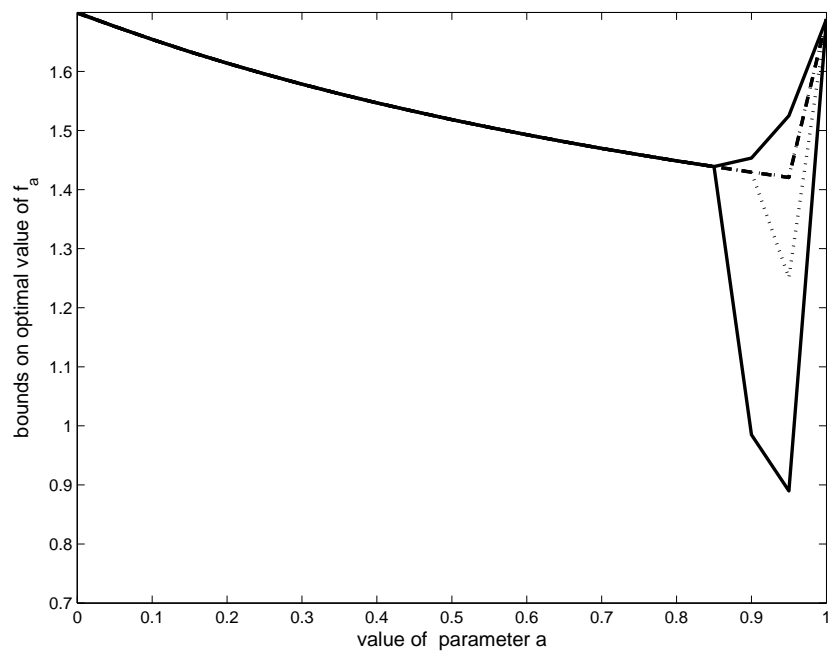

Figure 6.2: Upper and lower bounds for Pólya relaxations **a** (solid), **b** (dotted) and **c** (dashed).
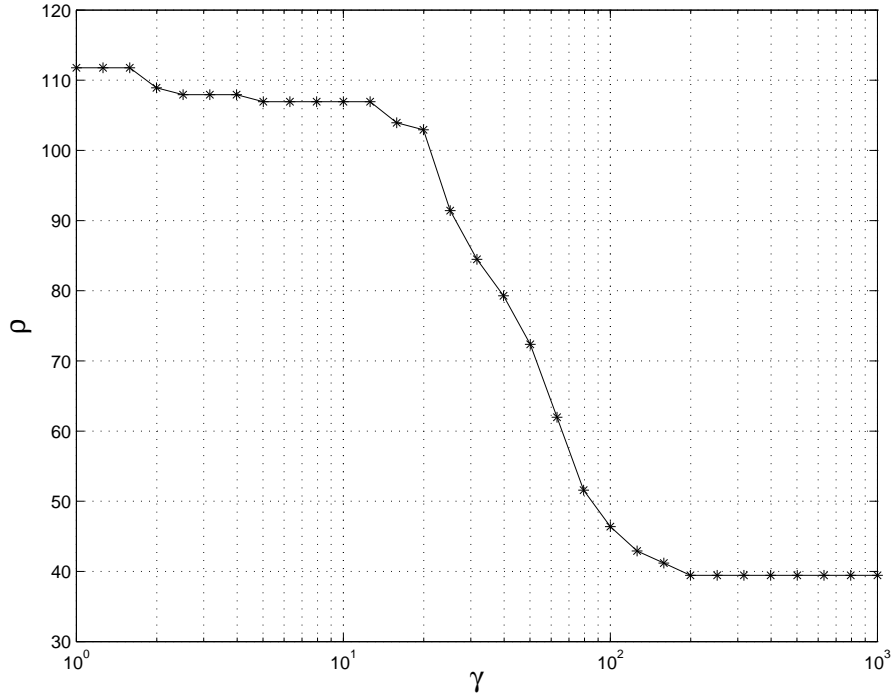
Figure 6.3: SOS lower bounds for the norm-bounded set.

where $m(\delta) = (1, \delta_1, \delta_2, \delta_3)$. Then the system is robustly stable if there exist $P_i = P_i^T \in \mathbb{R}^{4 \times 4}$, $i = 1, \ldots, 4$ such that

$$A(\delta)^T P(\delta) + P(\delta) A(\delta) + \sum_{k=1}^{3} \frac{\partial A(\delta)}{\partial \delta_k} \vartheta_k \prec 0$$

for all $\|\delta\| \leq \gamma$ and all $|\vartheta_k| \leq \rho$, $k = 1, 2, 3$ [46]. Hence with the definitions $x := \begin{pmatrix} \delta^T & \vartheta^T \end{pmatrix}^T$, $y := \begin{pmatrix} \operatorname{vec}(P_1)^T, \ldots, \operatorname{vec}(P_4)^T \end{pmatrix}^T$,

$$F(x, y) := A(\delta)^T P(\delta) + P(\delta) A(\delta) + \sum_{k=1}^{3} \frac{\partial A(\delta)}{\partial \delta_k} \vartheta_k,$$

$G = \operatorname{diag}(g_1, \ldots, g_4)$, $g_1(x) := \|\delta\|^2 - \gamma^2$ and $g_{1+i}(x) := |\vartheta_i| - \rho$, $i = 1, 2, 3$, feasibility of y in (6.1) implies robust stability.

We compute SOS relaxations of (6.1) with SOS bases $u_0(x) = \begin{pmatrix} 1 & x^T \end{pmatrix}^T$ and $u_i(x) = 1$, $i = 1, \ldots, 4$. Figure 6.3 shows the results. Note that the results cannot directly be compared with those in [98], [65] and [133], since we consider a norm-bounded instead of a polytopic set and the relaxations in those reference can only be applied to polytopic sets. This illustrates the additional flexibility of our framework, since it can be applied to any uncertainty set that admits a polynomial SDP description.

For comparison, we also computed bounds for the polytope $|\delta_k| \leq \gamma$, $k = 1, 2, 3$ that *can* be compared to [65], as shown in Figure 6.4. The figure shows that the bounds are similar[2].

---

[2]The values of $(\circ-)$ in the Figure 6.4 are adopted from [98]. They have been obtained by personal communication with T. Iwasaki.
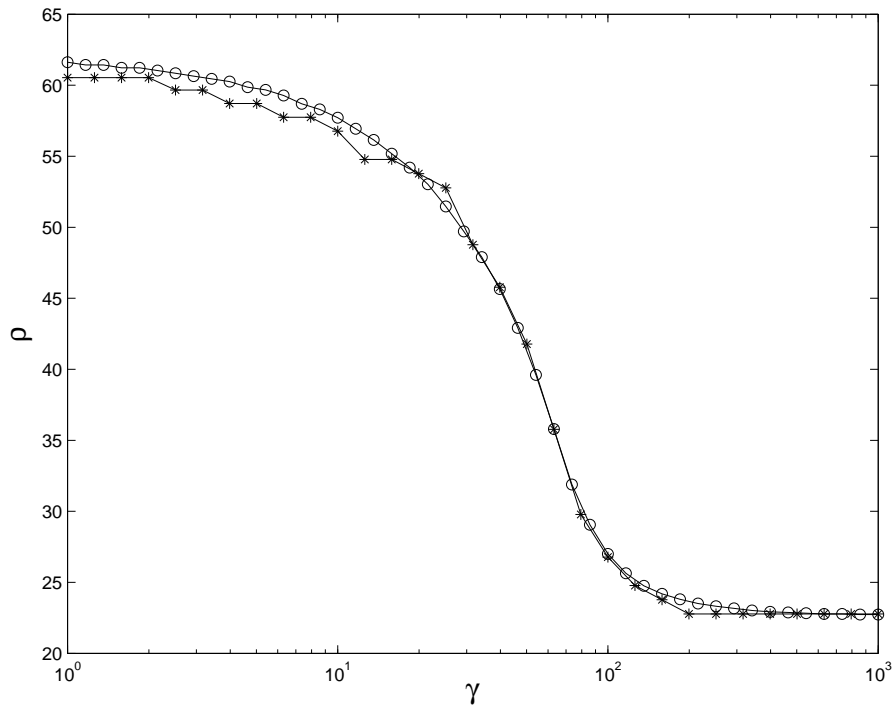
Figure 6.4: SOS lower bounds ($*-$) and Gahinet's lower bounds ($\circ-$) for the polytopic set.

6.7 Conclusions

For a general class of robust SDP problems with polynomial or rational dependence on the uncertainties, we have shown how to approximately compute upper bounds on the optimal value. These bounds are applicable to robust analysis problems with uncertainty regions described in terms of polynomial matrix inequalities. LMI regions and compact polytopes are special yet practically important instances of this class of uncertainty sets. Under the same constraint qualification as in Chapter 4 we have revealed how to construct a convergent sequence of relaxations with a guaranteed bi-quadratic growth of the number of variables in both $F(x)$ and $G(x)$. We have finally pointed out how one can systematically reduce the conservatism as present in the relaxations based on D-scalings or standard convex hull relaxations using the presented SOS decompositions. This concludes our work on sum-of squares relaxations. In the following chapters we proceed with the fixed-order controller synthesis problem, where our goal is the development of a controller synthesis technique that can be applied to design controllers for industrial servo-systems.

Chapter 7

Interior Point optimization

This and the following chapters focus on the development and implementation of fixed-order synthesis techniques for generalized plants with high McMillan degree, i.e. of degree 25 up until about 90. In this chapter we present an Interior Point (IP) algorithm and local optimality certificates, and apply both to a generalized plant with a McMillan degree of 27.

The success of IP methods to solve LMI problems is one of the key reasons to consider this technique to solve problems with BMIs arising from structured controller synthesis. Although the guaranteed efficiency of IP methods does not in general extend to BMIs, it is plausible that it exhibits practical good convergence properties for some problem instances. The numerical example presented at the end of this chapter gives some evidence to substantiate this statement.

In this chapter we present a nonlinear SDP approach to the $\mathcal{H}_\infty$ fixed-order controller synthesis problem. Interior Point methods have been applied to control problems by several authors. In the augmented Lagrangian method of Apkarian and Noll [4] the BMIs are replaced by LMIs and a bilinear equality constraint. This new problem is then solved by an augmented Lagrangian trust region method. In Goh *et. al.* [71] it has been pointed out that Interior Point methods may be promising, although no actual numerical implementations were discussed. Leibfritz and Mostafa [121] proposed an Interior Point trust region method for a special class of nonlinear Semi-Definite Programming (SDP) problems. The method is a sequential minimization method of a logarithmic barrier function subject to a nonlinear matrix equality constraint. To the best of our knowledge these methods have only been applied successfully to fixed-order controller synthesis for systems with small state dimensions $n = 8$. This is significantly smaller than the number of states 27 of the plant to which we apply our method in Section 7.4.

In Section 7.1 we explain the interior-point algorithm, which is based on the work of Jarre [105]. The non-convexity of the BMI constraints necessitates extra measures to guarantee local convergence, if compared to IP methods for LMIs. We do this using a curved line-search, which searches simultaneously for the optimal trust-region size and step length. The additional computational time that we spend on this line-search is often justified, since it reduces the number of computationally expensive evaluations of the Hessian of the barrier function.

To the best of our knowledge our paper [93] in 2003 is the first reported application of this algorithm to $\mathcal{H}_\infty$-optimal fixed-order controller synthesis[1]. Furthermore, our contri-

---

[1] In the paper of Kocvara *et. al.* [112], that appeared in 2004, $\mathcal{H}_\infty$-optimal controllers are computed by the optimization software PENNON. The software [113] is based on an augmented Lagrangian method.

butions to this algorithm are:

(i.) a transparent description of the Hessian of the barrier function that is used in the corrector step

(ii.) a computation method for the trust region steps using a fixed number of Sylvester equations

(iii.) an application of the method to a practical control design problem with a generalized plant of McMillan degree 27

(iv.) a derivation of optimality conditions as discussed in the sequel

Item (ii.) reduces the computation time for generalized plants with large McMillan degree. The corresponding Sylvester equations in Section 7.2 are derived by exploiting the control-theoretic characteristics in the problem.

Under mild conditions the algorithm converges to a point that satisfies the first order necessary optimality conditions. This motivates to check second order optimality conditions of the computed solution, which we derive in Section 7.3 in terms of the original BMI problem. These conditions were first published in our paper [93], are based on elementary arguments if compared to [182, 24] and are 'sharper' than those in [60].

In Section 7.4 we apply the method to synthesizing a controller for a 27$^{\text{th}}$ generalized plant model of an active suspension system. We compare its performance with two state-of-the-art fixed-order controller synthesis techniques: balanced controller reduction and the cone complementarity method as discussed in Section 3.1.1 and 3.2.1 respectively.

7.1   Curved line-search Interior Point method

7.1.1   Preliminaries

In this section we present a Curved Line-search Interior Point (CLIP) method to solve the fixed order $\mathcal{H}_\infty$ problem based on the work of Jarre [105]. In Section 2.3 it was shown how to convert the dynamic controller synthesis problem into a static output feedback (SOF) problem where the controller is the matrix

$$K := \left(\begin{array}{cc} A_K & B_K \\ C_K & D_K \end{array} \right).$$

Let us denote the number of variables in K by $m_{\text{c}} := (n_{\text{c}}+m_2)(n_{\text{c}}+p_2)$. The SOF problem is formulated as minimizing γ subject to

$$\mathcal{B}(\gamma, X, K) := -\left(\begin{array}{ccc} A_{\text{cl}}(K)^T X + X A_{\text{cl}}(K) & X B_{\text{cl}}(K) & C_{\text{cl}}(K)^T \\ B_{\text{cl}}(K)^T X_{\text{cl}} & -\gamma I & D_{\text{cl}}(K)^T \\ C_{\text{cl}}(K) & D_{\text{cl}}(K) & -\gamma I \end{array} \right) \succ 0 \qquad (7.1)$$

and

$$X \succ 0, \qquad (7.2)$$

where $(A_{\text{cl}}(K), B_{\text{cl}}(K), C_{\text{cl}}(K), D_{\text{cl}}(K))$ are defined in (2.8). We introduce the following constraints to restrict the variables in size:

$$X \prec \rho_X I_{n+n_c}, \quad \|K\|_{\text{F}} < \rho_K, \qquad (7.3)$$

114

where $\rho_X \in \mathbb{R}$ and $\rho_K \in \mathbb{R}$ are large positive constants. These extra constraints prevent numerical difficulties due to variables with very large values. ρ_X and ρ_K are chosen by a factor 1000 larger than the largest eigenvalue of the initial matrix X and the Frobenius norm of the initial matrix K, respectively.

The optimization problem considered in this chapter is

$$\begin{array}{ll} \text{minimize} & \gamma \\ \text{subject to} & \mathcal{B}(\gamma, X, K) \succ 0, \quad 0 \prec X \prec \rho_X I_{n+n_c}, \quad \|K\|_{\mathrm{F}} < \rho_K. \end{array} \tag{7.4}$$

Like in the previous chapters we combine the set of decision variables compactly into $x = (\gamma, X, K)$. The vector x is an element of the vector space $\mathcal{X}$:

$$\mathcal{X} := \left\{ x = (\gamma, X, K) \,\middle|\, \gamma \in \mathbb{R}, K \in \mathbb{R}^{(m_2+n_c) \times (p_2+n_c)}, \; X \in \mathcal{S}^{n+n_c} \right\}. \tag{7.5}$$

And the feasible set $\mathcal{G}$ of the minimization in (7.4) is:

$$\mathcal{G} := \left\{ x = (\gamma, X, K) \in \mathcal{X} \,\middle|\, \mathcal{B}(\gamma, X, K) \succ 0, \; 0 \prec X \prec \rho_X I, \; \|K\|_{\mathrm{F}} < \rho_K \right\}. \tag{7.6}$$

Due to its nice properties as described in Section 2.5.2, we employ the following logarithmic barrier function:

$$\phi(x, \mu) = \frac{\gamma}{\mu} - \log(\det(\mathcal{B}(\gamma, X, K))) - \log(\det(X)) - \log(\det(\rho_X I_{n+n_c} - X)) - \log(\rho_K - \|K\|_{\mathrm{F}}^2),$$

where $\mu \in \mathbb{R}$ is a barrier parameter as defined in Section 2.5.2. Since the domain of the log-barrier function is the interior of the feasible region $\mathcal{G}$, the Interior Point algorithm must be initialized with a feasible point. This issue is discussed in Section 7.1.6. The log-barrier function

$$\phi(x, \mu) = \frac{x_1 + x_2}{\mu} - \log\left(\det(-G(x))\right) \tag{7.7}$$

is illustrated in Figure 7.1, where $\mu = 1000$ and

$$G(x) := \begin{pmatrix} -1 + 3.7 x_1 x_2 & -x_1 \\ -x_1 & -3.7 + x_1^2 + x_2^2 \end{pmatrix} \preceq 0.$$

7.1.2 Algorithm outline

The method consists of an outer loop with two steps: a corrector and a predictor step. In the corrector step the barrier function is minimized for a fixed barrier parameter μ. In the predictor step the barrier parameter μ is reduced and x is modified. The minimization of the barrier function in the corrector step is the inner loop of the algorithm. We use x_k^l to denote the decision variable at the lth inner iteration and k^{th} outer iteration. The following steps, which will be made more precise in Sections 7.1.3 and 7.1.4, can therefore be distinguished:

- corrector: given (x_k^0, μ_k) generate a sequence x_k^l, $l \in \{1, 2, \ldots\}$ to iteratively minimize the barrier $\phi(x, \mu_k)$ with respect to x until an approximate minimum is reached,

- predictor: given (x_k^l, μ_k) compute a new pair (x_{k+1}^0, μ_{k+1}) such that the barrier parameter is reduced as much as possible, while at the same time x_{k+1}^0 is not 'too far' from the minimum of the barrier $\phi(x, \mu_{k+1})$.
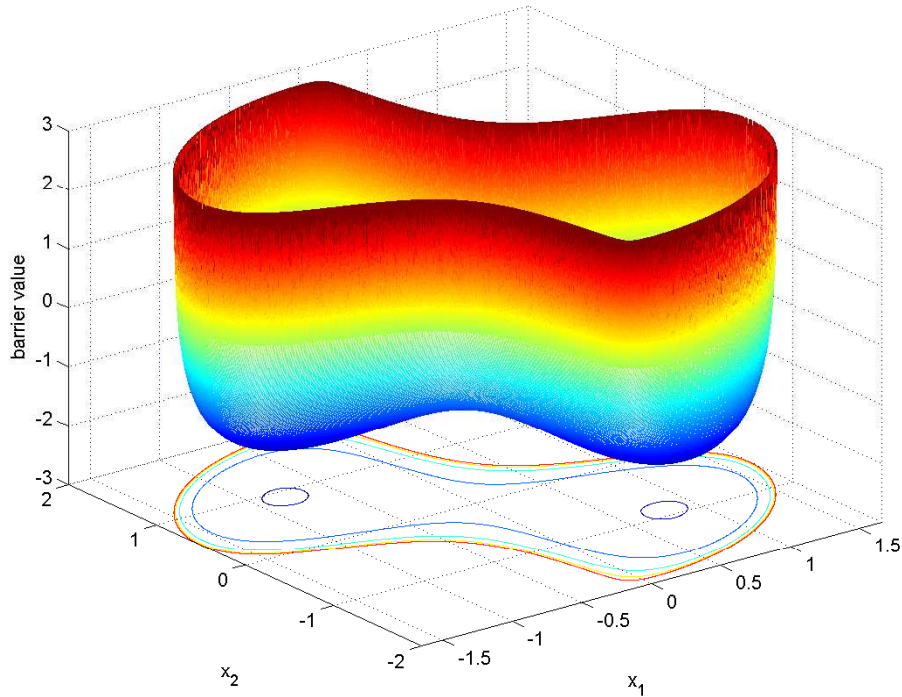
Figure 7.1: Log-barrier function for a non-convex set

Example 7.1 *In Figure 7.2 we illustrate these steps for the barrier function $\phi(x, \mu)$ as in (7.7). Starting with $\mu_1 = 100$ and $x_1 = (0, 0)$ we applied the following interior point scheme for $k = 1, \ldots, 19$*

- *minimize the barrier function $\phi(x_k, \mu_k)$ with respect to x with the algorithm as discussed in the next section. Denote the final iterate by x_{k+1}*

- *set $\mu_{k+1} = \frac{1}{2}\mu_k$*

This is a slightly simplified version of the IP algorithm that will be presented in Algorithm 7.2. The iterates x_k, $k = 1, \ldots, 20$ are shown in Figure 7.2. The infeasible region $\{x \in \mathbb{R}^2 | \ G(x) \succ 0\}$ is the solidly filled area. For the iterates with indices $k \in \{1, 2, 8, 10, 20\}$ the value of the barrier parameter μ_k is shown. In Figure 7.3 the same sequence x_k, $k = 1, \ldots, 20$ is shown, together with a sequence $\overline{x}_k$, $k = 1, \ldots, 20$ obtained in the same fashion, but with a slightly different initial vector $\overline{x}_0 = (-0.01, 0.01)$. This illustrates that for different initial points the Interior Point method may converge to completely different locally optimal points.

7.1.3 Corrector step

In the corrector step the barrier function is minimized for fixed $\mu = \mu_k$. The usual linear system of the centering Newton step, which minimizes a local quadratic approximation of the barrier function, is replaced by a trust region method. Direct application of Newton's method may fail because the problem is possibly nonconvex. For the trust region method the first and second order derivatives of the barrier function have to be known, as will be clarified in the sequel. They can be computed analytically, see Appendix A.4.1. In Section 7.2 we will present a method for computing the trust-region search steps using the inherent
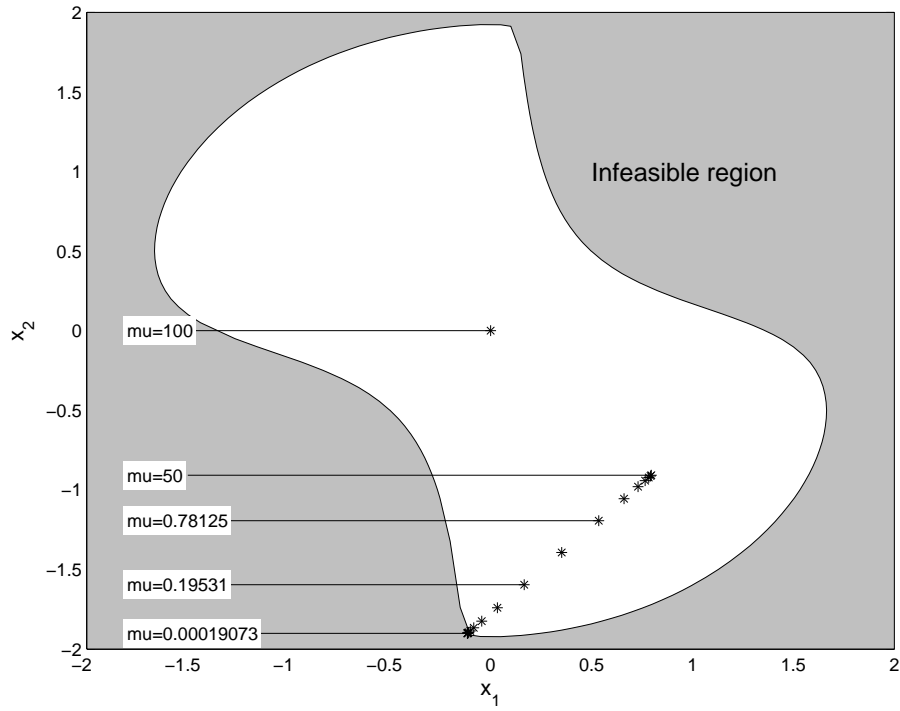
116

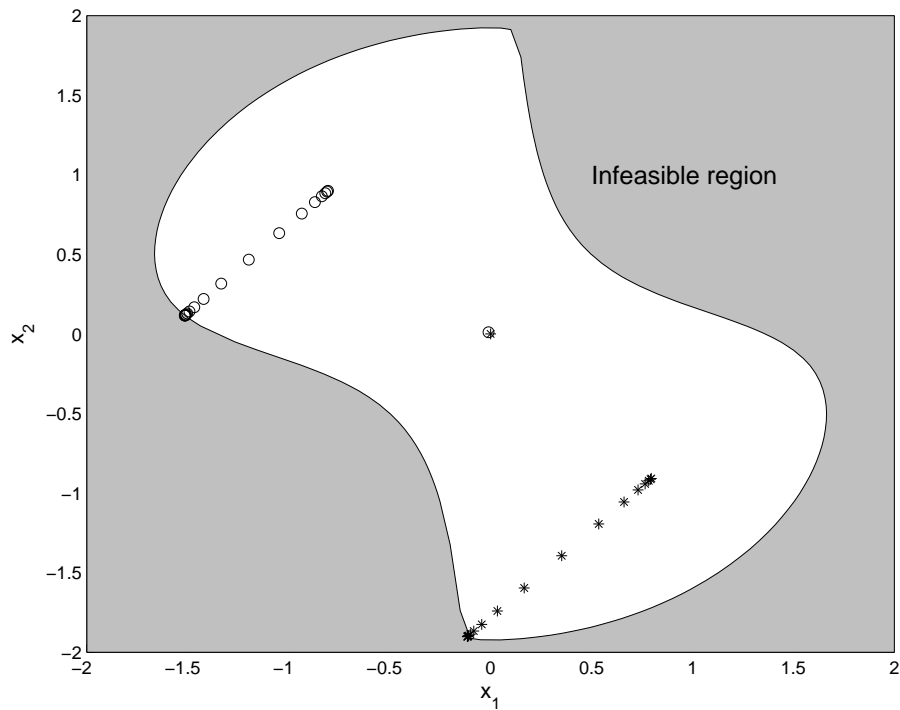Figure 7.2: Iterates after corrector steps and their barrier parameter value μ.

Figure 7.3: Iterates after corrector steps with initial guesses $x_0 = (0,0)$ ($*$) and $\overline{x}_0 = (-0.01, 0.01)$ ($\circ$).

Lyapunov-like structure of the underlying equations. In this section we restrict ourselves to the following scalarization technique. Stack the decision variables $x = (\gamma, X, K)$ into one vector $\tilde{x} := \mathrm{ve}(x) \in \mathbb{R}^m$, where $m = 1 + m_\mathrm{c} + \frac{1}{2}(1 + n + n_\mathrm{c})(n + n_\mathrm{c})$ and the linear bijection $\mathrm{ve} : \mathbb{R} \times \mathcal{S}^{n+n_\mathrm{c}} \times \mathbb{R}^{m_2+n_\mathrm{c} \times p_2+n_\mathrm{c}} \to \mathbb{R}^m$ is defined by:

$$\mathrm{ve}\,(\gamma, X, K) := \begin{pmatrix} \gamma \\ \mathrm{svec}(X) \\ \mathrm{vec}(K) \end{pmatrix} \tag{7.8}$$

and

$$\begin{aligned}
\mathrm{svec}(X) &:= \begin{bmatrix} X_{11}, X_{12}, \ldots, X_{1n}, X_{22}, X_{23}, \ldots, X_{2(n+n_\mathrm{c})}, \ldots, X_{(n+n_\mathrm{c})(n+n_\mathrm{c})} \end{bmatrix}^T, \\
\mathrm{vec}(K) &:= \begin{bmatrix} K_{11}, K_{21}, \ldots, K_{(n_\mathrm{c}+p_2)1}, K_{12}, K_{22}, \ldots, K_{(n_\mathrm{c}+p_2)2}, \ldots, K_{(n_\mathrm{c}+m_2)(n_\mathrm{c}+p_2)} \end{bmatrix}^T.
\end{aligned}$$

The mapping $\mathcal{B}$ and the barrier function ϕ in terms of the stacked variable $\tilde{x}$ are denoted by $\tilde{\mathcal{B}} := \mathcal{B} \circ \mathrm{ve}^{-1}$ and $\tilde{\phi} := \phi \circ \mathrm{ve}^{-1}$ respectively. For a given $\tilde{x}_k^l \in \mathbb{R}^N$, the algorithm computes the gradient $\nabla_{\tilde{x}} \tilde{\phi}(\tilde{x}_k^l, \mu_k)^T := \left[\partial_{\tilde{x}} \tilde{\phi}(\tilde{x}_k^l, \mu_k) \right]^T \in \mathbb{R}^N$ and the Hessian $\nabla_{\tilde{x}}^2 \tilde{\phi}(\tilde{x}_k^l, \mu_k) := \partial_{\tilde{x}}^2 \tilde{\phi}(\tilde{x}_k^l, \mu_k) \in \mathcal{S}^N$ of $\tilde{\phi}(\tilde{x}_k^l, \mu_k)$ using the Kronecker product as described in Section 7.2.3.

The Hessian may be indefinite, such that a Newton step may point into a 'wrong' (ascending) direction. To prevent wrong search steps we restrict them to a ball with radius ρ, the *trust-region*. We compute a search direction to minimize the barrier function $\tilde{\phi}$ by solving the following trust region problem:

$$\min_{s^T s \leq 2\rho^2} g^T s + \frac{1}{2} s^T H s, \tag{7.9}$$

where $H := \nabla_{\tilde{x}}^2 \tilde{\phi}(\tilde{x}_k^l, \mu_k)$, $g := \nabla_{\tilde{x}} \tilde{\phi}(\tilde{x}_k^l, \mu_k)$ and $\rho > 0$ denotes the trust region radius. This is a Quadratically constrained Quadratic Program (QQP). Using the S-procedure (as presented in Lemma 2.17) we conclude that the optimal value $p_\mathrm{opt} \leq 0$ of (7.9) equals that of the LMI problem

$$\begin{aligned}
&\text{supremize} && t \\
&\text{subject to} && \begin{pmatrix} H + \lambda I & g \\ g^T & -2t - \lambda\rho^2 \end{pmatrix} \succeq 0, \quad \lambda \geq 0, \quad \lambda \in \mathbb{R}
\end{aligned} \tag{7.10}$$

In Appendix A.4.4 it is shown that for every $\rho > 0$ the optimum of (7.10) is attained by unique $t(\rho)$ and $\lambda(\rho)$. Under mild conditions (see Jarre [105, 179]) $\lambda(\cdot)$ is continuous and for every trust region size $\rho \in (0, \infty)$ the search direction

$$\tilde{d}_\mathrm{c}(\lambda(\rho)) := - (H + \lambda(\rho)I)^{-1} g, \tag{7.11}$$

is the unique solution of the QQP in (7.9). If $\|g\| > 0$, this step direction has the following additional properties:

1. $\lim_{\rho \to 0} \frac{\tilde{d}_\mathrm{c}(\lambda(\rho))}{\|\tilde{d}_\mathrm{c}(\lambda(\rho))\|} = -\frac{g}{\|g\|}$ (steepest descent direction),

2. $\lim_{\rho \to \infty} \frac{\tilde{d}_\mathrm{c}(\lambda(\rho))}{\|\tilde{d}_\mathrm{c}(\lambda(\rho))\|} = -\frac{H^{-1}g}{\|H^{-1}g\|}$, if $H \succ 0$ (Newton step).

The optimal size of the trust region ρ is searched at each iteration, such that the algorithm can perform full Newton steps if the Hessian is positive definite. If this is not the case, the trust region must be sufficiently small to guarantee that the search direction is a descent step. For this purpose $\lambda(\rho)$ must be larger than or equal to the smallest negative eigenvalue of the Hessian, i.e., it must satisfy

$$\lambda(\rho) \geq \underline{\lambda} := \max\left\{0, -\lambda_{\min}(H)\right\},$$

where $\lambda_{\min}(H)$ denotes the smallest eigenvalue of H. Since there is a one-to-one correspondence between trust region size ρ and $\lambda(\rho)$ we can and will directly search over $\lambda \geq \underline{\lambda}$ instead of ρ.

Remark. This discussion reveals that the trust-region approach described above is nothing else than a (standard) Levenberg-Marquardt regularization of the Hessian (see e.g. [17]). However, one of the differences of our algorithm compared to conventional optimization schemes is the explicit search over the trust-region size, as discussed in the sequel.

In addition to the search over the optimal regularization parameter λ we search simultaneously for the best step-size $\alpha > 0$ along the direction $\tilde{d}_{\mathrm{c}}(\lambda)$. This leads to a search in the two variables λ and α:

$$(\lambda_*, \alpha_*) \in \arg\inf_{\lambda > \underline{\lambda},\ \alpha > 0} \tilde{\phi}\left(\tilde{x}_k^l + \alpha \tilde{d}_{\mathrm{c}}(\lambda), \mu_k\right). \tag{7.12}$$

Since the minimization in (7.12) has two decision variables λ and α, it is called a 'plane search' as opposed to a *line search* over a single decision variable. A plane search is numerically more expensive than a line search. In our experience the plane search often reduces the number of outer iterations if compared to the line search. The overall computation time often is smaller with the plane search, because the very expensive computation of the Hessian is done once every outer iteration. The next QQP iterate $\tilde{x}_k^{l+1}$ is equal to:

$$\tilde{x}_k^{l+1} = \tilde{x}_k^l + \alpha_* \tilde{d}_{\mathrm{c}}(\lambda_*). \tag{7.13}$$

where α_* and λ_* are given by (7.12).

7.1.4 Predictor step

At the predictor step one could simply reduce the barrier parameter μ in (7.7) and continue with the corrector step. We try to prevent that the corrector needs many QQP-steps after the update of μ, by computing a step in $\tilde{x}$ based on a convex approximation of the feasible set by a *Dikin ellipsoid* [104, 157]. The predictor direction $\tilde{d}_{\mathrm{p}}$ is defined as the direction minimizing γ over the Dikin ellipsoid. For a convex set $\mathcal{C}$ the Dikin ellipsoid at a point $\tilde{x}_*$ is given by:

$$E_\phi(\tilde{x}_*) = \left\{\tilde{x} \mid (\tilde{x} - \tilde{x}_*)^T \nabla^2 \tilde{\phi}(\tilde{x}_*)(\tilde{x} - \tilde{x}_*) \leq 1\right\}, \tag{7.14}$$

where $\tilde{\phi}(\cdot)$ is a barrier function for $\mathcal{C}$. The feasible set $\mathcal{G}$ given by Equation (7.6) and the barrier function $\tilde{\phi}$ are not necessarily convex, since $\tilde{\mathcal{B}}$ is a bilinear function. To resolve this, we replace the barrier by a convex one based on linearization of the constraints.

The derivative of $\mathcal{B}$ at $(\gamma_*, X_*, K_*) := x_* := \mathrm{ve}^{-1}(\tilde{x}_*)$ is

$$\partial\mathcal{B}(x_*)d_x := \mathcal{L}_\gamma(x_*)(d_\gamma) + \mathcal{L}_X(x_*)(d_X) + \mathcal{L}_K(x_*)(d_K), \tag{7.15}$$

where $\partial \mathcal{B}(x_*)d_x$ denotes the derivative of $\mathcal{B}$ at x_* in the direction $d_x := (d_\gamma, d_X, d_K)$. Furthermore, $\mathcal{L}_\gamma(x_*) : \mathbb{R} \to \mathcal{S}^{n+n_c+m_1+p_1}$, $\mathcal{L}_X(x_*) : \mathcal{S}^{n+n_c} \to \mathcal{S}^{n+n_c+m_1+p_1}$ and $\mathcal{L}_K(x_*) : \mathbb{R}^{(n_c+m_2)\times(n_c+p_2)} \to \mathcal{S}^{n+n_c+m_1+p_1}$ are defined by

$$\mathcal{L}_\gamma(x_*)d_\gamma := \begin{pmatrix} 0 & 0 & 0 \\ 0 & d_\gamma I_{m_1} & 0 \\ 0 & 0 & d_\gamma I_{p_1} \end{pmatrix} \tag{7.16}$$

$$\mathcal{L}_X(x_*)d_X := - \begin{pmatrix} I_{n+n_c} \\ 0 \\ 0 \end{pmatrix} d_X \begin{pmatrix} A_{\mathrm{cl}}(K_*)^T \\ B_{\mathrm{cl}}(K_*)^T \\ 0 \end{pmatrix}^T - \begin{pmatrix} A_{\mathrm{cl}}(K_*)^T \\ B_{\mathrm{cl}}(K_*)^T \\ 0 \end{pmatrix} d_X \begin{pmatrix} I_{n+n_c} \\ 0 \\ 0 \end{pmatrix}^T \tag{7.17}$$

$$\mathcal{L}_K(x_*)d_K := - \begin{pmatrix} X_*B_2 \\ 0 \\ D_{12} \end{pmatrix} d_K \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix}^T - \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix} d_K{}^T \begin{pmatrix} X_*B_2 \\ 0 \\ D_{12} \end{pmatrix}^T . \tag{7.18}$$

Let us linearize the constraints of (7.4) at $x_k^l = (\gamma_k^l, X_k^l, K_k^l)$:

$$\mathcal{B}(x_k^l) + \partial \mathcal{B}(x_k^l)d_x \succ 0, \quad X_k^l + d_X \succ 0, \quad X_k^l + d_X \prec \rho_X I_{n+n_c}, \quad 2\mathrm{Trace}(d_K{}^T K_k^l) < \rho_K^2. \tag{7.19}$$

The feasible set of (7.19) $\mathcal{G}_{\mathrm{lin}}$ is a *convex* approximation of $\mathcal{G}$. We therefore compute the Dikin ellipsoid of $\mathcal{G}_{\mathrm{lin}}$ to approximate $\mathcal{G}$ and use this set to find a suitable prediction step.

To this purpose, we use the following log-barrier function for $\mathcal{G}_{\mathrm{lin}}$:

$$\phi_{x_k^l}^{\mathrm{lin}}(d_x) := -\log\left(\det(\mathcal{B}(x_k^l) + \partial\mathcal{B}(x_k^l)d_x)\right) - \log\left(\det(X_k^l + d_X)\right) - $$
$$- \log\left(\rho_K^2 - 2\mathrm{Trace}(d_K{}^T K_k^l)\right) - \log\left(\det(\rho_X I - X_k^l - d_X)\right).$$

Let $\tilde{\phi}_{\tilde{x}_k^l}^{\mathrm{lin}}(\tilde{d}_x) := \phi_{\mathrm{ve}^{-1}(\tilde{x}_k^l)}^{\mathrm{lin}}(\mathrm{ve}^{-1}(\tilde{d}_x))$ be the corresponding barrier function in terms of the stacked variables $\tilde{x}_k^l$ and $\tilde{d}_x$. The predictor step direction $\tilde{d}_{\mathrm{p}}$ is then calculated as the direction minimizing:

$$\text{minimize } e_1^T \tilde{d}_{\mathrm{p}} \text{ over the Dikin ellipsoid: } \tilde{d}_{\mathrm{p}} \in E_{\tilde{\phi}_{\tilde{x}_k^l}^{\mathrm{lin}}}(0).$$

If $\nabla^2 \tilde{\phi}_{\tilde{x}_k^l}^{\mathrm{lin}}(0)$ is positive definite, the solution is given by

$$\tilde{d}_{\mathrm{p}} = -\frac{(\nabla^2 \tilde{\phi}_{\tilde{x}_k^l}^{\mathrm{lin}}(0))^{-1} e_1}{\sqrt{e_1^T (\nabla^2 \tilde{\phi}_{\tilde{x}_k^l}^{\mathrm{lin}}(0))^{-1} e_1}}.$$

The predictor step is then:

$$\tilde{x}_{k+1}^0 = \tilde{x}_k^l + \tilde{d}_{\mathrm{p}}.$$

The barrier parameter μ is updated as follows. Given a pair $(\tilde{x}_k^l, \mu_k)$ we set a target value $\mu_k^+ = \frac{\mu_k}{10}$. Then we compute the minimum $\tilde{x}_{k+1}^0$ of $\tilde{\phi}(\cdot, \mu_k^+)$ along the Dikin direction $\Delta\tilde{x}$ using a line search. We modify the target μ_k^+ by finding $\mu_{k+1} \le \frac{\mu_k}{2}$ such that $\|\nabla_{\tilde{x}}\tilde{\phi}(\tilde{x}_{k+1}^0, \mu_{k+1})\|$ is minimized. All this is done to reduce the number of QQP steps in the next corrector step.

7.1.5 Interior Point algorithm

The method can be summarized by the following algorithm adopted from [105].

Algorithm 7.2 *(Curved Line-search Interior Point method (CLIP))*
Initialization: let $\tilde{x} = \tilde{x}_0^0$, $\mu_0 = 1$. For $k = 0, 1, 2, \ldots$ do (outer loop):

1. *Corrector steps (inner loop):*
   Set $l = 0$. Do until convergence:

   (a) *Define $\tilde{d}_c(\lambda) = -(H + \lambda I_N)^{-1} g$ where $g = \nabla_{\tilde{x}} \tilde{\phi}(\tilde{x}_k^l, \mu_k)$ and $H = \nabla^2_{\tilde{x}} \tilde{\phi}(\tilde{x}_k^l, \mu_k)$.*

   (b) *Find the step lengths $\lambda_* \geq \max\left\{0, -\lambda_{\min}(\nabla^2_{\tilde{x}} \tilde{\phi}(\tilde{x}_k^l, \mu_k))\right\}$ and $\alpha_* > 0$ that minimize $\tilde{\phi}(\tilde{x}_k^l + \alpha \tilde{d}_c(\lambda), \mu_k)$.*

   (c) *Set $\tilde{x}_k^{l+1} = \tilde{x}_k^l + \alpha_* \tilde{d}_c(\lambda_*)$ and $l = l + 1$.*

2. *Predictor step:*

   (a) *Set $\tilde{d}_p = -\dfrac{(\nabla^2 \tilde{\phi}_{\tilde{x}_k^l}^{\lin}(0))^{-1} e_1}{\sqrt{e_1^T (\nabla^2 \tilde{\phi}_{\tilde{x}_k^l}^{\lin}(0))^{-1} e_1}}$.*

   (b) *Define $\mu_k^+ = \frac{\mu_k}{10}$.*

   (c) *Find the step length $t_* > 0$ that minimizes $\tilde{\phi}(\tilde{x}_k^l + t\tilde{d}_p, \mu_k^+)$ and set $\tilde{x}_{k+1}^0 = \tilde{x}_k^l + t_* \tilde{d}_p$.*

   (d) *Choose $\mu_{k+1} \in \arg\min_{\mu \leq \frac{\mu_k}{2}} \|\nabla \tilde{\phi}(\tilde{x}_{k+1}^0, \mu)\|_2$.*

Jarre [105] shows that the QQP steps in (7.13) with exact plane search return, under mild conditions, a point that satisfies the first order necessary conditions for optimality, i.e., the gradient of the barrier $\tilde{\phi}(\cdot, \mu_k)$ with respect to the first variable for fixed μ_k is zero at an accumulation point. Furthermore the outer loop of the Algorithm 7.2 converges to a point that satisfies the necessary first order optimality conditions of the original BMI problem of minimizing γ subject to (7.1), (7.2) and (7.3). If the Hessian is positive definite in one of the accumulation points (which implies local convexity), Jarre shows that the iterates converge superlinearly to this point[2].

7.1.6 Initialization

The IP algorithm requires an initial choice of the decision variables in the interior of the feasible region $\mathcal{G}$. Finding a feasible initial solution is called a *phase-one problem*. An initial stabilizing controller can often be found by closed-loop balanced reduction of the full order order $\mathcal{H}_\infty$-optimal controller, see Section 3.1.1.

If a stabilizing initial controller K_{init} is given, an initial guess $\tilde{x}_0^0$ can be constructed by solving a $\mathcal{H}_\infty$-norm analysis problem. Indeed for fixed $K = K_{\text{init}}$, minimizing γ subject to (7.1) and (7.2) is a feasible standard LMI problem. If $(\gamma_{\text{init}}, X_{\text{init}})$ is a solution to this problem, then $\tilde{x}_0^0 := \text{ve}(\gamma_{\text{init}}, X_{\text{init}}, K_{\text{init}})$ is a suitable initial guess, provided that $X_{\text{init}} \prec \rho I_{n+n_c}$ and $\|K\|_{\text{F}} < \rho_K$. As described in Section 2.2.4, finding a feasible pair $(\gamma_{\text{init}}, X_{\text{init}})$ by solving Riccati equations is often faster and more reliable.

---

[2]see Luenberger [125] for a definition of superlinear convergence.

If a stabilizing controller is not given and cannot be constructed, the phase-one problem

$$
\begin{aligned}
&\text{infimize} && t \\
&\text{subject to} && \mathcal{B}(\gamma, X, K) \succ -tI, && X \succ -tI \\
& && \rho_X I - X \succ -tI, && \|K\|_\mathrm{F}^2 - \rho_K < t
\end{aligned}
\tag{7.20}
$$

can also be solved by the IP algorithm, i.e., Algorithm 7.2. If its optimal value t_opt is strictly negative then the optimal solution of this problem is a feasible initial guess for the $\mathcal{H}_\infty$ problem. The minimization in (7.20) is a possibly non-convex BMI problem and there is no guarantee that the IP algorithm converges to the global optimum. The resulting value t_IP of the objective function at the final iteration might therefore be positive, even if the global optimal value t_opt is negative. In other words, if the IP algorithm is applied to the phase-one problem it might not find a stabilizing controller even though one exists.

7.1.7 Intermediate analysis steps

The kth corrector step is started with $x_k^0 = (\gamma_k^0, X_k^0, K_k^0)$ computed in the predictor step. An improved starting point can often be computed by the solution of the $\mathcal{H}_\infty$-analysis LMI problem for the fixed controller $K = K_k^0$: minimize γ subject to (7.1) and $X \succ 0$. The resulting pair γ_*, X_* together with the controller K_k^0 yields a candidate vector $x_* = (\gamma_*, X_*, K_k^0)$. If the corresponding value of the barrier function is smaller than that for x_{k+1}^0, we replace x_{k+1}^0 by x_* before we start the corrector step. This is an ad-hoc way to reduce the barrier function which works very well in practice.

7.1.8 Discussion

Under the conditions described in Section 7.1.5 the method converges superlinearly to a local optimum. This is intuitively appealing, since the corrector steps allow to compute (and perform) full Newton steps if the barrier function is locally convex. The algorithm can therefore behave like Newton's method close to local optima that have a positive definite Hessian. This is the motivation for spending much time to compute the full Hessian of the barrier function. As μ decreases, the optimal value in the corrector step gets closer and closer to the boundary of the feasible region, which implies that the local curvature of the barrier gets larger and the minimization problem of the barrier becomes increasingly ill-conditioned. For these ill-conditioned problems first-order methods like the conjugate gradient method or quasi-Newton methods [125] will converge very slowly. Furthermore if the underlying constrained problem has many inequalities of which only a few are almost active, first-order methods require many steps to gather enough information about the Hessian before they can make efficient search steps. This is even true if the constraints are convex [136]. For systems with high McMillan degree the number of constraints in the fixed-order synthesis problem, i.e. (7.1), (7.2) and (7.3) is indeed large. It is therefore reasonable to believe that the computation of the full Hessian is required to obtain good convergence of the corrector for small values of the barrier parameter μ.

The computation of the Hessian is expensive for fixed-order $\mathcal{H}_\infty$ synthesis for plants with high McMillan degree n, since the Hessian is a large matrix in $\mathcal{S}^{1+m_\mathrm{c}+\frac{1}{2}(n+n_\mathrm{c})(n+n_\mathrm{c}+1)}$ that is computed using Kronecker products as described in Section 7.2.3. However, based on the arguments above, it seems unwise to eliminate this evaluation of the Hessian totally. Instead, effort is made to reduce the number of times that it needs to be computed. As has been discussed, this is done by the introduction of the plane search, which requires some

more function evaluations than a one-dimensional search. The next iterate will in general have a lower value of the barrier function for a plane search than for a one-dimensional search. Spending some computation time in the plane search before computing a new Hessian will therefore be worth the effort for many optimization problems and in particular for the fixed-order $\mathcal{H}_\infty$-optimal control problem. In the next section we describe how the Hessian can be computed more efficiently for large n.

Since our algorithm converges locally to a point satisfying the first-order necessary conditions, it makes sense to check if the final iterate is indeed a local optimum. This can be done using LMI tests for second-order optimality as described in Section 7.3. Before we derive these optimality conditions, we will explain how we can compute the Newton direction.

7.2 Computation of the Newton direction

Let us fix a point $(\tilde{x}_k^l, \mu_k)$ at which we have determined the gradient $g := \nabla_{\tilde{x}}\tilde{\phi}(\tilde{x}_k^l, \mu_k)$ and the Hessian $H := \nabla_{\tilde{x}}^2\tilde{\phi}(\tilde{x}_k^l, \mu_k)$ of the barrier function. In the corrector step of Algorithm 7.2 we solve

$$(H + \lambda I)\tilde{d}_x = -g, \tag{7.21}$$

for various values of λ. Recall that if H is nonsingular the Newton direction for the barrier function is the solution to $H\tilde{d}_x = -g$. In Section 7.2.1 we explain that solving (7.21) or $H\tilde{d}_x = -g$ comes down to solving a set of non-standard Sylvester equations. In Section 7.2.2 we argue why this set of equations cannot be solved using standard solvers. Then we show how a solution can be found using Kronecker products in Section 7.2.3. Since this method is inefficient for large state dimension n, we present in Section 7.2.4 a more efficient method by decomposing the search for $\tilde{d}_x$ into a sequence of standard Sylvester equations.

7.2.1 Newton step as solution to three matrix equations

To compute a Newton step (d_X, d_K, d_γ) at $x = (\gamma, X, K)$ we introduce the mappings $\mathcal{L}_\gamma^*(x) : \mathcal{S}^{n+n_c+m_1+p_1} \to \mathbb{R}$, $\mathcal{L}_X^*(x) : \mathcal{S}^{n+n_c+m_1+p_1} \to \mathcal{S}^{n+n_c}$ and $\mathcal{L}_K^*(x) : \mathcal{S}^{n+n_c+m_1+p_1} \to \mathbb{R}^{(n_c+m_2)\times(n_c+p_2)}$

$$\mathcal{L}_\gamma^*(x)d_W := \mathrm{Trace}\left(d_W\begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{m_1} & 0 \\ 0 & 0 & I_{p_1} \end{pmatrix}\right) \tag{7.22}$$

$$\mathcal{L}_X^*(x)d_U := -\begin{pmatrix} A_{\mathrm{cl}}(K)^T \\ B_{\mathrm{cl}}(K)^T \\ 0 \end{pmatrix}^T d_U \begin{pmatrix} I_{n+n_c} \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} I_{n+n_c} \\ 0 \\ 0 \end{pmatrix}^T d_U \begin{pmatrix} A_{\mathrm{cl}}(K)^T \\ B_{\mathrm{cl}}(K)^T \\ 0 \end{pmatrix} \tag{7.23}$$

$$\mathcal{L}_K^*(x)d_V := -2\begin{pmatrix} C_2 & D_{21} & 0 \end{pmatrix} d_V \begin{pmatrix} XB_2 \\ 0 \\ D_{12} \end{pmatrix}$$

These mappings are *adjoints* [124] of $\mathcal{L}_\gamma(x)$, $\mathcal{L}_X(x)$ and $\mathcal{L}_K(x)$ respectively, i.e. the linear operators represented in the appropriate input and output bases by the transposes of the matrix-representations of $\mathcal{L}_\gamma(x)$, $\mathcal{L}_X(x)$ and $\mathcal{L}_K(x)$ respectively. This is explained in Appendix A.4.3. Additionally we introduce the following mappings $\mathcal{L}_{XK}(x)$:

$\mathbb{R}^{(n_c+m_2)\times(n_c+p_2)} \to \mathcal{S}^n$ and $\mathcal{L}_{KX}: \mathcal{S}^n \to \mathbb{R}^{p_2 \times m_2}$

$$\mathcal{L}_{XK}(x)(d_K) := B_2 d_K \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix}^T \mathcal{B}(x)^{-1} \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix}^T \mathcal{B}(x)^{-1} \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix} d_K^T B_2^T \tag{7.24}$$

and

$$\mathcal{L}_{KX}(x)(d_X) := 2 \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix}^T \mathcal{B}(x)^{-1} \begin{pmatrix} d_X B_2 \\ 0 \\ 0 \end{pmatrix}, \tag{7.25}$$

to write $\partial_K \partial_X \phi(\gamma, X, K, \mu)$ as

$$\partial_K \partial_X \phi(\gamma, X, K, \mu)(d_\gamma, d_K, d_X) = \mathrm{Trace}\left(\mathcal{B}(x)^{-1}\mathcal{L}_X(x)d_X \mathcal{B}(x)^{-1}\mathcal{L}_K(x)d_K\right) + $$
$$+ \mathrm{Trace}(\mathcal{L}_{XK}(x)(d_K)d_X),$$

see Appendix A.4.1 for a derivation. Solving (7.21) in terms of the matrix representation of $d_x = (d_\gamma, d_X, d_K)$ is equivalent to solving the following Sylvester equations for (d_γ, d_X, d_K):

$$\mathcal{L}_\gamma^*(x)\left(\mathcal{B}(x)^{-1}\left[\mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K) + \mathcal{L}_\gamma(x)(d_\gamma)\right]\mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1}\right) + \frac{1}{\mu} + \lambda d_\gamma = 0, \tag{7.26}$$

$$\mathcal{L}_X^*(x)\left(\mathcal{B}(x)^{-1}\left[\mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K) + \mathcal{L}_\gamma(x)(d_\gamma)\right]\mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1}\right) + \mathcal{L}_{XK}(x)(d_K) -$$
$$- X^{-1} + X^{-1}d_X X^{-1} + (\rho_X I - X)^{-1} + (\rho_X I - X)^{-1}d_X(\rho_X I_{n_X} - X)^{-1} + \lambda d_X = 0, \tag{7.27}$$

and

$$\mathcal{L}_K^*(x)\left(\mathcal{B}(x)^{-1}\left[\mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K) + \mathcal{L}_\gamma(x)(d_\gamma)\right]\mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1}\right) + \mathcal{L}_{KX}(x)(d_X) +$$
$$+ \frac{2}{\rho_K - \|K\|_F^2}(K + d_K)^T + \frac{4\mathrm{Trace}(K^T d_K)}{(\rho_K - \|K\|_F^2)^2}K^T + \lambda d_K = 0, \quad (7.28)$$

where $\mathcal{L}_\gamma(x)(d_\gamma)$, $\mathcal{L}_X(x)(d_X)$ and $\mathcal{L}_K(x)(d_K)$ are defined in (7.16), (7.17) and (7.18) respectively. See Appendix A.4.2 for a derivation of these equations.

The set of equations has a 'Lyapunov-like' characteristic. It is not straightforward to exploit this Lyapunov-like characteristics in computations, because the number of coefficients of each unknown (d_γ, d_K, d_X) is larger than that of the standard Sylvester or Lyapunov equations, as discussed in the next section.

7.2.2 Standard Sylvester techniques

Equations (7.26), (7.27) and (7.28) cannot be solved using the so-called 'standard Sylvester techniques' as described in e.g. [190, 66, 107, 108] since the number of terms is too large. The standard techniques can for instance solve equations of the type

$$F_1 d_X F_2 + F_3 d_X F_4 = F_0 \tag{7.29}$$

where $d_X \in \mathbb{R}^{n\times n}$ is the unknown and F_i, $i = 0, \ldots, 4$ are the data matrices, which are for simplicity all assumed to be elements of $\mathbb{R}^{n\times n}$. One of the well-known techniques is

an extension of the Bartels-Stewart algorithm as discussed in [66]. In this algorithm the coefficient matrices F_i, $i = 1, \ldots, 4$ are first transformed using orthogonal transformation matrices Q_1, Q_2, Z_1 and Z_2 into upper triangular and quasi-upper-triangular matrices using the QZ algorithm. Left and right multiplication by Q_1 and Q_2^T respectively transforms (7.29) to the following set of equivalent equations:

$$
\begin{aligned}
Q_1(F_1 d_X F_2 + F_3 d_X F_4)Q_2^T = & \\
= (Q_1 F_1 Z_1)(Z_1^T d_X Z_2)(Z_2^T F_2 Q_2^T) & + (Q_1 F_3 Z_1)(Z_1^T d_X Z_2)(Z_2^T F_4 Q_2^T) = \\
& = T_1 Y S_2^T + S_1 Y T_2^T = Q_1 F_0 Q_2^T
\end{aligned}
$$

where $T_1 = Q_1 F_1 Z_1$ and $T_2 = Q_2 F_4^T Z_2$ are upper triangular, $S_1 = Q_1 F_3 Z_1$ and $S_2 = Q_2 F_2^T Z_2$ are *quasi-upper-triangular* and $Y = Z_1^T d_X Z_2$, where a quasi-upper-triangular matrix is a block-upper-triangular matrix with 2×2 and/or 1×1 blocks on the diagonal. A solution for Y can be found by solving a sequence of 2×2 and/or 1×1 systems of equations, starting from the right lower corner of $Q_1 F_0 Q_2^T$. The worst-case computation time is $72n^3$ [66] floating point operations (flops). By a flop is meant a single elementary arithmetic operation like scalar addition, substraction, multiplication or division [72]. For large n, the number of flops is relatively small if compared to that for computing the Newton step using scalarization, as explained in the next section.

The technique described in this section cannot straightforwardly be applied to solve (7.26)-(7.28) [12]. In the next two sections we present two different ways to solve these equations. In Section 7.2.3 we ignore the matrix-structure and solve these equations by the usually not very efficient technique of scalarization, whereas in Section 7.2.4 we exploit the matrix structure and solve the equations by a (fixed) number of Sylvester equations.

7.2.3 Computation of trust regions steps by scalarization

This section discusses the solution of (7.26), (7.27) and (7.28) by stacking the variables γ, X and K into a large vector $\tilde{x}$ as in (7.8). The resulting equations are $(H + \lambda I)\tilde{d}_x = -g$, where H and g are the Hessian and gradient of $\tilde{\phi}(\tilde{x}, \mu)$ with respect to $\tilde{x}$ respectively. To simplify the presentation of H and g we introduce matrix representations of the mappings $\mathcal{L}_\gamma$, $\mathcal{L}_X$ etc. Since $\text{vec}(AXB) = (B^T \otimes A) \text{vec}(X)$ (see e.g. [30]), the linear mapping $\mathcal{L}_X(x)(\cdot)$ has the following matrix representation

$$
\text{vec}(\mathcal{L}_X(x)(d_X)) = L_X \text{vec}(d_X)
$$

where

$$
L_X := \begin{pmatrix} A_{\text{cl}}(K)^T \\ B_{\text{cl}}(K)^T \\ 0 \end{pmatrix} \otimes \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} A_{\text{cl}}(K)^T \\ B_{\text{cl}}(K)^T \\ 0 \end{pmatrix}.
$$

Remark. Although L_X depends on x, we omit this dependency in the notation for simplicity. Furthermore, we write $L_X \text{vec}(d_X)$ without brackets around $\text{vec}(d_X)$, since L_X is a matrix. To stress that $\mathcal{L}_X(x)(\cdot)$ is linear mapping and not a matrix, we do write $\mathcal{L}_X(x)(d_X)$ with brackets around d_X.

Similarly to L_X, the following matrices represent the corresponding mappings defined in Section 7.1.3:

$$L_\gamma := \text{vec}\left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{m_1} & 0 \\ 0 & 0 & I_{p_1} \end{pmatrix}\right),$$

$$L_\gamma^* := L_\gamma^T,$$

$$L_X^* := \begin{pmatrix} I & 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} A_{\text{cl}}(K) & B_{\text{cl}}(K) & 0 \end{pmatrix} + \begin{pmatrix} A_{\text{cl}}(K) & B_{\text{cl}}(K) & 0 \end{pmatrix} \otimes \begin{pmatrix} I & 0 & 0 \end{pmatrix},$$

$$L_K := \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix} \otimes \begin{pmatrix} XB_2 \\ 0 \\ D_{12} \end{pmatrix} + \begin{pmatrix} XB_2 \\ 0 \\ D_{12} \end{pmatrix} \otimes \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix} \Pi_{\text{tr}},$$

$$L_K^* := \begin{pmatrix} B_2^T X & 0 & D_{12}^T \end{pmatrix} \otimes \begin{pmatrix} C_2 & D_{21} & 0 \end{pmatrix},$$

$$L_{XK} := \begin{pmatrix} I & 0 & 0 \end{pmatrix} \mathcal{B}(x)^{-1} \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix} \otimes B_2 + B_2 \otimes \begin{pmatrix} I & 0 & 0 \end{pmatrix} \mathcal{B}(x)^{-1} \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix},$$

$$L_{KX} := B_2^T \otimes \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix}^T \mathcal{B}(x)^{-1} \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix},$$

where Π_{tr} is the permutation matrix that maps $\text{vec}(d_K)$ into $\text{vec}(d_K{}^T)$, i.e. $\text{vec}(d_K{}^T) = \Pi_{\text{tr}}\text{vec}(d_K)$. These definitions allow us to express the matrix representation of

$$d_X \mapsto \mathcal{L}_X^*(x)\left(\mathcal{B}(x)^{-1}\mathcal{L}_X(x)(d_X)\mathcal{B}(x)^{-1}\right)$$

as follows:

$$L_X^*(\mathcal{B}(x)^{-1} \otimes \mathcal{B}(x)^{-1})L_X.$$

This follows since

$$\text{vec}\left(\mathcal{L}_X^*(x)\left(\mathcal{B}(x)^{-1}\mathcal{L}_X(x)(d_X)\mathcal{B}(x)^{-1}\right)\right) = L_X^*\text{vec}\left(\mathcal{B}(x)^{-1}\mathcal{L}_X(x)(d_X)\mathcal{B}(x)^{-1}\right) =$$
$$= L_X^*\left(\mathcal{B}(x)^{-1} \otimes \mathcal{B}(x)^{-1}\text{vec}(\mathcal{L}_X(x)(d_X))\right) = L_X^*(\mathcal{B}(x)^{-1} \otimes \mathcal{B}(x)^{-1})L_X\text{vec}(d_X).$$

Similar results hold for the other mappings. The Hessian can be split up into the three separate terms

$$H = H_{\text{KYP}} + H_{\text{r}} + H_{\text{nc}},$$

where H_{KYP} is the Hessian corresponding to the LMI (2.11) in the KYP-lemma, i.e. Lemma 2.6. This Hessian is

$$H_{\text{KYP}} := \begin{pmatrix} L_\gamma^* & \Pi_{\text{svec}}^T L_X^* & L_K^* \end{pmatrix} (\mathcal{B}(x)^{-1} \otimes \mathcal{B}(x)^{-1}) \begin{pmatrix} L_\gamma \\ L_X \Pi_{\text{svec}} \\ L_K \end{pmatrix}, \tag{7.30}$$

where Π_{svec} maps a vectorized symmetric matrix $\text{vec}(X)$ into $\text{svec}(X)$, i.e. $\Pi_{\text{svec}}\text{vec}(X) = \text{svec}(X)$ for all $X = X^T$. H_{r} is the Hessian corresponding to the 'remaining' inequalities (7.2) and (7.3) and is given by

$$H_{\text{r}} := \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Pi_{\text{svec}}^T H_{\text{r}XX}\Pi_{\text{svec}} & 0 \\ 0 & 0 & H_{\text{r}KK} \end{pmatrix}, \tag{7.31}$$

where

$$H_{\mathrm{r}XX} := X^{-1} \otimes X^{-1} + (\rho_X I - X)^{-1} \otimes (\rho_X I_{n_X} - X)^{-1}$$

and

$$H_{\mathrm{r}KK} := \frac{2}{\rho_k - \|K\|_{\mathrm{F}}^2} I_{(n_c + m_2)(n_c + p_2)} + \frac{4}{(\rho_K - \|K\|_{\mathrm{F}}^2)^2} \mathrm{vec}(K^T) \mathrm{vec}(K^T)^T.$$

Finally H_{nc} is the indefinite part of the Hessian

$$H_{\mathrm{nc}} := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \Pi_{\mathrm{svec}}^T L_{XK} \\ 0 & L_{KX}\Pi_{\mathrm{svec}} & 0 \end{pmatrix}. \tag{7.32}$$

Remark that $H_{\mathrm{KYP}} \succeq 0$ and $H_{\mathrm{r}} \succeq 0$ but H_{nc} is not necessarily positive semi-definite. The solution (7.21) can be found by solving

$$(H + \lambda I)^{-1}\tilde{d}_x = (H_{\mathrm{KYP}} + H_{\mathrm{r}} + H_{\mathrm{nc}} + \lambda I)^{-1}\tilde{d}_x = \begin{pmatrix} -\frac{1}{\mu} + \mathcal{L}_\gamma^*(x)\,(\mathcal{B}(x)^{-1}) \\ \mathrm{svec}\,(\mathcal{L}_X^*(x)\,(\mathcal{B}(x)^{-1})) \\ \mathrm{vec}\,(\mathcal{L}_K^*(x)\,(\mathcal{B}(x)^{-1})) \end{pmatrix} \tag{7.33}$$

for $\tilde{d}_x$.

Updating the Hessian during the plane search

Since H is a symmetric $m \times m$ matrix where $m = 1 + m_{\mathrm{c}} + \frac{1}{2}(n + n_{\mathrm{c}})(n + n_{\mathrm{c}} + 1)$, solving (7.33) by Cholesky factorization for instance (assuming that $H + \lambda I \succ 0$) requires

$$\frac{1}{3}m^3 + 2m^2$$

flops in worst-case [72]. Fortunately it is not required to solve (7.33) in this fashion for each evaluation of $\tilde{d}_{\mathrm{c}}(\lambda)$ in (7.12) during the plane search, since we can circumvent that by the following technique inspired by the work of Jarre [105]. Compute a Schur decomposition $H = Q^T D Q$, with Q being orthogonal and D being diagonal. Remark that D is diagonal since H is symmetric. This costs at most $\frac{4}{3}m^3$ flops in worst-case [72]. Then

$$\tilde{d}_{\mathrm{c}}(\lambda) = -(H + \lambda I)^{-1}g = -(Q^T D Q + \lambda I)^{-1}g = -Q^T(D + \lambda I)^{-1}Qg \tag{7.34}$$

can easily be computed in about $5m^2$ flops [72]. The Schur decomposition of H has to be computed only once for each plane search. If $\tilde{d}_{\mathrm{c}}(\lambda)$ is evaluated for $N_{\mathrm{ls}} \in \mathbb{N}$ distinct values of λ in the plane search, then the total cost of this procedure is $\frac{4}{3}m^3 + N_{\mathrm{ls}}(2m^2 + 8m)$, whereas the Cholesky factorization approach requires $N_{\mathrm{ls}}(\frac{1}{3}m^3 + 2m^2)$. Hence the Schur decomposition is for moderate and large m already favorable if $N_{\mathrm{ls}} > 4$, if compared to solving (7.33) directly for each different value of λ.

Because m grows quadratically with the state dimension n, we still need to perform $\mathcal{O}(n^6)$ flops in a single correction step. In the following section we present a method to solve (7.21) that is, in general, computationally cheaper for large n. This is achieved by exploiting the control-theoretic characteristics.

7.2.4 Solving the Sylvester equations directly

The Newton step in d_X for the Bounded Real Lemma Inequality (7.1)

For simplicity, let us first consider the Sylvester equation

$$\mathcal{L}_X^*(x)\left(\mathcal{B}(x)^{-1}(\mathcal{L}_X(x)(d_X))\mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1}\right) = 0, \tag{7.35}$$

where $\mathcal{L}_X^*(x)(y)$ should be read as the linear mapping $\mathcal{L}_X^*(x)(\cdot)$ applied to y. Analysis of the existence of solutions of this equation is beyond the scope of this thesis. We refer the reader to [151] for more details on this topic. In vectorized form this equation is:

$$\begin{pmatrix} 0 \\ I \\ 0 \end{pmatrix} H_{\mathrm{KYP}} \begin{pmatrix} 0 & I & 0 \end{pmatrix} \mathrm{vec}(d_X) = -\begin{pmatrix} 0 \\ I \\ 0 \end{pmatrix} g,$$

where H_{KYP} is defined in (7.30). Note that this is the Sylvester equation corresponding to the Newton step for (7.1) in X for fixed K and γ. The solution method presented here is inspired by Vandenberghe, *et. al.* [194, 195], see also [200]. These authors consider an IP method to solve the LMI dual of the $\mathcal{H}_\infty$-analysis problem and some generalizations thereof. They compute the Newton steps in the IP method with a similar technique as described in this section, although controller synthesis was not addressed.

Since $\mathcal{B}(x)^{-1} \succ 0$ we can factorize it as $\mathcal{B}(x)^{-1} = R^T R$ by a Cholesky factorization, with upper triangular R partitioned as

$$R = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{pmatrix},$$

where $R_{11} \in \mathbb{R}^{n \times n}$, $R_{22} \in \mathbb{R}^{m_1 \times m_1}$ and $R_{33} \in \mathbb{R}^{p_1 \times p_1}$. Using these definitions we can simplify the left-hand side of (7.35) as

$$\mathcal{L}_X^*(x)\left(\mathcal{B}(x)^{-1}(\mathcal{L}_X(x)(d_X))\mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1}\right) =$$
$$= V\left(W^T d_X V + V^T d_X W\right) W^T + W\left(W^T d_X V + V^T d_X W\right) V^T =$$
$$= V d_Z W^T + W d_Z V^T$$

where

$$V := \begin{pmatrix} V_1 & V_2 \end{pmatrix} := \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}^T,$$

$$W := \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}^T = \begin{pmatrix} R_{11}^T & 0 \end{pmatrix},$$

$$d_Z := W^T d_X V + V^T d_X W.$$

Hence, the solution d_X can be found by first solving $V d_Z W^T + W d_Z V^T = -\mathcal{L}_X^*(x)(\mathcal{B}(x))$ for d_Z and then solving $W^T d_X V + V^T d_X W = d_Z$ for d_X. These are however *not* standard Sylvester equations, since the solution d_Z to $V d_Z W^T + W d_Z V^T = -\mathcal{L}_X^*(x)(\mathcal{B}(x))$ is not unique.

However, by solving $2(n + n_{\mathrm{c}})m_1$ Sylvester equations a solution can be found using the same idea as in [200]. Indeed let us define $Q := V_2 V_2^T d_X$ and use a (standard) basis

$\{E_1, E_2 \ldots, E_{(n+n_c)m_1}\}$, $E_i \in \mathbb{R}^{m_1 \times (n+n_c)}$, for Q, i.e. $Q(q) = \sum_{i=1}^{(n+n_c)m_1} q_i E_i$ with $q \in \mathbb{R}^{(n+n_c)m_1}$ and define $E_0 = 0$. Then (7.35) reduces to

$$V_1 \left(W_1^T d_X V_1 + V_1^T d_X W_1 \right) W_1^T + W_1 \left(W_1^T d_X V_1 + V_1^T d_X W_1 \right) V_1^T +$$
$$+ Q(q) W_1 W_1^T + W_1 W_1^T Q(q)^T = \mathcal{L}_X^*(x) \left(\mathcal{B}(x)^{-1} \right).$$

Now let us denote by d_{Xi} the solution to

$$V_1 \left(W_1^T d_X V_1 + V_1^T d_X W_1 \right) W_1^T + W_1 \left(W_1^T d_X V_1 + V_1^T d_X W_1 \right) V_1^T +$$
$$+ E_i W_1 W_1^T + W_1 W_1^T E_i^T = \mathcal{L}_X^*(x) \left(\mathcal{B}(x)^{-1} \right),$$

for each $i = 0, 1, \ldots, (n + n_c)m_1$. For each i, d_{Xi} can be computed by solving the following two standard Sylvester equations

$$V_1 d_{Zi} W_1^T + W_1 d_{Zi} V_1^T = -E_i W_1 W_1^T - W_1 W_1^T E_i - \mathcal{L}_X^*(x) \left(\mathcal{B}(x) \right) \tag{7.36}$$

and

$$V_1^T d_{Xi} W_1 + W_1^T d_{Xi} V_1 = d_{Zi}. \tag{7.37}$$

Once d_{Xi}, $i = 0, 1, \ldots, (n + n_c)m_1$, have been computed, solution to (7.35) is finally computed as

$$d_{X*} = d_{X0} + \sum_{i=1}^{(n+n_c)m_1} q_i^s d_{Xi},$$

where $q_i^s, \ldots q_{(n+n_c)m_1}^s$ are the solutions to

$$\sum_{i=1}^{(n+n_c)m_1} q_i E_i = V_2 V_2^T \left(d_{X0} + \sum_{i=1}^{(n+n_c)m_1} q_i d_{Xi} \right), \tag{7.38}$$

for $q_1, \ldots, q_{(n+n_c)m_1}$. Solving (7.38) is computationally cheap, since its number of unknowns is $(n + n_c)m_1$, which is a small number compared to the number of unknowns in (7.33) if $n + n_c$ is much larger than m_1.

The Newton step in all variables for the Bounded Real Lemma Inequality (7.1)

It seems not straightforward to solve

$$\mathcal{L}_\gamma^*(x) \left(\mathcal{B}(x)^{-1} \left[\mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K) + \mathcal{L}_\gamma(x)(d_\gamma) \right] \mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1} \right) = 0, \tag{7.39}$$
$$\mathcal{L}_X^*(x) \left(\mathcal{B}(x)^{-1} \left[\mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K) + \mathcal{L}_\gamma(x)(d_\gamma) \right] \mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1} \right) = 0, \tag{7.40}$$
$$\mathcal{L}_K^*(x) \left(\mathcal{B}(x)^{-1} \left[\mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K) + \mathcal{L}_\gamma(x)(d_\gamma) \right] \mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1} \right) = 0, \tag{7.41}$$

for (d_γ, d_X, d_K) in the same fashion as (7.35), due to dependence of these equations on all three unknowns d_γ, d_X and d_K. However, if the number of controller variables m_c is small (or K is a structured controller with only a few degrees of freedom), and if we write $d_K = \sum_{i=1}^{m_c} k_i E_i$, where $\{E_1, \ldots, E_{m_c}\}$ is a basis for $\mathbb{R}^{(n+n_c+m_2) \times (n+n_c+p_2)}$, a similar solution as for (7.35) is possible. For $E_0 = 0$ let d_{Xi} be the solution to

$$\mathcal{L}_X^*(x) \left(\mathcal{B}(x)^{-1} \left(\mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(E_i) \right) \mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1} \right) = 0, \tag{7.42}$$

for $i = 0, 1, \ldots, m_c$ and let $d_{X m_c + 1}$ be the solution to

$$\mathcal{L}_X^*(x) \left(\mathcal{B}(x)^{-1} \left(\mathcal{L}_X(x)(d_X) + \mathcal{L}_\gamma(x)(1) \right) \mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1} \right) = 0. \tag{7.43}$$

Then d_X defined by

$$d_X(k_i) = d_{X0} + \sum_{i=1}^{m_c} k_i d_{Xi} + d_\gamma d_{X m_c + 1}, \tag{7.44}$$

is a partial solution of (7.39)-(7.41), parameterized in terms of k_i and d_γ. After substitution of d_X using (7.44) and $d_K = \sum_{i=1}^{m_c} k_i E_i$ in (7.39) and (7.41), these equations have as only unknowns d_γ and k_i, $i = 0, 1, \ldots, m_c$. If the number of controller variables is small, these equations can be solved in reasonable computation time by e.g. scalarization. Remark that the number of these unknowns are independent of the McMillan degree n of the plant.

Non-convex terms and remaining terms

Now let us focus on solving

$$\left(H_{\mathrm{KYP}} + H_r + H_{\mathrm{nc}} + \lambda I \right) d_x = -g,$$

where H_{KYP}, H_r and H_{nc} are as defined in (7.30), (7.31) and (7.32), respectively.

If X renders (7.1) satisfied, then $X \succ 0$ is equivalent to the existence of a Y such that

$$(A + B_2 K C_2)^T Y + Y(A + B_2 K C_2) \prec 0 \text{ and } Y \succ 0. \tag{7.45}$$

Hence, let us replace $X \succ 0$ by (7.45). As the final steps towards an efficient solution to (7.39)-(7.41), let us ignore the non-convex part of the Hessian H_{nc}, the regularization term λ and the constraint $X \prec \rho_X I$. If we (again) write $d_K = \sum_{i=1}^{m_c} k_i E_i$, where $\{E_1, \ldots, E_{m_c}\}$ is a basis for $\mathbb{R}^{(n+n_c+m_2) \times (n+n_c+p_2)}$, $E_0 = 0$, then a partial solution of the Newton step in d_X can be expressed as (7.44), where d_{Xi}, $i = 0, \ldots, m_c + 1$ are the solutions to (7.42) and (7.43).

The search step in Y can be computed in a similar fashion using m_c Sylvester equations, see Appendix A.4.6. Concluding, by separating the $\mathcal{H}_\infty$ performance constraint and the stability constraint over two variables X and Y, we reduce the convexified Newton step equations for X to smaller subproblems that can be solved efficiently. See [121] for a different way to exploit the same idea.

Final remarks on the direct solution

Even though the set (7.26), (7.27) and (7.28) have an apparently simple matrix-structure, we have seen that is not very straightforward to solve these equations directly. We summarize here the steps to arrive at a set of standard Sylvester equations:

1. eliminate the part of the equation corresponding to the non-convex Hessian H_{nc} and set $\lambda = 0$,

2. introduce a basis for d_K,

3. split the constraint on the closed-loop $\mathcal{H}_\infty$-norm and the stability constraint into two equations with variables X and Y, respectively

4. introduce a basis for $Q = V_2 V_2^T d_X$

Table 7.1: Complexity of steps to compute d_γ, d_X and d_K, where $n_\mathcal{B} := n + n_c + m_1 + p_1$

step	Complexity (no. of flops in worst-case)	no. of steps
Cholesky factor of $\mathcal{B}(x)^{-1}$	$\frac{1}{3}n_\mathcal{B}^3 + 2n_\mathcal{B}$	1
Solving Sylvester eq. (7.36)	$72(n + n_c)^3$	$(n + n_c + 1)m_1 m_c$
Solving Sylvester eq. (7.37)	$72(n + n_c)^3$	$(n + n_c + 1)m_1 m_c$
Finding q_i in (7.38)	$\frac{2}{3}((n + n_c)m_1)^3 + 2((n + n_c)m_1)^2$	m_c
Finding search step d_Y in Y	$72m_c(n + n_c)^3$	1

5. split the equation $\mathcal{L}_X^*(x)\,(\mathcal{B}(x)^{-1}(\mathcal{L}_X(x)(d_X))\mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1}) = 0$ into:

$$\mathcal{L}_X^*(x)\left(d_Z - \mathcal{B}(x)^{-1}\right) = 0$$

and

$$\mathcal{B}(x)^{-1}(\mathcal{L}_X(x)(d_X))\mathcal{B}(x)^{-1} = d_Z.$$

This allows us to compute a convexified Newton step by solving $2n_p(1 + (n + n_c)m_1)$ standard Sylvester equations. The complexity of the main steps are shown in Table 7.1, where, as mentioned before, $m_c = (n_c + m_2)(n_c + p_2)$ denotes the number of controller variables. Minor steps that are not dominant in the complexity analysis have been omitted for simplicity. From the table we can read off that the total number of flops in worst-case is

$$N(n, n_c, m_1, p_1, m_2, p_2) := \frac{n_\mathcal{B}^3}{3} + 2n_\mathcal{B} + 2m_c(72(n_{cl}^4 + n_{cl})m_1 + \frac{(n_{cl}m_1)^3}{3} + (n_{cl}m_1)^2) + 36n_{cl}^3,$$

where $n_\mathcal{B} := n + n_c + m_1 + p_1$ (which equals the number of columns in $\mathcal{B}(\gamma, X, K)$ in (7.1)) and $n_{cl} = n + n_c$ is the total number of states in the closed-loop plant. $N(n, n_c, m_1, p_1, m_2, p_2)$ is a 4th order polynomial in the state dimension n, if the other variables sizes are considered fixed. In Section 7.2.3 it has been clarified that the number of flops in worst-case of the trust-region step using scalarization is $\mathcal{O}(n^6)$, if all other variable sizes are considered fixed. Hence for large (state dimension) n and small (numbers of controller variables and generalized disturbance signals) m_c and m_1, the approach presented in this section may be favorable to scalarization.

This concludes our results on the trust region step computation, which completes the IP algorithm for fixed-order synthesis. Since the algorithm is guaranteed to converge to a point satisfying first-order necessary conditions, it is useful to verify second-order local optimality after running this algorithm. In the next section we present first and second order optimality conditions and the corresponding numerically computable tests for the fixed order synthesis problem. These results can also be applied to any other nonlinear Semi-Definite Program with twice differentiable objective and constraint functions.

7.3   Optimality conditions

In this section we recap the usual first order necessary conditions and provide a rather elementary derivation of the most general second order necessary and sufficient conditions for optimality for the BMI problem

$$\begin{array}{ll} \text{infimize} & \gamma \\ \text{subject to} & (7.1), (7.2) \text{ and } (7.3). \end{array} \tag{7.46}$$

These are local optimality conditions, which contrast with the global optimality certificates obtained with the SOS relaxations in Chapters 4 and 5. One expects that local optimality is in general easier to verify computationally and we will see in this section that this is indeed true.

Recall that the decision variable is $x = (\gamma, X, K)$ and the feasible set is $\mathcal{G}$ as defined in (7.6). The objective function is denoted by $f(x) = \gamma$ and the constraint functions by

$$G_1(x) = -\mathcal{B}(\gamma, X, K), \ G_2(x) = -X, \ G_3(x) = X - \rho_X I, \ g_4(x) = \|K\|_\mathrm{F}^2 - \rho_K,$$

which are collected in the usual fashion as $G = \mathrm{diag}(G_1, G_2, G_3, g_4)$. Since the feasible set $\mathcal{G}$ is open and the objective γ is linear, optimal values are not attained. For the analysis in this section we therefore replace the strict inequalities $G(x) \prec 0$ by the non-strict ones $G(x) \preceq 0$, as we did in Chapter 4. Weierstrass' Theorem (as for instance described in [124]) implies that if the feasible set is nonempty, then the global minimum of $\inf_{G(x) \preceq 0} f(x)$, where $f(x) := \gamma$, is attained.

Note that $G_i(x) \preceq 0$ for $i = 1, 2, 3$ are semi-definite constraints. The functions f and G are C^∞ on the interior of the feasible set $\mathcal{G}$, and the first and second order (directional) derivatives of G at x (in directions d or d_1, d_2 in $\mathcal{X}$) are denoted by $\partial G(x)$ and $\partial^2 G(x)$ (or $\partial G(x)d$ and $\partial^2 G(x)(d_1, d_2)$) respectively. Explicit formulas are given in Appendix A.4.5.

With $\mathcal{X}$ as defined in (7.5), the non-strict version of the BMI problem just amounts to minimizing $f(x)$ over the feasible set

$$\mathcal{G} = \{x \in \mathcal{X} |\ G(x) \preceq 0\}. \tag{7.47}$$

For our analysis f and G are required to be merely twice continuously differentiable (instead of C^∞). In the next sections we present conditions for $x \in \mathcal{X}$ to be a local optimal point for this problem, see Section 2.5.5 for the definition of such points. The results are extensions of optimality conditions for optimization problems with scalar inequality constraints, as described in standard texts, e.g. [125, 17, 102].

Example 7.3 *We illustrate for a scalar problem some of the most important situations that can occur at a candidate local point in Figure 7.4. We consider the minimization of $f(x)$ over the feasible region $\mathcal{G} := \{x \mid g_1(x) \leq 0, g_2(x) \leq 0\}$, i.e. the optimization problem*

$$\inf_{g_1(x) \leq 0, g_2(x) \leq 0} f(x),$$

$x = (x_1, x_2) \in \mathbb{R}^2$, $f(x) = x_2$ $g_1(x) = -x_1$, for four different choices of the constraint $g_2(x)$. At the critical point $x = (0,0)$ we have the following cases:

- *with $g_2(x) = x_1^3 + x_1 - x_2$ the first order sufficient conditions are satisfied (left upper subplot '**a**')*

- *with $g_2(x) = x_1^2 - x_2$ the second order sufficient conditions are satisfied (right upper subplot '**b**')*

- *with $g_2(x) = -x_1^3 - x_2$ the second order necessary conditions are satisfied, but the point $(0,0)$ is not optimal (left lower subplot '**c**')*

- *with $g_2(x) = x_1^3 - x_2$ the point $(0,0)$ is optimal, but the second order sufficient conditions are violated (right lower subplot '**d**').*
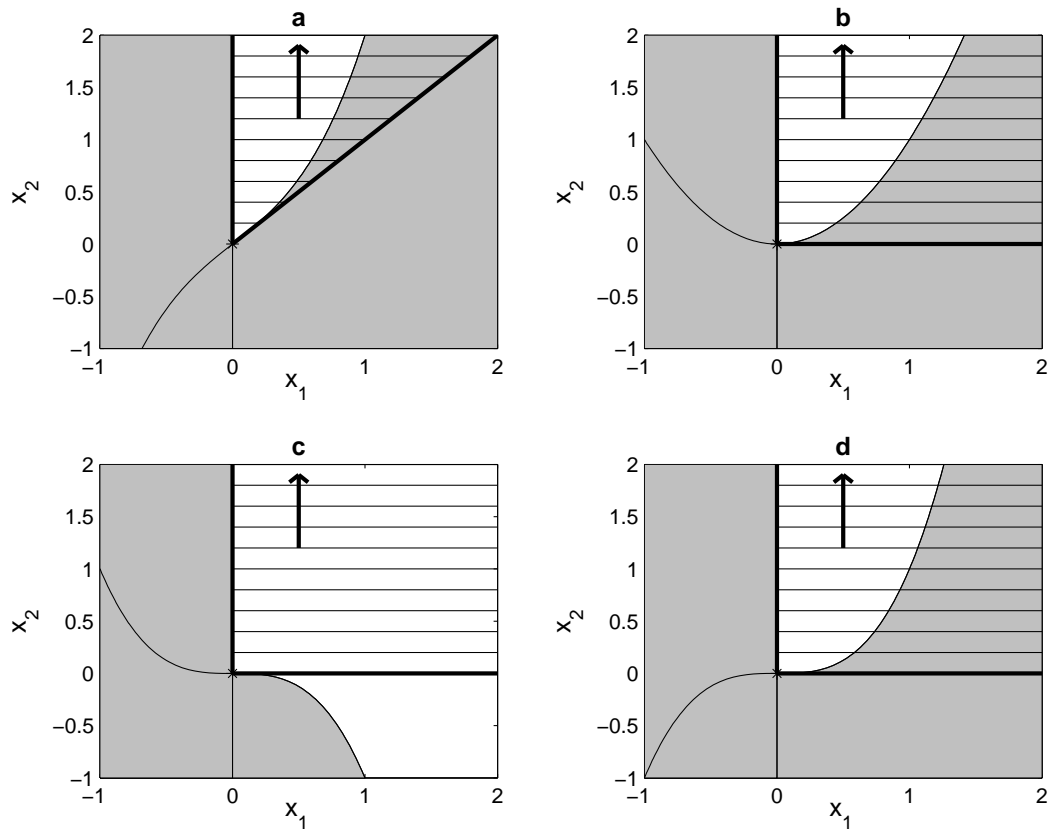
Figure 7.4: Optimality conditions for four different situations, as described on page 132. Shown are the gradient of f (arrow), the infeasible region $\mathbb{R}^2 \backslash \mathcal{G}$ (solidly filled region) and the tangent cone (horizontally shaded area) at $x = 0$.

7.3.1 First order necessary optimality conditions

To derive the optimality conditions, we rely on the usual concept of a *tangent cone*.

Definition 7.4 - *Tangent cone.* *The tangent cone $T_{\mathcal{G}}(x_*)$ of the feasible set $\mathcal{G}$ at $x_* \in \mathcal{G}$ is the union of the zero vector together with the set of all vectors $d \in \mathcal{X}$ that can be represented as $d = \lambda s$ for some $\lambda > 0$ such that there exists a sequence s_ν with*

$$s_\nu \neq 0, \ \lim_{\nu \to \infty} s_\nu = 0, \ \lim_{\nu \to \infty} \frac{s_\nu}{\|s_\nu\|} = s \text{ and } x_* + s_\nu \in \mathcal{G} \text{ for all } \nu = 1, 2, \dots \qquad (7.48)$$

Lemma 7.5 *With $\epsilon > 0$ let $z : [0, \epsilon] \to \mathbb{R}^n$ be any curve with $z(0) = x_* \in \mathcal{G}$ and $z(t) \in \mathcal{G}$ for $t \in [0, \epsilon]$. Furthermore let it have a right-derivative $\partial_+ z(0) \neq 0$ at $t = 0$. Then $\partial_+ z(0)$ is contained in $T_{\mathcal{G}}(x_*)$.*

Proof. For the sequence chosen as $s_\nu := z(\epsilon \frac{1}{\nu}) - x_*$ we immediately observe $\lim_{\nu \to \infty} s_\nu = 0$, $x + s_\nu \in \mathcal{G}$ for all $\nu = 1, 2, \dots$. Furthermore

$$\frac{s_\nu}{\|s_\nu\|} = \frac{z(\epsilon \frac{1}{\nu}) - x_*}{\|z(\epsilon \frac{1}{\nu}) - x_*\|} = \frac{\frac{z(\mu) - x_*}{\mu}}{\left\|\frac{z(\mu) - x_*}{\mu}\right\|} = \frac{\frac{z(\mu) - z(0)}{\mu}}{\left\|\frac{z(\mu) - z(0)}{\mu}\right\|},$$

with $\mu = \frac{\epsilon}{\nu}$. Hence

$$\lim_{\nu \to \infty} \frac{s_\nu}{\|s_\nu\|} = \lim_{\mu \to 0, \mu > 0} \frac{\frac{z(\mu) - z(0)}{\mu}}{\left\|\frac{z(\mu) - z(0)}{\mu}\right\|} = \frac{\partial_+ z(0)}{\|\partial_+ z(0)\|},$$

where the latter equality follows from the definition of first order right-derivatives. $\blacksquare$

Note that for all $x_* \in \mathcal{G}$, $T_{\mathcal{G}}(x)_*$ is a closed cone (which is allowed to be empty if x_* is isolated). That $T_{\mathcal{G}}(x)$ is a cone (as defined in Section 2.5.1) is trivial from the definition. Closure is shown in the following lemma.

Lemma 7.6 *For all $x_* \in \mathcal{G}$, $T_{\mathcal{G}}(x_*)$ is a closed set.*

Proof. Based on a proof in [17]. Let $x_* \in \mathcal{G}$ be arbitrary and d_k, $k = 1, 2, \dots$ be a sequence such that $d_k \in T_{\mathcal{G}}(x_*)$ for all $k = 1, 2, \dots$ with limit $d := \lim_{k \to \infty} d_k$. To prove that $T_{\mathcal{G}}(x)$ is closed we show that $d \in T_{\mathcal{G}}(x_*)$. Since $d_k \in T_{\mathcal{G}}(x_*)$ there exist for all $k = 1, 2, \dots$ sequences s_ν^k satisfying $s_\nu^k \neq 0$ for all $\nu = 1, 2, \dots$ and such that

$$s_\nu^k \neq 0, \ \lim_{\nu \to \infty} s_\nu^k = 0, \ \lim_{\nu \to \infty} \frac{s_\nu^k}{\|s_\nu^k\|} = s^k \text{ and } x_* + s_\nu^k \in \mathcal{G} \text{ for all } \nu = 1, 2, \dots,$$

which is (7.48) for $s_\nu = s_\nu^k$. Hence, for such sequences s_ν^k there exist for each k a large enough index ν_k such that $\|s_{\nu_k}\| < \frac{1}{k}$ and

$$\left\| \frac{s_{\nu_k}^k}{\|s_{\nu_k}^k\|} - \frac{d_k}{\|d_k\|} \right\| < \frac{1}{k}.$$

For these choice of indices ν_k we therefore conclude that

$$\lim_{k\to\infty} s_{\nu_k}^k = 0 \text{ and } \lim_{k\to\infty} \left\| \frac{s_{\nu_k}^k}{\|s_{\nu_k}^k\|} - \frac{d_k}{\|d_k\|} \right\| = 0.$$

Obviously $x + s_{\nu_k}^k \in \mathcal{G}$ for all k. Furthermore, for all k we have (by the triangle inequality of the norm)

$$\left\| \frac{s_{\nu_k}^k}{\|s_{\nu_k}^k\|} - \frac{d}{\|d\|} \right\| \leq \left\| \frac{s_{\nu_k}^k}{\|s_{\nu_k}^k\|} - \frac{d_k}{\|d_k\|} \right\| + \left\| \frac{d_k}{\|d_k\|} - \frac{d}{\|d\|} \right\|,$$

and by combining the preceding two relations with the fact that $\lim_{k\to\infty} d_k = d$ we obtain

$$\lim_{k\to\infty} s_{\nu_k}^k = 0 \text{ and } \lim_{k\to\infty} \left\| \frac{s_{\nu_k}^k}{\|s_{\nu_k}^k\|} - \frac{d}{\|d\|} \right\| = 0.$$

Hence we conclude that $d \in T_{\mathcal{G}}(x)$

■

The definition of the tangent cone is motivated by the following necessary condition for local optimality.

Theorem 7.7 *Let x_* be a local minimum point of f on $\mathcal{G}$. Then*

$$\partial f(x_*)d \geq 0 \text{ for all } d \in T_{\mathcal{G}}(x_*). \tag{7.49}$$

Proof. If $d = 0$ then (7.49) trivially holds true. Choose an element $d \neq 0$ of the tangent cone and a corresponding sequence s_ν satisfying $d = \lambda s$ for some $\lambda \geq 0$ and (7.48). Local optimality of x_* and (7.48) imply

$$\frac{f(x_* + s_\nu) - f(x_*)}{\|s_\nu\|} \geq 0 \text{ for all sufficiently large } \nu. \tag{7.50}$$

By the definition of first order derivatives we infer

$$\lim_{\nu\to\infty} \frac{f(x_* + s_\nu) - f(x_*)}{\|s_\nu\|} = \lim_{\nu\to\infty} \partial f(x_*) \frac{s_\nu}{\|s_\nu\|} = \partial f(x_*)s. \tag{7.51}$$

Combining (7.50) and (7.51) implies

$$\partial f(x_*)d = \partial f(x_*)(\lambda s) = \lambda \partial f(x_*)(s) \geq 0.$$

■

In a similar fashion one can prove the following first-order sufficient condition for local optimality.

Theorem 7.8 *Let $x_* \in \mathcal{G}$ satisfy*

$$\partial f(x_*)d > 0 \text{ for all } d \in T_{\mathcal{G}}(x_*), \ d \neq 0.$$

Then x_* is a strict local minimum point of f on $\mathcal{G}$.

Proof. Suppose x_* is not a strictly local optimum. Then for each ν there exists a point $x_\nu \in \mathcal{G}$, $x_\nu \neq x_*$ with $\|x_* - x_\nu\| \leq \frac{1}{\nu}$ and $f(x_\nu) \leq f(x_*)$. If we define the sequence $s_\nu := x_\nu - x_*$ then obviously $\lim_{\nu \to \infty} \|s_\nu\| = 0$ and $x_* + s_\nu \in \mathcal{G}$ for all $\nu = 1, 2, \ldots$. Furthermore

$$f(x_* + s_\nu) \leq f(x_*) \text{ for all } \nu = 1, 2, \ldots. \tag{7.52}$$

Observe that $x_\nu \neq x_*$ implies $\|s_\nu\| > 0$ for all ν such that $\frac{s_\nu}{\|s_\nu\|}$ is well-defined for all ν. Since $\frac{s_\nu}{\|s_\nu\|}$ is bounded it admits a convergent subsequence. We can assume that the sequence itself converges with limit $\lim_{\nu \to \infty} \frac{s_\nu}{\|s_\nu\|} = s$. This reveals that $s \in T_\mathcal{G}(x_*)$. To show $s \neq 0$ we use continuity of $\|\cdot\|$ to conclude

$$\|s\| = \left\| \lim_{\nu \to \infty} \frac{s_\nu}{\|s_\nu\|} \right\| = \lim_{\nu \to \infty} \left\| \frac{s_\nu}{\|s_\nu\|} \right\| = 1.$$

By the definition of derivatives we infer

$$\partial f(x_*)s = \lim_{\nu \to \infty} \frac{f(x_* + s_\nu) - f(x_*)}{\|s_\nu\|} \leq 0, \tag{7.53}$$

where the inequality follows from (7.52). Hence we have found a nonzero $s \in T_\mathcal{G}(x_*)$ with $\partial f(x_*)s \leq 0$, a contradiction.

∎

Although conceptually very simple, the derived optimality conditions are not necessarily straightforward to verify computationally, because in general it is not simple to describe the tangent cone. Under a mild constraint qualification the description of $T_\mathcal{G}(x)$ becomes very simple, as clarified in the sequel.

For this purpose we introduce the concept of the *linearization cone* of G at x:

$$L_G(x) := \left\{ d \in \mathcal{X} \mid u^T [\partial G(x)d]u \leq 0 \text{ for all } u \in \mathrm{Ker}(G(x)) \right\}.$$

Note that the linearization cone depends on G, i.e. on the *representation* of the feasible set $\mathcal{G}$, whereas the tangent cone only depends on the set $\mathcal{G}$ itself. Just by using the first order optimality condition it is elementary to see that $T_\mathcal{G}(x) \subset L_G(x)$, which is subject of the following lemma

Lemma 7.9 *For arbitrary $x_* \in \mathcal{G}$ the inclusion $T_\mathcal{G}(x_*) \subset L_G(x_*)$ holds true.*

Proof. Take an arbitrary $d \in T_\mathcal{G}(x_*)$ and $u \in \mathrm{Ker}(G(x_*))$ and observe that the mapping $g_u(x) := u^T G(x)u$ attains its maximum 0 on $\mathcal{G}$ at $x = x_*$. Hence the first order necessary conditions imply that $\partial g_u(x_*)d = u^T [\partial G(x_*)d] u \leq 0$, since d is in the tangent cone. This implies $d \in L_G(x_*)$.

∎

The converse inclusion is not necessarily true, as illustrated in the next example.

Example 7.10 *This example is adopted from [110]. For*

$$G(x) := \begin{pmatrix} -1 + x_2 - (1 - x_1)^3 & 0 \\ 0 & 1 - x_2 \end{pmatrix}$$

136
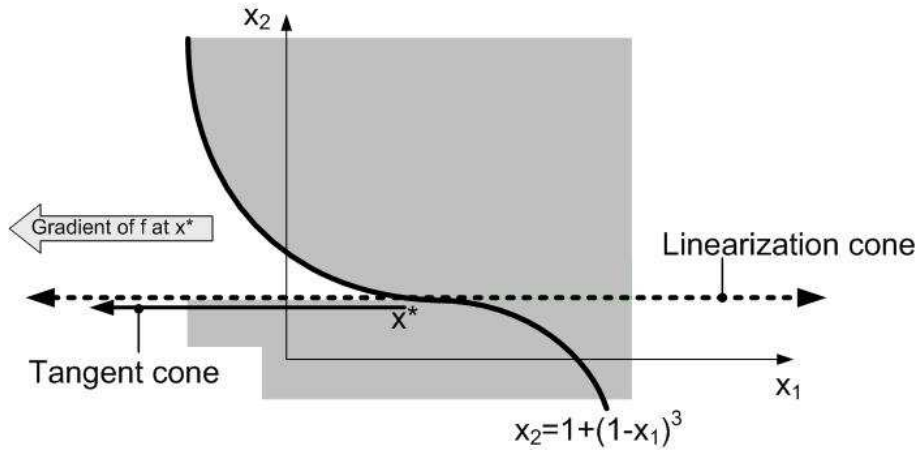
Figure 7.5: Example that illustrates difference between tangent and linearization cone. Shown are the gradient of f (grey arrow), the constraint function $G_{(1,1)}(x) = 0$ (solid line), the infeasible region (shaded area), the tangent cone (solid arrow) and the linearization cone (dashed arrow).

the tangent cone at $x_* = (1,1)$ is

$$T_{\mathcal{G}}(x_*) := \left\{ \begin{pmatrix} \lambda \\ 0 \end{pmatrix} \mid \lambda \in \mathbb{R}, \lambda \geq 0 \right\},$$

whereas the linearization cone is

$$L_G(x_*) := \left\{ \begin{pmatrix} \lambda \\ 0 \end{pmatrix} \mid \lambda \in \mathbb{R} \right\}.$$

These cones as well as the feasible set $\{x \in \mathbb{R}^2 \mid G(x) \preceq 0\}$ are depicted in Figure 7.5.

As a consequence, the sufficient optimality condition in terms of the linearization cone at x_* is violated, even though x_* is an optimal point. To illustrate this consider $\inf_{G(x) \preceq 0} f(x)$ with $f(x) = -x_1$. The global optimum is attained at x_*, such that the first order necessary condition in terms of the tangent cone holds true. However, if we replace $T_{\mathcal{G}}(x_*)$ by $L_G(x_*)$ in (7.49) the condition is violated.

Fortunately, under the aforementioned Constraint Qualification (CQ), which very often holds true in practice, the tangent and linearization cones are equal. To simplify the presentation of this CQ at x_*, let us choose a nonsingular matrix $U_* = (V_* \ W_*)$ such that the columns of V_* span the kernel of $G(x_*)$.

Theorem 7.11 *Let the following analogue of the standard Mangasarian-Fromowitz Constraint Qualification (MFCQ) be satisfied: there exists some $d_0 \in \mathcal{X}$ with*

$$u^T[\partial G(x_*)d_0]u < 0 \ \text{for all } u \neq 0 \ \text{with } u \in Ker(G(x_*)). \tag{7.54}$$

Then $T_{\mathcal{G}}(x_*) = L_G(x_*)$.

Proof. Let d_0 be a vector that satisfies (7.54). Then (7.54) obviously implies

$$V_*^T[\partial G(x_*)d_0]V_* \prec 0. \tag{7.55}$$

137

We will first show that $d_0 \in T_{\mathcal{G}}(x_*)$. If $d_0 = 0$ then $d_0 \in T_{\mathcal{G}}(x_*)$, so let us assume $d_0 \neq 0$. To prove $d_0 \in T_{\mathcal{G}}(x_*)$ we will consider the sequence $s_\nu := \frac{1}{\nu}d_0$. Define

$$\epsilon_\nu := \frac{G(x_* + \frac{1}{\nu}d_0) - G(x_*) - \partial G(x_*)\frac{d_0}{\nu}}{\|\frac{1}{\nu}d_0\|}$$

such that $\lim_{\nu\to\infty} \epsilon_\nu = 0$. By the definition of V_* we have $G(x_*)V_* = 0$ such that

$$\frac{U_*^T G(x_* + \frac{1}{\nu}d_0)U_*}{\|\frac{1}{\nu}d_0\|} = U_*^T \epsilon_\nu U_* + \frac{\nu}{\|d_0\|}\begin{pmatrix} 0 & 0 \\ 0 & W_*^T G(x_*)W_* \end{pmatrix} +$$

$$+ \frac{\nu}{\|d_0\|}\begin{pmatrix} V_*^T\left[\partial G(x_*)\frac{d_0}{\nu}\right]V_* & V_*^T\left[\partial G(x_*)\frac{d_0}{\nu}\right]W_* \\ W_*^T\left[\partial G(x_*)\frac{d_0}{\nu}\right]V_* & W_*^T\left[\partial G(x_*)\frac{d_0}{\nu}\right]W_* \end{pmatrix} =$$

$$= U_*^T \epsilon_\nu U_* + \frac{1}{\|d_0\|}\begin{pmatrix} V_*^T\left[\partial G(x_*)d_0\right]V_* & V_*^T\left[\partial G(x_*)d_0\right]W_* \\ W_*^T\left[\partial G(x_*)d_0\right]V_* & \nu W_*^T G(x_*)W_* + W_*^T\left[\partial G(x_*)d_0\right]W_* \end{pmatrix}.$$

By feasibility of x_* we conclude $W_*^T G(x_*)W_* \prec 0$. This combined with (7.55) implies for large enough ν,

$$\begin{pmatrix} V_*^T\left[\partial G(x_*)d_0\right]V_* & V_*^T\left[\partial G(x_*)d_0\right]W_* \\ W_*^T\left[\partial G(x_*)d_0\right]V_* & \nu W_*^T G(x_*)W_* + W_*^T\left[\partial G(x_*)d_0\right]W_* \end{pmatrix} \prec 0.$$

This follows from a Schur complement argument and the fact that the largest eigenvalue of the negative definite term $\nu W_*^T G(x_*)W_*$ can be made arbitrary small for large enough ν. Hence the sequence s_ν satisfies, for all large ν, the conditions of the tangent cone (7.48) such that $\|d_0\| \lim_{\nu\to\infty} s_\nu = d_0 \in T_{\mathcal{G}}(x_*)$.

Now choose an arbitrary $d_l \in L_G(x_*)$. Then $d(\tau) := (1-\tau)d_l + \tau d_0$ for arbitrary $\tau \in (0,1]$ is a vector satisfying (7.54) since for any $u \neq 0$ with $u \in \text{Ker}(G(x_*))$

$$u^T\left[\partial G(x_*)d(\tau)\right]u = \tau \underbrace{u^T\left[\partial G(x_*)d_0\right]u}_{<0 \text{ by (7.54)}} + (1-\tau)\underbrace{u^T\left[\partial G(x_*)d_l\right]u}_{\leq 0 \text{ by } d_l \in L_G(x_*)} < 0.$$

Since any vector that satisfies (7.54) is in $T_{\mathcal{G}}(x_*)$ (as shown in the first part of this proof), we conclude $d(\tau) \in T_{\mathcal{G}}(x_*)$ for all $\tau \in (0,1]$. Finally due to the fact that $T_{\mathcal{G}}(x_*)$ is closed, $d_l = \lim_{\tau\to 0,\tau>0} d(\tau) \in T_{\mathcal{G}}(x_*)$. Since $d_l \in L_G(x_*)$ was arbitrary we conclude $L_G(x_*) \subset T_{\mathcal{G}}(x_*)$. Combining this with Lemma 7.9 yields $L_G(x_*) = T_{\mathcal{G}}(x_*)$ ∎

Example 7.12 *(Example 7.10 continued.)* For

$$G(x) := \begin{pmatrix} -1 + x_2 - (1-x_1)^3 & 0 \\ 0 & 1 - x_2 \end{pmatrix},$$

we show that $x_* = (1,1)$ does not satisfy the MFCQ. Observe that

$$\partial G(x)d := \begin{pmatrix} d_2 + 3d_1(1-x_1)^2 & 0 \\ 0 & -d_2 \end{pmatrix}.$$

138

Since $G(x_*) = 0$ we can choose $V_* = I_2$. To satisfy (7.54) we need to find a $d_0 :=$ $\begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix}^T$ *such that*

$$V_*^T [\partial G(x_*)d_0] V_* = \begin{pmatrix} \delta_2 & 0 \\ 0 & -\delta_2 \end{pmatrix} \prec 0,$$

which is obviously impossible.

If the MFCQ holds at x_*, and if the columns of V_* span the kernel of $G(x_*)$, the necessary optimality condition translates into the SDP

$$\begin{aligned} \text{minimize} \quad & \partial f(x_*)d \\ \text{subject to} \quad & d \in \mathcal{X}, \quad V_*^T[\partial G(x_*)d]V_* \leq 0 \end{aligned}$$

having optimal value zero. Since this SDP is strictly feasible, its dual also has optimal value zero. Consequently, there exists some $Z \succeq 0$ with $\partial f(x_*)d + \text{Trace}[(V_* Z V_*^T) \partial G(x_*)d] = 0$ for all $d \in \mathcal{X}$. Observe that $\Lambda := V_* Z V_*^T \succeq 0$ and $\Lambda G(x_*) = V_* Z V_*^T G(x_*) = 0$ by definition of V_*. This immediately leads to the standard first order necessary optimality conditions: There exists a Lagrange multiplier Λ with

$$\partial f(x_*) + \partial_x z(x_*, \Lambda) = 0, \quad \Lambda \succeq 0, \quad \Lambda G(x_*) = 0, \tag{7.56}$$

where $z(x, \Lambda) := \text{Trace}(\Lambda G(x))$. Let us denote the set of all matrices Λ satisfying (7.56) by $\mathbf{\Lambda}_*$.

Remark. The difference between the first order necessary and sufficient conditions seem to be small, since a non-strict inequality for the necessary condition is replaced by a strict one. In practice the gap between these conditions is large, in the sense that for many SDP problems encountered in practice the first order sufficient conditions do not hold at locally optimal points. To be able to verify local optimality for these cases we formulate second order conditions in the next section.

7.3.2 Second order conditions

The second order conditions for SDPs are somewhat more involved than for usual scalar constraints. Using $U_* = (V_*\ W_*)$ we infer that $G(x) \preceq 0$ if and only if $U_*^T G(x) U_* \preceq 0$ if and only if

$$G_{\mathrm{r}}(x) := [V_*^T G(x) V_*] - [V_*^T G(x) W_*](W_*^T G(x) W_*)^{-1}[W_*^T G(x) V_*] \preceq 0, \tag{7.57}$$

whenever $x \in \mathcal{X}$ satisfies $W_*^T G(x) W_* \prec 0$. This motivates to define

$$\mathcal{W} := \left\{ x \in \mathcal{X} \mid W_*^T G(x) W_* \prec 0 \right\} \quad \text{and} \quad \mathcal{G}_{\mathrm{r}} := \left\{ x \in \mathcal{W} \mid G_{\mathrm{r}}(x) \preceq 0 \right\}$$

such that

$$x_* \in \mathcal{W} \cap \mathcal{G} = \mathcal{G}_{\mathrm{r}}. \tag{7.58}$$

Since $\mathcal{W}$ is open and $x_* \in \mathcal{W}$, a whole neighborhood of x_* lies in $\mathcal{W}$. Hence local optimality of x_* on $\mathcal{G}$ is equivalent to its local optimality on $\mathcal{W} \cap \mathcal{G}$ and, since $\mathcal{W} \cap \mathcal{G} = \mathcal{G}_{\mathrm{r}}$, it is also equivalent to local optimality of x_* on $\mathcal{G}_{\mathrm{r}}$. Hence x_* is a local minimal point of f on $\mathcal{G}$ if and only if x_* is a local minimal point of f on $\mathcal{G}_{\mathrm{r}}$. As the essential difference, since $G_{\mathrm{r}}(x_*) = 0$, all constraints $G_{\mathrm{r}}(x) \preceq 0$ are 'binding' at x_*, whereas this is not necessarily

true for $G(x) \preceq 0$.

We now clarify that the first order necessary optimality conditions for minimizing f on $\mathcal{G}$ and on $\mathcal{G}_\mathrm{r}$ are fully equivalent. First observe that

$$
\begin{aligned}
\partial G_\mathrm{r}(x_*)d &= V_*^T \left[\partial G(x_*)d\right] V_* - \left(V_*^T [\partial G(x_*)d]W_*\right)\left(W_*^T G(x)W_*\right)^{-1}\left(W_*^T G(x)V_*\right) - \\
&\quad - \left(V_*^T G(x)W_*\right)\left(W_*^T G(x)W_*\right)^{-1}\left(W_*^T \left[\partial G(x_*)d\right]W_*\right)\left(W_*^T G(x)W_*\right)^{-1}\left(W_*^T G(x)V_*\right) - \\
&\quad - \left(V_*^T G(x)W_*\right)\left(W_*^T G(x)W_*\right)^{-1}\left(W_*^T \left[\partial G(x_*)d\right]V_*\right) = V_*^T \left[\partial G(x_*)d\right] V_*,
\end{aligned}
$$

where the most right equality follows since $G(x_*)V_* = 0$. This allows us to relate the Lagrange multipliers $\boldsymbol{\Lambda}_*$ corresponding to the constraint $G(x_*) \preceq 0$ to the set

$$
\boldsymbol{\Sigma}_* := \left\{\Sigma \succeq 0 : \ \partial f(x_*) + \mathrm{Trace}(\Sigma\, \partial G_\mathrm{r}(x_*)) = 0\right\},
$$

of the Lagrange multipliers related to $G_\mathrm{r}(x)$ in the following simple fashion.

Lemma 7.13 *The set of Lagrange multipliers $\boldsymbol{\Lambda}_*$ is related to $\boldsymbol{\Sigma}_*$ by $\boldsymbol{\Lambda}_* = V_*\boldsymbol{\Sigma}_*V_*^T$ and $\boldsymbol{\Sigma}_* = V_*^+\boldsymbol{\Lambda}_*(V_*^+)^T$ with the Moore-Penrose [94] left inverse V_*^+ of V_*. In particular $\boldsymbol{\Lambda}_* \neq \emptyset$ iff $\boldsymbol{\Sigma}_* \neq \emptyset$.*

Proof. For an arbitrary $\Sigma \in \boldsymbol{\Sigma}_*$ we infer

$$
\begin{aligned}
\partial f(x_*) + \mathrm{Trace}(V_*\Sigma V_*^T \partial G(x_*)) &= \\
&= \partial f(x_*) + \mathrm{Trace}(\Sigma V_*^T \partial G(x_*)V_*) = \\
&= \partial f(x_*) + \mathrm{Trace}(\Sigma \partial G_\mathrm{r}(x_*)) = 0,
\end{aligned}
$$

$V_*\Sigma V_*^T \succeq 0$ and $V_*\Sigma V_*^T G(x) = 0$ such that $V_*\Sigma V_*^T \in \boldsymbol{\Lambda}_*$ and hence $V_*\boldsymbol{\Sigma}_*V_*^T \subset \boldsymbol{\Lambda}_*$. On the other hand, for an arbitrary $\Lambda \in \boldsymbol{\Lambda}_*$, $\Lambda G(x) = 0$ implies that Λ can be decomposed into $\Lambda = V_*\Sigma V_*^T$ for some unique $\Sigma \succeq 0$. Furthermore $\Lambda \in \boldsymbol{\Lambda}_*$ implies

$$
\begin{aligned}
0 = \partial f(x_*) + \mathrm{Trace}(\Lambda\, \partial G(x_*)) &= \\
&= \partial f(x_*) + \mathrm{Trace}(V_*\Sigma V_*^T \partial G(x_*)) = \\
&= \partial f(x_*) + \mathrm{Trace}(\Sigma \partial G_\mathrm{r}(x_*))
\end{aligned}
$$

such that we conclude $\Sigma \in \boldsymbol{\Sigma}_*$. Since $\Lambda \in \boldsymbol{\Lambda}_*$ is arbitrary we infer $\boldsymbol{\Lambda}_* \subset V_*\boldsymbol{\Sigma}_*V_*^T$. Summarizing we have shown $V_*\boldsymbol{\Sigma}_*V_*^T = \boldsymbol{\Lambda}_*$. Since $V_*^+V_* = I$ this furthermore implies

$$
\boldsymbol{\Sigma}_* = V_*^+\left(V_*\boldsymbol{\Sigma}_*V_*^T\right)(V_*^+)^T = V_*^+\boldsymbol{\Lambda}_*(V_*^+)^T.
$$

∎

All this motivates to formulate the second order conditions in terms of the (reduced) Lagrangian

$$
l_\mathrm{r}(x, \Sigma) = f(x) + \mathrm{Trace}(\Sigma\, G_\mathrm{r}(x)).
$$

Let us now continue under the assumption that $\boldsymbol{\Sigma}_*$ is nonempty (without any hypotheses

on constraint qualifications). After fixing some $\Sigma_* \in \mathbf{\Sigma}_*$, we introduce the new constraint set

$$\mathcal{G}_* := \left\{ x \in \mathcal{W} \mid u^T G_r(x) u \leq 0 \text{ for all } u \in \mathrm{Ker}(\Sigma_*) \text{ and } G_r(x)\Sigma_* = 0 \right\}.$$

This definition is motivated by

$$f(x) = l_r(x, \Sigma_*) \text{ for all } x \in \mathcal{G}_*, \tag{7.59}$$

which will be essential in proving the following two results.

Lemma 7.14 *If x_* is a local minimal point of f on $\mathcal{G}$, then it is also a local minimum point of $l_r(\cdot, \Sigma_*)$ on $\mathcal{G}_*$.*

Proof. Choose an open neighborhood $\mathcal{U}$ of x_* such that x_* is a minimal point of f on $\mathcal{U} \cap \mathcal{G}$. The set $\mathcal{V} := \mathcal{W} \cap \mathcal{U}$ has the property $x_* \in \mathcal{V} \cap \mathcal{G}_* \subset \mathcal{V} \cap \mathcal{G}_r = \mathcal{V} \cap \mathcal{G}$ due to (7.58) and $\mathcal{V} \subset \mathcal{W}$. Hence x_* is also a minimal point of f on $\mathcal{V} \cap \mathcal{G}_*$. Now we use (7.59) in order to infer that the same property holds for the Lagrangian. ∎

Theorem 7.15 (Second order necessary optimality condition) *Suppose that x_* is a local minimal point of f on $\mathcal{G}$. Then*

$$\partial_x^2 l_r(x_*, \Sigma_*)(d, d) \geq 0 \ \ \text{for all} \ \ d \in T_{\mathcal{G}_*}(x_*).$$

Proof. Represent $d \in T_{\mathcal{G}_*}(x_*)$ as $d = \lambda s$ with $\lambda \geq 0$, $\lim_{\nu \to \infty} s_\nu / \|s_\nu\| = s$ where $s_\nu \neq 0$ for all $\nu = 1, 2, \ldots$ and $\lim_{\nu \to \infty} \|s_\nu\| = 0$ and $x_* + s_\nu \in \mathcal{G}_*$. We infer $\lim_{\nu \to \infty} \epsilon_\nu = 0$ for

$$\epsilon_\nu := \frac{l_r(x_* + s_\nu, \Sigma_*) - l_r(x_*, \Sigma_*) - \partial_x l_r(x_*, \Sigma_*)s_\nu - 0.5\, \partial_x^2 l_r(x_*, \Sigma_*)(s_\nu, s_\nu)}{\|s_\nu\|^2}.$$

We make use of $\partial_x l_r(x_*, \Sigma_*) = 0$ by the definition of Σ_* to get

$$\frac{l_r(x_* + s_\nu, \Sigma_*) - l_r(x_*, \Sigma_*)}{\|s_\nu\|^2} = \epsilon_\nu + 0.5\, \partial_x^2 l_r(x_*, \Sigma_*)\left(\frac{s_\nu}{\|s_\nu\|}, \frac{s_\nu}{\|s_\nu\|} \right).$$

At this point we exploit $x_*, x_* + s_\nu \in \mathcal{G}_*$ to infer

$$l_r(x_* + s_\nu, \Sigma_*) = f(x_* + s_\nu) \geq f(x_*) = l_r(x_*, \Sigma_*)$$

for all large ν by local optimality. Taking the limit implies $\partial_x^2 l_r(x_*, \Sigma_*)(s, s) \geq 0$ and hence $\partial_x^2 l_r(x_*, \Sigma_*)(d, d) \geq 0$. ∎

If we define the linearization cone of $\mathcal{G}_*$ at x_* as

$$L_*(x_*) = \left\{ d \in \mathcal{X} : \ u^T \left[\partial G_r(x_*)d \right] u \leq 0 \text{ for all } u \in \mathrm{Ker}(\Sigma_*) \text{ and } [\partial G_r(x_*)d]\Sigma_* = 0 \right\}, \tag{7.60}$$

it is again simple to show that $T_{\mathcal{G}_*}(x_*) \subset L_*(x_*)$.

Lemma 7.16 *The inclusion $T_{\mathcal{G}_*}(x_*) \subset L_*(x_*)$ holds true in general.*

Proof. Take arbitrary $d \in T_{\mathcal{G}_*}(x_*)$, $u \in \mathrm{Ker}(\Sigma_*)$ and $v, w \in \mathbb{R}^q$, where q is the number of rows of $G_\mathrm{r}(x)$. The mapping $g(x) := u^T G_\mathrm{r}(x)u + v^T G_\mathrm{r}(x)\Sigma_* w$ satisfies $g(x) \le 0$ for all $x \in \mathcal{G}_*$ and $g(x_*) = 0$, which implies that g attains a maximum at $x = x_*$ on $\mathcal{G}_*$. By Theorem 7.7 this implies

$$\partial g(x)d = u^T \left[\partial G_\mathrm{r}(x_*)d\right] u + v^T \left[\partial G_\mathrm{r}(x_*)d\right] \Sigma_* w \le 0.$$

Since u, v and w were arbitrary we infer

$$u^T \left[\partial G_\mathrm{r}(x_*)d\right] u \le 0 \text{ for all } u \in \mathrm{Ker}(\Sigma_*) \text{ and } [\partial G_\mathrm{r}(x)d]\Sigma_* = 0.$$

Therefore $d \in L_*(x_*)$.

∎

We will show that $T_{\mathcal{G}_*}(x_*) = L_*(x_*)$ if the following two conditions are satisfied:

$$\mathcal{X} \ni d \mapsto [\partial G_\mathrm{r}(x_*)d]\Sigma_* \in \{Y\Sigma_* : Y = Y^T\} \text{ is surjective} \tag{7.61}$$

and

$$\exists d_0 \in \mathcal{X} \text{ with } \left[\partial G_\mathrm{r}(x_*)d_0\right] \Sigma_* = 0 \text{ and } u^T \left[\partial G_\mathrm{r}(x_*)d_0\right] u < 0 \text{ for all } u \in \mathrm{Ker}(\Sigma_*) \backslash \{0\}. \tag{7.62}$$

To show $T_{\mathcal{G}_*}(x_*) = L_*(x_*)$ we additionally need the following auxiliary result.

Lemma 7.17 *Consider a continuously differentiable function $h : \mathbb{R}^m \mapsto \mathbb{R}^n$ with $x_* \in \mathbb{R}^m$ such that $h(x_*) = 0$ and $\partial h(x_*)$ is surjective. Then for each $d \in \mathrm{Ker}(\partial h(x_*))$ there exist an $\epsilon > 0$ and a continuously differentiable curve $c : (-\epsilon, \epsilon) \to \mathbb{R}^m$ with $c(0) = x_*$ such that $h(c(t)) = 0$ for all $t \in (-\epsilon, \epsilon)$ and $\frac{d}{dt}c(0) = d$.*

Proof. See [125].

∎

We are now in the position to prove $T_{\mathcal{G}_*}(x_*) = L_*(x_*)$ under the hypotheses (7.61) and (7.62).

Theorem 7.18 *If (7.61) and (7.62) hold true then $T_{\mathcal{G}_*}(x_*) = L_*(x_*)$.*

Proof. Let d_0 be a vector that satisfies (7.62). We will first show that $d_0 \in T_{\mathcal{G}}(x_*)$. If $d_0 = 0$ then $d_0 \in T_{\mathcal{G}}(x_*)$, so let us assume $d_0 \neq 0$. To prove $d_0 \in T_{\mathcal{G}}(x_*)$ we will construct a sequence s_ν that satisfies (7.48). Let us choose a orthogonal matrix $U_{\Sigma_*} = \begin{pmatrix} V_{\Sigma_*} & W_{\Sigma_*} \end{pmatrix}$ such that the columns of V_{Σ_*} span the kernel of Σ_*. Furthermore, define

$$h_*(x) := \begin{pmatrix} \mathrm{vec}\left(V_{\Sigma_*}^T G_\mathrm{r}(x)W_{\Sigma_*}\right) \\ \mathrm{svec}\left(W_{\Sigma_*}^T G_\mathrm{r}(x)W_{\Sigma_*}\right) \end{pmatrix},$$

where svec is as defined in (7.9). The definition of h_* is motivated by

$$h_*(x) = 0 \text{ if and only if } G_\mathrm{r}(x)\Sigma_* = 0 \tag{7.63}$$

and (7.61) implies that $\partial h_*(x_*)$ is surjective.

Now observe that (7.62) implies $[\partial G_{\mathrm{r}}(x_*)d_0]\,\Sigma_* = 0$ and hence $\partial h_*(x_*) = 0$. Combined with surjectivity of $\partial h_*(x_*)$ and Lemma 7.17 we conclude that there exists an $\epsilon > 0$ and a continuously differentiable curve $c : (-\epsilon, \epsilon) \to \mathbb{R}^m$, such that $h(c(t)) = 0$ for all $t \in (-\epsilon, \epsilon)$ and $\frac{d}{dt}c(0) = d_0$. Since $\frac{d}{dt}c(0) = d_0 \neq 0$ there exists a sequence t_ν such that $\lim_{\nu \to \infty} t_\nu = 0$ and

$$t_\nu \in (-\epsilon, \epsilon) \text{ and } c(t_\nu) \neq c(0) = x_* \text{ for all } \nu.$$

If we therefore choose $s_\nu = c(t_\nu) - x_*$, then

$$\lim_{\nu \to \infty} \frac{s_\nu}{\|s_\nu\|} = \lim_{\nu \to \infty} \frac{c(t_\nu) - x_*}{\|c(t_\nu) - x_*\|} = \lim_{\nu \to \infty} \frac{\frac{c(t_\nu) - c(0)}{t_\nu}}{\left\| \frac{c(t_\nu) - c(0)}{t_\nu} \right\|} = \frac{\frac{d}{dt}c(0)}{\left\| \frac{d}{dt}c(0) \right\|} = \frac{d_0}{\|d_0\|} = \lambda d_0,$$

for $\lambda = \frac{1}{\|d_0\|}$. Furthermore $h(c(t)) = 0$ for all $t \in (-\epsilon, \epsilon)$ implies that $G_{\mathrm{r}}(x_* + s_\nu)\Sigma_* = G_{\mathrm{r}}(c(t_\nu))\Sigma_* = 0$ for all ν.

Finally, let us show that for all large enough ν

$$u^T \partial G_{\mathrm{r}}(x_*)(s_\nu)u < 0 \text{ for all } u \in \mathrm{Ker}(\Sigma_*)\backslash\{0\}$$

or equivalently

$$V_{\Sigma_*}^T G_{\mathrm{r}}(x_* + s_\nu)V_{\Sigma_*} \prec 0. \tag{7.64}$$

We show that (7.64) holds true for all large enough ν using similar arguments as in Theorem 7.11. For this purpose define $R(t) := V_{\Sigma_*}^T G_{\mathrm{r}}(c(t))V_{\Sigma_*}$ and

$$\epsilon_\nu := \frac{R(t_\nu) - R(0) - \partial R(0)(t_\nu)}{t_\nu}$$

such that $\lim_{\nu \to \infty} \epsilon_\nu = 0$. Since $G_{\mathrm{r}}(x_*) = 0$ we have $R(0) = V_{\Sigma_*}^T G_{\mathrm{r}}(x_*)V_{\Sigma_*} = 0$ and by the chain rule of differentiation we conclude

$$\partial R(0) = V_{\Sigma_*}^T \left[\partial G_{\mathrm{r}}(x_*)\frac{d}{dt}c(0) \right] V_{\Sigma_*} = V_{\Sigma_*}^T \left[\partial G_{\mathrm{r}}(x_*)d_0 \right] V_{\Sigma_*}.$$

This implies that

$$\frac{1}{t_\nu} \left(V_{\Sigma_*}^T G_{\mathrm{r}}(x_* + s_\nu)V_{\Sigma_*} \right) = \frac{R(t_\nu)}{t_\nu} = \epsilon_\nu + \frac{1}{t_\nu}R(0) + \frac{1}{t_\nu} \left[\partial R(0)t_\nu \right] = \epsilon_\nu + \frac{1}{t_\nu}V_{\Sigma_*}^T \left[\partial G_{\mathrm{r}}(x_*)d_0 \right] V_{\Sigma_*}$$

Since $V_{\Sigma_*}^T \left[\partial G_{\mathrm{r}}(x_*)d_0 \right] V_{\Sigma_*} \prec 0$ by (7.62) we infer that for all large enough ν (7.64) holds true.

In summary, the sequence s_ν satisfies for all large enough ν all the conditions in (7.48) such that $\lim_{\nu \to \infty} \frac{s_\nu}{\|s_\nu\|} = \lambda d_0 \in T_{\mathcal{G}}(x_*)$ and hence $d_0 \in T_{\mathcal{G}}(x_*)$.

Now choose an arbitrary $d_l \in L_*(x_*)$. Then $d(\tau) := (1 - \tau)d_l + \tau d_0$ for arbitrary $\tau \in (0, 1]$ is a vector that satisfies (7.62) because

$$[\partial G_{\mathrm{r}}(x_*)d(\tau)]\,\Sigma_* = \tau \underbrace{[\partial G_{\mathrm{r}}(x_*)d_0]\,\Sigma_*}_{=0 \text{ by } (7.62)} + (1 - \tau)\underbrace{[\partial G_{\mathrm{r}}(x_*)d_l]\,\Sigma_*}_{=0 \text{ by } d_l \in L_*(x_*)} = 0$$

and for arbitrary $u \in \mathrm{Ker}(\Sigma_*)\backslash\{0\}$

$$u^T \left[\partial G_{\mathrm{r}}(x_*)d(\tau)\right] u = \tau \underbrace{u^T \left[\partial G_{\mathrm{r}}(x_*)d_0\right] u}_{<0 \text{ by } (7.62)} + (1-\tau) \underbrace{u^T \left[\partial G_{\mathrm{r}}(x_*)d_l\right] u}_{\leq 0 \text{ by } d_l \in L_*(x_*)} < 0.$$

Hence $d(\tau)$ for $\tau \in (0,1]$ is contained in $T_{\mathcal{G}_*}(x_*)$. Finally, due to the fact that $T_{\mathcal{G}_*}(x_*)$ is closed, $d_l = \lim_{\tau \to 0, \ \tau > 0} d(\tau) \in T_{\mathcal{G}_*}(x_*)$. Since $d_l \in L_*(x_*)$ was arbitrary we conclude $L_*(x_*) \subset T_{\mathcal{G}_*}(x_*)$. Combining this with Lemma 7.16 yields $L_*(x_*) = T_{\mathcal{G}_*}(x_*)$. ∎

Under these constraint qualifications the second order condition can be formulated with $T_{\mathcal{G}_*}(x_*)$ replaced by $L_*(x_*)$. This shows that the second order necessary conditions are ideally close to the following sufficient conditions for local optimality which do not require any constraint qualification whatsoever.

Theorem 7.19 (Second order sufficient optimality condition) *Suppose at $x_* \in \mathcal{G}$ there exists $\Sigma_* \in \mathbf{\Sigma}_*$ such that*

$$\partial_x^2 l_r(x_*, \Sigma_*)(d,d) > 0 \quad \text{for all} \quad d \in L_*(x_*) \setminus \{0\}.$$

Then x_* is a strict local minimum point of f on $\mathcal{G}$.

Proof. If x_* is not strictly locally optimal, there exists a sequence $s_\nu \neq 0$ with $s_\nu \to 0$, $x + s_\nu \in \mathcal{G}$, $s_\nu / \|s_\nu\| \to s \neq 0$ and $f(x_* + s_\nu) \leq f(x_*)$. Clearly

$$\frac{f(x_* + s_\nu) - f(x_*) - \partial f(x_*)s_\nu}{\|s_\nu\|} \to 0,$$

and

$$\frac{G_{\mathrm{r}}(x_* + s_\nu) - G_{\mathrm{r}}(x_*) - \partial G_{\mathrm{r}}(x_*)s_\nu}{\|s_\nu\|} \to 0.$$

For large ν we infer $x_* + s_\nu \in N \cap \mathcal{G}$ and hence $x_* + s_\nu \in \mathcal{G}_{\mathrm{r}}$ or $G_{\mathrm{r}}(x + s_\nu) \leq 0$. Since $G_{\mathrm{r}}(x_*) = 0$ we infer $\partial f(x_*)s \leq 0$ and $\partial G_{\mathrm{r}}(x_*)s \leq 0$ by taking the limit. Since $\Sigma_* \geq 0$ we have $\mathrm{Trace}(\Sigma_*[\partial G_{\mathrm{r}}(x_*)s]) \leq 0$. On the other hand we exploit $\partial f(x_*)s + \mathrm{Trace}(\Sigma_*[\partial G_{\mathrm{r}}(x_*)s]) = 0$ to see $\mathrm{Trace}(\Sigma_*[\partial G_{\mathrm{r}}(x_*)s]) \geq 0$. Hence in fact $\mathrm{Trace}(\Sigma_*[\partial G_{\mathrm{r}}(x_*)s]) = 0$ and thus even $[\partial G_{\mathrm{r}}(x_*)s]\Sigma_* = 0$. Therefore $s \in L_*(x_*)$. By hypothesis we conclude $\partial_x^2 l_r(x_*, \Sigma_*)(s,s) > 0$ which leads to a contradiction as follows. Since l_{r} is C^2 as a function of x we infer

$$\frac{l_{\mathrm{r}}(x_* + s_\nu, \Sigma_*) + q_\nu}{\|s_\nu\|^2} \to 0.$$

where

$$q_\nu := -l_{\mathrm{r}}(x_*, \Sigma_*) - \partial_x l_{\mathrm{r}}(x_*, \Sigma_*)(s_\nu) - \partial_x^2 l_{\mathrm{r}}(x_*, \Sigma_*)(s_\nu, s_\nu)$$

We now use $l_{\mathrm{r}}(x_* + s_\nu, \Sigma_*) = f(x_* + s_\nu) + \mathrm{Trace}(\Sigma_* G_{\mathrm{r}}(x_* + s_\nu)) \leq f(x_* + s_\nu) \leq f(x_*) = f(x_*) + \mathrm{Trace}(\Sigma_* G_{\mathrm{r}}(x_*)) = l_{\mathrm{r}}(x_*, \Sigma_*)$. Due to $\partial_x l_{\mathrm{r}}(x_*, \Sigma_*) = 0$ we arrive at

$$\partial_x^2 l_{\mathrm{r}}(x_*, \Sigma_*)(s_\nu, s_\nu)/\|s_\nu\|^2 \leq 0$$

which leads to $\partial_x^2 l_{\mathrm{r}}(x_*, \Sigma_*)(s,s) \leq 0$ by taking the limit, a clear contradiction.

∎

Table 7.2: Parameters of two mass system

variable	unit	value
m_1	kg	0.2
m_2	kg	0.2
k	$\frac{N}{m}$	300
c	$\frac{Ns}{m}$	0.1

7.3.3 Discussion of the optimality conditions

We have derived in an elementary fashion the first and second order conditions in terms of tangent and linearization cones under weak hypotheses. We believe that the presented approach is considerably more elementary than those in [182, 24] and their references, and it still leads, in contrast to [60], to very 'sharp' conditions.

In this thesis the main purpose of the second order conditions is to verify at least local optimality of a computed reduced order controller. Unfortunately, for the problem under investigation the second order sufficient conditions cannot be satisfied since $\{x \in \mathcal{X}: f(x) = \gamma\} \cap \mathcal{G}$ never has isolated points. This simply follows from the fact that one can change x by a coordinate change in the controller state without modifying $f(x)$. A precise analysis of this phenomenon and its effects on the CLIP algorithm are the subject of Chapter 8.
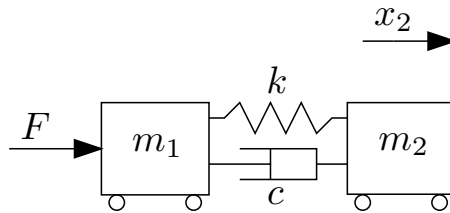
Figure 7.6: One degree-of-freedom mechanical motion system with two masses connected by a spring and a damper.

Example 7.20 *We illustrate the optimality conditions with a fixed-order $\mathcal{H}_\infty$-optimal controller synthesis problem for a two-mass system as illustrated in Figure 7.6* [3]. *The parameters of this system are given in Table 7.2*

The transfer function from the force F to the position x_2 is given by

$$G(s) = \frac{s + 3000}{2s^2(s^2 + s + 3000)}.$$

A Bode plot of this transfer function is shown in Figure 7.7. We have designed a full-order $\mathcal{H}_\infty$-optimal controller using the four-block design that will be discussed in full detail in Section 9.3. The generalized plant has McMillan degree $n = 7$ and is given in Appendix E. The optimal full order controller is shown in Figure 7.8. Its closed-loop $\mathcal{H}_\infty$-norm is 1.78. We reduced this controller to one with McMillan degree $n_c = 4$ by closed-loop balanced

---

[3]This example is inspired by personal communication with E. van de Meché.
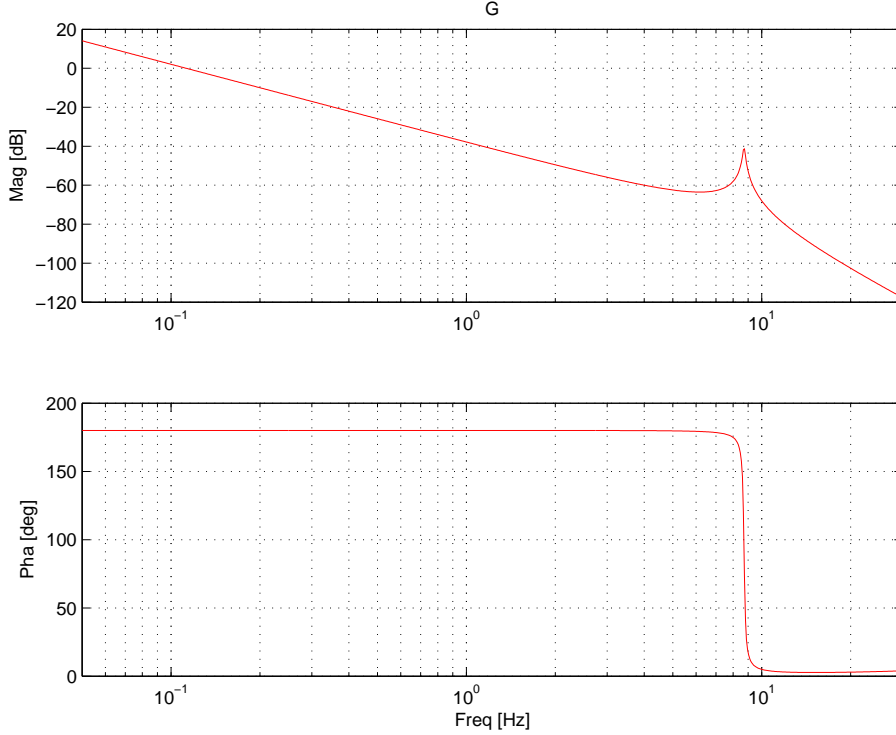
Figure 7.7: Bode plot of open-loop system G

truncation, as discussed in Section 3.1.1. The resulting controller has a closed-loop $\mathcal{H}_\infty$-norm of 1.84. Based on this controller, a Riccati solution for X has been calculated, as explained in Section 7.1.6. This gives a barrier value $\phi(x_{init}) = 21.9$ with $x_{init} := (\gamma_{init}, K_{init}, X_{init})$. We optimized the controller with the Interior Point method, where x_{init} has been used as initial point. The resulting optimized controller results in a closed-loop $\mathcal{H}_\infty$-norm of 1.79. The reduced and optimized controller are shown in Figure 7.8 in red and green, respectively.

We verify the optimality conditions for the final iterate $x_f = (\gamma_f, X_f, K_f)$ of the interior-point algorithm for the optimization problem $\inf_{G(x) \preceq 0} f(x)$, where $x := (\gamma, X, K)$, $f(x) := \gamma$ and $G(x) := diag(-X, -\mathcal{B}(\gamma, X, K))$.

To test the first-order optimality conditions, we constructed matrices V_f that approximately span the kernel of $G(x_f)$. We can not exactly find a bases of $Ker(G(x_f))$ to V_*, due to the round-off errors in the computations. Therefore, we computed V_f by first calculating an eigenvalue decomposition of X. Then we constructed a matrix V_X consisting of all eigenvectors corresponding to eigenvalues that are smaller than $10^{-6}\rho_{sp}(X)$, where $\rho_{sp}(\cdot)$ denotes the spectral radius of a matrix. Similarly we constructed $V_\mathcal{B}$ of all eigenvectors corresponding to eigenvalues of $\mathcal{B}(\gamma, X, K)$ that are smaller than $10^{-6}\rho_{sp}(\mathcal{B}(\gamma, X, K))$. Finally, V is constructed by diagonal augmentation of these matrices: $V = diag(V_X, V_\mathcal{B})$.

By solving a feasibility LMI problem in Matlab LMILAB, we found a vector d with $V^T \partial G(x_f) dV \prec 0$, which implies that G approximately satisfies the MFCQ at x_f, where 'approximately' refers to the approximate construction of the vectors that span the kernel of $G(x_f)$.

To test the first-order necessary optimality conditions we solved the problem

$$\inf_{\Lambda \succeq 0, \ \Lambda G(x_f) = 0} \|\partial_x h(x_f, \Lambda)\|,$$

146
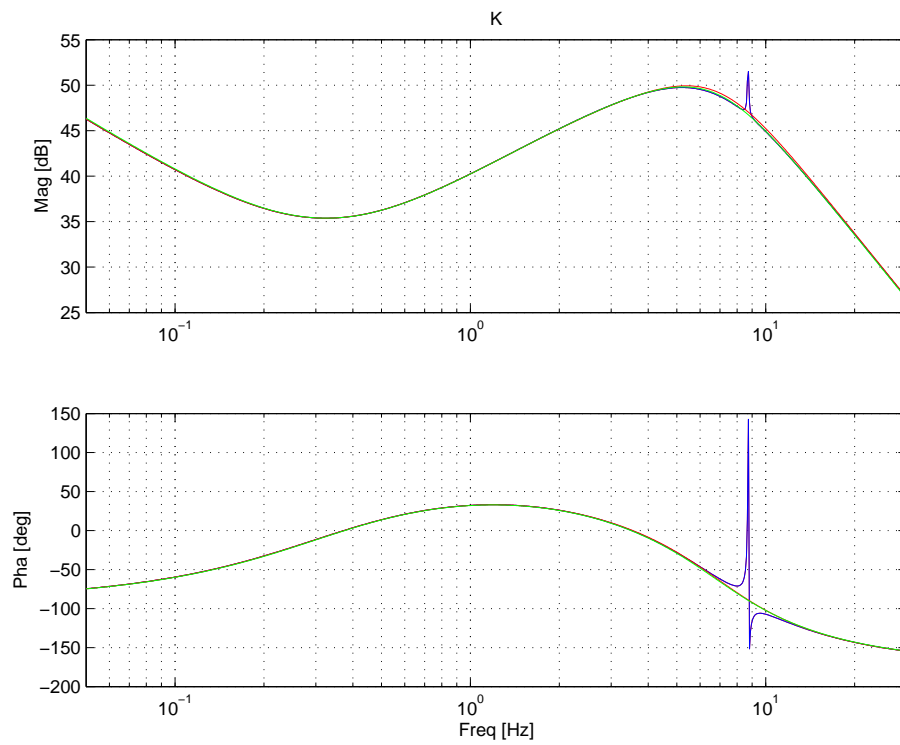
Figure 7.8: Bode plot of full order (blue), reduced (red) and optimized controller (green)
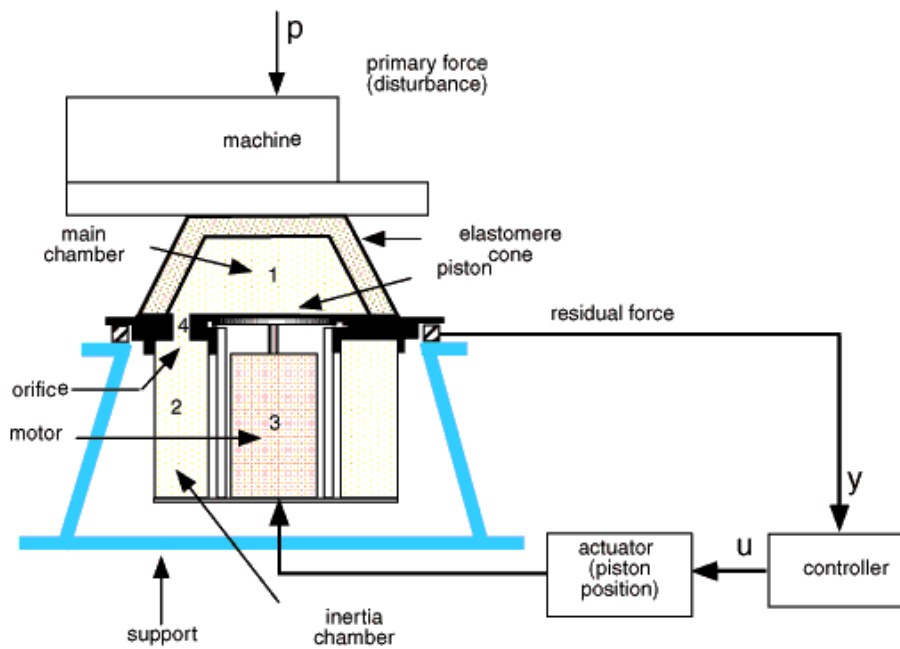
Figure 7.9: Active suspension system

where $h(x, \Lambda) := f(x) + Trace\left(\Lambda \partial G(x)\right)$ *The optimal value is* $3.2 \cdot 10^{-12}$, *which is approximately attained at* Λ_f. Λ_f *is block-diagonal and allows the partitioning*

$$\Lambda_f = \begin{pmatrix} \Lambda_{f,X} & 0 \\ 0 & \Lambda_{f,\mathcal{B}} \end{pmatrix},$$

where $\Lambda_{f,X} \in \mathcal{S}^{n+n_c}$, $\Lambda_{f,X} \in \mathcal{S}^{n+n_c+m_1+p_1}$, $n + n_c = 10$ *and* $n + n_c + m_1 + p_1 = 14$. *We computed an approximation of the linearization cone* $L_*(x_f)$, *as defined in* (7.60). *For simplicity, we discard the extra condition related to* $Ker(\Lambda_f)$, *i.e. instead of* (7.60) *we compute the subspace*

$$L_{subspace}(x_f) = \left\{ d \in \mathcal{X} : \ [\partial G_r(x_f)d]\Sigma_f = 0 \right\}.$$

Since $L_*(x) \subset L_{subspace}(x)$ *for all* $x \in \mathcal{X}$, *the sufficient condition in Theorem 7.19 holds true if* L_* *is replaced by* $L_{subspace}$. *We compute* $L_{subspace}(x_f)$ *by computing the vectors that approximately span the kernel of the matrix* $M_{L_{subspace}}$ *that represents the mapping*

$$\mathcal{X} \ni d \rightarrow [\partial G_r(x_f)d] \Lambda_{subspace}.$$

This span is approximately determined by a singular value decomposition $M_{L_{subspace}} = U_{L_{subspace}} \Sigma V_{L_{subspace}}$: *it consists of all column vectors of* $V_{L_{subspace}}$ *that correspond to singular values that are smaller than* $10^{-8} \|M_{L_{subspace}}\|$.

We tested a second order sufficient condition for optimality by computing the eigenvalues of the matrix representation $H_{L_{subspace}}$ *of the mapping*

$$L_{subspace}(x_f) \ni d \rightarrow \partial_x^2 l_r(x_f, \Sigma_f)(d, d)$$

Since $H_{L_{subspace}} \in \mathcal{S}^{34}$ *is symmetric, all the eigenvalues of* $H_{L_{subspace}}$ *are real-valued. If they are furthermore all strictly positive, Theorem 7.19 implies that* x_f *is locally optimal. Eight of the 34 eigenvalues of* $H_{L_{subspace}}$ *are smaller than* 10^{-12}, *the most negative one equal to* $-5.9 \cdot 10^{-12}$. *Hence, the sufficient condition is violated. This example will be continued in Section 8.4.3 in Chapter 8, where optimality will be tested for the same problem, but with a different parametrization with a smaller number of decision variables.*

Remark. The outcome of the optimality conditions strongly depends on the accuracy of the LMI solver that is used and the accuracy of the calculation of the vectors that span the bases of the kernels of G, Λ etc.

7.4   Application

7.4.1   Controller design for an active suspension system

The IP method is applied to the reduced order synthesis of an active suspension system, see Figure 7.9. This design, together with the IP algorithm and the optimality conditions have been published in [93, 90]. The reader is referred to [118] for details on the active suspension system. The system has been a benchmark for the synthesis of fixed-order controllers. A special issue [117] of the European Journal of Control has been devoted to this benchmark, which contains a nice collection of low-order controller designs for the system. The experimental performance of these controllers, including the controller designed in this section, are compared in it.
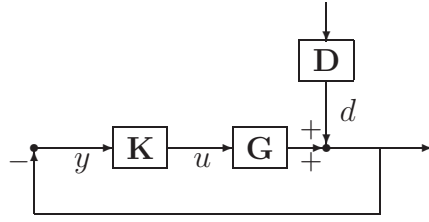
Figure 7.10: Closed-loop system of the active suspension system
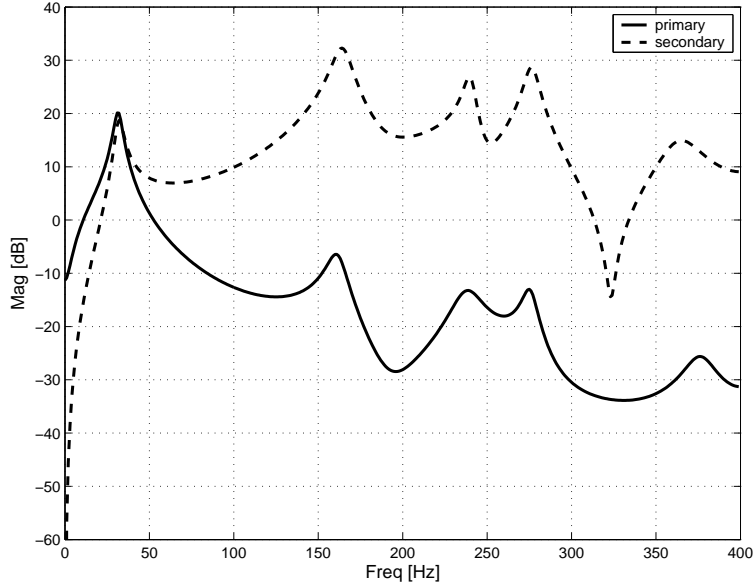
Figure 7.11: Open loop primary and secondary path frequency response

The structure of the closed-loop system is shown in Figure 7.10. The control input u drives the position of the piston. The measured output is the residual force. The transfer function between the excitation of the shaker and the residual force is called the primary path or the 'disturbance model' **D**. The system **G** is in this setting called the 'secondary path'. Its transfer function $G(z)$ maps the control input into the residual force. The control objective is to compute a linear discrete-time controller which minimizes the residual force around the first and second vibration modes of the primary path (disturbance) model **D** and which distributes the amplification of the disturbances over the higher frequencies. We choose the sampling frequency at $f_s = 800$Hz, which is also the sampling frequency of the identified discrete-time models. The frequency response of the primary path $D(z)$ and secondary path $G(z)$ are shown in Figure 7.11. Note that a linear frequency axis is used in these and the plots presented in the sequel. For such an axis is chosen in this chapter, since the performance specifications for the benchmark system in [117] are specified with straight lines on a Bode magnitude plot with linear frequency axis.

The control objective is represented in terms of constraints on the closed loop sensitivity $S(z) = \frac{1}{1+G(z)K(z)}$ and controller sensitivity $K(z)S(z) = \frac{-K(z)}{1+G(z)K(z)}$ in the frequency domain, as shown in Figures 7.12 and 7.13 respectively.

We compute the controllers by $\mathcal{H}_\infty$ synthesis in discrete-time using a four-block design, as was depicted in Figure 2.2. This four-block interconnection structure is repeated in
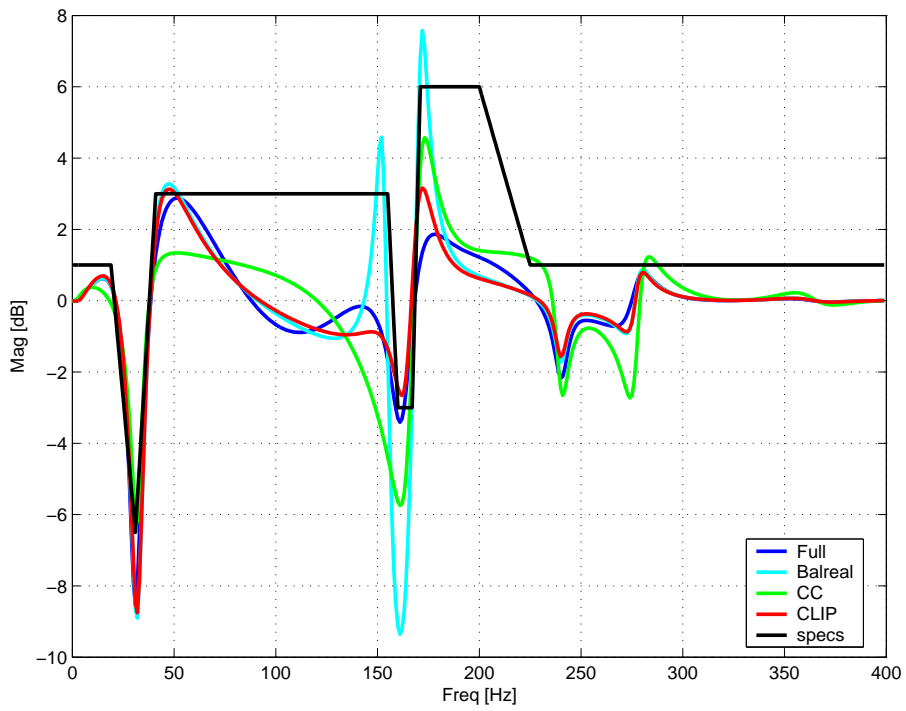
Figure 7.12: Sensitivity
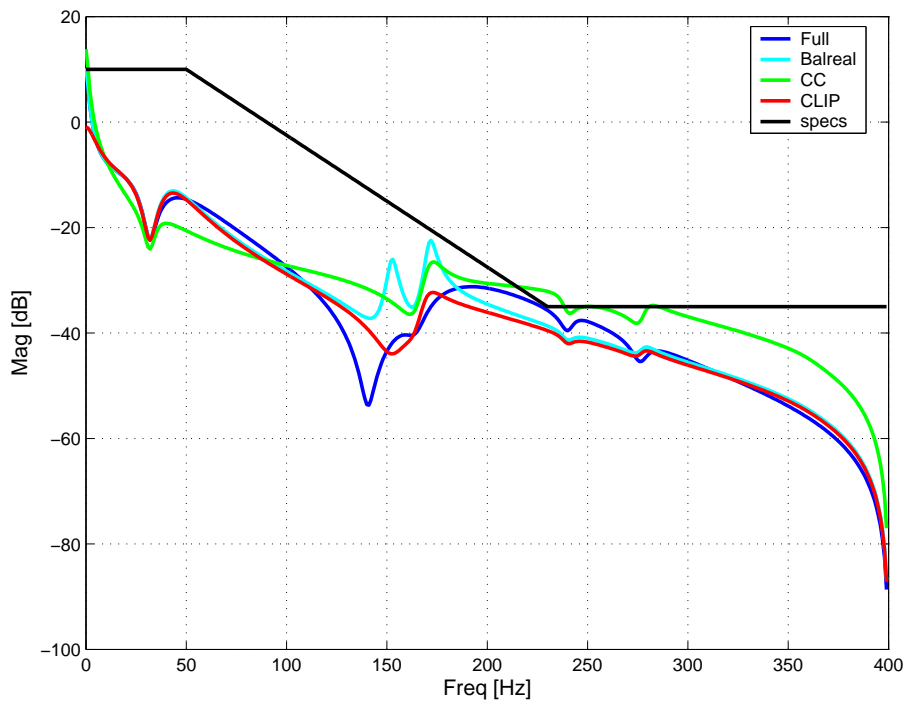
Figure 7.13: Controller Sensitivity
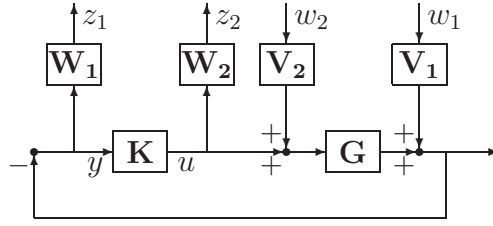
Figure 7.14: Four-block controller set-up

Figure 7.14 for the reader's convenience. We used a four-block design, since this leads in our experience to better results for this system (i.e. a closed-loop that better satisfies the performance criteria) than an S/KS [183] design. A further motivation of this choice and detailed explanation of the design procedure is postponed to Section 9.3. An identified 17th order discrete-time model $G(z)$ of the secondary path $\mathbf{G}$ is used. The transfer function $P(z)$ of the plant $\mathbf{P}$ is given by:

$$P = \begin{pmatrix} -W_1V_1 & -W_1GV_2 & -W_1G \\ 0 & 0 & W_2 \\ -V_1 & -GV_2 & -G \end{pmatrix} \tag{7.65}$$

and the closed loop is given by

$$P = \begin{pmatrix} W_1SV_1 & W_1SGV_2 \\ W_2KSV_1 & W_2KSGV_2 \end{pmatrix}$$

where $S(z) = \frac{1}{1+K(z)G(z)}$ is the sensitivity and $V_1(z)$, V_2, W_1 and $W_2(z)$ are the weights in $\mathcal{RH}_\infty$ The transfer functions of the weights are obtained by full order synthesis tuning. They are given by:

$$
\begin{aligned}
V_1(z) &= \frac{0.575z^4 - 1.207z^3 + 1.205z^2 - 0.75z + 0.2335}{z^4 - 2.498z^3 + 3.022z^2 - 2.336z + 0.8914} \\
W_1(z) &= 1 \\
W_2(z) &= \frac{8.952z^6 - 17.52z^5 + 14.16z^4 - 6.046z^3 + 1.437z^2 - 0.1799z + 0.009276}{z^6 + 0.8845z^5 + 0.4842z^4 + 0.1615z^3 + 0.03581z^2 + 0.004889z + 0.0002892} \\
V_2(z) &= 0.01
\end{aligned}
$$

The frequency responses of V_1 and W_2 are given in Figures 7.15 and 7.16 respectively. The 6th order weight W_2 suppresses large control action at high frequencies (larger than 200Hz). The weight V_1 is a low order approximation of the transfer function of the disturbance model $D(z)$. It is a 4th order filter with reversed notches at 31.5Hz and 160Hz, which are the most dominant vibration modes of the primary path. The order of the plant is 27: 17 are attributed to the identified model of the secondary path G and 10 to the weighting functions. For both full and fixed-order design the plant is bi-linearly transformed into a continuous-time model using Tustin's approximation $z \rightarrow \frac{1+\frac{1}{2}sh}{1-\frac{1}{2}sh}$, where $h = \frac{1}{f_s} = 1.25$ms is the sampling period. The controller synthesis is performed in continuous time and the controller is transformed back into discrete time. The frequency response of the controllers are shown in Figure 7.17 where the following abbreviations are used:

- **Full**: full order controller design. The controller order is 28.

- **Balreal**: fixed-order controller design by a posteriori reduction as explained in Section 3.1. The controller order is 6.
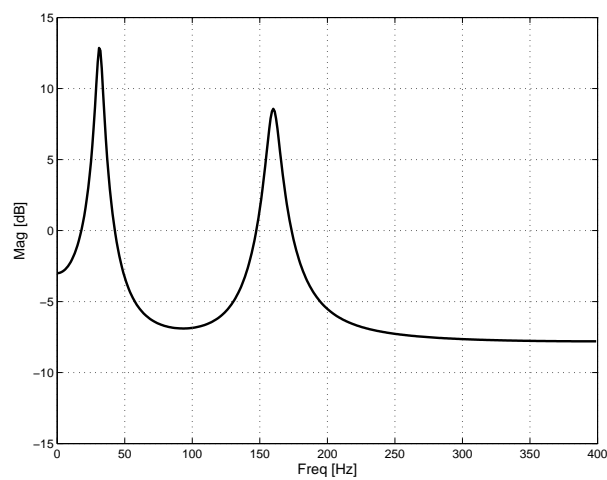
Figure 7.15: Weight $V_1(z)$: amplitude frequency response.
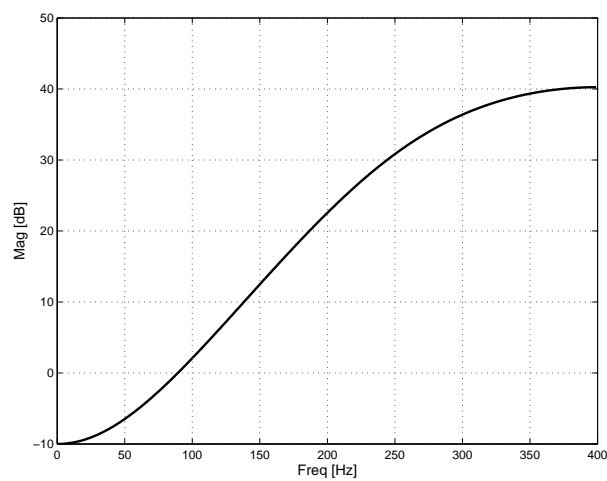
Figure 7.16: Weight $W_2(z)$: amplitude frequency response.

- **CC**: fixed-order controller design by the cone complementarity method as explained in Section 3.2.1. The controller order is 6.

- **CLIP**: fixed-order controller design by the curved line-search Interior Point method as explained in Section 7.1. The controller order is 6.

The computed fixed-order controllers are all of order $n_c = 5$. Because it is required that the controller gain is equal to zero at frequency $0.5f_s$, an extra term $\frac{z+1}{z}$ is added to all controllers after the fixed-order design. The displayed fixed-order controllers all have therefore McMillan degree 6. The extra term $\frac{z+1}{z}$ does not change the transfer function of the controllers much in the important frequency range (0-300[Hz]), such that the performance loss due its a posteriori addition is probably small. The controllers are implemented at sampling frequency $f_s = 800$Hz. Their closed loop performance is shown in Figures 7.12 and 7.13 respectively. The transfer functions of the fixed-order controllers are given in Appendix B. The maximum singular value of the closed loop plant $\bar{\sigma}\left(F_l(P(e^{j\omega h}), K(e^{j\omega h}))\right)$ is plotted in Figure 7.18 as function of the frequency ω, where P is the transfer function of the plant as defined in (7.65) and $h = 1.25$ms the sampling period. The full order $\mathcal{H}_\infty$-optimal controller (denoted by 'Full') yields a closed-loop performance of $\|\mathcal{F}_l(P, K)\|_\infty \approx 2.48$, where $\|\cdot\|_\infty$ denotes the discrete-time $\mathcal{H}_\infty$-norm. The full order controller has been reduced to a 5th order controller (denoted by 'Balreal') by closed loop weighted controller reduction. The $\mathcal{H}_\infty$-norm of the weighted closed-loop is 3.41 for this controller. Although the pole-pair at 160Hz is present in both full-order and Balreal controller, the shape of the amplitude of the frequency response around this frequency is very different. This results in large sensitivity peaks slightly below and above 160Hz. These peaks are also the reason for the increase of the weighted closed-loop $\mathcal{H}_\infty$-norm, as is clear from Figure 7.18. This controller served as starting point for the Curved Line-search Interior Point method, as presented in Section 7.1. By this method we computed a controller (denoted by 'CLIP') with weighted closed-loop $\mathcal{H}_\infty$-norm 2.51, which is very close to that of the full order controller. From the magnitude of the frequency response of the controller in Figure 7.17 it is clear that the full order controller exhibits many small peaks that are likely to be irrelevant for the closed loop behavior. (The phase of the frequency response of the controllers is shown for completeness in Appendix B.) The CLIP controller has a resonant pole-pair at about 160Hz respectively, close to the second resonant mode. It does not have a pole-pair at 31.5Hz to suppress the first vibration mode, because the open-loop plant has already a mode there. The resonances of the plant above 200Hz are not incorporated in the CLIP controller. This does not lead to much performance loss with respect to the full-order controller, apparently because these resonances are not so important for the suppression at the vibration modes. The closed-loop interconnection of the plant with the controller (denoted by 'CC') computed by the cone complementarity method has $\mathcal{H}_\infty$-norm 2.63. Its shape at high frequencies (> 200Hz) is remarkably different from the other controllers. It also violates the specifications on the controller sensitivity at high frequencies, see Figure 7.13. The high-frequency content can probably be suppressed with a different weight W_2. We have chosen however, to use the same weights for all controllers.

The performance and computation time are shown in Table 7.4. The number of decision variables in the Interior Point optimization is $1 + \frac{1}{2}(n + n_c)(n + n_c + 1) + m_c = 565$. Closed-loop experiments have been performed on the experimental setup in Grenoble with the controller computed with the CLIP method. The disturbance force is a Pseudo-Random Binary Signal (PRBS) applied to the system by a shaker. Estimates of the square root of the power spectral densities of the sensitivity and controller sensitivity are shown in Figures
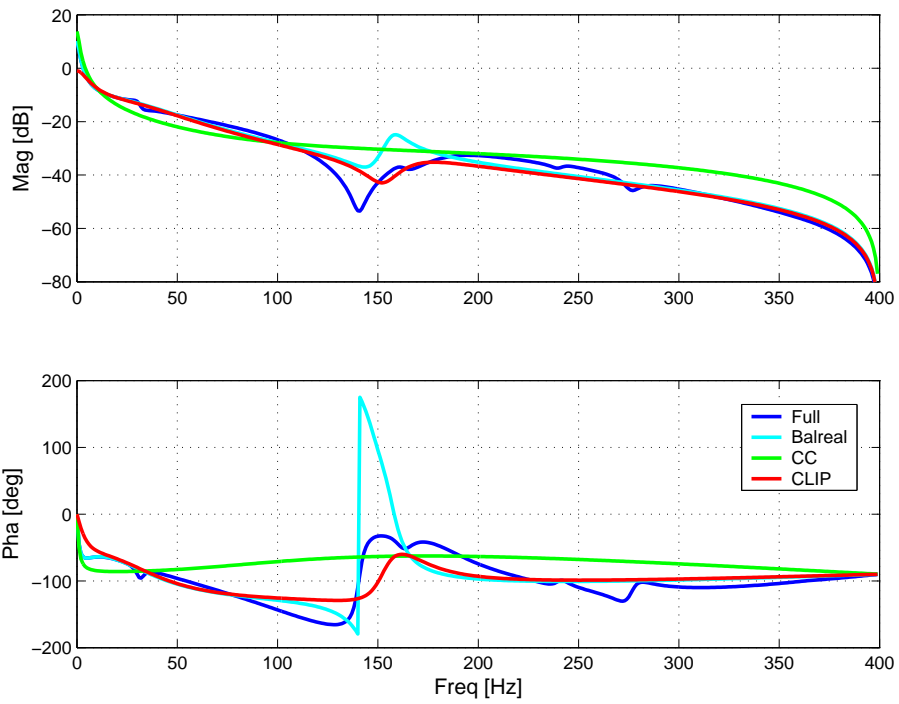
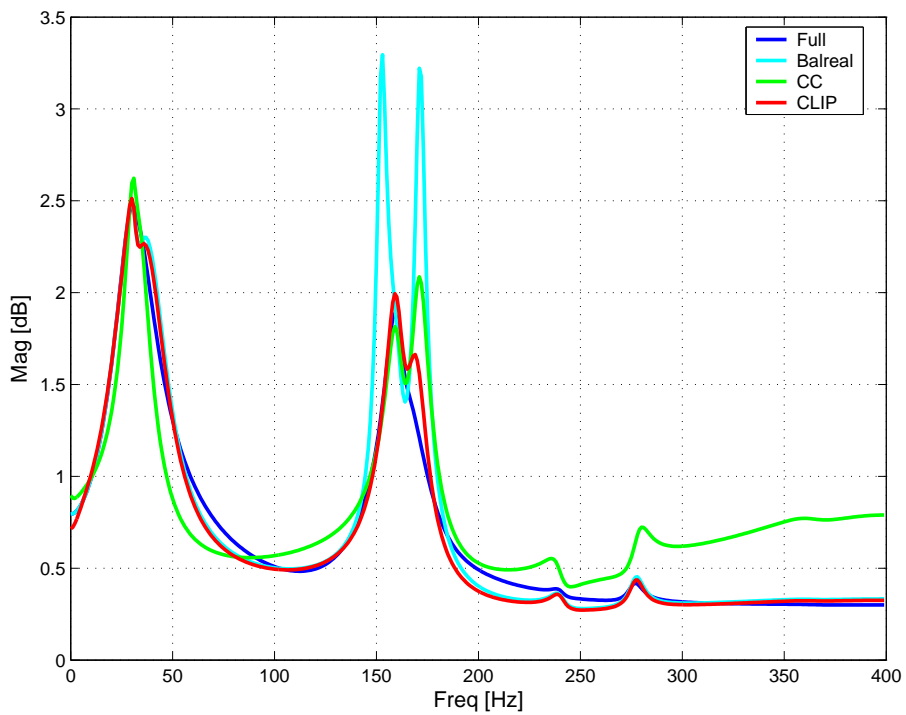Figure 7.17: Controllers: frequency response

Figure 7.18: Maximum singular value of the closed loop plant

154

7.19 and 7.20, which illustrates the good performance of the controller. The resonances at 37Hz and 160Hz are sufficiently suppressed and the control effort satisfies the specifications. A slight violation of the specification is observed around 180Hz, 230Hz and at a few higher frequencies.
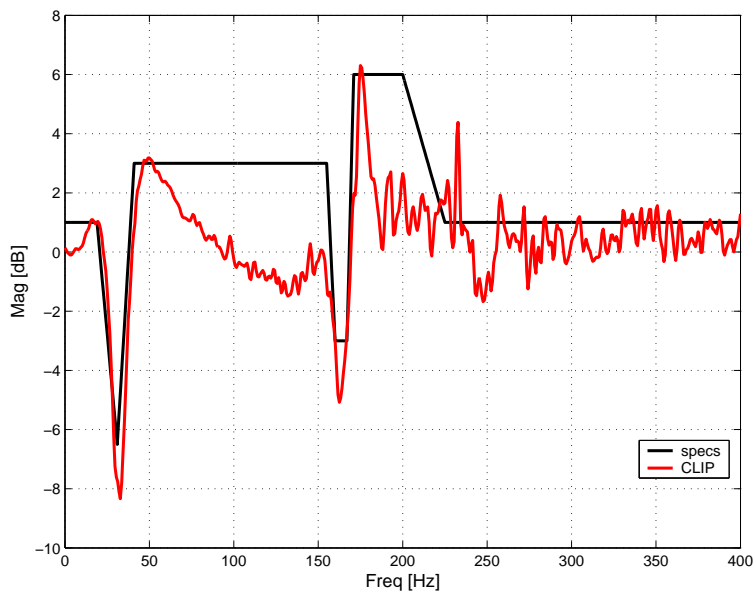

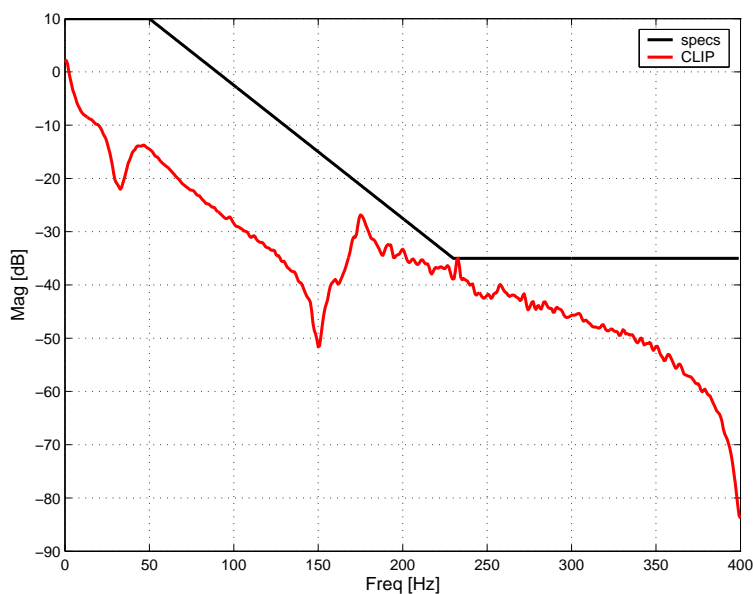Figure 7.19: Experimental results: estimated square root spectral density of sensitivity

Figure 7.20: Experimental results: estimated square root spectral density of controller sensitivity

7.4.2 Discussion

Due to its relatively low computational effort, the closed loop balanced reduction method is an efficient method for tuning of the weighting functions for reduced order $\mathcal{H}_\infty$ synthesis. Because the method requires only the solution of two Lyapunov equations if the full

Table 7.3: Distribution of computation time in CLIP algorithm

component	relative comp. time
Function evaluations	**2.7** %
Intermediate analysis steps	**16.6**%
Trust region:	**71.1**%
Hessian computation	15.8 %
Schur decomposition $H = Q^T D Q$	33.5%
Solve (7.34)	21.8%

order controller is given, it is very efficient. The closed loop performance of the resulting controllers can also be quickly verified. The method is therefore very fast. Even the closed loop performance for all controllers of order 1 up to full order can be computed within a few seconds in this application. A disadvantage of the method is that stability of the closed loop with the reduced controller is not necessarily preserved and that there is no guarantee of optimality in any respect.

We have shown that the CLIP method converges to a feasible solution to the BMI problem for $\mathcal{H}_\infty$-synthesis that satisfies the first order necessary conditions. This contrasts with the other methods, for which no converge to a local optimal solution can be proven. Furthermore we have provided numerically verifiable tests for the first and second order necessary and sufficient optimality conditions. We have verified that the first order necessary conditions are indeed satisfied by the solution. The computation time of 8 min 11 sec is reasonable[4] and significantly smaller than for the Cone Complementarity algorithm. The distribution of the computation time over the different tasks is shown in Table 7.3. The trust region steps take 71.1% of the total computation time. The Hessian 16.6% of the computation is spent on the intermediate analysis steps by LMI optimization as discussed in Section 7.1.7.

The number of decision variables in the CLIP algorithm is large, since it contains the Lyapunov variable X which has $\frac{1}{2}(n + n_c)(n + n_c + 1)$ variables. Apparently this is the price we have to pay to formulate the problem using the Bounded Real Lemma, i.e. with the constraints (7.1) and (7.2). The main advantage of this formulation is smooth dependence of the objective and constraints on the decision variables. The IP algorithm heavily exploits this smoothness by computing search directions based on first and second-order derivatives.

The controller K_{Balreal} served in the application as a good initial controller for the CLIP method. The CLIP method yields a controller with closed-loop $\mathcal{H}_\infty$-norm that is only slightly larger than the closed-loop $\mathcal{H}_\infty$-norm of the full order controller. The performance of the controller has also been evaluated in terms of the original control objectives, not only in terms of the closed-loop $\mathcal{H}_\infty$-norm. The experimental results have illustrated that the CLIP controller performs also well in this respect, i.e. it adequately suppresses the resonance frequencies of the suspension system.

The closed-loop $\mathcal{H}_\infty$-performance of the CC controller is in between that of K_{Balreal} and K_{CLIP}. An advantage of the method is that the computationally most expensive step of the method, step 2 in Algorithm 3.3, can be solved with standard software, i.e. an LMI-solver. However, for systems with large state dimensions the LMI problems in this step have too many decision variables and constraints to be solved efficiently. This will be illustrated in Chapter 9 for another LMI-based method, the XK-iteration.

---

[4]Due to more efficient implementation the computation time is much smaller than reported in [93]

The interested reader is referred to [58] for an application of a genetic algorithm to our four-block design as discussed in this section.

Table 7.4: Performance and computation time of controllers

Controller	γ	Computation time
Full order	2.4759	11.9sec
Balanced reduced	3.4051	12.8sec
CLIP	2.5065	8min11sec
CC	2.6302	14min33sec

7.5   Conclusion

In this chapter we have presented an algorithm for structured controller synthesis based on Interior Point Optimization. Local convergence is guaranteed using a trust-region, which gives a partial solution to the problem posed in ① of Section 2.8. The algorithm is applicable to every polynomial SDP.

To reduce the computational complexity, we have *exploited the control-theoretic characteristics* in the Newton step computation. The result is an algorithm that is able to solve controller design problems in industrial practice, as is illustrated with the design example for the active suspension system. This chapter therefore has also contributed to an answer of the question posed in ③ of Section 2.8.

Finally, we have developed local optimality certificates for general SDP problems and for the fixed-order control problem in particular. The presented LMI conditions are good means to verify the guaranteed local convergence of the IP algorithm. The results furthermore contribute to an answer to ② in Section 2.8, in the sense that they provide a computationally cheap alternative to the global certificates presented in Chapters 4 and 5. As mentioned on page 145 a reduced parametrization is needed to be able to verify strict optimality of dynamic controllers. This is the subject of the next chapter.

As final remark, we would like to point to the parallel with the SOS relaxations discussed in Chapter 4 that were also applicable to every polynomial SDP, and where the exploitation of its control-theoretic characteristics in Chapter 5 also reduced the computational complexity of the algorithm.

Chapter 8

Over-parametrization of state-space controllers

The state-space representation of a transfer function of a dynamic controller is not unique in general. This may introduce over-parameterization in fixed-order $\mathcal{H}_\infty$-optimal controller synthesis algorithms and absence of *verifiably locally optimal* controllers, by which we mean controllers that satisfy the second order sufficient optimality conditions as have been presented in the previous chapter. A more important and often overlooked drawback of over-parametrization is that it complicates the computation of accurate search directions and hampers the speed of convergence of optimization algorithms.

In this chapter we analyze those negative effects of over-parametrization on the Interior Point algorithm of the previous chapter and present a method to resolve them.

If the controller is parameterized in a different fashion than in state-space coordinates, over-parametrization does not necessarily occur. An example is the parametrization of a SISO controllers in terms of the coefficients of the numerator and denominator polynomial of the elements of its transfer function, and the controllable and observable canonical forms. These parametrizations are sensitive to round-off errors [128], and are therefore not suitable to synthesize fixed-order controllers for systems with large state dimension. Reduced parametrization of all transfer functions $G \in \mathcal{RH}_\infty$ with $\|G\|_\infty < 1$ have been presented in [185] and [162]. However, it seems not straightforward to apply these parameterizations to fixed-order controller synthesis in state-space coordinates.

Over-parametrization due to similarity transformations is a fundamental problem in optimization with state-space realizations. In system identification this has been recognized by several authors [198, 127, 78]. To the best of our knowledge, and surprisingly, there has not been any contribution on over-parametrization in the context of controller synthesis.

In this chapter we discuss the effects of over-parametrization on the Interior Point method as has been presented in the previous chapter. Although focus is on this particular optimization routine, many of the results are applicable to other algorithms as well. We analyze how optimizing over a reduced search space may alleviate the negative effects of over-parameterization. Since the optimal $\mathcal{H}_\infty$-controller transfer function is often not unique, we cannot guarantee a priori the existence of verifiable optimal points in the reduced parameter space. The best we can achieve is that, if a locally uniquely optimal controller transfer function exists, its reduced state-space coordinates are also verifiably locally optimal. We derive conditions on the reduced parameterization that guarantee this.

Secondly, we present a reduced parametrization of the controller that

- satisfies for almost all controllers the conditions mentioned above and

- covers with a single coordinate system all (possible MIMO) controller transfer functions with a McMillan degree not larger than a fixed number n_c.

For an $\mathcal{H}_\infty$-optimal controller design example we show that optimizing in a reduced space requires less corrector steps in the IP algorithm and yields a better final closed-loop $\mathcal{H}_\infty$-norm. The proposed parametrization therefore strengthens the optimality certificates presented in the previous chapter, and improves the IP algorithm as well.

The outline of this chapter is as follows. In Section 8.1 we explain why the decision variables in the Interior Point-method are over-parameterized. In Section 8.2 we describe the effects of over-parametrization on the verifiability of the sufficient optimality conditions. Furthermore, we present the above mentioned conditions to make the existence of verifiable optimal points possible. We illustrate the conditions with the canonical controllable form in Section 8.3 for SISO controllers. Finally, in Section 8.4 we present the reduced parametrization for MIMO controllers as mentioned above and apply it to a control design example.

8.1 Reasons for over-parametrization

We explain in this section that similarity transformations of the controller result in over-parametrization of the minimization of the barrier function in the corrector step of the Interior Point algorithm presented in the previous chapter. The state-space controller with McMillan degree n_c can be represented with the matrix

$$K := \begin{pmatrix} A_K & B_K \\ C_K & D_K \end{pmatrix},$$

where $A_K \in \mathbb{R}^{n_c \times n_c}$, $B_K \in \mathbb{R}^{n_c \times p_2}$, $C_K \in \mathbb{R}^{m_2 \times n_c}$ and $D_K \in \mathbb{R}^{m_2 \times p_2}$. We assume in this chapter that $n_c > 1$. Let us denote the number of variables in K by $m_c := (n_c + m_2)(n_c + p_2)$. The $\mathcal{H}_\infty$-optimal control problem is formulated as minimizing γ subject to

$$\mathcal{B}(\gamma, X, K) := - \begin{pmatrix} A_{cl}(K)^T X + X A_{cl}(K) & X B_{cl}(K) & C_{cl}(K)^T \\ B_{cl}(K)^T X_{cl} & -\gamma I & D_{cl}(K)^T \\ C_{cl}(K) & D_{cl}(K) & -\gamma I \end{pmatrix} \succ 0 \qquad (8.1)$$

and

$$X \succ 0, \qquad (8.2)$$

where $(A_{cl}(K), B_{cl}(K), C_{cl}(K), D_{cl}(K))$ are defined in (2.8). For simplicity we do not take the constraints in (7.3) into account. These constraints were introduced in the previous chapter to bound the decision variables.

Recall that at each corrector step of the IP algorithm the log-barrier function

$$\phi_r(x, \mu) := \frac{\gamma}{\mu} - \log\left(\det\left(\mathcal{B}(\gamma, X, K)\right)\right) - \log(\det(X)) \qquad (8.3)$$

is minimized with respect to $x := (\gamma, X, K)$ for fixed barrier parameter $\mu > 0$. The value of the barrier function is invariant under certain transformations of the decision variables. To show this, we consider a similarity transformation of the controller states $T z_k = x_k$,

where T is a matrix in $\mathbb{R}^{n_c \times n_c}$ and $\det(T) = 1$. This change of coordinates of the controller states transforms K into

$$K^{\text{trans}}(K, T) := \begin{pmatrix} T & 0 \\ 0 & I \end{pmatrix}^{-1} K \begin{pmatrix} T & 0 \\ 0 & I \end{pmatrix} =: \begin{pmatrix} A_K^{\text{trans}}(K, T) & B_K^{\text{trans}}(K, T) \\ C_K^{\text{trans}}(K, T) & D_K^{\text{trans}}(K, T) \end{pmatrix}. \quad (8.4)$$

The corresponding closed matrices are

$$\begin{pmatrix} \bar{T}^{-1} A_{\text{cl}}(K) \bar{T} & \bar{T}^{-1} B_{\text{cl}}(K) \\ C_{\text{cl}}(K) \bar{T} & D_{\text{cl}}(K) \end{pmatrix},$$

where:

$$\bar{T} = \begin{pmatrix} I & 0 \\ 0 & T \end{pmatrix}. \quad (8.5)$$

Furthermore $\mathcal{B}(\gamma, X, K)$ as defined in (8.1) transforms likewise into:

$$-\begin{pmatrix} (\bar{T}^{-1} A_{\text{cl}}(K) \bar{T})^T X + X (\bar{T}^{-1} A_{\text{cl}}(K) \bar{T}) & X \bar{T}^{-1} B_{\text{cl}}(K) & (C_{\text{cl}}(K) \bar{T})^T \\ (\bar{T}^{-1} B_{\text{cl}}(K))^T X & -\gamma I & D_{\text{cl}}(K)^T \\ C_{\text{cl}}(K) \bar{T} & D_{\text{cl}}(K) & -\gamma I \end{pmatrix}. \quad (8.6)$$

Since $\det(T) = 1$ the transformed controller $K^{\text{trans}}(K, T)$ together with

$$X^{\text{trans}}(X, T) := \bar{T}^T X \bar{T}, \quad (8.7)$$

have the same value of the barrier function ϕ_r in (8.3) as the original variables. To see this, define

$$x^{\text{trans}}(x, T) := (\gamma, \text{svec}(X^{\text{trans}}(X, T)), \text{vec}(K^{\text{trans}}(K, T))) \quad (8.8)$$

and observe that

$$\phi_r(x^{\text{trans}}(x, T), \mu) =$$
$$= \frac{\gamma}{\mu} - \log\left(\det\left[\begin{pmatrix} \bar{T}^T & 0 \\ 0 & I \end{pmatrix} \mathcal{B}(\gamma, X, K) \begin{pmatrix} \bar{T}^T & 0 \\ 0 & I \end{pmatrix}\right]\right) - \log(\det(\bar{T}^T X \bar{T})) = \phi_r(x).$$

For later reference we derive now the first- and second-order derivatives[1] of x^{trans}. The derivative with respect to $x = (\gamma, X, K)$ is easily derived to be

$$\partial_x x^{\text{trans}}(x, T)(d_x) = \left(d_\gamma, \partial_1 X^{\text{trans}}(X, T)(d_X), \partial_K K^{\text{trans}}(K, T)(d_K)\right), \quad (8.9)$$

where

$$\partial_K K^{\text{trans}}(K, T)(d_K) = \begin{pmatrix} T^{-1} & 0 \\ 0 & I_{m_2} \end{pmatrix} d_K \begin{pmatrix} T & 0 \\ 0 & I_{p_2} \end{pmatrix}$$

$$\partial_X X^{\text{trans}}(X, T)(d_X) = \begin{pmatrix} I_n & 0 \\ 0 & T \end{pmatrix}^T d_X \begin{pmatrix} I_n & 0 \\ 0 & T \end{pmatrix} \quad (8.10)$$

The second derivative of $x^{\text{trans}}(x, T)$ with respect to x is zero, i.e. $\partial_x^2 x^{\text{trans}}(x, T)(d_{x1}, d_{x2}) = 0$ for all d_{x1} and d_{x2}. To get the derivative with respect to T, consider a small perturbation d_T. Then, a first order approximation of $x^{\text{trans}}(x, T)$ is

$$\left(\gamma, \begin{pmatrix} 0 & 0 \\ 0 & T + d_T \end{pmatrix}^T X \bar{T} + \bar{T}^T X \begin{pmatrix} 0 & 0 \\ 0 & d_T \end{pmatrix}, K^{\text{trans}}(K, T + d_T)\right),$$

---

[1] See e.g. [5] for some useful derivatives of standard matrix-valued functions.

161

where

$$K^{\text{trans}}(K, T+d_T) := \begin{pmatrix} T & 0 \\ 0 & I_{m_2} \end{pmatrix}^{-1} K \begin{pmatrix} T+d_T & 0 \\ 0 & I_{p_2} \end{pmatrix} + \begin{pmatrix} -T^{-1}d_T T^{-1} & 0 \\ 0 & 0 \end{pmatrix} K \begin{pmatrix} T & 0 \\ 0 & I_{p_2} \end{pmatrix}$$

such that the first derivative with respect to T is given by

$$\partial_T x^{\text{trans}}(x, T)(d_T) = \left(0, \partial_T X^{\text{trans}}(K, T)(d_T), \partial_T K^{\text{trans}}(K, T)(d_T)\right) \tag{8.11}$$

where

$$\partial_T X^{\text{trans}}(K, T)(d_T) = \begin{pmatrix} 0 & 0 \\ 0 & d_T \end{pmatrix}^T X \begin{pmatrix} I_n & 0 \\ 0 & T \end{pmatrix} + \begin{pmatrix} I_n & 0 \\ 0 & T \end{pmatrix}^T X \begin{pmatrix} 0 & 0 \\ 0 & d_T \end{pmatrix} \tag{8.12}$$

$$\partial_T K^{\text{trans}}(K, T)(d_T) = \begin{pmatrix} \partial_T A_K^{\text{trans}}(K, T)(d_T) & \partial_T B_K^{\text{trans}}(K, T)(d_T) \\ C_K d_T & 0 \end{pmatrix}$$

with

$$\partial_T A_K^{\text{trans}}(K, T)(d_T) = -T^{-1}d_T T^{-1}A_K T + T^{-1}A_K d_T \tag{8.13}$$

$$\partial_T B_K^{\text{trans}}(K, T)(d_T) = -T^{-1}d_T T^{-1}B_K. \tag{8.14}$$

The second derivative with respect to T is

$$\partial_T^2 x^{\text{trans}}(x, T)(d_{T1}, d_{T2}) = \left(0, \partial_T^2 X^{\text{trans}}(K, T)(d_{T1}, d_{T2}), \partial_T^2 K^{\text{trans}}(K, I_{n_c})(d_{T1}, d_{T2})\right) \tag{8.15}$$

where

$$\partial_T^2 X^{\text{trans}}(X, T)(d_{T1}, d_{T2}) = \begin{pmatrix} I_n & 0 \\ 0 & d_{T1} \end{pmatrix}^T X \begin{pmatrix} I_n & 0 \\ 0 & d_{T2} \end{pmatrix} + \begin{pmatrix} I_n & 0 \\ 0 & d_{T2} \end{pmatrix}^T X \begin{pmatrix} I_n & 0 \\ 0 & d_{T1} \end{pmatrix}$$

$$\partial_T^2 K^{\text{trans}}(K, T)(d_{T1}, d_{T2}) = \begin{pmatrix} \partial_T^2 A_K^{\text{trans}}(K, T)(d_{T1}, d_{T2}) & \partial_T^2 B_K^{\text{trans}}(K, T)(d_{T1}, d_{T2}) \\ 0 & 0 \end{pmatrix}$$

with

$$\partial_T^2 A_K^{\text{trans}}(K, T)(d_{T1}, d_{T2}) = T^{-1}d_{T2}T^{-1}d_{T1}T^{-1}A_K T + T^{-1}d_{T1}T^{-1}d_{T2}T^{-1}A_K T - \\ -T^{-1}d_{T1}T^{-1}A_K d_{T2} - T^{-1}d_{T2}T^{-1}A_K d_{T1}$$

and

$$\partial_T^2 B_K^{\text{trans}}(K, T)(d_{T1}, d_{T2}) = T^{-1}d_{T2}T^{-1}d_{T1}T^{-1}B_K + T^{-1}d_{T1}T^{-1}d_{T2}T^{-1}B_K. \tag{8.16}$$

Finally, we partition d_K as

$$d_K = \begin{pmatrix} d_{A_K} & d_{B_K} \\ d_{C_K} & d_{D_K} \end{pmatrix},$$

where $d_{A_K} \in \mathbb{R}^{n_c \times n_c}$, $d_{B_K} \in \mathbb{R}^{n_c \times p_2}$, $d_{C_K} \in \mathbb{R}^{m_2 \times n_c}$ and $d_{D_K} \in \mathbb{R}^{m_2 \times p_2}$. Then $\partial_x \partial_T x^{\text{trans}}(x, T)$ is given by:

$$\partial_x \partial_T x^{\text{trans}}(x, T)(d_x, d_T) = \left(0, \partial_X \partial_T X^{\text{trans}}(X, T)(d_X, d_T)\right), \partial_K \partial_T K^{\text{trans}}(K, T)(d_K, d_T) \tag{8.17}$$

where

$$\partial_X \partial_T X^{\text{trans}}(X,T)(d_X, d_T) = \begin{pmatrix} 0 & 0 \\ 0 & d_T \end{pmatrix}^T d_X \begin{pmatrix} I_n & 0 \\ 0 & T \end{pmatrix} + \begin{pmatrix} I_n & 0 \\ 0 & T \end{pmatrix}^T d_X \begin{pmatrix} 0 & 0 \\ 0 & d_T \end{pmatrix}$$

$$\partial_K \partial_T K^{\text{trans}}(K,T)(d_K, d_T) = \begin{pmatrix} \partial_K \partial_T A_K^{\text{trans}}(K,T)(d_K, d_T) & \partial_K \partial_T B_K^{\text{trans}}(K,T)(d_K, d_T) \\ d_{C_K} d_T & 0 \end{pmatrix}$$

with

$$\partial_K \partial_T A_K^{\text{trans}}(K,T)(d_K, d_T) = T^{-1} d_T T^{-1} d_{A_K} T + T^{-1} d_{A_K} d_T$$
$$\partial_K \partial_T B_K^{\text{trans}}(K,T)(d_K, d_T) = T^{-1} d_T T^{-1} d_{B_K}.$$

8.2 Effects of over-parametrization on optimality conditions

8.2.1 Optimality conditions for barrier problem

We will explain in this section why over-parametrization may cause numerical problems in the optimization, and makes it impossible to verify the sufficient optimality conditions at optimal points. We do this in a slightly more general setting than just considering Interior Point minimization. For this purpose ϕ_r is replaced by f in this section and we consider

$$\inf_{x \in \mathcal{G}} f(x), \tag{8.18}$$

where $\mathcal{G}$ is an open subset of a vector space $\mathcal{X}$ of dimension m. We describe the over-parametrization by a $\mathcal{C}^2$-mapping $g : \mathcal{X} \times \Omega \to \mathcal{X}$ where $\Omega := \{y \in \mathcal{Y} | \ c(y) = 0\}$ is a subset of a vector space $\mathcal{Y}$ of dimension $n < m$ and c is a $\mathcal{C}^1$-mapping. We use the following notational shorthand for the image of the derivative of g at (x_*, y_*), if the domain is restricted to the kernel of $\partial c(y_*)$:

$$\text{Im}\left(\partial_y g(x_*, y_*)|_{\text{Ker}(\partial c(y_*))}\right) := \{\partial_y g(x_*, y_*)d \in \mathcal{X}| \ d \in \text{Ker}\left(\partial c(y_*)\right) \subset \mathcal{Y}\}$$

To describe the over-parametrization we need the following assumption.

Assumption 8.1 *The mapping $g : \mathcal{X} \times \Omega \to \mathcal{X}$ has the following properties:*

(i.) for all $x \in \mathcal{G}$ there exists some $\hat{y} \in \Omega$ such that $g(x, \hat{y}) = x$

(ii.) $\partial_x g(x, y)$ is invertible for all $x \in \mathcal{G}$ and $y \in \Omega$

(iii.) f and g satisfy

$$g(x, y) \in \mathcal{G} \ and \ f(g(x, y)) = f(x) \ for \ all \ x \in \mathcal{G}, y \in \Omega \tag{8.19}$$

(iv.) for all $x_* \in \mathcal{G}$, $y_* \in \Omega$ there exists a $d \in Im\left(\partial_y g(x_*, y_*)|_{Ker(\partial c(y_*))}\right) \backslash \{0\}$

Example 8.2 *We elucidate Assumption 8.1 with the corrector step of the Interior Point method for structured $\mathcal{H}_\infty$-optimal controller synthesis for $n_c > 1$. $f(x)$ is in this case given by the barrier function in (8.3) and $x = (\gamma, X, K)$. With $\mathcal{X} = \mathbb{R} \times \mathbb{R}^{p_1 + n_c \times m_1 + n_c} \times \mathcal{S}^{n+n_c}$ the domain of f is $\mathcal{G} := \{x = (\gamma, X, K) \in \mathcal{X} | \ X \succ 0, \ \mathcal{B}(\gamma, X, K) \succ 0\}$. Let $\mathcal{Y} = \mathbb{R}^{n_c \times n_c}$,*

$c(y) := \det(y) - 1$ such that Ω is the set $\{T \in \mathbb{R}^{n_c \times n_c} | \det(T) = 1\}$. With $x = (\gamma, X, K)$ let the mapping g be defined by

$$g(x, y) := x^{trans}(x, y), \tag{8.20}$$

where x^{trans} is as in (8.8). All items in Assumption 8.1 are satisfied in this case. Indeed, with $\hat{y} := I_{n_c} \in \Omega$, (i.) is satisfied for all x. Secondly, (ii.) follows from (8.9). (iii.) is true since $\phi_r(x^{trans}(x, T)) = \phi_r(x)$ for all x and T with $\det(T) = 1$. To show (iv.), consider $x_* = (\gamma_*, X, K) \in \mathcal{G}$, and $y_* = T \in \Omega$. In the sequel we construct a d_T that satisfies

- $Trace(d_T) = 0$

- $\partial_T X^{trans}(K, T)(d_T) \neq 0$ in (8.12) is nonzero

These properties imply that $\partial_y g(x_*, y_*)(d_T) \in Im\left(\partial_y g(x_*, y_*)|_{Ker(\partial c(y_*))}\right) \backslash \{0\}$. To simplify the exposition, we denote the right lower block of dimension $n_c \times n_c$ of X by X_K, i.e.

$$X_K := \begin{pmatrix} 0 & I_{n_c} \end{pmatrix} X \begin{pmatrix} 0 \\ I_{n_c} \end{pmatrix};$$

then the right lower $n_c \times n_c$ block of (8.12) reads as

$$\begin{pmatrix} 0 & I_{n_c} \end{pmatrix} \partial_T X^{trans}(K, T)(d_T) \begin{pmatrix} 0 \\ I_{n_c} \end{pmatrix} = d_T^T X_K T + T^T X_K d_T.$$

Let us first suppose that $Trace(X_K^{-1} T^{-T}) = 0$; then obviously $d_T := X_K^{-1} T^{-T}$ does the job, since $Trace(d_T) = 0$ and $d_T^T X_K T + T^T X_K d_T = 2 I_{n_c}$, such that $\partial_T X^{trans}(K, T)(d_T) \neq 0$. Now suppose $Trace(X_K^{-1} T^{-T}) \neq 0$. This implies that at least one element on the diagonal of $X_K^{-1} T^{-T}$ is nonzero, i.e. there exists a $k \in \{1, \ldots, n_c\}$, such that $e_k^T X_K^{-1} T^{-T} e_k \neq 0$, where e_k is the kth basis vector in $\mathbb{R}^{n_c}$. Now if we choose $d_T := X_K^{-1} T^{-T} \left(I_{n_c} - \frac{Trace(X_K^{-1} T^{-T})}{e_k^T X_K^{-1} T^{-T} e_k} e_k e_k^T \right)$, then

$$Trace(d_T) = Trace(X_K^{-1} T^{-T}) - \frac{Trace(X_K^{-1} T^{-T})}{e_k^T X_K^{-1} T^{-T} e_k} Trace(X_K^{-1} T^{-T} e_k e_k^T) = 0.$$

Furthermore

$$d_T^T X_K T + T^T X_K d_T = 2 I_{n_c} - 2 \frac{Trace(X_K^{-1} T^{-T})}{e_k^T X_K^{-1} T^{-T} e_k} e_k e_k^T,$$

which is nonzero, since $n_c > 1$. Hence d_T has the desired properties, which concludes the proof of (iv.).

Remark. It seems not very difficult to mold the over-parametrization due to group structure in Sum-Of-Squares optimization as described in [67] into our framework. This is, however, beyond the scope of this thesis work.

Since $\mathcal{G}$ is open, a first order necessary optimality condition for an arbitrary $x_* \in \mathcal{G}$ is

$$\partial f(x_*) d = 0, \text{ for all } d \in \mathcal{X}, \tag{8.21}$$

as has been described in Section 7.3. If x_* satisfies (8.21), we call it a *critical point*.

Lemma 8.3 *Suppose Assumption 8.1 is true. If $x_* \in \mathcal{G}$ is an arbitrary critical point of (8.18), then $g(x_*, y_*)$ is also a critical point of (8.18) for arbitrary $y_* \in \Omega$.*

Proof. Differentiating (8.19) at x_*, y_* with respect to x yields

$$\partial f\left(g(x_*, y_*)\right) \partial_x g(x_*, y_*) = \partial f(x_*).$$

Since $\partial f(x_*) = 0$ and $\partial_x g(x_*, y_*)$ is invertible by Assumption 8.1 we conclude

$$\partial f(g(x_*, y_*)) = 0.$$

$\blacksquare$

Remark. The notation $\partial f(g(x_*, y_*))$ has been used in the proof of Lemma 8.3, by which we mean the derivative of f evaluated at $x = g(x_*, y_*)$.

The following result shows that for a critical point $x_* \in \mathcal{G}$ and arbitrary $y \in \Omega$ that dimensions of the kernels of f at $x = x_*$ and at $x = g(x_*, y_*)$ are equal. It will be used in the proofs of the theorems that follow.

Lemma 8.4 *If x_* is a critical point then for arbitrary $y_* \in \Omega$ the following holds true:*

$$\dim\left(Ker\left(\partial^2 f(x_*)\right)\right) = \dim\left(Ker\left(\partial^2 f(g(x_*, y_*))\right)\right)$$

Proof. Differentiating (8.19) at x_*, y_* twice with respect to x yields for arbitrary $d \in \mathcal{X}$

$$
\begin{aligned}
\partial^2 f(x_*)(d, d) &= \\
&= \partial^2 f\left(g(x_*, y_*)\right)\left(\partial_x g(x_*, y_*)d, \partial_x g(x_*, y_*)d\right) + \underbrace{\partial f\left(g(x_*, y_*)\right)\left(\partial_x^2 g(x_*, y_*)(d, d)\right)}_{=0 \text{ by Lemma 8.3}} \\
&= \partial^2 f\left(g(x_*, y_*)\right)\left(\partial_x g(x_*, y_*)d, \partial_x g(x_*, y_*)d\right).
\end{aligned}
$$

Since $\partial_x g(x_*, y_*)$ is invertible, the result follows.

$\blacksquare$

Critical points that are optimal satisfy the second order necessary conditions

$$\partial^2 f(x_*)(d, d) \geq 0, \text{ for all } d \in \mathcal{X}. \tag{8.22}$$

If x_* is a critical point, then the following second order conditions are sufficient for (local) optimality of x_* for (8.18):

$$\partial^2 f(x_*)(d, d) > 0 \text{ for all } d \in \mathcal{X} \text{ with } d \neq 0. \tag{8.23}$$

If there exists a g that satisfies Assumption 8.1, (8.23) will never be satisfied for a critical point of (8.18). This is the content of the following theorem.

Theorem 8.5 *Suppose Assumption 8.1 is true. If $x_* \in \mathcal{G}$ is an arbitrary critical point of (8.18), then there exists a $d \in \mathcal{X}$, $d \neq 0$ such that*

$$\partial^2 f(x_*)(d, d) = 0. \tag{8.24}$$

Hence there is no critical point satisfying (8.23). More specifically, for any $\hat{x} \in \mathcal{G}$ and $\hat{y} \in \Omega$ with $g(\hat{x}, \hat{y}) = x_*$ every

$$d \in Im\left(\partial_y g(\hat{x}, \hat{y})|_{Ker(\partial c(\hat{y}))}\right)$$

satisfies (8.24)

Proof. Let x_* be a critical point, such that (8.21) is satisfied. Since $g(\hat{x}, \hat{y}) = x_*$ we conclude

$$\partial f(g(\hat{x}, \hat{y})) = 0. \tag{8.25}$$

Differentiating (8.19) evaluated at $x = \hat{x}$, $y = \hat{y}$ once with respect to y implies

$$\partial f(g(\hat{x}, \hat{y}))\partial_y g(\hat{x}, \hat{y})d = 0 \text{ for all } d \in Ker\left(\partial c(\hat{y})\right).$$

Differentiating once again with respect to y implies for all $d_1, d_2 \in Ker\left(\partial c(\hat{y})\right)$

$$\partial^2 f(g(\hat{x}, \hat{y}))(\partial_y g(\hat{x}, \hat{y})d_1, \partial_y g(\hat{x}, \hat{y})d_2) + \partial f(g(\hat{x}, \hat{y}))(\partial_y^2 g(\hat{x}, \hat{y})(d_1, d_2)) = 0. \tag{8.26}$$

Due to (8.25) the second term on the left-hand side is zero. Hence (8.26) implies

$$\partial^2 f(g(\hat{x}, \hat{y}))(\partial_y g(\hat{x}, \hat{y})d_1, \partial_y g(\hat{x}, \hat{y})d_2) = 0 \text{ for all } d_1, d_2 \in Ker\left(\partial c(\hat{y})\right)$$

such that for all $d \in Im\left(\partial_y g(\hat{x}, \hat{y})|_{Ker(\partial c(\hat{y}))}\right)$ we conclude

$$\partial^2 f(x_*)(d, d) = \partial^2 f(g(\hat{x}, \hat{y}))(d, d) = 0,$$

which implies (8.24). Due to (iv.) in Assumption 8.1, this hence holds for some $d \neq 0$, which implies that (8.23) cannot hold true. This finishes the proof. ∎

Remark. It is easy to show that also for the original BMI problem

$$\begin{array}{ll} \text{infimize} & \gamma \\ \text{subject to} & \mathcal{B}(\gamma, X, K) \succeq 0 \text{ and } X \succeq 0 \end{array} \tag{8.27}$$

the second order sufficient conditions as derived in Theorem 7.19 of Section 7.3 *cannot hold true* for any feasible point. Consider an arbitrary feasible point $x_* = (\gamma, X, K)$ and the function

$$z(t) := x^{\text{trans}}\left(x_*, \frac{1}{\det(I + tR)}(I + tR)\right)$$

for an arbitrary but not identically zero matrix $R \in \mathbb{R}^{n_c \times n_c}$, $\|R\| < 1$. The function $z(t)$ satisfies $z(0) = x_*$ and $z(1) \neq z(0)$. Furthermore, $(\gamma_t, X_t, K_t) := z(t)$ satisfies

$$\mathcal{B}(\gamma_t, X_t, K_t) \succeq 0 \text{ and } X_t \succeq 0$$

for all $t \in [0, 1]$. This implies that x_* does not satisfy the sufficient optimality conditions for (8.27).

8.2.2   Reduced parametrization

The problems with verifying optimality as discussed in the previous section can often be resolved by re-parametrization of the decision variables. For this purpose consider a continuously twice differentiable function $h : \Psi \subset \mathcal{Z} \to \mathcal{X}$, where Ψ is an open subset of a vector space $\mathcal{Z}$ of dimension $\dim(\mathcal{Z}) = \dim(\mathcal{X}) - \dim(\mathcal{Y})$ and where $h(z)$ satisfy the following additional hypothesis:

$$\text{For each } x \in \mathcal{X} \text{ there exist } y \in \Omega, z \in \Psi \text{ with } h(z) = g(x, y). \tag{8.28}$$

Based on this hypothesis we conclude that

$$\inf_{x \in \mathcal{X}} f(x) = \inf_{z \in \Psi} f(h(z)). \tag{8.29}$$

In other words, the minimization with the reduced parametrization z yields the same optimal value as the original problem.

Example 8.6 *To illustrate that re-parametrization can improve the convergence of minimization algorithms, we consider (inspired by [105]) the objective function*

$$f(x_1, x_2, x_3) := (x_1 - 1)^2 + \gamma((1 + x_2)x_3^2 - x_1^2)^2.$$

We describe the over-parametrization by

$$g(x, y) := \begin{pmatrix} x_1 & x_2 y_1^2 + y_1^2 - 1 & x_3 y_2 \end{pmatrix}^T,$$

for all $y = \begin{pmatrix} y_1 & y_2 \end{pmatrix}^T \in \Omega := \{y \in \mathbb{R}^2| \ c(y) = 0\}$ with $c(y) = y_1 y_2 - 1$. $g(x, y)$ satisfies all items in Assumption 8.1, if $\mathcal{G} = \{x \in \mathbb{R}^3| \ x_3 \neq 0\}$. Indeed $g(x, 1) = x$ and

$$f(g(x, y)) = (x_1 - 1)^2 + \gamma((1 + x_2 y_1^2 + y_1^2 - 1)(x_3 y_2)^2 - x_1^2)^2 = f(x)$$

for all $x \in \mathcal{G}$ and $y \in \Omega$. Differentiating $g(x, y)$ with respect to x yields

$$\partial_x g(x, y) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & y_1^2 & 0 \\ 0 & 0 & y_2 \end{pmatrix},$$

which is invertible for all $x \in \mathcal{X}$ and $y \in \Omega$. Finally

$$\partial_y g(x, y) = \begin{pmatrix} 0 & 2x_2 y_1 + 2y_1 & 0 \\ 0 & 0 & x_3 \end{pmatrix}$$

is nonzero for all $x \in \mathcal{G}$. Hence (iv.) in Assumption 8.1 is satisfied.

We consider the reduced parametrization $h(z_1, z_2) = (z_1, z_2, 1)^T$. For an arbitrary $x = (x_1, x_2, x_3)^T$ choose $z = (x_1, x_2 x_3^2 + x_3^2 - 1)^T$ and $y = (x_3, \frac{1}{x_3})$ to render (8.28) satisfied, which directly follows from

$$g(x, y) = \begin{pmatrix} x_1 & x_2 x_3^2 + x_3^2 - 1 & \frac{x_3}{x_3} \end{pmatrix} = h(z).$$

We have used the trust-region optimization algorithm in Section 7.1.3 to solve $\inf_{x \in \mathcal{G}} f(x)$ and $\inf_{z \in \Psi} f(h(z))$, where $\Psi = \mathbb{R}^2$. To obtain a fair comparison we choose initial values

Table 8.1: Performance and computation time of controllers

	over-parameterized (inf $f(x)$)			reduced parametrization (inf $f(h(z))$)		
γ	it	$f(x_*)$	cond(H)	it	$f(x_*)$	cond(H)
1	7	5.487e-24	3.916e+012	5	6.769e-10	33.969
100	6	0.771	4.117e+018	12	4.009e-13	2508
10000	75	2.707e-13	4.509e9	43	1.827e-14	2.5e+5

such that $h(z^{init}) = x^{init}$ by choosing $x^{init} = (-1, -2, 1)$ and $z^{init} = (-1, -2)$ respectively. The optimal values of $\inf_{x \in \mathcal{G}} f(x)$ and $\inf_{z \in \Psi} f(h(z))$ are attained at $x = (1, 0, 1)$ and $z = (1, 0)$, respectively. In Table 8.1 the number of iterations, the final objective function and the condition number of the Hessian at the optimum is given. For $\gamma = 100$ the optimization of f gets stuck at $f(x) = 0.77$ due to slow convergence, whereas the reduced parametrization reaches the value $4 \cdot 10^{-013}$, which is very close to the optimal value zero. For $\gamma = 10000$ the optimization does converge to an almost optimal value, but this requires 75 iterations instead of the 43 iterations needed with the reduced parametrization. In Figure 8.1 the function f is shown for fixed $x_1 = 1$ and $\gamma = 1$ and $x_2, x_3 \in [-1, 1]$. The optimal points are shown as the two solid curves and the dashed curve is the set $\{(h(z), f(h(z))) | z \in \Psi\}$. The optimum of $\inf_{z \in \Psi} f(h(z))$ is attained at $z_* = (1, 0)$, which is the point where one of the solid curves and the dashed curve intersect. The corresponding optimal point for $\inf_{x \in \mathcal{G}} f(x)$ is $x_* = h(z_*) = (1, 0, 1)$.
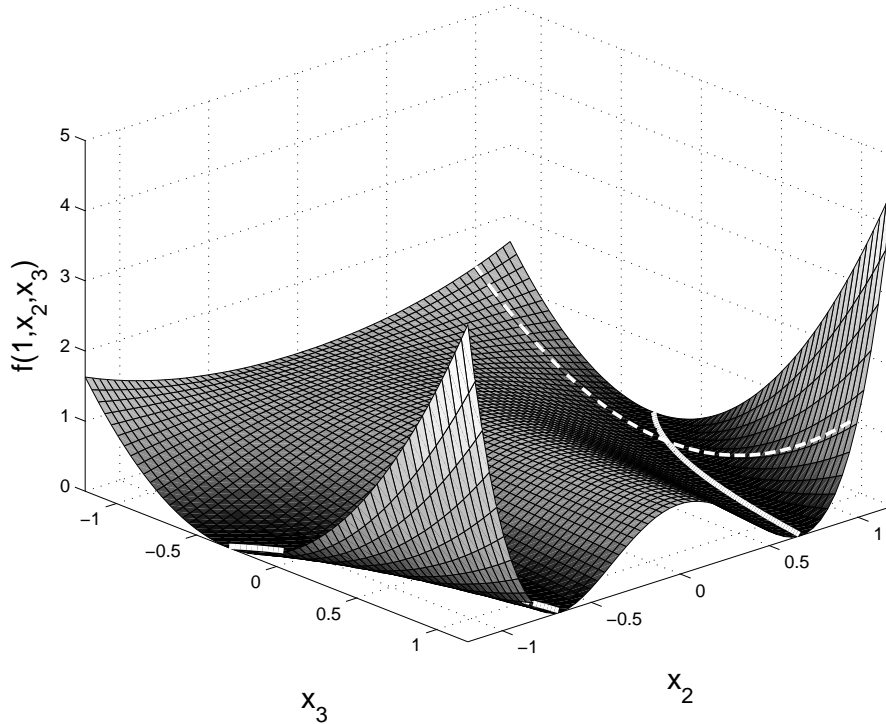
Figure 8.1: $f(1, x_2, x_3)$ for $\gamma = 1$. The optimal points are shown as solid white curves and the set $\{(h(z), f(h(z))) | z \in \Psi\}$ as dashed curve.

Remark. Since the trust-region method applied in Example 8.6 is tailor-made for ill-conditioned nonconvex problems, its convergence to points with a singular Hessian will often be faster than for less sophisticated methods. The difference in convergence between full and reduced parametrization is therefore likely to be even more pronounced if we had used for instance a gradient method.

8.2.3 Optimality conditions for reduced parametrizations

To analyze the optimality conditions of Section 7.3 for the minimization problem with the reduced parametrization, let us define $r(\cdot) := f(h(\cdot))$, such that

$$\partial r(z)d = \partial f(h(z))(\partial h(z)d)$$

and

$$\partial^2 r(z)(d_1, d_2) = \partial^2 f(h(z))\left((\partial h(z)d_1, \partial h(z)d_2\right) + \partial f(h(z))\left(\partial^2 h(z)(d_1, d_2)\right). \qquad (8.30)$$

Necessary conditions

It is easy to see that the necessary optimality conditions of the original problem imply that the necessary optimality conditions for the reduced parametrization hold true. This is the content of the following lemma.

Lemma 8.7 *If a point $x_* \in \mathcal{G}$ satisfies the necessary conditions (8.21) and (8.22), then for any z_* and y_* with $h(z_*) = g(x_*, y_*)$, z_* satisfies the first order necessary conditions for $\inf_{z \in \Psi} r(z)$.*

Proof. By Lemma 8.3 we conclude $\partial f(g(x_*, y_*)) = 0$ such that

$$\partial r(z_*) = \partial f(h(z_*))\partial h(z_*) = \partial f(g(x_*, y_*))\partial h(z_*) = 0.$$

Differentiating (8.19) twice at x_*, y_* with respect to x implies for all $d \in \mathcal{X}$

$$\partial^2 f(g(x_*, y_*))(\partial_x g(x_*, y_*)d, \partial_x g(x_*, y_*)d) + \partial f(g(x_*, y_*))(\partial_x^2 g(x_*, y_*)(d, d)) = \partial^2 f(x_*)(d, d). \qquad (8.31)$$

Since $\partial f(g(x_*, y_*)) = 0$ and $\partial^2 f(x_*) \succeq 0$ by local optimality of x_* (i.e. by (8.21) and (8.22)) we infer $\partial^2 f(g(x_*, y_*)) \left(\partial_x g(x_*, y_*)d, \partial_x g(x_*, y_*)d\right) \geq 0$ for all $d \in \mathcal{X}$. By invertibility of $\partial_x g(x_*, y_*)$ (due to (ii.) in Assumption 8.1) we conclude $\partial^2 f(g(x_*, y_*)) \succeq 0$. Hence for all $\delta \in \mathcal{Z}$

$$\partial^2 r(z_*)(\delta, \delta) = \partial^2 f(h(z_*))(\partial h(z_*)\delta, \partial h(z_*)\delta) + \partial f(h(z_*))(\partial^2 h(z_*)(\delta, \delta)) =$$
$$= \partial^2 f(h(z_*))(\partial h(z_*)\delta, \partial h(z_*)\delta) = \partial^2 f(g(x_*, y_*))(\partial h(z_*)\delta, \partial h(z_*)(\delta)) \geq 0 \quad (8.32)$$

$\blacksquare$

Sufficient conditions: towards verifiably locally optimal controllers

In the sequel we analyze under which conditions re-parametrization enables to verify optimality in terms of the sufficient optimality conditions as discussed in Section 7.3 of the previous chapter. Let x_* be a critical point of (8.18) and let $y_* \in \Omega$ and z_* be such that $g(x_*, y_*) = h(z_*)$. For verifiability of the sufficient conditions at z_* for $\inf_{z \in \Psi} r(z)$, it is obvious that $\partial h(z_*)$ must have full column rank. Furthermore h must satisfy at $z = z_*$:

$$\mathrm{Im}\left(\partial h(z_*)\right) \cap \mathrm{Im}\left(\partial_y g(x_*, y_*)|_{\mathrm{Ker}(\partial c(y_*))}\right) = \{0\}. \tag{8.33}$$

Only if (8.33) is satisfied, it is *not impossible* to verify the sufficient condition for $\inf_{z \in \Psi} r(z)$ at z_*. However $\partial h(z_*)$ having full column rank and (8.33) are *not* sufficient. To illustrate this we consider the fixed-order $\mathcal{H}_\infty$-optimal synthesis of a SISO controller, where $K_* = \begin{pmatrix} A_K & B_K \\ C_K & D_K \end{pmatrix}$ is a locally optimal controller. Suppose that arbitrary close to K_* there exists for each $\epsilon > 0$ another optimal K_0 with the same McMillan degree but a different transfer function and such the $\|K_* - K_0\| < \epsilon$, where $\| \cdot \|$ is the matrix 2-norm. Then it is obvious that the sufficient optimality do not hold, even though we use for instance the controller canonical form to re-parametrize the controllers.

The following condition together with $\partial h(z_*)$ full column rank and (8.33) *are* sufficient for verifiable optimality of z_* :

$$\dim\left(\mathrm{Im}\left(\partial_y g(x_*, y_*)|_{\mathrm{Ker}(\partial c(y_*))}\right)\right) = \dim\left(\mathrm{Ker}\left(\partial^2 f(x_*)\right)\right). \tag{8.34}$$

These observations are summarized in the following theorem.

Theorem 8.8 *Suppose x_* satisfies the first-order and second-order necessary optimality conditions for $\inf_{x \in \mathcal{G}} f(x)$ and let $g(x_*, y_*) = h(z_*)$ and $r(\cdot) := f(h(\cdot))$. Furthermore let f and g satisfy Assumption 8.1. Then z_* is verifiably locally optimal for $\inf_{z \in \Psi} r(z)$ if the following conditions hold true*

(i.) (8.33)

(ii.) (8.34)

(iii.) $\partial h(z_*)$ *has full column rank.*

Proof. Suppose (i.), (ii.) and (iii.) hold true. By Lemma 8.3 we conclude that z_* satisfies the second order necessary conditions for $\inf_{z \in \Psi} r(z)$. It remains to prove that $\partial^2 r(z_*) \succ 0$. From (8.32) we know that $\partial^2 r(z_*) \succeq 0$, so it remains to show that $\partial^2 r(z_*)$ is nonsingular. To prove this by contradiction, suppose there exists a $d \in \mathcal{Z}$, $d \neq 0$ such that

$$\partial^2 r(z_*)(d, d) = 0. \tag{8.35}$$

Since x_* is a critical point by assumption, (8.21) holds. This implies by Theorem 8.5 that

$$\mathrm{Im}\left(\partial_y g(x_*, y_*)|_{\mathrm{Ker}(\partial c(y))}\right) \subset \mathrm{Ker}\left(\partial^2 f(x_*)\right).$$

Hence to contradict (ii.) it suffices to construct a nonzero

$$d_1 \in \mathrm{Ker}\left(\partial^2 f(x_*)\right) \setminus \left(\mathrm{Im}\left(\partial_y g(x_*, y_*)|_{\mathrm{Ker}(\partial c(y_*))}\right)\right).$$

Since x_* is a critical point, Lemma 8.3 implies

$$\partial_x f(h(z_*)) = \partial_x f(g(x_*, y_*)) = 0. \tag{8.36}$$

Then (8.30) reads as

$$\underbrace{\partial^2 r(z_*)(d, d)}_{=0 \text{ by } (8.35)} = \partial^2 f(h(z_*)) \left(\partial h(z_*)d, \partial h(z_*)d\right) + \underbrace{\partial f(h(z_*)) \left(\partial^2 h(z_*)(d, d)\right)}_{=0 \text{ by } (8.36)},$$

which implies

$$\partial^2 f(h(z_*)) \left(\partial h(z_*)d, \partial h(z_*)d\right) = 0.$$

Since $\partial h(z_*)$ has full column rank (by (iii.)) we infer $d_1 := \partial h(z_*)d \neq 0$. Since $d_1 \in \mathrm{Im}\,(\partial h(z_*))$, (i.) implies

$$d_1 \notin \mathrm{Im}\left(\partial_y g(x_*, y_*)|_{\mathrm{Ker}(\partial c(y_*))}\right).$$

Hence

$$d_1 \in \mathrm{Ker}(\partial^2 f(x_*)) \setminus \left(\mathrm{Im}(\partial_y g(x_*, y_*))|_{\mathrm{Ker}(\partial c(y_*))}\right)$$

is the required contradicting vector.

∎

Remark. The condition on the dimension of the subspaces as given in (8.34) can be interpreted as follows. Suppose x_* satisfies the necessary optimality conditions. Then (8.34) implies that all vectors d in the kernel of the Hessian matrix $\partial^2 f(x_*)$ are in the image of $\partial_y g(x_*, y_*)$. In other words, x_* is not locally verifiable optimal only because of the over-parametrization described by g. In terms of the fixed-order $\mathcal{H}_\infty$-optimal control problem this occurs for an optimal controller K_* if every controller close to K_*, but with a different transfer function, has a strictly larger optimal value of the barrier function.

The following theorem is a statement on the converse implication of Theorem 8.8.

Theorem 8.9 *Suppose x_* satisfies the first-order and second-order necessary optimality conditions for $\inf_{x \in \mathcal{X}} f(x)$ and let $g(x_*, y_*) = h(z_*)$ and $r(\cdot) := f(h(\cdot))$. Furthermore let f and g satisfy Assumption 8.1. If*

$$\dim\left(Im\left(\partial h(z_*)\right)\right) + \dim\left(Im\left(\partial_y g(x_*, y_*)|_{Ker(\partial c(\hat{y}))}\right)\right) = \dim(\mathcal{X}) \tag{8.37}$$

then violation of at least one of the items (i.)-(iii.) in Theorem 8.8 implies that z_* is not verifiably locally optimal for $\inf_{z \in \Psi} r(z)$.

Proof. Since x_* is a critical point, Lemma 8.3 implies that $g(x_*, y_*)$ is also a critical point, i.e.

$$\partial_x f(h(z_*)) = \partial_x f(g(x_*, y_*)) = 0. \tag{8.38}$$

Hence for arbitrary $d \in \mathcal{Z}$ we infer from (8.30) that

$$\partial^2 r(z_*)(d, d) = \partial^2 f(h(z_*)) \left(\partial h(z_*)d, \partial h(z_*)d\right) + \underbrace{\partial f(h(z_*)) \left(\partial^2 h(z_*)(d, d)\right)}_{=0 \text{ by } (8.38)} =$$

$$= \partial^2 f(h(z_*)) \left(\partial h(z_*)d, \partial h(z_*)d\right). \tag{8.39}$$

Suppose (i.) in Theorem 8.8 does not hold true, i.e. there exists a $d \neq 0$ such that $d_1 := \partial h(z_*)d \in \text{Im}\left(\partial_y g(x_*, y_*)|_{\text{Ker}(\partial c(y_*))}\right)$. By (8.39) we conclude

$$\partial^2 r(z_*)(d, d) = \partial^2 f(g(x_*, y_*))\left(\partial h(z_*)d, \partial h(z_*)d\right) = \partial^2 f(g(x_*, y_*))(d_1, d_1) = 0, \quad (8.40)$$

where the last equality follows from Theorem 8.5. Equation (8.40) obviously implies that the second order sufficient conditions for $\inf_{z \in \Psi} r(z)$ are violated at z_*.

Secondly, suppose that (i.) in Theorem 8.8 is true, but (ii.) does not hold. Observe that (i.) and (8.37) imply

$$\text{Im}\left(\partial h(z_*)\right) \oplus \text{Im}\left(\partial_y g(x_*, y_*)|_{\text{Ker}(\partial c(\hat{y}))}\right) = \mathcal{X}. \quad (8.41)$$

Due to Theorem 8.5 and since (ii.) is not true,

$$\dim\left(\text{Im}\left(\partial_y g(x_*, y_*)|_{\text{Ker}(\partial c(y_*))}\right)\right) < \dim\left(\text{Ker}\left(\partial^2 f(x_*)\right)\right).$$

Lemma 8.4 implies that

$$\dim\left(\text{Im}\left(\partial_y g(x_*, y_*)|_{\text{Ker}(\partial c(y_*))}\right)\right) < \dim\left(\text{Ker}\left(\partial^2 f(g(x_*, y_*))\right)\right)$$

as well. Hence, there exists a $d \notin \text{Im}\left(\partial_y g(x_*, y_*)|_{\text{Ker}(\partial c(y_*))}\right)$ with

$$\partial^2 f(g(x_*, y_*))d = 0. \quad (8.42)$$

Due to (8.41), d can be decomposed as $d = d_1 + d_2$ where $d_1 \in \text{Im}\left(\partial_y g(x_*, y_*)|_{\text{Ker}(\partial c(y_*))}\right)$ and $d_2 \in \text{Im}\left(\partial h(z_*)\right) \neq 0$. Theorem 8.5 implies $\partial^2 f(g(x_*, y_*))d_1 = 0$, such that by (8.42) we conclude

$$0 = \partial^2 f(g(x_*, y_*))d = \partial^2 f(g(x_*, y_*))d_1 + \partial^2 f(g(x_*, y_*))d_2 = \partial^2 f(g(x_*, y_*))d_2. \quad (8.43)$$

Since $d_2 \in \text{Im}\left(\partial h(z_*)\right)$ there exists a d_3 such that $d_2 = \partial h(z_*)d_3$. By (8.39) we infer

$$\partial^2 r(z_*)(d_3, d_3) = \partial^2 f(h(z_*))(\partial h(z_*)d_3, \partial h(z_*)d_3) = \partial^2 f(g(x_*, y_*))(d_2, d_2) = 0,$$

where the utmost right equality follows from (8.43). Again we infer that the second order sufficient conditions for $\inf_{z \in \Psi} r(z)$ are violated at z_*.

Finally suppose that $\partial h(z_*)$ does not have full column rank, i.e. there exist a $d \neq 0$ with $\partial h(z_0)d = 0$. Then (8.39) implies that

$$\partial^2 r(z_*)(d, d) = \partial^2 f(h(z_*))(\partial h(z_*)d, \partial h(z_*)d) = \partial^2 f(h(z_*))(0, 0) = 0,$$

which again contradicts the second order sufficient conditions for $\inf_{z \in \Psi} r(z)$. ∎

Example 8.10 *We analyze the conditions described in this section for the optimal point $x_* = (1, 0, 1)$, $y_* = (1, 1)$ and $z_* = (1, 0)$ in Example 8.6. Firstly, observe that $\partial h(z)$ has full column rank for all z. Secondly $\partial f(x_*) = 0$ and*

$$\partial^2 f(x_*) = \begin{pmatrix} 2 + 8\gamma & -4\gamma & -8\gamma \\ -4\gamma & 2\gamma & 4\gamma \\ -8\gamma & 4\gamma & 8\gamma \end{pmatrix},$$

with eigenvalues $\lambda_1 = 0$, $\lambda_2(\gamma) = 1 + 9\gamma + \sqrt{1 - 2\gamma + 81\gamma^2}$ and $\lambda_3(\gamma) = 1 + 9\gamma - \sqrt{1 - 2\gamma + 81\gamma^2}$. Since $\lambda_2(\gamma) > 0$ and $\lambda_3(\gamma) > 0$ for all $\gamma > 0$, the Hessian $\partial^2 f(x_*)$ is positive semi-definite for all $\gamma > 0$ and

$$\dim\left(ker(\partial^2 f(x_*))\right) = 1 \text{ for all } \gamma > 0. \tag{8.44}$$

Now let us consider an arbitrary $\gamma > 0$. Since

$$\partial_y g(x_*, y_*) = \begin{pmatrix} 0 & 0 \\ 2 & 0 \\ 0 & 1 \end{pmatrix}$$

and $\partial c(y_*) = \begin{pmatrix} 1 & 1 \end{pmatrix}$, we infer

$$Im\left(\partial_y g(x_*, y_*)|_{Ker(\partial c(y_*))}\right) = \left\{ \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix} t \middle|\ t \in \mathbb{R} \right\}. \tag{8.45}$$

Furthermore

$$\partial h(z_*) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

such that (8.33) follows. (8.45) implies

$$\dim\left(Im\left(\partial_y g(x_*, y_*)|_{Ker(\partial c(y_*))}\right)\right) = 1.$$

This and (8.44) imply that (8.34) is true. Theorem 8.8 therefore implies that z_* is a verifiable local optimal point for $\inf_{z \in \Psi} r(z)$ for $r(\cdot) = f(h(\cdot))$. Indeed $r(z) := (z_1 - 1)^2 + \gamma(1 + z_2 - z_1^2)^2$ satisfies $\partial r(z_*) = 0$ and

$$\partial^2 r(z_*) = \begin{pmatrix} 2 + 8\gamma & -4\gamma \\ -4\gamma & 2\gamma \end{pmatrix} \succ 0 \text{ for all } \gamma > 0.$$

This implies that z_* is verifiable optimal. Since $\gamma > 0$ was arbitrary, we infer that for all $\gamma > 0$, z_* is verifiable optimal.

The conditions in Theorem 8.8 can also geometrically be illustrated using Figure 8.1, for $\gamma = 1$. The white solid curve in the figure through x_* is the collection of points

$$\{(g(x_*, y), f(g(h(z_*), y)))|\ y \in \Omega\}.$$

We observe from the figure that (8.33) holds true, since the tangents of this white solid curve and the white dashed curve at x_* are not the same. Furthermore it can be seen that the dimension of the kernel of the Hessian at x_* and the dimension of $Im\left(\partial_y g(x_*, y_*)|_{Ker(\partial c(y_*))}\right)$ are both 1, such that (8.34) holds. Therefore, Theorem 8.8 implies that z_* is verifiable optimal, which is also clearly visible in the figure.

This concludes the analysis on the existence of locally optimal verifiable points for optimization problems with over-parametrization. In the next sections we will use these results to analyze two parametrizations for the fixed-order controller synthesis problem. We start with a parametrization based on the controllable canonical form for SISO controllers

8.3 Reduced parametrizations based on controllable canonical form

Well-known controller parametrizations for SISO controllers are the *controllable* and *observable canonical* forms [62, 123]. In this section we discuss a reduced parametrization on the basis of the controllable canonical form. It will be verified that this parametrization eliminates the non-uniqueness due to similarity transformation of the controller states, if the parametrization is applied to the corrector step of the interior-point method for fixed-order $\mathcal{H}_\infty$-optimal control problem. Recall from Section 8.1 that $x = (\gamma, X, K)$, $\mathcal{X} = \mathbb{R} \times \mathcal{S}^{n+n_c} \times \mathbb{R}^{m_2+n_c \times p_2+n_c}$ and

$$\mathcal{G} := \{x = (\gamma, X, K) \in \mathcal{X} \mid X \succ 0, \ \mathcal{B}(\gamma, X, K) \succ 0\}.$$

As illustrated in Example 8.2, the non-uniqueness can be described with

$$g(x, y) := x^{\text{trans}}(x, y), \tag{8.46}$$

where x^{trans} is as defined in (8.8) and $\Omega = \{T \in \mathbb{R}^{n_c \times n_c} \mid \det(T) = 1\}$. A controller in controllable canonical form has the following structure

$$\begin{pmatrix} A_K & B_K \\ C_K & D_K \end{pmatrix} = \left(\begin{array}{ccccc|c} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & & 0 \\ \vdots & & \ddots & \ddots & \vdots & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ -a_{n_c} & -a_{n_c-1} & -a_{n_c-2} & \cdots & -a_1 & 1 \\ \hline c_1 & c_2 & c_3 & \cdots & c_{n_c} & d \end{array} \right) =: h_{\text{ctr}}(v), \tag{8.47}$$

where $a_i \in \mathbb{R}$, $c_i \in \mathbb{R}$, $i = 1, \ldots, n_c$, $d \in \mathbb{R}$ and $v = \begin{pmatrix} a_1 & \ldots & a_{n_c} & c_1 & \ldots & c_{n_c} & d \end{pmatrix}^T$. We will analyze the conditions discussed in the previous section for this parametrization. For this purpose we choose

$$h(z) = (\gamma, X, h_{\text{ctr}}(v)) \tag{8.48}$$

with $X \in \mathcal{S}^{n+n_c}$ and

$$z = \begin{pmatrix} \gamma \\ \text{svec}(X) \\ v \end{pmatrix}.$$

Obviously $h(z)$ is injective. We show that (8.33) holds true for every triple (x_0, y_0, z_0) with $x_0 \in \mathcal{G}$, $y_0 \in \Omega$, $z_0 \in \mathcal{Z} = \mathbb{R}^{1 + \frac{1}{2}(n+n_c)(n+n_c+1)+2n_c+1}$ satisfying $g(x_0, y_0) = h(z_0)$. For this purpose we define $x_1 := g(x_0, y_0)$ such that $g(x_1, I) = g(x_0, y_0)$. In Appendix A.5.2 it is shown that $\text{Im}(\partial g(x_1, I)) = \text{Im}(\partial g(x_0, y_0))$. Hence, we can assume that $y_0 = I$ without loss of generality. This implies that with $x_0 =: (\gamma_0, X_0, K_0)$, K_0 has the structure as (8.47). Assume that $d = (d_\gamma, d_K, d_X)$ is in the images of $\partial h(z_0)$ and of $\partial_y g(x_0, y_0)$. $d \in \text{Im}(\partial_y g(x_0, I))$ implies $d_\gamma = 0$ and

$$d_K = \left(\begin{array}{c|c} -d_T A_K + A_K d_T & -d_T e_{n_c} \\ \hline c d_T & 0 \end{array} \right), \tag{8.49}$$

for some $d_T \in \mathbb{R}^{n_c \times n_c}$ and where A_K has the structure as in (8.47), while e_{n_c} is the n_c^{th} basis vector in $\mathbb{R}^{n_c}$. On the other hand, $d \in \text{Im}(\partial h(z_0))$ implies

$$d_K = \left(\begin{array}{cc|c} 0_{(n_c-1) \times (n_c)} & & 0_{(n_c-1) \times 1} \\ d_a^T & & 0_{1 \times 1} \\ \hline d_c & & 0_{1 \times 1} \end{array} \right), \tag{8.50}$$

for some $d_a \in R^{n_c}$, where for clarity we show the dimensions of the zero elements. Combining (8.49) and (8.50) implies

$$\left(\begin{array}{c|c} -d_T A_K + A_K d_T & -d_T e_{n_c-1} \\ \hline c d_T & 0 \end{array} \right) = \left(\begin{array}{c|c} 0_{(n_c-1)\times(n_c)} & 0_{(n_c-1)\times 1} \\ d_a{}^T & 0_{1\times 1} \\ \hline d_c & 0_{1\times 1} \end{array} \right). \qquad (8.51)$$

The right upper $n_c \times 1$ sub-block of this equation implies $d_T e_{n_c} = 0$. d_T can therefore be partitioned as $d_T = \left(\begin{array}{cc} d_{T11} & 0 \end{array} \right)$, where $d_{T11} \in \mathbb{R}^{(n_c)\times(n_c-1)}$. The left upper $n_c \times n_c$ sub-block of (8.51) implies $d_{T11} = 0$, as is shown in Appendix A.5.1. We therefore conclude that d_T and hence $d = 0$, such that (8.33) follows.

Theorem 8.8 therefore implies that the reduced parametrization eliminates the non-uniqueness due to similarity transformations, for each $x_0 \in \mathcal{G}$ for which there exist $y_0 \in \Omega$, $z_0 \in \mathcal{Z}$ such that $g(x_0, y_0) = h(z_0)$. However, this parametrization is numerically not easy to work with, due to the often large differences in the magnitude of the coefficients that appear in $h(z)$. In our experience this parametrization is therefore not suited to be used in controller optimization, unless the controller order is very small. We propose a novel reduced parametrization for SISO and MIMO controllers that does have (for almost all controllers) the desired properties in the next section.

8.4 MIMO controller parametrization

In this section we propose a surjective parametrization for the optimization of multivariable controllers by the Interior Point method. For the controller A_K-matrix we choose the following parametrization:

$$A_{\text{par}}(a) = \left(\begin{array}{cccccccc} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & \cdots & \cdots & a_{1,(n_c-1)} & a_{1,n_c} \\ a_{2,1} & a_{1,1} & a_{2,3} & a_{2,4} & \cdots & \cdots & a_{2,(n_c-1)} & a_{2,n_c} \\ 0 & 0 & a_{3,3} & a_{3,4} & \cdots & \cdots & a_{3,(n_c-1)} & a_{3,n_c} \\ 0 & 0 & a_{4,3} & a_{3,3} & \cdots & \cdots & a_{4,(n_c-1)} & a_{4,n_c} \\ 0 & 0 & 0 & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & a_{(n_c-1),(n_c-1)} & a_{(n_c-1),n_c} \\ 0 & 0 & 0 & 0 & \cdots & 0 & a_{n_c,(n_c-1)} & a_{(n_c-1),(n_c-1)} \end{array} \right), \qquad (8.52)$$

where $a = (a_{1,1}, a_{1,2}, \ldots a_{n_c,n_c})^T \in \mathbb{R}^N$, $N = \frac{1}{2} n_c(n_c + 1)$. Note that the diagonal elements in the 2-by-2 blocks of $A_{\text{par}}(a)$ are equal.

8.4.1 Orthogonal transformations into (8.52)

The purpose of this section is to show that every matrix in $\mathbb{R}^{n_c \times n_c}$ can be turned into the form of (8.52) by an orthogonal similarity transformation, i.e. for any $B \in \mathbb{R}^{n_c \times n_c}$ there exist an orthogonal matrix Q, $Q^T Q = I_{n_c}$ and $a = (a_{1,1}, a_{1,2}, \ldots a_{n_c,n_c})^T \in \mathbb{R}^N$ such that

$$Q^T B Q = A_{par}(a).$$

First, it is shown in for instance [72] that every matrix in $\mathbb{R}^{n_c \times n_c}$ can be transformed by an orthogonal transformation into the real Schur form, i.e. for any $B \in \mathbb{R}^{n_c \times n_c}$ there exist

an orthogonal matrix $Q_{\text{schur}} \in \mathbb{R}^{n_c \times n_c}$, $Q_{\text{schur}}^T Q_{schur} = I_{n_c}$ and $r = (r_{1,1}, r_{1,2}, \ldots r_{n_c,n_c})^T$ such that

$$Q_{\text{schur}}^T B Q_{\text{schur}} = R_{schur}(r) = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & r_{1,4} & \cdots & \cdots & r_{1,(n_c-1)} & r_{1,n_c} \\ r_{2,1} & r_{2,2} & r_{2,3} & r_{2,4} & \cdots & \cdots & r_{2,(n_c-1)} & r_{2,n_c} \\ 0 & 0 & r_{3,3} & r_{3,4} & \cdots & \cdots & r_{3,(n_c-1)} & r_{3,n_c} \\ 0 & 0 & r_{4,3} & r_{4,4} & \cdots & \cdots & r_{4,(n_c-1)} & r_{4,n_c} \\ 0 & 0 & 0 & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & r_{(n_c-1),(n_c-1)} & r_{(n_c-1),n_c} \\ 0 & 0 & 0 & 0 & \cdots & 0 & r_{n_c,(n_c-1)} & r_{(n_c),(n_c)} \end{pmatrix}. \tag{8.53}$$

Secondly, observe that the zero structure in (8.53) is invariant under the transformations with the following class of block-diagonal orthogonal matrices

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} & 0 & 0 & 0 & \cdots & 0 & 0 \\ q_{2,1} & q_{2,2} & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & q_{3,3} & q_{3,4} & 0 & \cdots & 0 & 0 \\ 0 & 0 & q_{4,3} & q_{4,4} & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & q_{(n_c-1),(n_c-1)} & q_{(n_c-1),n_c} \\ 0 & 0 & 0 & 0 & \cdots & 0 & q_{n_c,(n_c-1)} & q_{(n_c),(n_c)} \end{pmatrix}, \text{ where } Q^T Q = I. \tag{8.54}$$

Therefore it suffices to show that for any matrix $B \in \mathbb{R}^{2 \times 2}$:

$$B = \begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{pmatrix}$$

there exists an orthogonal matrix $Q \in \mathbb{R}^{2 \times 2}$ and real numbers $a_{1,1}, a_{1,2}, a_{2,1} \in \mathbb{R}$ such that

$$Q^T B Q = A_{\text{par}}(a) = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{1,1} \end{pmatrix}. \tag{8.55}$$

To this purpose consider a family of orthogonal matrices parameterized by

$$Q(\phi) := \begin{pmatrix} \sin(\phi) & \cos(\phi) \\ -\cos(\phi) & \sin(\phi) \end{pmatrix} \tag{8.56}$$

and suppose that ϕ is such that

$$Q(\phi)^T B Q(\phi) = A_{\text{par}}(a) = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{1,1} \end{pmatrix}. \tag{8.57}$$

From (8.57) we conclude that

$$\sin^2(\phi)b_{1,1} + \cos^2(\phi)b_{2,2} - \cos(\phi)\sin(\phi)(b_{2,1} + b_{1,2}) =$$
$$= \cos^2(\phi)b_{1,1} + \sin^2(\phi)b_{2,2} + \cos(\phi)\sin(\phi)(b_{2,1} + b_{1,2}),$$

which is equivalent to

$$(\sin^2(\phi) - \cos^2(\phi))(b_{1,1} - b_{2,2}) = 2\cos(\phi)\sin(\phi)(b_{2,1} + b_{1,2}).$$

176

Using standard trigoniometic relations we conclude

$$\cos(2\phi)(b_{2,2} - b_{1,1}) = \sin(2\phi)(b_{2,1} + b_{1,2}). \tag{8.58}$$

Hence (8.57) is satisfied if we choose ϕ as

$$\begin{aligned}
\phi &= \tfrac{1}{4}\pi, & \text{if } b_{2,1} + b_{1,2} = 0 \\
&= \tfrac{1}{2}\tan^{-1}\left(\frac{-b_{1,1} + b_{2,2}}{b_{2,1} + b_{1,2}}\right), & \text{otherwise.}
\end{aligned}$$

If we therefore select for each ith diagonal 2×2 block of R_{schur} the appropriate angle ϕ_i, $i = 1, \ldots, \frac{n_c}{2}$ as follows

$$\begin{aligned}
\phi_i &= \tfrac{1}{4}\pi, & \text{if } r_{(i+1),i} + r_{i,(i+1)} = 0 \\
&= \tfrac{1}{2}\tan^{-1}\left(\frac{-r_{i,i} + r_{(i+1),(i+1)}}{r_{(i+1),i} + r_{i,(i+1)}}\right), & \text{otherwise,}
\end{aligned}$$

we can transform R_{schur} to the format in (8.52) by similarity transformation with

$$Q_{\text{blk}} := \begin{pmatrix} Q(\phi_1) & 0 & \cdots & 0 \\ 0 & Q(\phi_2) & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & Q(\phi_{(\frac{n_c}{2})}) \end{pmatrix}.$$

Hence $Q := Q_{\text{schur}} Q_{\text{blk}}$ is the matrix that transforms B into $Q^T B Q = A_{\text{par}}(a)$ for some $a = (a_{1,1}, a_{1,2}, \ldots a_{n_c, n_c})^T \in \mathbb{R}^N$.

8.4.2 Reduced parametrization for controller optimization

We propose the following parametrization for the optimization of fixed-order controllers

$$h(z) = \left(\gamma, \begin{pmatrix} X_{\text{P}} & U \\ U^T & \beta I_{n_c} \end{pmatrix}, \begin{pmatrix} A_{\text{par}}(a_{1,1}, a_{1,2}, \ldots, a_{n_c, n_c}) & B_K \\ C_K & D_K \end{pmatrix} \right) \tag{8.59}$$

where A_{par} is as in (8.52), $X_{\text{P}} \in \mathcal{S}^n$ and

$$z = \left(\gamma, \text{svec}(X_{\text{P}})^T, \text{vec}(U)^T, \beta, a_{1,1}, a_{1,2}, \ldots, a_{n_c, n_c}, \text{vec}(B_K)^T, \text{vec}(C_K)^T, \text{vec}(D_K)^T \right)^T. \tag{8.60}$$

Obviously $\partial h(z_*)$ has full column rank for all $z_* \in \mathcal{Z} := \mathbb{R}^{\frac{1}{2}n_c(n_c+1)+1+N}$, where $N := \frac{1}{2}n(n+1) + (n+m_2+p_2)n_c + m_2 p_2$. Furthermore the range of the mapping $q(z, T) := x^{\text{trans}}(h(z), T)$ contains all (γ, X, K) in the feasible domain $\mathcal{G}$, or more precisely

$$\mathcal{G} \subset \left\{ q(z, T) \in \mathbb{R} \times \mathcal{S}^{n+n_c} \times \mathbb{R}^{(n_c+m_2) \times (n_c+p_2)} \big| z \in \mathcal{Z}, \ X_{\text{P}} \succ 0, \ \beta > 0, \ \det(T) = 1 \right\}.$$

We show this by constructing for an arbitrary $x_0 = (\gamma_0, X_0, K_0) \in \mathcal{G}$ a z_0 and T such that $x_0 = x^{\text{trans}}(h(z_0), T)$. First, transform the controller states by a similarity matrix

$$T_1 := \begin{pmatrix} I_n & 0 \\ 0 & \det(R)R^{-1}, \end{pmatrix}$$

where R is the Cholesky factor of

$$X_K := \begin{pmatrix} 0 & I_{n_c} \end{pmatrix} X_0 \begin{pmatrix} 0 \\ I_{n_c} \end{pmatrix},$$

i.e. $X_K = R^T R$. This transformation yields

$$x^{\text{trans}}(x, T_1) = \left(\gamma, \begin{pmatrix} \tilde{X}_{\text{P}} & \tilde{U} \\ \tilde{U}^T & \det(R)^2 I \end{pmatrix}, \tilde{K}\right),$$

for some $\tilde{X}_{\text{P}} \in \mathcal{S}^n$ and $\tilde{U} \in \mathbb{R}^{n \times n_c}$. Next, using the procedure of the previous section an orthogonal Q can be constructed such that $Q^T A_K Q = A_{\text{par}}(a)$ for some $a = (a_{1,1}, a_{1,2}, \ldots, a_{n_c,n_c})$. Now we compute the the transformation matrix $T = T_1 T_2$, where T_2 is given by

$$T_2 := \begin{pmatrix} I_n & 0 \\ 0 & Q \end{pmatrix}.$$

This transformation matrix satisfies

$$x^{\text{trans}}(x_0, T) = \left(\gamma, \begin{pmatrix} X_{\text{P}} & U \\ U^T & , \beta I_{n_c} \end{pmatrix}, \begin{pmatrix} A_{\text{par}}(a_{1,1}, a_{1,2}, \ldots, a_{n_c,n_c}) & B_K \\ C_K & D_K \end{pmatrix}\right) = h(z_0),$$

for some X_{P}, U, β, a, B_K, C_K, D_K and z_0. We observe that T^{-1} is the desired transformation matrix satisfying $x^{\text{trans}}(h(z_0), T^{-1}) = x_0$. Hence we conclude that $(z, T) \mapsto x^{\text{trans}}(h(z), T)$ maps onto $\mathcal{G}$. Next, we show that (8.33) is true, if K satisfies the additional assumption:

$$\begin{pmatrix} I_{n_c} & 0 \end{pmatrix} K \begin{pmatrix} I_{n_c} & 0 \end{pmatrix}^T \in \mathcal{T}^{n_c \times n_c}, \tag{8.61}$$

where $\mathcal{T}^{n_c \times n_c} \subset \mathbb{R}^{n_c \times n_c}$ is the set of matrices $F \in \mathbb{R}^{n_c \times n_c}$ with the following properties:

1. F has distinct eigenvalues,

2. consider a real Schur form R of F. Then each diagonal block $R_{i,i} \in \mathbb{R}^{2 \times 2}$ does not have the form:

$$R_{i,i} = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \tag{8.62}$$

Theorem 8.11 *Let g and h be given by (8.20) and (8.59) respectively. Let $x_0 \in \mathcal{G}, y_0 \in \Omega, z_0 \in \mathcal{Z}$ be such that $g(x_0, y_0) = h(z_0) =: (\gamma, X, K)$, $K = \begin{pmatrix} A_K & B_K \\ C_K & D_K \end{pmatrix}$ with $A_K \in \mathcal{T}^{n_c \times n_c}$ for an even number n_c. Then (8.33) holds true.*

Proof. Consider a $d_x := (d_\gamma, d_X, d_K) \in \text{Im}\,(\partial h(z_0)) \cap \text{Im}\,(\partial_y g(x_0, y_0)|_{\text{Ker}(\partial c(y))})$. $d_x \in \text{Im}\,(\partial h(z_0))$ implies that

$$d_X := \left(\begin{array}{c|c} d_{X_{\text{P}}} & d_U \\ \hline d_U^T & d_\beta I_{n_c} \end{array}\right)$$

and

$$d_K = \left(\begin{array}{c|c} d_{A_K} & d_{B_K} \\ \hline d_{C_K} & d_{D_K} \end{array}\right),$$

for some $d_{X_{\text{P}}} \in \mathcal{S}^n$, $d_U \in \mathbb{R}^{n \times n_c}$ and $d_\beta \in \mathbb{R}$ and state space matrices $(d_{A_K}, d_{B_K}, d_{C_K}, d_{D_K})$, where

$$d_{A_K} = A_{\text{par}}\left(d_{a1,1}, d_{a1,2}, \ldots, d_{an_c,n_c}\right),$$

for some $d_{a1,1}, d_{a1,2}, \ldots, d_{an_c,n_c}$. On the other hand, $d_x \in \text{Im}\,(\partial_y g(x_0, y_0)|_{\text{Ker}(\partial c(y))})$ implies that there exists a d_T such that $\text{Trace}(d_T) = 0$ and

$$d_{A_K} = -d_T A_K + A_K d_T \tag{8.63}$$

178

and

$$\begin{pmatrix} 0 & I_{n_c} \end{pmatrix} d_X \begin{pmatrix} 0 \\ I_{n_c} \end{pmatrix} = \beta d_T + \beta d_T{}^T.$$

By combining these facts we conclude that $-d_T A_K + A_K d_T = A_{\text{par}} \left(d_{a1,1}, d_{a1,2}, \ldots, d_{an_c,n_c} \right)$ and $d_T + d_T{}^T = \frac{d_\beta}{\beta} I$. The latter implies that $d_T = \frac{d_\beta}{\beta} I + d_{T \text{skew}}$ for some skew-symmetric $d_{T \text{skew}}$. Then we can rewrite (8.63) as

$$d_{A_K} = -d_T A_K + A_K d_T = -\left(\frac{d_\beta}{\beta} + d_{T \text{skew}} \right) A_K + A_K \left(\frac{d_\beta}{\beta} + d_{T \text{skew}} \right) =$$
$$= -d_{T \text{skew}} A_K + A_K d_{T \text{skew}}. \quad (8.64)$$

From (8.52) we know that A_K admits the partition

$$A_K = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{(1,\frac{1}{2}n_c)} \\ 0 & A_{2,2} & \cdots & A_{(2,\frac{1}{2}n_c)} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & A_{(\frac{1}{2}n_c,\frac{1}{2}n_c)} \end{pmatrix}, \quad (8.65)$$

where $A_{i,j}$, $i,j = \left\{ 1, 2, \ldots, \frac{1}{2}n_c \right\}$ are 2×2 blocks. Let us partition $d_{T \text{skew}}$ accordingly as

$$d_{T \text{skew}} = \begin{pmatrix} d_{T1,1} & d_{T1,2} & \cdots & d_{T(1,\frac{1}{2}n_c)} \\ d_{T2,1} & d_{T2,2} & \cdots & d_{T(2,\frac{1}{2}n_c)} \\ \vdots & & \ddots & \vdots \\ d_{T\frac{1}{2}n_c,1} & \cdots & d_{T(\frac{1}{2}n_c,\frac{1}{2}n_c-1)} & d_{T(\frac{1}{2}n_c,\frac{1}{2}n_c)} \end{pmatrix}.$$

The left lower 2×2-block of (8.64) reads as

$$-d_{T(\frac{1}{2}n_c,1)} A_{1,1} + A_{(\frac{1}{2}n_c,\frac{1}{2}n_c)} d_{T(\frac{1}{2}n_c,1)} = 0.$$

This Sylvester equation has a unique solution $d_{T(\frac{1}{2}n_c,1)} = 0$, since the spectra of A_{11} and $A_{(\frac{1}{2}n_c,\frac{1}{2}n_c)}$ are disjoint by the assumption that $A_K \in \mathcal{T}^{n_c \times n_c}$.

$d_{T\frac{1}{2}n_c,1} = 0$ implies that the second left lower block reduces to

$$-d_{T(\frac{1}{2}n_c,2)} A_{2,2} + A_{(\frac{1}{2}n_c,\frac{1}{2}n_c)} d_{T(\frac{1}{2}n_c,2)} = 0$$

and since the spectra of $A_{(1,1)}$ and $A_{(\frac{1}{2}n_c,\frac{1}{2}n_c)}$ are also disjoint, we conclude $d_{T(\frac{1}{2}n_c,2)} = 0$. This reasoning can be continued to show that $d_{T \text{skew}}$ has the same block-upper triangular structure as A_K in (8.65). To prove this by induction, assume for some $i \in \{2, \ldots, \frac{1}{2}n_c\}$ and some $j < i$ that all elements of d_T left-lower to (i,j) are zero, more precisely that

$$d_{T(k,l)} = 0 \text{ for all } (k,l) \in \left(\{i, i+1 \ldots, \frac{1}{2}n_c\} \times \{1, 2, \ldots, j\} \right) \setminus (i,j). \quad (8.66)$$

Then $(e_i \otimes I_2)^T (8.64)(e_j \otimes I_2)$ reads as

$$-d_{T(i,j)} A_{j,j} + A_{i,i} d_{T(i,j)} = 0.$$

Since the spectra of $A_{i,i}$ and $A_{j,j}$ are disjoint, we conclude $d_{Ti,j} = 0$. Because the induction hypothesis (8.66) holds true for $i = \frac{1}{2}n_c$ and $j = 1$, we conclude

$$d_{T(i,j)} = 0 \text{ for all } i \in \{2, \ldots, \frac{1}{2}n_c\} \text{ and } j \in \{1, \ldots, i-1\}.$$

179

This implies that d_T is block upper triangular. Since $d_{T\text{skew}}$ is skew-symmetric as well, we conclude that it is even block diagonal with the only possible nonzero blocks of size 2×2. Suppose one such a block, say $d_{T(i,i)}$ for some $i \in \{1, 2, \ldots, \frac{1}{2}n_c\}$ is nonzero. Since it is skew-symmetric, it can be written as

$$d_{T(i,i)} = \begin{pmatrix} 0 & t \\ -t & 0 \end{pmatrix}$$

for some $t \in \mathbb{R}$. Since $A_{i,i}$ is a diagonal sub-block of A_K, it admits the following partitioning:

$$A_{i,i} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{1,1} \end{pmatrix}$$

for some $a_{1,1}$, $a_{1,2}$ and $a_{2,1}$. Furthermore $a_{1,2} + a_{2,1} \neq 0$, since $A_K \in \mathcal{T}^{n_c \times n_c}$ by assumption. Hence (8.64) implies

$$(e_i \otimes I_2)^T d_{A_K}(e_i \otimes I_2) = -d_{Ti,i}A_{i,i} + A_{i,i}d_{Ti,i} = \begin{pmatrix} -t(a_{2,1} + a_{1,2}) & 0 \\ 0 & t(a_{2,1} + a_{1,2}) \end{pmatrix}$$

Since

$$(e_i \otimes I_2)^T d_{A_K}(e_i \otimes I_2) = \begin{pmatrix} d_{A_K(2i-1,2i-1)} & d_{A_K(2i-1,2i)} \\ d_{A_K(2i,2i-1)} & d_{A_K(2i-1,2i-1)} \end{pmatrix},$$

we infer

$$-t(a_{2,1} + a_{1,2}) = t(a_{2,1} + a_{1,2})$$

This latter equation implies $t = 0$, since $a_{1,2} + a_{2,1} \neq 0$ (by $A_K \in \mathcal{T}^{n_c \times n_c}$). Hence we conclude $d_T = 0$, which completes the proof.

∎

By counting the dimensions of the subspaces $\text{Im}(\partial h(z_0))$ and $\text{Im}\left(\partial_y g(x_0, y_0)|_{\text{Ker}(\partial c(y_0))}\right)$ we infer that (8.37) holds true. Indeed the number of elements in $a = (a_{1,1}, a_{1,2}, \ldots, a_{n_c,n_c})$ is $\frac{1}{2}n_c(n_c + 1)$, $\beta \in \mathbb{R}$ and the number of variables in X_P, U, B_K, C_K and D_K is in total $N := \frac{1}{2}n(n + 1) + (n + m_2 + p_2)n_c + m_2 p_2$, such that $\dim(\text{Im}\partial h(z_0)) = \frac{1}{2}n_c(n_c + 1) + 1 + N$. If $x_0 = (\gamma, X, K) \in \mathcal{G}$ with K satisfying (8.61) and $y \in \Omega$, the dimension of $\text{Im}\left(\partial_y g(x_0, y_0)|_{\text{Ker}(\partial c(y_0))}\right)$ is $n_c^2 - 1$. This is true, since $\partial_y g(x_0, y_0)d = 0$ for some $d \in \text{Ker}(\partial c(y_0))$ implies $d = 0$, (as follows from the proof of Theorem 8.11) and $\dim(\text{Ker}(\partial c(y_0))) = 1$. Hence

$$\dim(\text{Im}\partial h(z_0)) + \dim \text{Im}(\partial_y g(x_0, y_0)) =$$
$$= n_c^2 + \frac{1}{2}n_c(n_c + 1) + \frac{1}{2}n(n + 1) + (n + m_2 + p_2)n_c + m_2 p_2 =$$
$$= \frac{1}{2}(n + nc)(n + n_c + 1) + (n_c + m_2)(n_c + p_2) = \dim(\mathcal{X}),$$

which implies (8.37).

Summarizing, since the proposed parametrization satisfies the properties above, we conclude that it

- parameterizes all stabilizing multi-variable controllers of order n_c

Table 8.2: Barrier values after 100 trust region steps for two parametrizations

Parametrization	full transformed	reduced
Final barrier value $\phi(x^{100})$	-27.093	-35.672
Final closed-loop $\mathcal{H}_\infty$-norm	1.7993	1.7838

and for almost all controllers (i.e. all controllers with $A_K \in \mathcal{T}^{n_c \times n_c}$, where $\mathcal{T}^{n_c \times n_c}$ is the set of matrices is as defined on page 178)

- eliminates the non-uniqueness due to similarity transformations

- and hence eliminates the failure of optimality tests due to these transformations.

8.4.3 Application

We analyze the convergence of the IP algorithm and the optimality conditions for the two-mass system as presented in Example 7.20, where we compare the full and reduced parametrization. We consider the first corrector step in the Interior Point algorithm and compare the convergence for a full and reduced parametrization. The initial set of decision variables is x_{init}, as has also been used in Example 7.20. Using the transformation $x_{\text{trans}} := g(x_{\text{init}}, y)$ in (8.8), we transformed the initial guess to the form in (8.59). We minimized the barrier function with the Interior Point method discussed in the previous chapter for two different parametrizations:

- full parametrization with the transformed variables (denoted by full transformed),

- reduced parametrization with $h(z)$ as in (8.59) (denoted by 'reduced')

The final values of the objective function after 100 trust-region steps are shown in Table 8.2. We observe that the final barrier value of the reduced parametrization is better than of the full parametrization, which illustrates the improved convergence that can be achieved with the reduced parametrization.

To evaluate the effect of the parametrization on the optimality conditions, we performed for the reduced parametrization 500 instead of 100 IP steps, in order to be in an (almost) optimal point. The resulting vector z_* has an objective value of $f(h(z_*)) = -36$. The smallest eigenvalue is $-1.4 \cdot 10^{-5}$, so z_* is not verifiably locally optimal. The number of eigenvalues of the Hessian H smaller than $10^{-16}\|H\|$ is 5. For the full parametrization, the Hessian $\partial^2 f(_*)$ evaluated at $x_* = h(z_*)$ has 13 eigenvalues smaller than 10^{-8} and its smallest eigenvalue is -2.91.

The controllers have been optimized further with both parametrizations, i.e. the IP-method has been applied with a full and reduced parametrization with starting points x_{trans} and $h(x_{\text{trans}})$, respectively. The following slight modifications have been applied:

- the simplified version of the predictor step as discussed in Section 7.1.4 has been used

- no intermediate analysis steps as described in Section 7.1.7 are performed.

The resulting decision variables are denoted by $x_{\text{full,red}}$ and $z_{\text{opt,red}}$, respectively. The closed-loop $\mathcal{H}_\infty$-norm of the final controller $K_{\text{opt,red}}$ corresponding to $z_{\text{opt,red}}$ for the reduced parametrization is 1.78, slightly better than that of the full parametrization 1.80,

and better than 1.79 obtained in Example 7.20. Note that the slight difference between the full optimization in this example and in Example 7.20 is due to the difference in initial decision variables and the slightly different predictor step.

Using the approach as described in Example 7.20, we evaluated the first and second order conditions for $z_{\mathrm{opt,red}}$ and $x = h(z_{\mathrm{opt,red}})$. The MFCQ as well as the first order necessary conditions are satisfied for both parameterizations. To evaluate the second order sufficient conditions, we have plotted in Figure 8.2 the eigenvalues of the reduced Hessian matrix $H_{L_{\mathrm{opt}}}$ as defined in Example 7.20. All eigenvalues smaller than 10^{-25} have been shown as being equal to 10^{-25}. For the reduced parametrization one eigenvalue of the
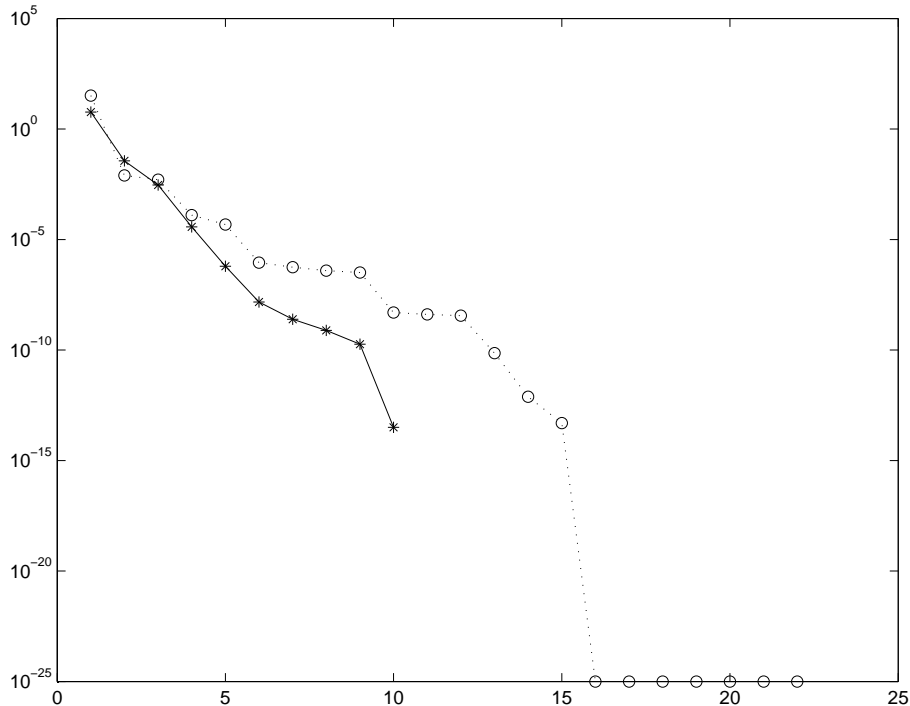
Figure 8.2: Eigenvalues of the Hessian matrix $H_{L_{\mathrm{opt}}}$ for the full parametrization $(\cdots\circ\cdots)$ and the reduced parametrization $(-*-)$. All eigenvalues smaller than 10^{-25} are shown as being equal to 10^{-25}.

Hessian matrix $H_{L_{\mathrm{opt}}}$ is smaller than 10^{-12}, such that unfortunately the sufficient condition is not convincingly satisfied. However, one small eigenvalue is a clear reduction compared to the seven almost zero eigenvalues of $H_{L_{\mathrm{opt}}}$ for the full parametrization.

8.4.4 Conclusions on reduced parametrization

In the previous section we have shown that $h(z)$ in (8.59) satisfies all hypotheses in Theorem 8.8 for all $x \in \mathcal{G}$ where K satisfies (8.61). This theorem implies that for these x, the reduced parametrization eliminates the non-uniqueness due to similarity transformations for almost all controllers. Furthermore we can describe through $h(z)$ *all* controller transfer functions with a single parametrization. This is in contrast to multi-variable extensions of controller canonical parametrizations [123], where this not possible in general.

Finally, we have illustrated the application of the parametrization to fixed order $\mathcal{H}_\infty$-optimal controller synthesis by interior-point optimization with an example. The parametriza-

tion improved the convergence of the corrector step of the IP algorithm and yielded slightly better final optimal closed-loop $\mathcal{H}_\infty$-norm, if compared to the final norm with the standard parametrization.

The proposed parametrization enables to verify local optimality of controllers.

8.5   Conclusion

The theoretical arguments and the controller design example presented in this chapter show that the proposed parametrization

- improves the Interior Point optimization algorithm and

- strengthens the local optimality conditions discussed in the previous chapter.

This completes our developments on research questions ① and ②, as have been posed on page 40.

The results in this chapter are a first step towards controller optimization oriented parameterizations. The analysis and the parametrization may be applied to other controller optimization algorithms and other optimization problems as well. Further developments in this area may for instance be an analysis of the sensitivity of the controller parameterizations to round-off errors during numerical optimization.

Chapter 9

Application to a wafer stage

One of the goals of this thesis project is to develop a fixed-order controller synthesis algorithm that can be used for controller design for industrial servo-systems. To evaluate the functionality of the presented algorithms for this purpose, we will discuss in this chapter the $\mathcal{H}_\infty$-optimal fixed-order controller design for a prototype of a wafer stage at Philips Applied Technologies. This system has approximately locally linear dynamics (although with quite some position dependence) and significant interaction beyond the required bandwidth. The interaction hampers achieving good performance for separate SISO controller designs. This motivates to apply $\mathcal{H}_\infty$-optimal controller synthesis based on a multi-variable model of the system. After a short description of the experimental set-up in Section 9.1, a model of the wafer stage is derived in Section 9.2 by closed-loop identification. Secondly, a four-block full-order $\mathcal{H}_\infty$-optimal controller design for the wafer stage is presented in Section 9.3. The resulting generalized plant is employed for designing and optimizing fixed-order controllers in Section 9.4. Both 3×3 MIMO and SISO controllers with structural constraints on their McMillan degree have been designed with several of the algorithms presented in previous chapters. The results allow us to compare the different algorithms in terms of convergence and computation time and their applicability to this type of servo-system control problems. On the basis of the numerical experience obtained by this exercise, we will present some guidelines that help the control designer to make the fixed-order optimization less prone to numerical errors.

A selection of the optimized controllers has been implemented on the system. Results of these experiments will be shown in Section 9.5. They illustrate how fixed-order controller optimization can be employed to improve disturbance rejection of controllers obtained by order reduction of full order $\mathcal{H}_\infty$-controllers, and how this can contribute to a better time-domain performance. Finally, in Section 9.6 we draw some conclusions.

9.1 System description

The system under consideration is a wafer stage prototype designed and constructed at Philips Applied Technologies in Eindhoven, the Netherlands. It serves as prototype for a new generation wafer-scanners. To protect the interests of the manufacturer we keep the explanation rather short. Furthermore, the time and frequencies are normalized with respect to a certain reference value. As explained in Chapter 1, the wafer stage is a part of a machine for IC-production based on photolithography. The figure with the setup was presented in Chapter 1 and is repeated in Figure 9.1 for the reader's convenience.

The wafer stage performs the motion of the wafers during exposure. In the machine an
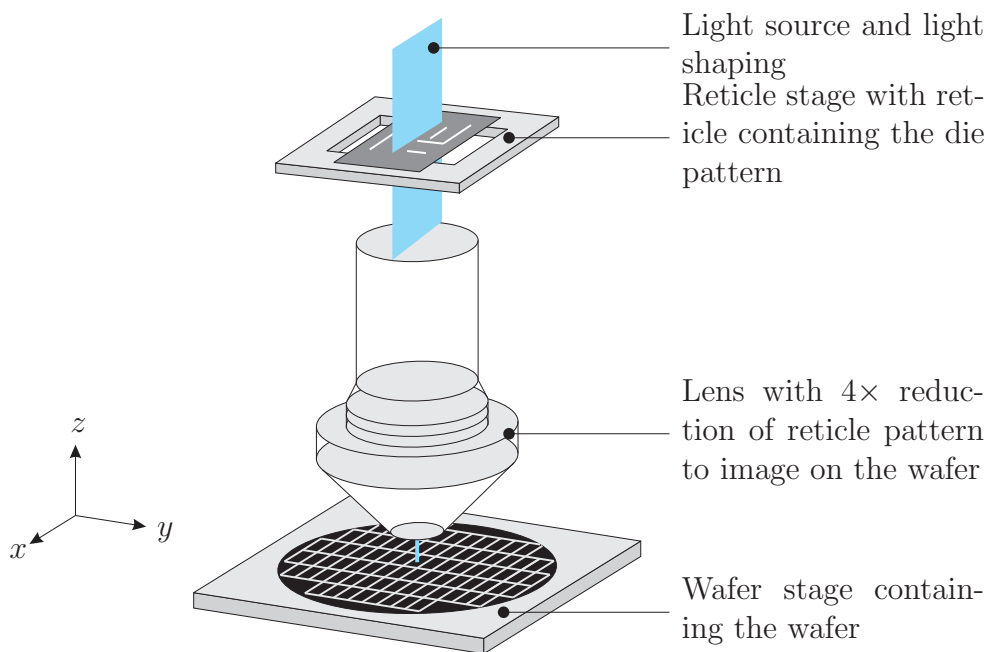
Figure 9.1: Representation of the basic layout of a wafer scanner.
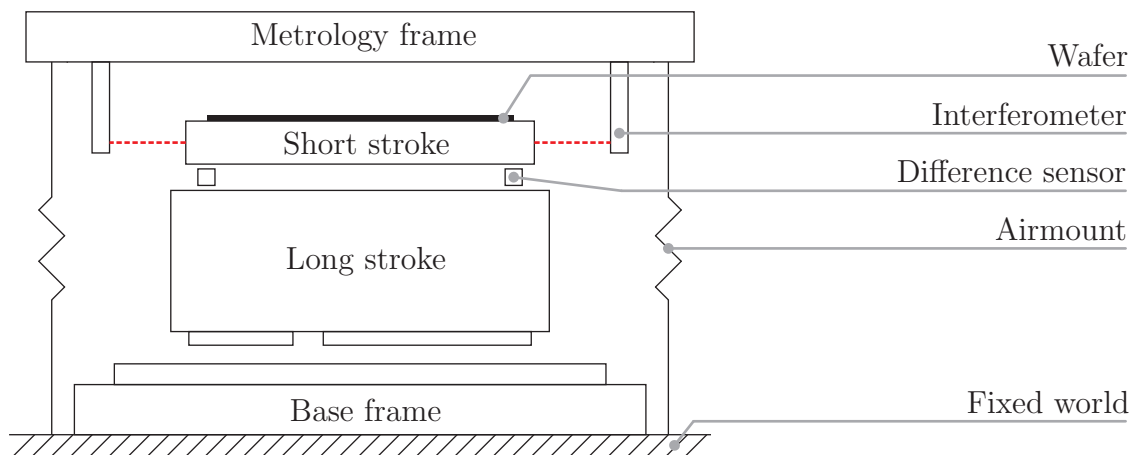
Figure 9.2: A schematic representation of the wafer stage carrier and measurement system.

optical system projects the IC-pattern on the wafer. The optical system is situated above the wafer stage in the complete wafer scanner, but at the time when experiments were performed, this system was not mounted onto the set-up.

Actuation and sensors

In the vertical direction the wafer stage and actuators are lifted with an electromagnetic gravity compensator, such that the system can move without friction in all directions. The wafer stage is actuated with *two strokes*. The *long-stroke actuator* performs the large motions. The *short-stroke actuator* is mounted on top of it, as is shown in Figure 9.2. It can only make small motions (with respect to the long-stroke), but has a much higher accuracy than the long stroke. Both actuators have 6 degrees of freedom. The order of magnitudes of their strokes and their required error bounds are given in Table 9.1. Since the highest demands on accuracy are on the short-stroke, we focus on the controller design

Table 9.1: Orders of magnitude of the strokes and errors of the long and short strokes.

	order of magnitude of the stroke	order of magnitude of the errors
Long stroke	1m	$1\mu m = 10^{-6}m$
Short stroke	1mm	$1nm = 10^{-9}m$

of this Lorentz-actuated system. The considered short-stroke is schematically depicted in Figure 9.3.

Coordinate transformations are used for the actuators and sensors, to obtain both controls and measurements in an orthogonal coordinate system with principal axes **x**, **y** and **z**. The scans are performed in the direction **y**. We focus on the translations, because the rotations are easily brought within the specifications with simple SISO controllers and their interaction with the translations can be ignored. The position and translations of
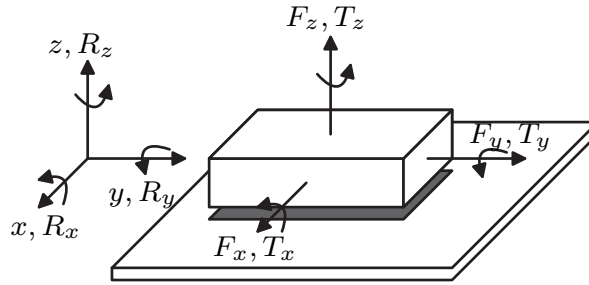
Figure 9.3: Degrees of freedom in the wafer stage.

the wafer are measured with six laser interferometers, with a resolution (and accuracy) in the order of tenths of nm. In addition the system has capacitive position sensors of lower accuracy which are used during start-up. Because of the low accuracy of these sensors, they are not used to control the short-stroke.

System specifications

The performance requirements of the system are in the order of 5nm for all three translations. The advanced mechatronic design of the system enables to achieve these tight specifications. As a result of this elaborate design, the wafer stage moves without friction, backlash and very little hysteresis so the dynamics of this system are approximately linear, with some slight position dependence. This position dependence is considered in more detail in [76] for a similar wafer stage prototype system.

Due to the relatively stiff construction, the resonance frequencies are high. To suppress excitation of the system by floor vibrations, the system is suspended with air mounts.

Controller synthesis method

The approximately linear dynamics of the system enable the control engineer to base the controller design on linear system analysis. To reach the high performance requirements, a high closed-loop bandwidth is needed. In this high-frequency region the interaction is significant. All this makes the system amenable for $\mathcal{H}_\infty$-optimal controller design. We design a single, fixed controller for the whole horizontal operating plane. A further improvement

could be reached by taking the position dependence explicitly into account using *Linear Parameter Varying* (LPV) controller synthesis (see e.g. [210], [46]). This is beyond the scope of this thesis. We refer the reader to [76] for an application of LPV control to a wafer stage.

Control architecture

The system is controlled with dedicated software. This software allows us to implement controllers in state-space format. Controllers designed in continuous time are discretised using the *First Order Hold* method. The controllers are implemented on six motion control boards at a sampling frequency of 30Hz (in terms of the scaled frequency axis)[1]. Measurements of the interferometers can be collected at the same sampling frequency. These measurements are used to identify the system in the next section, as well as for closed-loop control.

9.2 System identification

As a first step towards the controller design we have obtained an estimate of the *Frequency Response Function (FRF)* using an identification experiment. This experiment has to be performed in closed-loop, since the uncontrolled wafer stage would drift too much and make mechanical contact with the end of the strokes during an experiment. The controllers that were used for the experiments were three manually tuned low-bandwidth SISO PID-like controllers.

9.2.1 Experiments for system identification

Figure 9.4 shows the closed-loop set-up, where all signals are in discrete-time and the Digital to Analog (DA) and AD conversion are contained in $\mathbf{G}$. The discrete-time signals are the control outputs u, the measurement output y, the excitation signal n and the error signal e. They are vectors that comprise the three translations and are partitioned such that

$$
\begin{aligned}
u &= \left(\begin{array}{ccc} u_{\mathbf{x}} & u_{\mathbf{y}} & u_{\mathbf{z}} \end{array}\right)^T \\
y &= \left(\begin{array}{ccc} y_{\mathbf{x}} & y_{\mathbf{y}} & y_{\mathbf{z}} \end{array}\right)^T \\
n &= \left(\begin{array}{ccc} n_{\mathbf{x}} & n_{\mathbf{y}} & n_{\mathbf{z}} \end{array}\right)^T \\
e &= \left(\begin{array}{ccc} e_{\mathbf{x}} & e_{\mathbf{y}} & e_{\mathbf{z}} \end{array}\right)^T
\end{aligned}
$$

In three different experiments an excitation signal has been injected into one of the three directions $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. In the first experiment the actuation in $\mathbf{x}$-direction is excited, such that $n_{\mathbf{x}}$ is nonzero and $n_{\mathbf{y}} = n_{\mathbf{z}} = 0$. The excitation signal $n_{\mathbf{x}}$ is *pink noise*, i.e. a (pseudo-)random low-pass filtered white noise sequence. This low pass filter has a first order roll-off and a cut-off frequency 3Hz, which is a tenth of the sampling frequency. The standard deviation of the pink noise signal (in Newton) is $\frac{20}{3}$ N. In the other two experiments the $\mathbf{y}$- and $\mathbf{z}$-motors are excited in a similar fashion. At all three experiments the three measured positions y and measured control outputs u are logged at a sampling frequency of 30Hz. Each experiment is subdivided into 20 periods of about 550 seconds

---

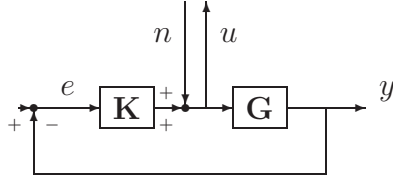[1]The unscaled sampling frequency is in orders of magnitude of kHz.

Figure 9.4: Closed loop system of Alpha tool wafer stage prototype

(in scaled time units). After each period the measurement data are saved on a hard-disk of the PC that is connected to the controller boards. The controller software computes an approximate *Frequency Response Function* of the open-loop system. We define *FRF estimate* of the open-loop system as a mapping $\tilde{G} : \Omega \to \mathbb{C}^{r \times r}$, where $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ is a finite number of frequency points. The number of degrees of freedom r is in this case equal to 3. The open-loop FRF estimate $\tilde{G}$ is distilled from the closed-loop measurement data as follows. First, the signals u, e and n are discrete-time Fourier transformed [87] to $\hat{u}(j\omega)$ $\hat{e}(j\omega)$ and $\hat{n}(j\omega)$, for a finite number of frequencies $\omega \in \Omega := \{\omega_1, \omega_2, \dots, \omega_N\}$. These are $N = 4889$ equidistant frequency points $\omega_k = 0.0494 + (k-1)0.00183$Hz, $k = 1, \dots, N$ such that $\omega_N = 9$Hz. An FRF estimate of the input sensitivity $S_{\text{inp}}(j\omega) := (I + K(j\omega)G(j\omega))^{-1}$, i.e. the transfer function from n to u, is computed as

$$\tilde{S}_{\text{inp}}(j\omega) := \left(\begin{array}{ccc} \frac{\hat{u}_1(j\omega)}{\hat{n}_1} & \frac{\hat{u}_2(j\omega)}{\hat{n}_2} & \frac{\hat{u}_3(j\omega)}{\hat{n}_3} \end{array} \right) \text{ for all } \omega \in \Omega,$$

where u_1 and n_1 denote the responses of the 1st experiment, i.e. where the **x**-motor has been excited[2]. Similarly u_2 and n_2 are the responses for the excited **y**-motor and u_3 and n_3 those for an excitation of the motor in **z**-direction. Since $u_i(j\omega)$ $i = 1, 2, 3$ are vectors in $\mathbb{C}^3$, $\tilde{S}_{\text{inp}}(j\omega)$ is a 3×3 complex matrix for all $\omega \in \Omega$. Similarly, an FRF-estimate of the process sensitivity $H(j\omega) := G(j\omega)(I + K(j\omega)G(j\omega))^{-1}$ is obtained by

$$\tilde{H}(j\omega) := \left(\begin{array}{ccc} \frac{\hat{e}_1(j\omega)}{\hat{n}_1} & \frac{\hat{e}_2(j\omega)}{\hat{n}_2} & \frac{\hat{e}_3(j\omega)}{\hat{n}_3} \end{array} \right), \text{ for all } \omega \in \Omega$$

The FRF estimate $\tilde{G} : \Omega \to \mathbb{C}^{r \times r}$ with $r = 3$ of the transfer function of the open-loop physical system (without controller) is finally computed as follows:

$$\tilde{G}(j\omega) = \tilde{H}(j\omega)\tilde{S}(j\omega)^{-1} \text{ for all } \omega \in \Omega.$$

This FRF is shown in Figure 9.5.

The rigid-body dynamics are clearly visible as double integrators in the diagonal elements of the system. Furthermore we see that the model is fairly well decoupled at low frequencies below 0.6Hz, since the magnitudes of the off-diagonal entries are small compared to the diagonal ones. This approximate decoupling of the rigid-body dynamics has been realized in the software by multiplying the controller outputs by a static matrix (i.e. static decoupling). At the diagonal entries the dynamics introduced by flexibilities are clearly visible from around 3Hz.

The FRF-estimate of the dynamics in **x**-direction is denoted by $\tilde{G}_{\mathbf{xx}}(j\omega)$. The corresponding FRF-data are shown in the left upper block of the figure. $\tilde{G}_{\mathbf{xy}}(j\omega)$ is the FRF-estimate of the dynamics from the **y** actuator to the measured **x**-position etc. In

[2]Actually a full 6×6 MIMO FRF is identified, where a posteriori the 3×3 subsystem **x**/**y**/**z** is split off.
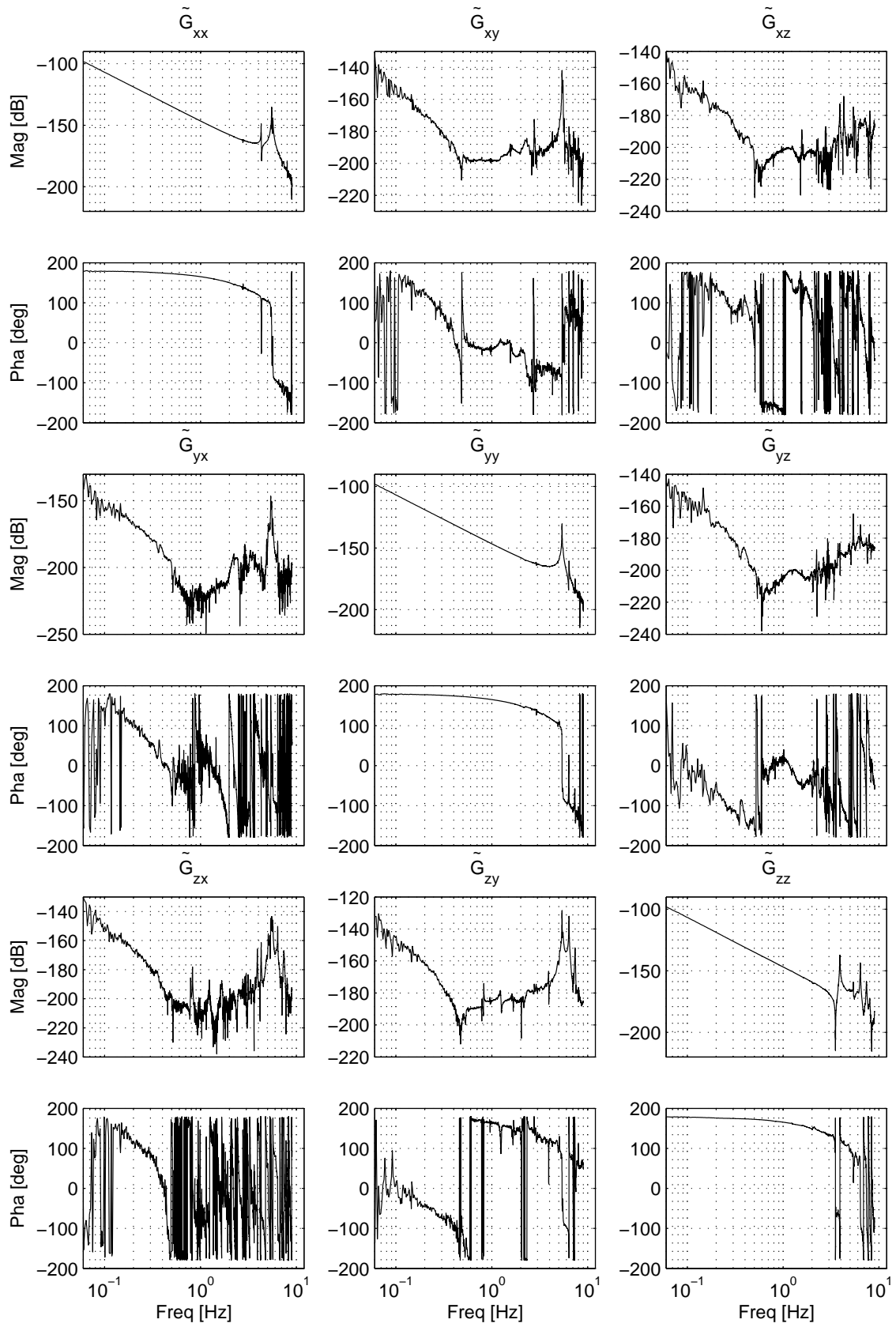
Figure 9.5: Estimated FRF $\widetilde{G}$ of the 3×3 MIMO open-loop system.
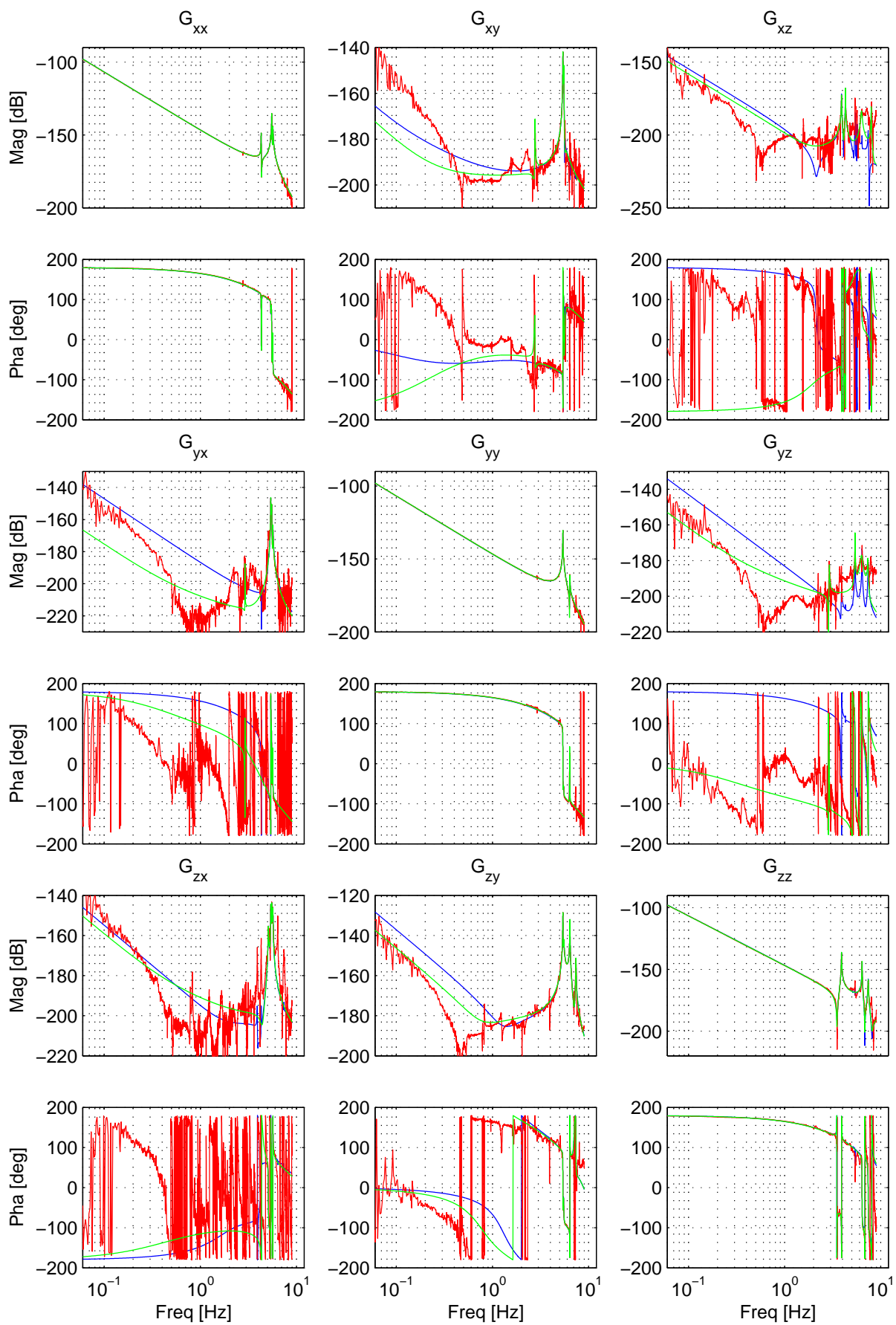
Figure 9.6: Estimated FRF $\tilde{G}$ (red), 82th order model $G^{82^{\text{nd}}}$ (green) and reduced 32th order model G (blue)

191

$\tilde{G}_{\mathbf{xx}}(j\omega)$ we observe two resonance pairs around 4.28Hz and 5.53Hz. The gain of $\tilde{G}_{\mathbf{xx}}(j\omega)$ rolls off with -4 slope after 5.69Hz.

$\tilde{G}_{\mathbf{yy}}(j\omega)$ is the transfer function from force on the **y**-motor to measured position in **y**-direction. The -4 slope starts in this entry of the FRF at around 5.4Hz. In the frequency response of the **z**-direction (i.e. $\tilde{G}_{\mathbf{zz}}(j\omega)$) we observe a complex zero pair followed by a complex pole-pair. The FRF estimates of the off-diagonal elements are much less reliable, due to the lower signal-to-noise ratio. There are peaks in $\tilde{G}_{\mathbf{xy}}(j\omega)$, $\tilde{G}_{\mathbf{yy}}(j\omega)$ and $\tilde{G}_{\mathbf{zy}}(j\omega)$ at around 5.4Hz, which indicates that they are possibly from the same source. They are probably caused by a flexibility that is excited by the **y**-motor and can be measured in all directions.

Remark. Formally speaking the FRF estimates do not have poles and zeros, such that the above terminology is not fully correct. We use these terms here, because the transfer functions identified in Section 9.2.3 will have poles and zeros at the corresponding frequencies.

9.2.2   Rigid-body dynamics and time delay

To improve the fit of a model on the FRF, we first remove the rigid body dynamics and time delay.

Rigid body dynamics

The rigid body dynamics of the system can be modelled by double integrators. These integrators make the low-frequent gains of the diagonal elements of the FRF very large, if compared to the high-frequent dynamics. To reduce the effect of round-off errors on the modelling of the high-frequent dynamics, we multiply the FRF by $(j\omega)^2 I$ for all $\omega \in \Omega$, i.e. the frequency response of the inverse of the rigid body dynamics.

Time delay

Due to computational delays in the real-time processor and Zero-Order Hold controller implementation, the open-loop system has a time delay. The delay is estimated using the FRF, by considering the phase lag at frequency $\omega_d = 0.6$Hz. We assume that the phase lag of the open-loop physical system at this frequency is only due to the rigid body dynamics and the delay time τ_d. Since the rigid body dynamics introduce a phase lag of 180 deg, the total phase at frequency ω_d can be expressed as

$$\phi = -180 - 360\tau_d\omega_d,$$

where ϕ and ω_d are in degrees and (scaled frequency unit) Hz respectively. The estimated phase lag for the diagonal elements of the FRF at $\omega_d = 0.6$Hz is shown in Table 9.2. On the basis of these estimates, we estimated the delay to be about 41.7ms. In our experience the quality of the model is improved if the phase-contribution of the delay is first removed from the data, before fitting transfer functions on the data. A *Padé approximation* of the delay is added afterwards to the identified model.

The 'elimination' of the rigid-body dynamics and delays results in the following Modified FRF:

$$\tilde{G}_{\mathrm{M}}(j\omega) = \left((j\omega)^2 e^{\tau_d j\omega} I\right) \tilde{G}(j\omega) \text{ for all } \omega \in \Omega$$

Table 9.2: Estimated delay times at frequency $\omega_d = 0.6$Hz for the three SISO FRF estimates.

	x	**y**	**z**
phase ϕ [deg]	-188.76	-188.74	-188.52
estimated time delay [ms]	40.6	40.5	39.5

After having obtained an identified and Reduced state space model G_{MR} of this Modified FRF, we multiply its transfer by $\frac{1}{s^2}p(s)I$ to obtain a model of the wafer stage system:

$$G(s) = \left(\frac{1}{s^2}p(s)I \right) G_{\text{MR}}(s) \tag{9.1}$$

where $p(s)$ is a SISO second order Padé approximation of $e^{-\tau_d s}$.

9.2.3 Identification of a state-space model

A state-space model $\mathbf{G}_{\text{M}}$ with transfer function $G_{\text{M}}(s)$ has been derived from the FRF estimate $\tilde{G}_{\text{M}}$ with the least squares approximation algorithm for SISO systems as described in [178]. We do this entry by entry, i.e. we fit each entry $G_{\text{M}}(s)_{\mathbf{xx}}$, $G_{\text{M}}(s)_{\mathbf{xy}}$, ..., $G_{\text{M}}(s)_{\mathbf{zz}}$ separately. For each entry the smallest model order has been chosen that resulted in a transfer function that well enough interpolates the estimated FRF $\tilde{G}_{\text{M}}$ at Ω, if judged by visual inspection of the amplitude and phase of the frequency response. After obtaining models for all entries in this fashion, they have been combined into an $r \times r$ stable transfer function matrix $G_{\text{M}}(s)$ with McMillan degree 70. Based on this transfer function matrix, an 82nd-order model has been computed: $G^{82^{\text{nd}}}(s) := \left(\frac{1}{s^2}p(s)I \right) G_{\text{M}}(s)$. The frequency response of $G^{82^{\text{nd}}}(s)$ and estimated FRF are shown in Figure 9.6 in green and red respectively.

Figure 9.7 shows the relative difference between the measured FRF $\tilde{G}$ and the 82nd order model computed G as

$$\begin{pmatrix} \frac{\left| G^{82^{\text{nd}}}_{\mathbf{xx}}(i\omega)-\tilde{G}_{\mathbf{xx}}(i\omega) \right|}{\left| G^{82^{\text{nd}}}_{\mathbf{xx}}(i\omega) \right|} & \frac{\left| G^{82^{\text{nd}}}_{\mathbf{xy}}(i\omega)-\tilde{G}_{\mathbf{xy}}(i\omega) \right|}{\left| G^{82^{\text{nd}}}_{\mathbf{xy}}(i\omega) \right|} & \frac{\left| G^{82^{\text{nd}}}_{\mathbf{xz}}(i\omega)-\tilde{G}_{\mathbf{xz}}(i\omega) \right|}{\left| G^{82^{\text{nd}}}_{\mathbf{xz}}(i\omega) \right|} \\ \frac{\left| G^{82^{\text{nd}}}_{\mathbf{yx}}(i\omega)-\tilde{G}_{\mathbf{yx}}(i\omega) \right|}{\left| G^{82^{\text{nd}}}_{\mathbf{yx}}(i\omega) \right|} & \frac{\left| G^{82^{\text{nd}}}_{\mathbf{yy}}(i\omega)-\tilde{G}_{\mathbf{yy}}(i\omega) \right|}{\left| G^{82^{\text{nd}}}_{\mathbf{yy}}(i\omega) \right|} & \frac{\left| G^{82^{\text{nd}}}_{\mathbf{yz}}(i\omega)-\tilde{G}_{\mathbf{yz}}(i\omega) \right|}{\left| G^{82^{\text{nd}}}_{\mathbf{yz}}(i\omega) \right|} \\ \frac{\left| G^{82^{\text{nd}}}_{\mathbf{zx}}(i\omega)-\tilde{G}_{\mathbf{zx}}(i\omega) \right|}{\left| G^{82^{\text{nd}}}_{\mathbf{zx}}(i\omega) \right|} & \frac{\left| G^{82^{\text{nd}}}_{\mathbf{zy}}(i\omega)-\tilde{G}_{\mathbf{zy}}(i\omega) \right|}{\left| G^{82^{\text{nd}}}_{\mathbf{zy}}(i\omega) \right|} & \frac{\left| G^{82^{\text{nd}}}_{\mathbf{zz}}(i\omega)-\tilde{G}_{\mathbf{zz}}(i\omega) \right|}{\left| G^{82^{\text{nd}}}_{\mathbf{zz}}(i\omega) \right|} \end{pmatrix}, \text{ for all } \omega \in \Omega,$$

We observe that at the diagonal entries both phase and amplitude of the estimated FRF and the model $G^{82^{\text{nd}}}(s)$ match quite well. The off-diagonal entries are badly estimated at low frequencies. It has been mentioned in the previous section that the estimated FRF is not very reliable in the low-frequency range, which (partially) explains the difficulty to obtain a good fit. It also implies that further improvement of the off-diagonal elements does not necessarily improve the model much. For multi-variable controller design a good fit of the off-diagonal elements is required at frequencies around the desired bandwidth, which is 0.9Hz for **x** and **y** and 0.48Hz for **z**. Based on visual inspection we conclude that around these frequencies the fit seems accurate enough. Many of the resonances that are important for control are between 3Hz and 7Hz. To assess the quality of the fit of the
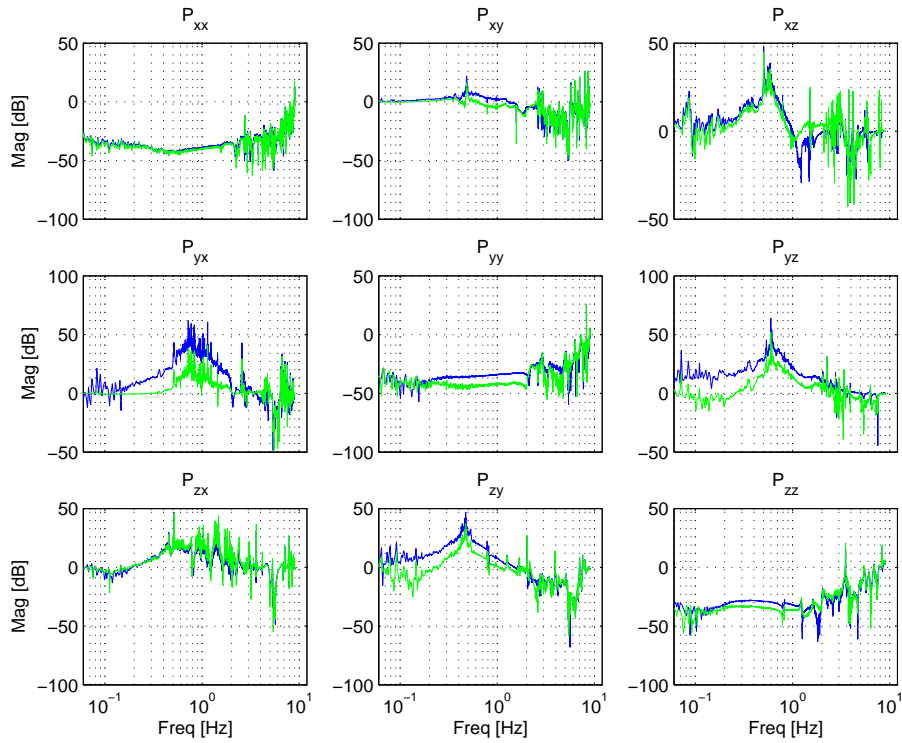
Figure 9.7: Relative difference between the estimated FRF $\tilde{G}$ and following two models: 82th order model $G^{82^{\text{nd}}}$ (green) and reduced 32th order model G (blue).

important resonances, a magnified version of Figure 9.6 is shown in Figure 9.8 between these frequencies. This figure illustrates that several modes of the off-diagonal elements are fitted well. For instance, the mode at 5.4Hz in the second column of the transfer function matrix is captured in the model. Furthermore $G_{\mathbf{xy}}^{82^{\text{nd}}}$ has a resonance at 5.35Hz, which is included in the model, and the peaks at 5.4Hz, 5.52Hz and 5.66Hz of $G_{\mathbf{yx}}^{82^{\text{nd}}}$ and 3.86Hz and 4.3Hz of $G_{\mathbf{xz}}^{82^{\text{nd}}}$ are captured by the model.

Remark. To cover the mismatch between the FRF and the model, the closed loop performance is not only analyzed on the model, but also on the FRF. This analysis is presented in Section 9.4.2.

Remark. Subspace identification in the frequency domain is an alternative technique for obtaining a state-space model fitting on the measured FRF. We have tried this technique [129], but the resulting model was not better than the 82th order model and the reduced model presented in the next section, both shown in Figure 9.6, if judged by visual inspection of the frequency response.

9.2.4 Model reduction

It is not unlikely that the frequency estimates of the different entries of $G^{82^{\text{nd}}}$ have common poles and zeros, because flexibilities in the mechanical structure may have effect on more than one physical direction, or may be excited by more than one actuator. Due to modelling inaccuracies these common resonances do not have the same pole-locations exactly, but
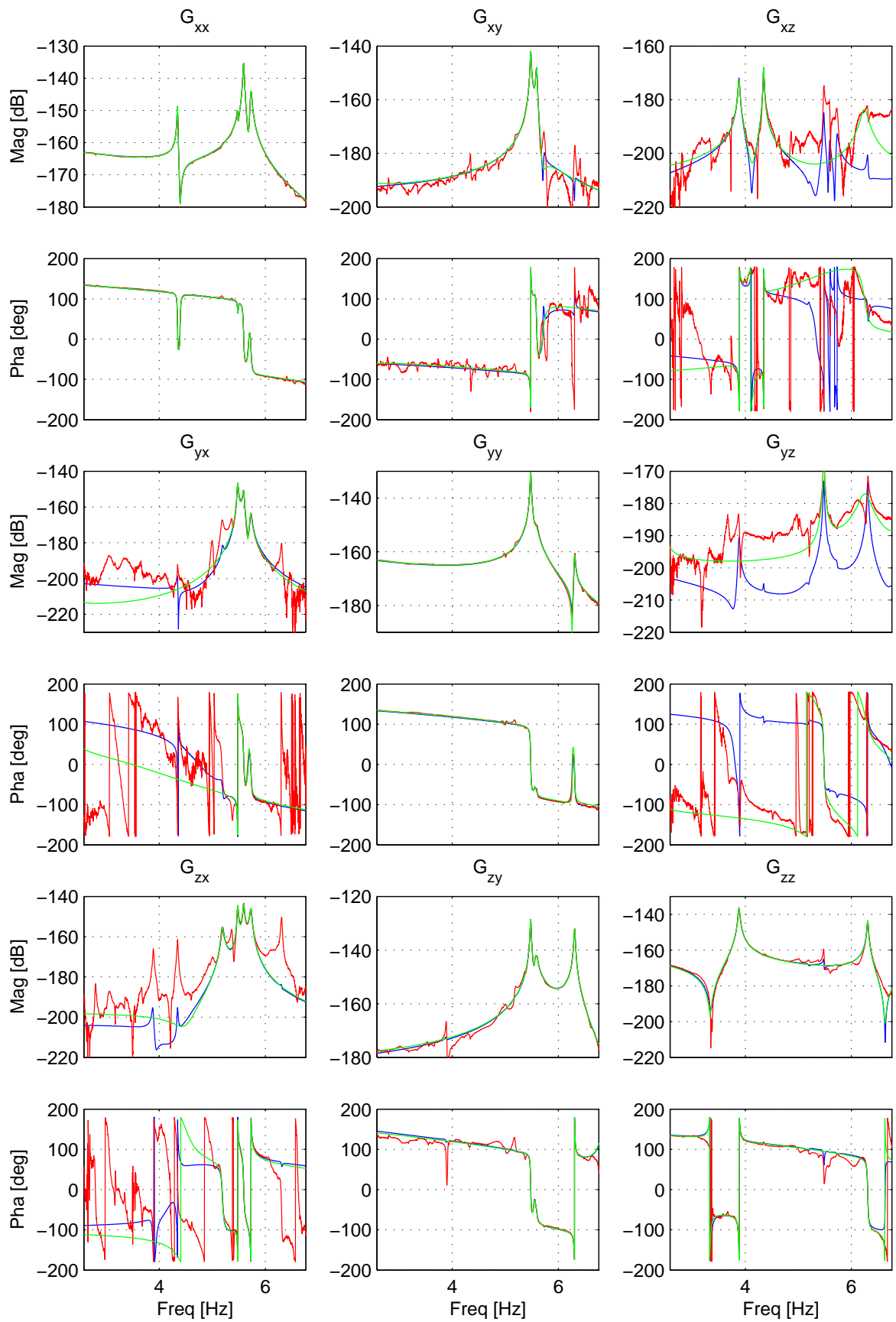
Figure 9.8: Estimated FRF $\tilde{G}$ (red), 82th order model $G^{82^{\text{nd}}}$ (green) and reduced 32th order model G (blue), zoomed in at the frequency region between 3Hz and 7Hz

they will be close to each other in terms of damping and natural frequency. For instance, the poles around 5.4Hz in the second column of the transfer function seem to result from a single dynamic mode.

As discussed in the previous section, a realization of the transfer function G_{M} is obtained by combining the realizations of the entries. Common poles of these entries may give rise to a non-minimal realization or a realization with some Hankel Singular Values (HSVs) that are very small if compared to the largest HSV. Elimination of the corresponding states in the balanced realization will not change the transfer function much, due to the bound on the $\mathcal{H}_\infty$-norm of the truncation error as explained in Lemma 3.1. At the same time, this reduction simplifies the controller optimization, because the computation time of the fixed-order algorithms is large for higher order models. The stable 70th order model $G_{\mathrm{M}}(s)$ has therefore been reduced by balanced residualization as explained in Section 3.1.1. With order $n = 20$ we have obtained a reasonable fit $G_{\mathrm{MR}}(s)$ of the estimated FRF. Finally, the integrators and Padé approximations are added to the reduced model $G_{\mathrm{MR}}(s)$ resulting in $G(s)$ as given by (9.1).

The transfer function is given in blue in Figures 9.6 and 9.8. The McMillan degree of the transfer function is 32. $G(s)$ is computed using (9.1). The frequency response of $G(s)$ and estimated FRF are shown in Figure 9.6 in green and red respectively. Figure 9.7 shows the relative difference between the measured FRF $\tilde{G}$ and the reduced 32th order model G computed as

$$
\begin{pmatrix}
\dfrac{\left|G_{\mathbf{xx}}(i\omega) - \tilde{G}_{\mathbf{xx}}(i\omega)\right|}{|G_{\mathbf{xx}}(i\omega)|} & \dfrac{\left|G_{\mathbf{xy}}(i\omega) - \tilde{G}_{\mathbf{xy}}(i\omega)\right|}{|G_{\mathbf{xy}}(i\omega)|} & \dfrac{\left|G_{\mathbf{xz}}(i\omega) - \tilde{G}_{\mathbf{xz}}(i\omega)\right|}{|G_{\mathbf{xz}}(i\omega)|} \\[2ex]
\dfrac{\left|G_{\mathbf{yx}}(i\omega) - \tilde{G}_{\mathbf{yx}}(i\omega)\right|}{|G_{\mathbf{yx}}(i\omega)|} & \dfrac{\left|G_{\mathbf{yy}}(i\omega) - \tilde{G}_{\mathbf{yy}}(i\omega)\right|}{|G_{\mathbf{yy}}(i\omega)|} & \dfrac{\left|G_{\mathbf{yz}}(i\omega) - \tilde{G}_{\mathbf{yz}}(i\omega)\right|}{|G_{\mathbf{yz}}(i\omega)|} \\[2ex]
\dfrac{\left|G_{\mathbf{zx}}(i\omega) - \tilde{G}_{\mathbf{zx}}(i\omega)\right|}{|G_{\mathbf{zx}}(i\omega)|} & \dfrac{\left|G_{\mathbf{zy}}(i\omega) - \tilde{G}_{\mathbf{zy}}(i\omega)\right|}{|G_{\mathbf{zy}}(i\omega)|} & \dfrac{\left|G_{\mathbf{zz}}(i\omega) - \tilde{G}_{\mathbf{zz}}(i\omega)\right|}{|G_{\mathbf{zz}}(i\omega)|}
\end{pmatrix} , \text{ for all } \omega \in \Omega.
$$

The figures illustrate that the 'dominant' dynamics of the diagonal and off-diagonal elements are fairly well captured by the reduced-order model. For instance, the resonance at 5.35Hz of $G_{\mathbf{xy}}$ is included in model. The peak of the resonance at 2.77Hz in $G_{\mathbf{xy}}$ is 'smaller', and this may be the explanation for its presence and absence in the high-order model and the 32th order model, respectively. The transfer function G of Mc-Millan degree 32 will be used for controller synthesis in the next section.

9.3 Four-block $\mathcal{H}_\infty$-optimal controller design

We compute the controllers by $\mathcal{H}_\infty$ synthesis using a four-block design, as depicted in Figure 9.9. The closed-loop plant $M(s)$ is given by
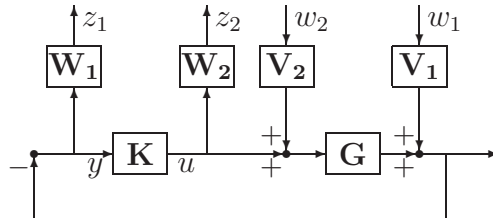
Figure 9.9: Four-block controller set-up

$$
M := \left(\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array} \right) = - \left(\begin{array}{c|c} W_1 S V_1 & W_1 S G V_2 \\ \hline W_2 K S V_1 & W_2 K S G V_2 \end{array} \right) \tag{9.2}
$$

196

where $S = (I + GK)^{-1}$. This implies that the four-block design procedure allows us to shape, although not independently, the loops of the following closed loop transfer functions:

1. the sensitivity $S = (I + GK)^{-1}$,

2. the process sensitivity SG,

3. the controller sensitivity KS and

4. KSG, which is in the SISO case equal to the complementary sensitivity $T = GKS$.

Apart from the possibility to shape four transfer functions, an advantage of the four-block design is that pole zero cancellations are not so likely to occur than with two-block designs such as S/KS or S/T as for instance explained in [183]. See [199] and the references therein for some more detailed considerations on the choice for the four-block problem. The reader is referred to this latter paper and the dissertation [45] for four-block $\mathcal{H}_\infty$-designs applied to predecessors of the wafer stage prototype discussed in this chapter.

To achieve good disturbance rejection we aim at bandwidths of around 0.9Hz for **x** and **y** and 0.48Hz for **z**. The choice of the weighting filters V_1, V_2, W_1 and W_2 is based on [199], where an $\mathcal{H}_\infty$-optimal controller design technique is presented for a wafer stage. For completeness we give a short explanation of this procedure.

9.3.1 Choice of weighting filters

The weighting filters are diagonal and chosen on the basis of SISO loop-shaping arguments. The diagonal entries of the weighting filters are parameterized using the following parameters:

- the target bandwidth f^{BW}, defined here for a SISO system as the frequency where the open-loop gain GK first crosses the 0dB line (or in other words the cross-over frequency).

- the frequency f^I. Up to this frequency the controller has integral action to suppress low frequent disturbances,

- the frequency f^R. Beyond this frequency the controller must have a -2 roll-off, to avoid amplification of sensor noise and highly saturating actuators and to avoid performance deterioration due to unmodelled high-frequent dynamics.

f^I is typically chosen five times smaller than the corresponding target bandwidth f^{BW} whereas f^R is usually chosen five times larger than f^{BW}. These few design parameters have a clear interpretation which makes the design easier and more intuitive.

To enforce zero steady state error, W_1 is chosen as the diagonal augmentation of first order filters, each with a single pole at $s = 0$:

$$W_1(s) = D \begin{pmatrix} \frac{s+2\pi f_{\mathbf{x}}^I}{2s} & 0 & 0 \\ 0 & \frac{s+2\pi f_{\mathbf{y}}^I}{2s} & 0 \\ 0 & 0 & \frac{s+2\pi f_{\mathbf{z}}^I}{2s} \end{pmatrix} \tag{9.3}$$

where $D = \mathrm{diag}(d_{\mathbf{x}}, d_{\mathbf{y}}, d_{\mathbf{z}})$ is a diagonal matrix whose diagonal nonzero entries $d_{\mathbf{x}}$, $d_{\mathbf{y}}$ and $d_{\mathbf{z}}$ are used to scale different units to comparable orders of magnitude and to address differences in specifications.

W_2 has high gain at high frequencies to enforce the controller to have roll-off. By tuning we concluded that first-order roll-off for $\mathbf{x}$ and $\mathbf{y}$ and second order roll-off for $\mathbf{z}$ work best. W_2 is therefore chosen as

$$W_2(s) = D \begin{pmatrix} w_{\mathbf{x}}(s) & 0 & 0 \\ 0 & w_{\mathbf{y}}(s) & 0 \\ 0 & 0 & w_{\mathbf{z}}(s) \end{pmatrix}, \tag{9.4}$$

where

$$w_{\mathbf{x}}(s) := k_{\mathbf{x}}^R \alpha_{\mathbf{x}} \frac{s + (2\pi f_{\mathbf{x}}^R)}{s + (\alpha_{\mathbf{x}} 2\pi f_{\mathbf{x}}^R)},$$

$$w_{\mathbf{x}}(s) := k_{\mathbf{y}}^R \alpha_{\mathbf{y}} \frac{s + (2\pi f_{\mathbf{y}}^R)}{s + (\alpha_{\mathbf{y}} 2\pi f_{\mathbf{y}}^R)},$$

$$w_{\mathbf{z}}(s) := k_{\mathbf{z}}^R \alpha_{\mathbf{z}}^2 \frac{s^2 + 4\pi \beta_{\mathbf{z}} f_{\mathbf{z}}^R s + (2\pi f_{\mathbf{z}}^R)^2}{s^2 + 4\pi \nu_{\mathbf{z}} \alpha_{\mathbf{z}} f_{\mathbf{z}}^R s + (\alpha_{\mathbf{z}} 2\pi f_{\mathbf{z}}^R)^2}.$$

$\alpha_{\mathbf{x}}$, $\alpha_{\mathbf{y}}$ and $\alpha_{\mathbf{z}}$ are the ratios between the eigenfrequencies of the zeros and poles and are all set equal to 100. The damping ratios $\beta_{\mathbf{z}}$ and $\nu_{\mathbf{z}}$ of the numerator and denominator respectively of $w_{\mathbf{z}}$ are set to 0.7.

To illustrate the choice of the scaling factors $k_{\mathbf{x}}^R$, $k_{\mathbf{y}}^R$ and $k_{\mathbf{z}}^R$ we consider the SISO design for the $\mathbf{x}$ direction with closed-loop plant

$$M_{\mathbf{x}} := \left(\begin{array}{c|c} M_{11\mathbf{x}} & M_{12\mathbf{x}} \\ \hline M_{21\mathbf{x}} & M_{22\mathbf{x}} \end{array} \right) := - \left(\begin{array}{c|c} W_{1,\mathbf{x}} S_{\mathbf{xx}} V_{1,\mathbf{x}} & W_{1,\mathbf{x}} S_{\mathbf{xx}} G_{\mathbf{xx}} V_{2,\mathbf{x}} \\ \hline W_{2,\mathbf{x}} K_{\mathbf{xx}} S_{\mathbf{xx}} V_{1,\mathbf{x}} & W_{2,\mathbf{x}} K_{\mathbf{x}} S_{\mathbf{xx}} G_{\mathbf{xx}} V_{2,\mathbf{x},} \end{array} \right) \tag{9.5}$$

where $S_{\mathbf{xx}}$, $G_{\mathbf{xx}}$ and $K_{\mathbf{xx}}$ are the sensitivity, system and controller in $\mathbf{x}$-direction and $V_{i,\mathbf{x}}$, $W_{i,\mathbf{x}}$, i.e. are the left upper elements of V_i and W_i, respectively, $i = 1, 2$. $k_{\mathbf{x}}^R$ is chosen such that at the target bandwidth the norms of the first row and second row of $M_{\mathbf{x}}(s)$ in (9.5) have the same order of magnitude. Observe from (9.5) that this is realized if $|W_{1,\mathbf{x}}(2\pi j f_{BW_{\mathbf{x}}})|$ is approximately equal to $|W_{2,\mathbf{x}}(2\pi j f_{BW_{\mathbf{x}}}) K_{\mathbf{xx}}(2\pi j f_{BW_{\mathbf{x}}})|$. Furthermore, since for a SISO system the loop gain is order one in magnitude at the bandwidth, we infer that $|G_{\mathbf{xx}}(2\pi j f_{BW_{\mathbf{x}}})|$ approximately equals $\frac{1}{|K_{\mathbf{xx}}(2\pi j f_{BW_{\mathbf{x}}})|}$. Hence we achieve that $M_{\mathbf{x}}(2\pi j f_{BW_{\mathbf{x}}})$ has rows with approximately equal magnitude if we choose

$$k_{\mathbf{x}}^R = \frac{1}{2} |G_{\mathbf{xx}}(2\pi j f_{BW_{\mathbf{x}}})|.$$

Likewise we choose

$$k_{\mathbf{y}}^R = \frac{1}{2} |G_{\mathbf{yy}}(2\pi j f_{BW_{\mathbf{y}}})|,$$

$$k_{\mathbf{z}}^R = \frac{1}{2} |G_{\mathbf{zz}}(2\pi j f_{BW_{\mathbf{z}}})|.$$

Using similar arguments we choose

$$V_2 = \begin{pmatrix} |G_{\mathbf{xx}}(2\pi j f_{BW_{\mathbf{x}}})|^{-1} & 0 & \\ 0 & |G_{\mathbf{yy}}(2\pi j f_{BW_{\mathbf{y}}})|^{-1} & 0 \\ 0 & 0 & |G_{\mathbf{zz}}(2\pi j f_{BW_{\mathbf{z}}})|^{-1} \end{pmatrix} V_1 \tag{9.6}$$

198

Table 9.3: Parameters of 4-block design parameters.

Parameter	x	y	z
f_{BW}	0.9	0.9	0.48
f^I	0.18	0.18	0.16
α	100	100	100
β	-	-	0.7
ν	-	-	0.7
d	1	4	1

to achieve that the first and second column of $M_{\mathbf{x}}\left(2\pi j f_{BW_{\mathbf{y}}}\right)$ have norms with equal orders of magnitude and using similar considerations for $\mathbf{y}$ and $\mathbf{z}$. Finally, we choose $V_1 = D^{-1}$, to enforce a bound on the peak of the sensitivity of about 2. Indeed since

$$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix} W_1(s)S(s)V_1(s) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = d_{\mathbf{x}}\frac{s+2\pi f_{\mathbf{x}}^I}{2s}S_{\mathbf{xx}}(s)\frac{1}{d_{\mathbf{x}}} = \frac{s+2\pi f_{\mathbf{x}}^I}{2s}S_{\mathbf{xx}}(s),$$

we observe that $\|M_{\mathbf{x}}(i\omega)\| < 1$ for all $\omega \in \mathbb{R}$ implies $\|S_{\mathbf{xx}}(i\omega)\| < 2$, $\omega \in \mathbb{R}$, which is the desired bound on the sensitivity peak.

Observe that the diagonal scaling matrix D is present in W_1 and W_2 and its inverse in V_1 and V_2. As a result, the scaling D does not affect the diagonal entries of M_{ij}, $i, j = 1, 2$.

In Table 9.3 we have collected the parameters, whose choice has been motivated above. By tuning we determined that $D = \mathrm{diag}(1, 4, 1)$ yields a good design.

Remark. The presented weighting functions are diagonal, since nonzero off-diagonal weighting functions are hard to interpret and tune. This chapter illustrates that norm-based MIMO controller synthesis can be done with diagonal weights, which is in our opinion one of the main advantages of this design technique.

9.3.2   Towards a generalized plant

The open-loop generalized performance outputs z_1 and z_2 and control inputs y are related to the generalized disturbances w_1 and w_2 and controls u by

$$\begin{pmatrix} z_1 \\ z_2 \\ \hline y \end{pmatrix} = P \begin{pmatrix} w_1 \\ w_2 \\ \hline u \end{pmatrix}$$

where

$$P = \begin{pmatrix} -W_1V_1 & -W_1GV_2 & -W_1G \\ 0 & 0 & W_2 \\ \hline -V_1 & -GV_2 & -G \end{pmatrix}$$

P is however not a generalized plant, since it is not stabilizable. Indeed the weighting filter W_1 contains unstable poles at $s = 0$, which cannot be stabilized by the controller. One way to overcome this, is to shift these poles or all poles of P slightly into the left-half plane [187, 186]. An often relevant drawback of this approach is that the resulting generalized
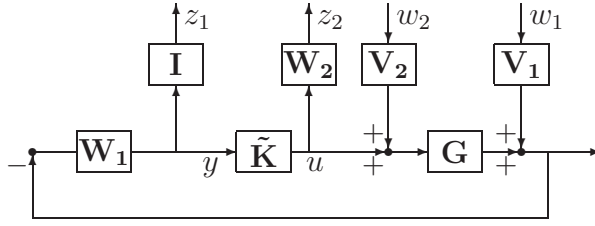
Figure 9.10: Four-block controller set-up with W_1 in the loop.

plant has uncontrollable poles close to the imaginary axis. These poles are then also present in the closed-loop plant and are relatively very close to the imaginary axis if compared to other closed-loop poles. In our numerical experience this makes the fixed-order synthesis prone to large numerical errors that often slow down the convergence rate. Moreover, the LMI-based synthesis algorithms are frequently not applicable to such systems, because the LMI algorithms often fail to determine solutions of feasible subproblems. Scaling of variables may definitely improve the accuracy of the results of the algorithms, but the effect of large differences in eigenvalues cannot easily be alleviated by scaling.

On the other hand the shift of the poles cannot be chosen too large, since the resulting controller will not necessarily satisfy the original specifications, even if the controller poles are shifted back. For this reason it is important to avoid uncontrollable poles very close to the imaginary axis.

An alternative way to enforce that the controller has integral action is based on the well-known *Internal Model Principle*. It has the advantage that it does not require to shift the uncontrollable poles into the left-half plane. In the next section we will see that this makes the fixed-order controller optimization for the wafer stage numerically easier. The poles at $s = 0$ are pulled into the loop [25], as illustrated in Figure 9.10, where the controller K is replaced by $\tilde{K} := KW_1^{-1}$. The resulting plant $\tilde{P}$ can be composed as:

$$
\tilde{P} = \left(\begin{array}{cc|c}
-W_1 V_1 & -W_1 G V_2 & -W_1 G \\
0 & 0 & W_2 \\
\hline
-W_1 V_1 & -W_1 G V_2 & -W_1 G
\end{array} \right)
$$

If G does not have zeros at $s = 0$, it is clear from Figure 9.10 that the poles of W_1 at $s = 0$ can be stabilized by the controller $\tilde{K}$, such that perturbation of these poles is not required to satisfy the stabilizability and detectability conditions for $\mathcal{H}_\infty$-synthesis in Assumption 2.4. After synthesis of $\tilde{K}$, a controller K is reconstructed using $K = \tilde{K} W_1$.

9.3.3 Full-order synthesis

The McMillan degree of the MIMO generalized plant $\tilde{P}$ is 39, where

- 32 orders are in the system

- 3 are due to the integral action in W_1 in (9.3)

- 4 are needed for the roll-offs in W_2 in (9.4)

Indeed, the transfer functions $w_{\mathbf{x}}$ and $w_{\mathbf{y}}$ in W_2 each have McMillan degree 1 and $w_{\mathbf{z}}$ has degree 2.

SISO controllers for all 3 DOFs as well as a single MIMO controller have been designed and their frequency responses are shown in Figure 9.11 in red and blue, respectively.

200

Table 9.4: Closed-loop $\mathcal{H}_\infty$-norms of optimal full order controllers.

SISO/MIMO	Controller	Closed-loop $\mathcal{H}_\infty$-norm
SISO	for **x**-direction	1.956
SISO	for **y**-direction	1.959
SISO	for **z**-direction	1.913
MIMO	diagonal	8.172
MIMO	full	1.962

Table 9.5: The maximum of the real values of the poles of closed-loop with full-order controller for two different design techniques.

direction	Max. real value of closed-loop poles designed	
	by shifting	by internal model principle
x	$-5.65 \cdot 10^{-5}$	-0.0412
y	$-5.65 \cdot 10^{-5}$	-0.0850
z	$-5.65 \cdot 10^{-5}$	-0.0590
MIMO	$-5.65 \cdot 10^{-5}$	-0.0412

The closed-loop $\mathcal{H}_\infty$-norms are shown in Table 9.4, together with that of the MIMO controller constructed by diagonal augmentation of the SISO controllers for the **x**-, **y**- and **z**-directions. The closed-loop $\mathcal{H}_\infty$-norm of the 3×3 controller consisting of diagonal augmentation of the SISO **x**-, **y**- and **z**-controllers is 8.17, denoted by 'diagonal' in the table. This is significantly larger than the performance 1.96 of the MIMO optimized full order controller as shown in the table.

In Table 9.5 the maximum of the real values of the poles of the closed-loop system are given for the two design techniques as discussed above, one by shifting all the eigenvalues of the plant with $\epsilon = -5.7 \cdot 10^{-5}$ and the second with the internal model principle. The minimum of the real values of the poles is for all directions in the order of -10^5. From the table we see that the shifted closed-loops are about a factor 100 closer to the imaginary axis than the closed-loop systems obtained using the internal model principle. This ratio is larger for smaller $\epsilon > 0$. The optimal full-order closed-loop $\mathcal{H}_\infty$-norms appear to be the same, irrespective of the method (shifting or internal model principle) that is applied. However, in Section 9.4.1 we will illustrate that fixed-order controller optimization for a generalized plant based on the Internal Model Principle often yields a controller with a better closed-loop $\mathcal{H}_\infty$-norm than for a plant with shifted eigenvalues.

The controllers are shown in Figure 9.11, where the SISO controllers are depicted in each corresponding diagonal element. The diagonal elements of the SISO and MIMO controllers are similar in the low frequent region, but they differ significantly around 6Hz. The SISO controllers have a PID-like shape of the magnitude, combined with notches and inverse notches at higher frequencies. Especially at higher frequencies the non-diagonal elements of the controller have significant gain, if compared to the diagonal elements.

The closed-loop sensitivities of the MIMO and diagonally augmented SISO controllers are shown in Figure 9.12. The sensitivity peaks of about 2.15 (i.e. 6.5dB) is slightly above the target of 2. The frequency responses of the other three closed loop transfer functions
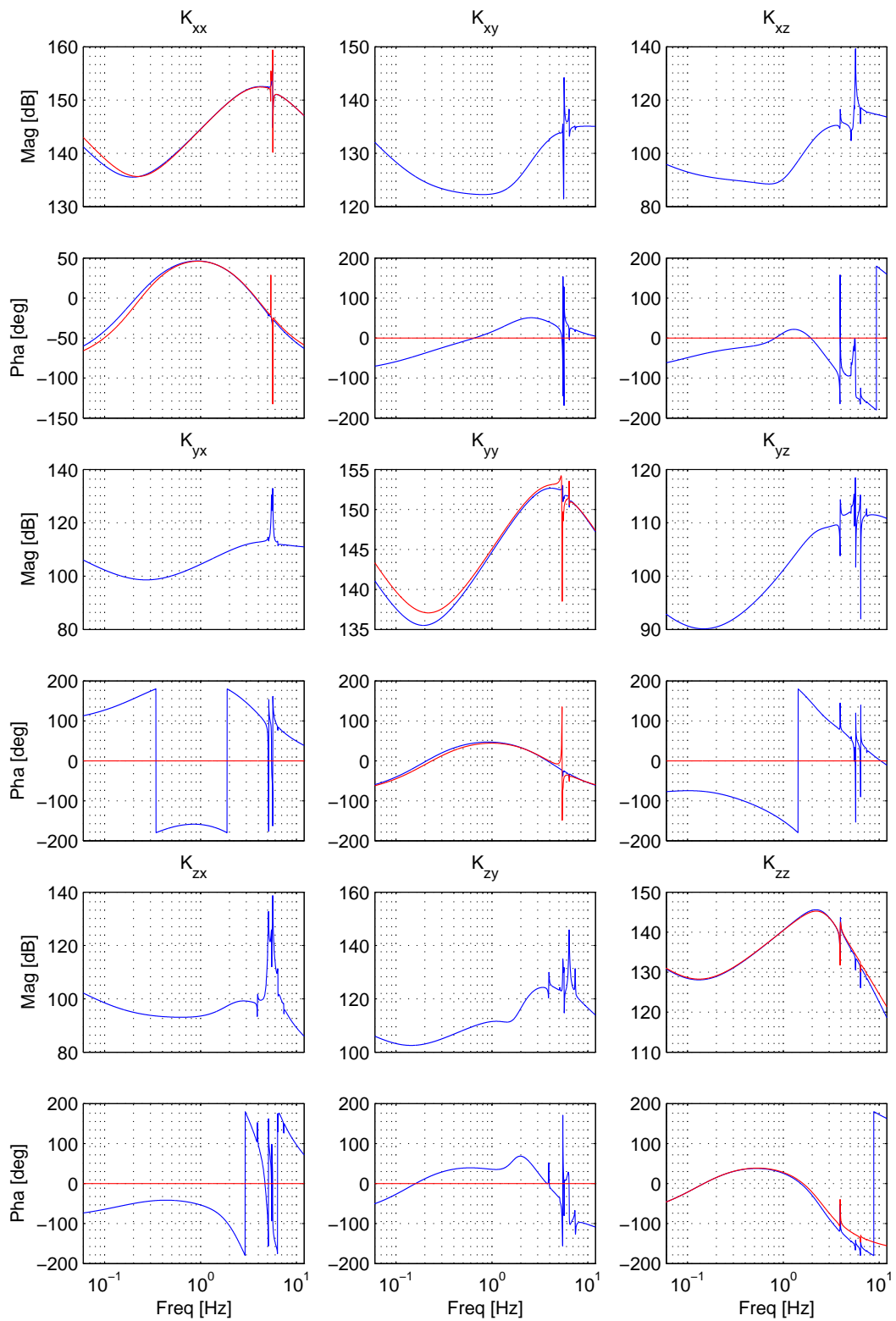
Figure 9.11: Frequency response of full order controllers. Blue: full MIMO controller, Red: diagonally augmented SISO controllers
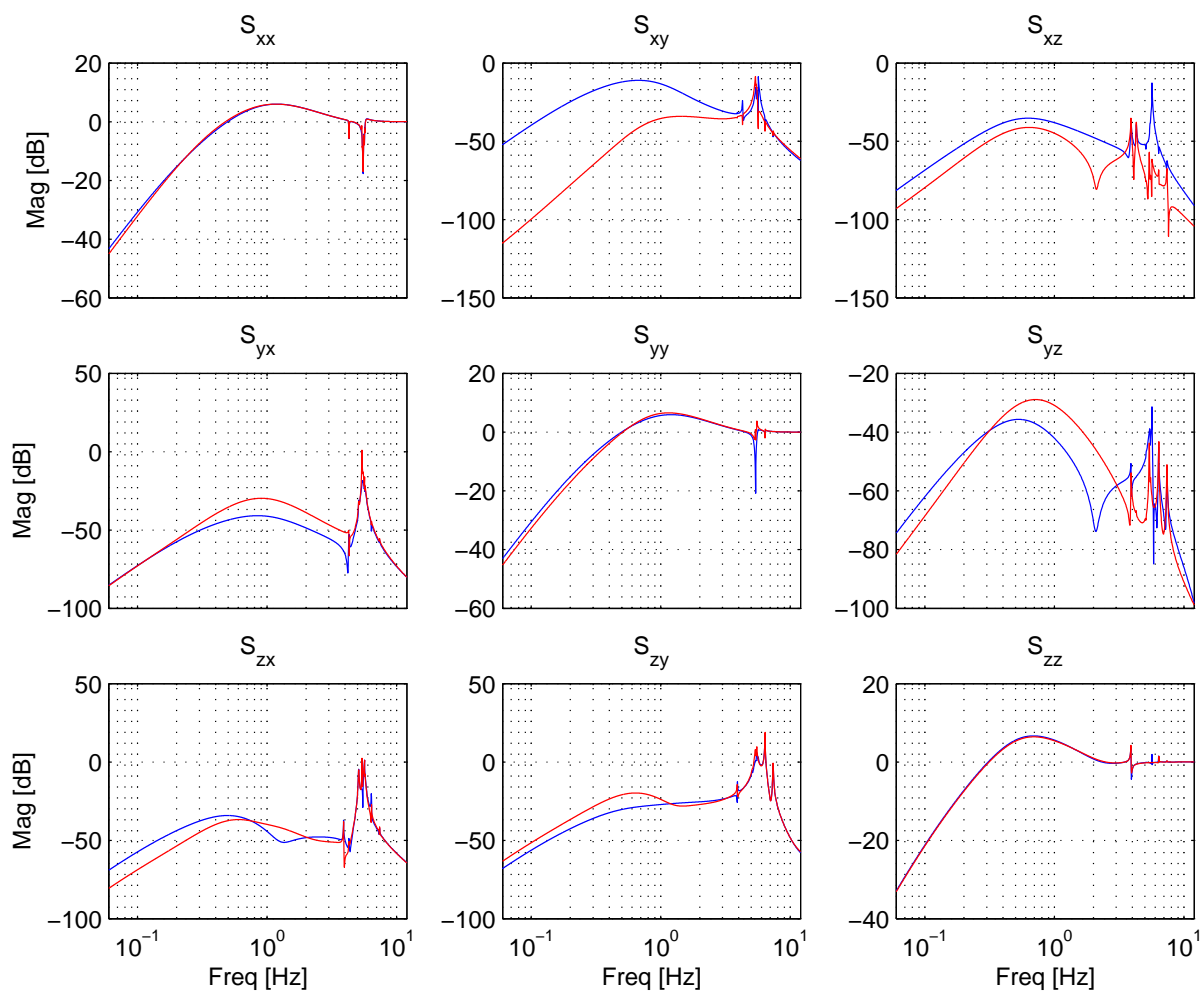
Figure 9.12: Frequency response of the closed-loop sensitivity of the loop with the MIMO plant model and two controllers: the full-order MIMO controller (blue) and diagonally augmented SISO controllers (red).

that are shaped with the four-block design are shown in Appendix C.2.

9.3.4 Unstable controller poles

All full order controllers are unstable due to the integral action. Besides these poles at $s = 0$ the SISO controllers for the directions $\mathbf{x}$ and $\mathbf{y}$ and the MIMO controller have single unstable pole-pairs with real values of 0.359, 0.396 and 0.0211 respectively. Controllers with unstable poles that have large real values are not desired from a practical viewpoint. In [46] an implementation scheme for a CD-player is discussed, of controllers with unstable poles apart from the pole for integral action. However, such controllers are not desirable in commercial wafer stages. We therefore aim at the synthesis of controllers, with unstable poles only due to integral action and with the other poles stable. Observe that the integrators are removed from the controller $\tilde{K}$ described above. This implies that the desired property of the controllers is assured if we search for a stable $\tilde{K}$, i.e. if we solve a fixed-order strong stabilization problem. In most of the fixed-order algorithms discussed in this thesis an extra constraint can be added to enforce stability of the controller poles. In particular this can be done for all algorithms used for optimization in the next section. However, it turned out not to be necessary to enforce these stability constraints explicitly. The optimized controllers are all stable, if we start the algorithms with stable initial controllers. How to obtain stable initial controllers is also explained in the next section.

9.4 Fixed-order controller synthesis

Using controller reduction by balanced closed-loop residualisation, as discussed in Section 3.1.3, we have constructed SISO controllers of McMillan degree 3, 3, 4 for the $\mathbf{x}$-, $\mathbf{y}$- and $\mathbf{z}$-direction respectively. With the same technique, we have reduced the full-order MIMO controller to a McMillan degree of 11. The choice of the degrees is based on a trade-off between McMillan degree and performance. This has been done by evaluation of the closed-loop $\mathcal{H}_\infty$-performance of the reduced controllers of all orders, i.e. static up to full order controllers. This is illustrated for the MIMO controller in Figure 9.13. This figure shows in black the closed-loop $\mathcal{H}_\infty$ norms of the controllers obtained by balanced reduction of the full order controller. The smallest controller order is 3, since if $\tilde{K}$ is static, $K = \tilde{K}W_1$ is a third order controller. The largest controller order is equal to 42. Of these 42 controller states 39 are from the full-order controller $\tilde{K}$ and 3 from W_1. The controllers up to order 11 are not stabilizing, and they are therefore depicted as bars of magnitude 20.

Almost all controllers are unstable. We have obtained stable controllers by eliminating the unstable controller modes in the following fashion. We perform a similarity transformation on a state-space representation of the controller such that

$$A_K = \left(\begin{array}{cc} A_K^{\mathrm{s}} & 0 \\ 0 & A^{\mathrm{u}} \end{array} \right)$$

where A_K^{s} is Hurwitz and truncate the states corresponding to A^{u}. The resulting closed-loop performances of these stable controllers are shown as white bars[3] in Figure 9.13. For

---

[3]To be able to compare the performance of the controllers, we have plotted the performance of the original reduced controller and its performance after truncation of the unstable part next to each other, although the stabilized controllers might have smaller order due to the truncation. Due to this choice of presentation, the stabilized controllers may have smaller McMillan degree than shown in the figure.
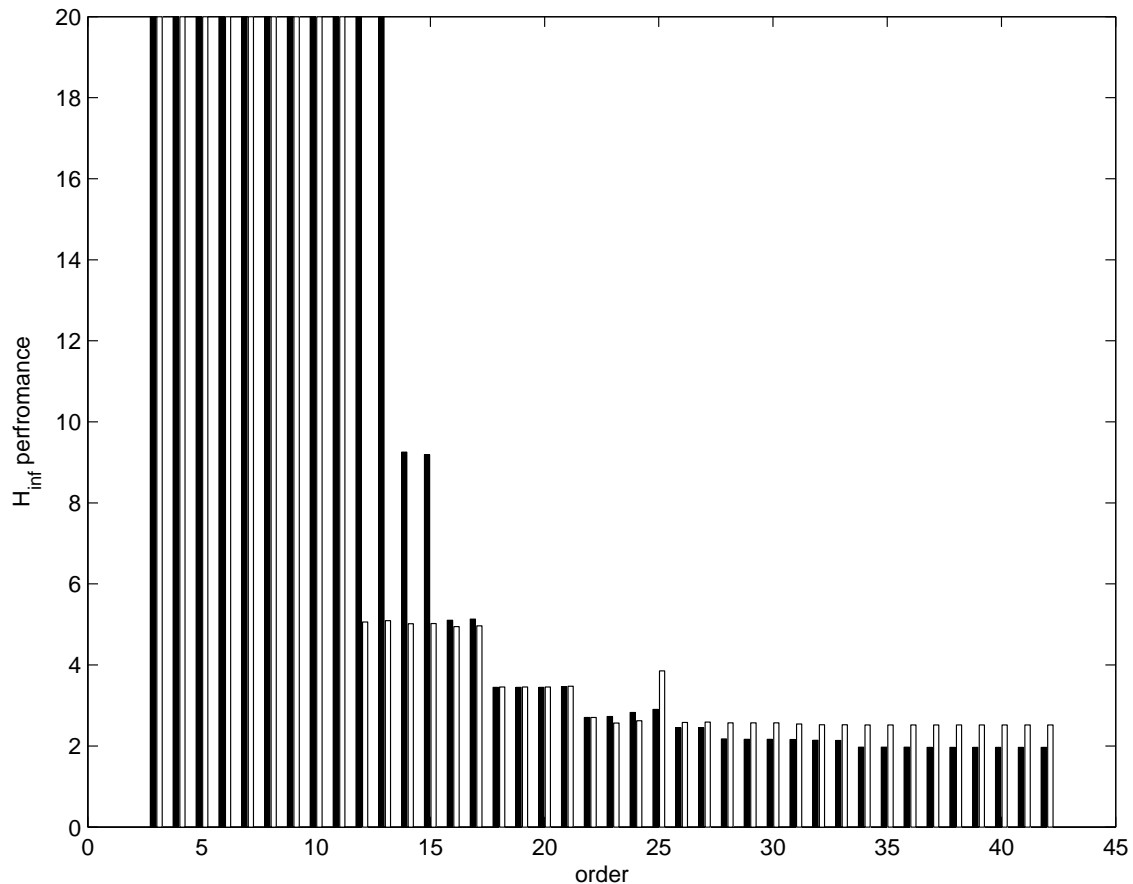
Figure 9.13: Closed-loop $\mathcal{H}_\infty$-norm for reduced MIMO controllers K of order 3 up to 42. Depicted are reduced controllers (black) and reduced and stabilized controllers (white).

the controllers of order 25 or larger, the stabilized controllers have larger closed-loop $\mathcal{H}_\infty$-norm as the original reduced controllers. For several lower order controllers the stabilized controllers appear to have better performance. The closed-loop $\mathcal{H}_\infty$-norm of the controller obtained by truncating the 15$^{\text{th}}$ order controller is 5.02. Since four poles are truncated, its McMillan degree is 11. This controller will be used for further optimization.

Stable SISO controllers of McMillan degree 3, 3, 4 respectively have been computed in a similar fashion. The McMillan degrees of the SISO controllers are the minimal that one can expect, since for a reasonably small closed-loop $\mathcal{H}_\infty$-norm one needs integral action, a lead-lag filter to create stability margin and a 1$^{\text{st}}$ order roll-off for $\mathbf{x}$ and $\mathbf{y}$ and a 2$^{\text{nd}}$ order roll-off for $\mathbf{z}$.

9.4.1 Controller optimization with several algorithms

We have optimized the SISO controllers $\mathbf{x}$ and $\mathbf{y}$ and $\mathbf{z}$ with constraints on their McMillan degree of 3, 3 and 4, respectively. The MIMO controller was restricted to McMillan degree 11. All these controllers have been optimized with the following algorithms:

- the MDS method as described in Section 3.4, denoted in this chapter by 'Simplex algorithm'.

- the Interior Point (IP) algorithm as presented in Chapter 7

- the XK algorithm as discussed in Section 3.2.2

Table 9.6: $\mathcal{H}_\infty$-norms of the closed-loop SISO generalized plants in three directions and of the closed-loop MIMO generalized plant, for all optimized controllers. Also shown are the computation times spent on the controller designs in seconds.

Controller	Full		Reduction		Simplex		Interior Point		XK-iteration	
	$\mathcal{H}_\infty$--norm	comp. time	$\mathcal{H}_\infty$--norm	comp. time	$\mathcal{H}_\infty$ -norm	comp. time	$\mathcal{H}_\infty$--norm	comp. time	$\mathcal{H}_\infty$--norm	comp. time
SISO, **x**	1.956	2.047	2.598	4.703	2.307	12.824	2.117	521.28	2.122	62.625
SISO, **y**	1.959	1.000	2.560	2.219	2.167	16.081	2.089	558.19	2.151	66.39
SISO, **z**	1.913	1.157	6.128	2.484	2.423	23.077	2.099	1013.0	-	-
MIMO	1.964	9.190	5.021	9.562	2.9612	108.16	3.1724	65231	-	-

We have modified the Simplex algorithm to speed up its convergence, as is discussed in Appendix D. All three algorithms have been initialized with the reduced controller. The reduced controller is obtained by closed-loop balanced residualization of a full-order controller with a slightly larger γ than optimal. This improves the quality (measured in terms of the closed-loop $\mathcal{H}_\infty$-norm) of the reduced controller significantly in comparison with the reduced optimal full-order controller. To make the optimization less prone to round-off errors, the state-space representation of both controller and plant have been balanced.

The closed-loop $\mathcal{H}_\infty$ performances of the optimized controllers are given in Table 9.6, together with their computation times. For the SISO controllers the $\mathcal{H}_\infty$-norm of the corresponding closed-loop SISO generalized plant is displayed in the table. For the MIMO controller the $\mathcal{H}_\infty$-norm of the closed-loop MIMO generalized plant is shown in the table.

The closed-loop balanced reduction is the fastest amongst all algorithms for fixed-order controller design. However, the closed-loop performance is worse than for the other algorithms. In particular, the closed-loop $\mathcal{H}_\infty$-norm of 6.13 of the reduced **z**-controller is very large. The final closed-loop $\mathcal{H}_\infty$-norms of the other algorithms are comparable for all 3 SISO controllers and acceptable, if compared to the full order performance. The IP method yields the best (in terms of closed-loop $\mathcal{H}_\infty$-norm) fixed-order SISO controllers and their performances are close to the full order closed-loop $\mathcal{H}_\infty$-norms. The XK-iteration algorithm could not compute a controller for the **z**-direction, because the initial LMI was infeasible. This must probably be attributed to numerical errors in the LMI optimization, since the initial controller is in fact stabilizing.

The IP-method and Simplex method computed MIMO controllers with a closed-loop $\mathcal{H}_\infty$-norm of 3.169 and 3.04 respectively, which is a large reduction if compared with the performance of the initial controller of 5.02. The closed-loop sensitivities of these controllers are shown in Figure 9.14. Apart from a slightly larger peak of the sensitivity of the IP optimized controller in the diagonal elements are the closed-loop sensitivities

For completeness the full, reduced and optimized MIMO controllers are shown in Figure C.1 in Appendix C.1. Figure 9.15 shows the optimized controllers for all orders up to full order. Again, controllers that are not stabilizing are depicted by a bar with $\mathcal{H}_\infty$-performance level 20. The performance of the initial controllers obtained by reduction are also shown in the plot. The plot reveals that the 17th order optimized controller has a good performance, i.e. of 2.4. The figure also shows that the algorithm does not necessarily compute globally optimal controllers, since for instance the closed-loop $\mathcal{H}_\infty$-norm of the 25th order controller is larger than of the 17th order controller.
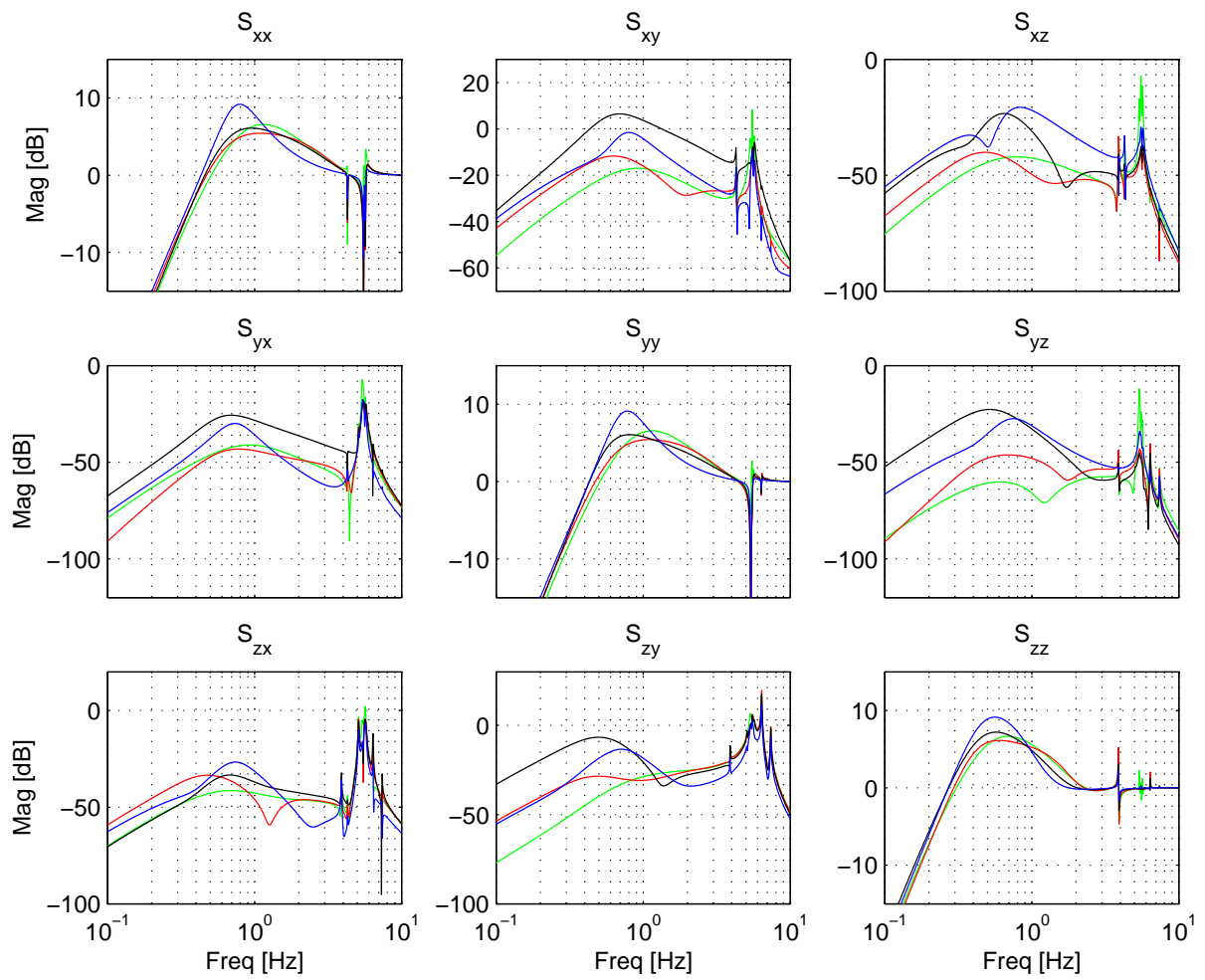
Figure 9.14: Frequency response of closed-loop sensitivity of the MIMO controllers: full order in green, reduced in red, IP optimized in blue and Simplex optimized in black.
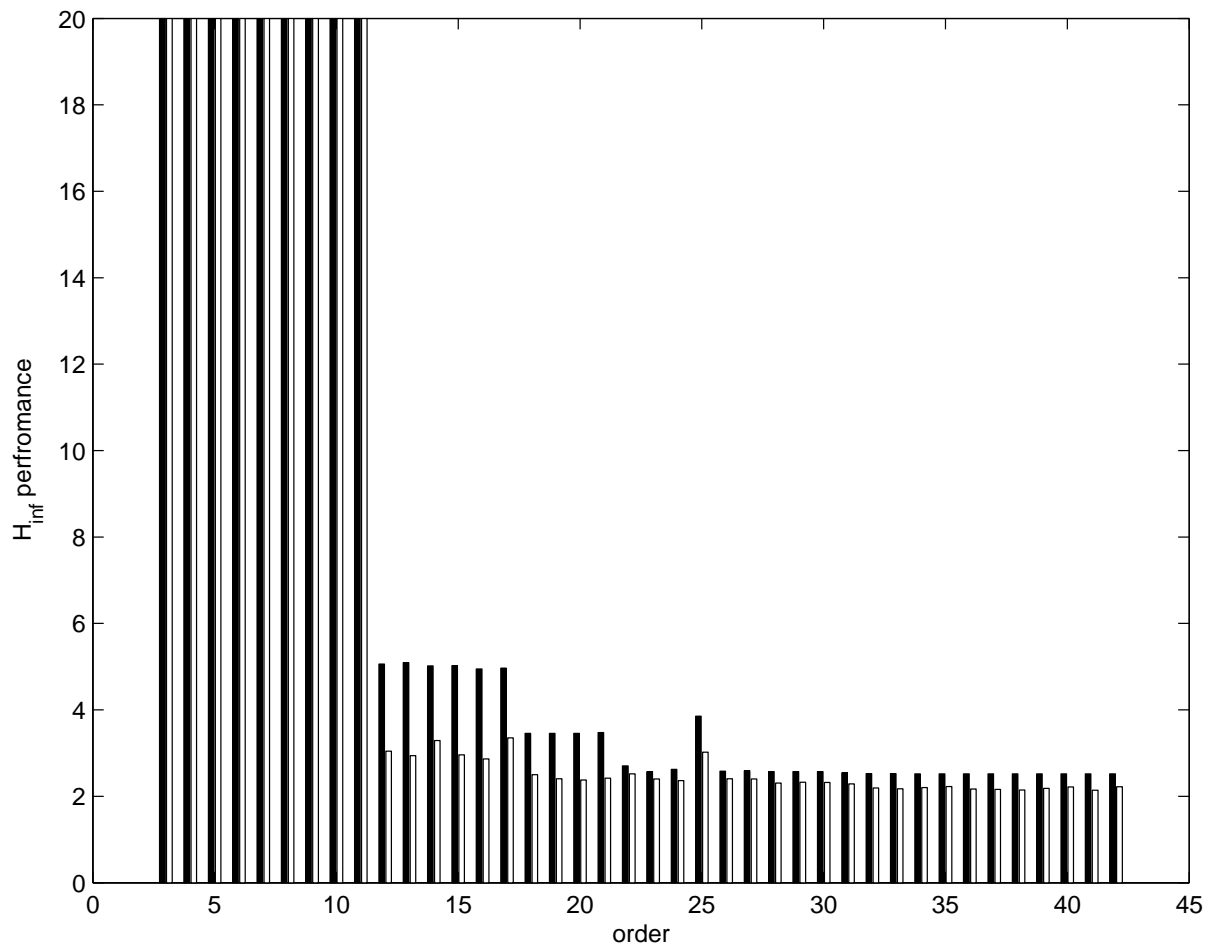
Figure 9.15: Closed-loop $\mathcal{H}_\infty$-norm for reduced MIMO controllers K of order 3 up to 42. Depicted are reduced and stabilized controllers (black) and Simplex optimized controllers (white).

Table 9.7: Comparison of optimization with generalized plant with shifted eigenvalues and with plant based on the Internal Model Principle (IMP): closed-loop $\mathcal{H}_\infty$-norms of optimized controllers and the computation times in seconds for the **x**-direction.

Plant	Full		Reduction		Simplex		Interior Point		XK-iteration	
	$\mathcal{H}_\infty$--norm	comp. time	$\mathcal{H}_\infty$--norm	comp. time	$\mathcal{H}_\infty$-norm	comp. time	$\mathcal{H}_\infty$--norm	comp. time	$\mathcal{H}_\infty$--norm	comp. time
IMP	1.956	1.312	2.598	2.938	2.307	10.328	2.117	533.44	2.122	62.188
Shifting	1.956	1.062	3.727	2.313	2.068	31.625	2.364	293.95	3.723	19.5

The other algorithms could not optimize the MIMO controller, due to limited memory and/or computation speed of the computer.

For the **x**-direction, Table 9.7 illustrates the difference between optimization with the shifted generalized plant and the plant based on the Internal Model Principle (IMP). The results for the shifted plant are the closed-loop $\mathcal{H}_\infty$-norms of the shifted plant $\tilde{P}$, together with the corresponding optimized controller, i.e. the controller has not been back-shifted. The full order performance is approximately the same for both plants; the shift of the poles is apparently small enough. The performance of the reduced controllers is very different: the controllers made with the IMP plant have much smaller closed-loop $\mathcal{H}_\infty$-norm. All optimized controllers with the IMP plant have better performance than those computed with the shifted plant. A notable exception is the controller optimized with the Simplex algorithm. The advantage of using the IMP plant is even more convincingly illustrated with the **y**-direction: the reduced controller is unstable for the shifted plant, such that all optimization algorithms could not be initialized with this controller. These results are therefore not shown in the table.

To illustrate that the Simplex model can also handle higher order models, we have optimized a reduced controller for the four-block generalized plant with the (unreduced) 70$^{\text{th}}$ order model. This plant has McMillan degree 89. The full order controller for this plant has been reduced by closed-loop balanced truncation to a 15$^{\text{th}}$-order controller with a closed-loop $\mathcal{H}_\infty$-norm of 12.7. This controller leading to a closed-loop $\mathcal{H}_\infty$-norm of 3.5 is found by the Simplex algorithm in 499 seconds. Nevertheless, the closed-loop norm is significantly larger than the full order performance of 1.96.

9.4.2 Controllers for implementation

The reduced SISO controllers and those that have been optimized by the Simplex algorithm have been implemented on the set-up. These controllers are shown together with the full-order controller in Figures 9.16, 9.17 and 9.18 for the directions **x**, **y** and **z** respectively. We observe from the figures that the magnitude of the reduced SISO controllers have a PID-like shape, with poles at $s = 0$ and a complex pole-pair.

Table 9.9 shows the achieved bandwidth, gain and phase margin and sensitivity peaks of the reduced and optimized controllers. The bandwidths of the **x** and **y** controllers are below the target bandwidth of 0.9Hz, but seem satisfactory. The bandwidth of the **z** controllers are quite close to the target bandwidth of 0.48. The phase and gain margins are good for all designs. The reduced controller in the **z**-direction has a large sensitivity

Table 9.8: Pole locations of reduced and optimized controllers

Reduced controller			Optimized controller		
x	**y**	**z**	**x**	**y**	**z**
$-16.4 + 20.2j$	$-15.8 + 21.3j$	-27.1	$-17.2 + 18.8j$	$-18.1 + 17.3j$	-21.8
$-16.4 - 20.2j$	$-15.8 - 21.3j$	$-6.41 + 11.8j$	$-17.2 - 18.8j$	$-18.1 - 17.3j$	$-6.61 + 10.1j$
0	0	$-6.41 - 11.8j$	0	0	$-6.61 - 10.1j$
		0			0

Table 9.9: Gain, phase margins of the controllers

Item	Unit	Reduced controller			Optimized controller		
		x	**y**	**z**	**x**	**y**	**z**
Bandwidth	Hz	0.789	0.762	0.492	0.755	0.796	0.462
Gain Margin	dB	8.34	8.72	7.16	8.72	8.21	7.817
Phase Margin	deg	30.2	29.9	30.6	31.3	30.0	27.4
Sensitivity Peak	dB	5.95	5.89	9.63	5.71	6.27	6.97

peak of 9.6, much larger than the target peak of 6dB.

At Philips Applied Technologies it is common practice to evaluate the frequency response of the closed-loop of the controllers with the measured FRF before implementation. With this plot we can evaluate the effects of the modelling errors introduced by the frequency fit and model reduction. The closed loop sensitivities are shown in Figure 9.19 and 9.20 for the reduced and optimized controllers, respectively.

The diagonal elements of the closed-loop system are almost identical. The off-diagonal elements differ much around 0.6Hz. This could be expected from the large difference in the off-diagonal elements of the open-loop FRF and the model around this frequency. The relatively large differences in elements $S_{\mathbf{xz}}$, $S_{\mathbf{yx}}$ and $S_{\mathbf{yz}}$ around 0.6Hz are not so harmful, since the magnitude of the sensitivity is small at that frequency.

In Section 9.2.4 it was mentioned that the resonance at 2.77Hz in $S_{\mathbf{xy}}$ was eliminated in the reduction step. In closed loop this mode yields a peak of -11.4dB of the closed-loop with the FRF, whereas it is -35dB in the model.

9.4.3 Conclusion on fixed-order controller synthesis

With the IP algorithm, the Simplex method and the XK-iteration we have computed fixed-order SISO controllers with satisfactory closed-loop $\mathcal{H}_\infty$-norms. The algorithms can compute SISO controllers with satisfactory performance, except for the computation of a **z**-controller by XK-iteration. The SISO controllers have been optimized by the IP algorithm to values that are within 103% of the full-order closed-loop $\mathcal{H}_\infty$-norm, as is clear from Table 9.6.

The IP algorithm has optimized the full MIMO controller of McMillan degree 11 to a performance of 3.17, where the initial controller has closed-loop $\mathcal{H}_\infty$-norm of 5.021. The generalized plant in this fixed-order control problem is 39. In our experience the IP
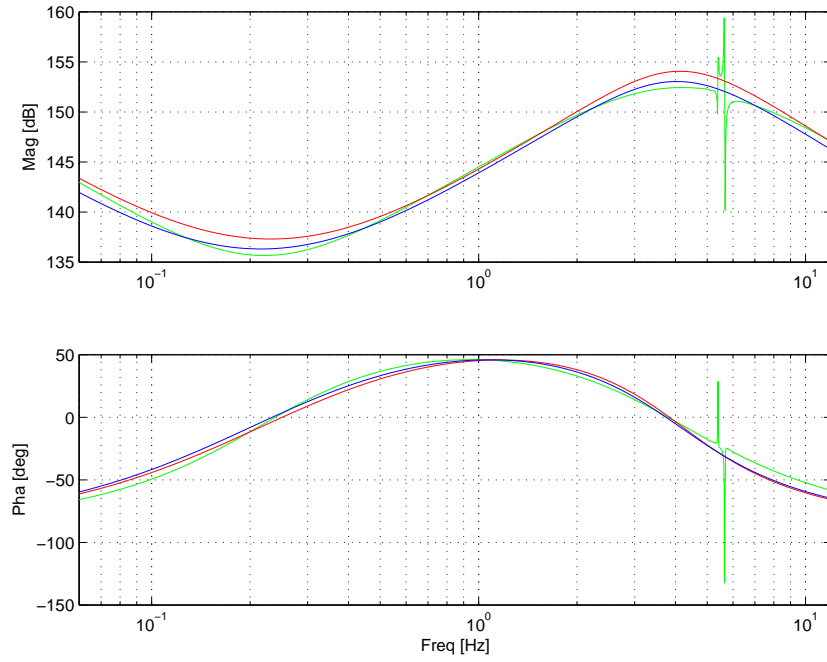
Figure 9.16: Frequency response of the controllers in **x** direction: full order in green, reduced in red and fixed-order optimized in blue.
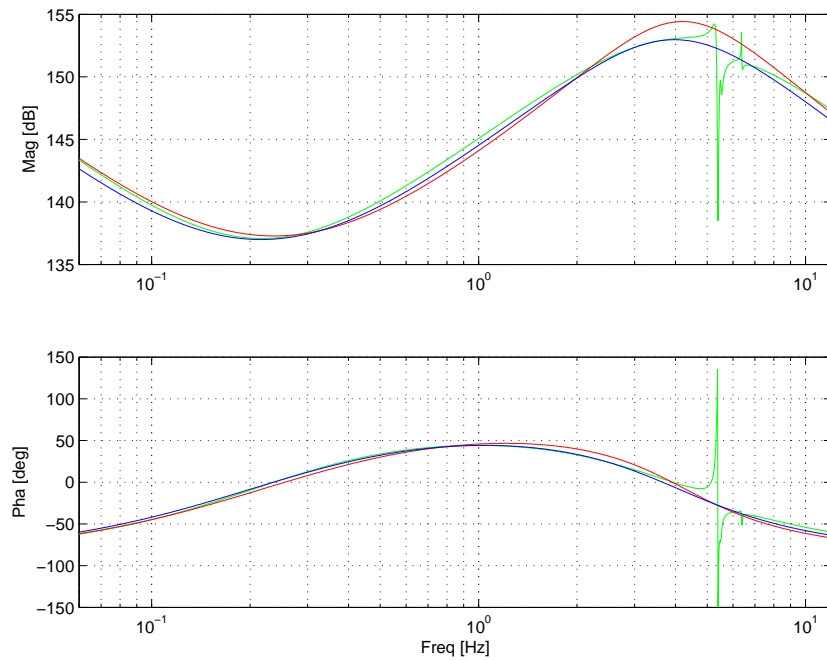
Figure 9.17: Frequency response of the controllers in **y** direction: full order in green, reduced in red and fixed-order optimized in blue.
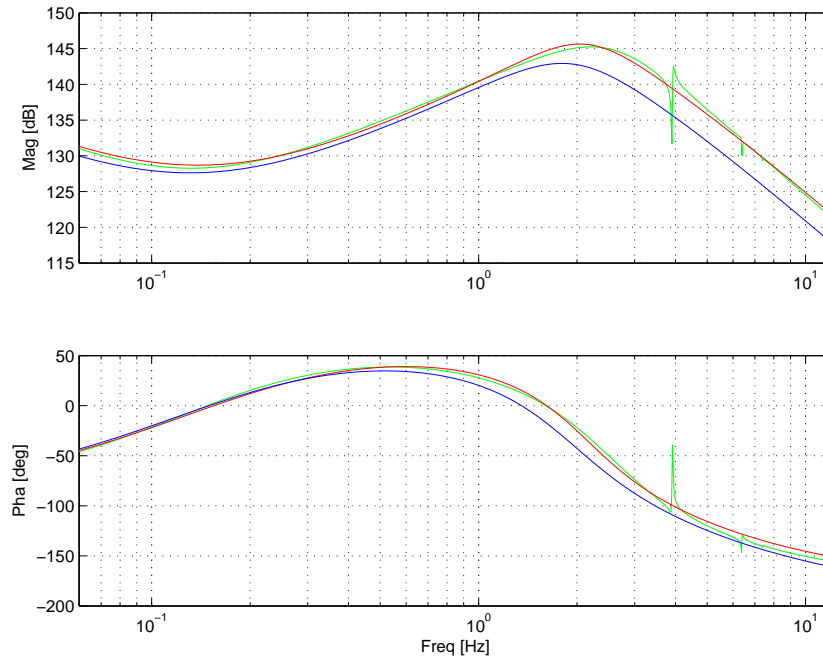
Figure 9.18: Frequency response of the controllers in **z** direction: full order in green, reduced in red and fixed-order optimized in blue.

algorithm is successful for plants with McMillan degree up to about $40 - 50$, since its computation time gets rather long. The XK-iteration was not able to compute MIMO controllers within reasonable computation time. The Simplex algorithm has optimized the full, low-order MIMO controller within $173s$. Although the controllers optimized by the IP and Simplex method have a much better closed-loop $\mathcal{H}_\infty$-norm than the initial controller obtained by closed-loop reduction, their sensitivity peak is rather large.

The computation times of the Simplex algorithm are in our experience short enough to use this algorithm in an iterative design procedure for practical controller design. Tuning the weighting functions of the generalized plant, doing a full order $\mathcal{H}_\infty$-optimal controller design, reduction and optimization can indeed be performed within half a minute for the SISO plants discussed in the chapter and within 4 minutes for the 39th order MIMO plant. Once the weighting functions are fixed, we suggest to use the IP algorithm to compute the final controller.

The closed-loop sensitivity of the optimized controller is quite good, except for the peak at element $S_{\mathbf{zy}}$ at 5.57Hz of 6.33dB. This peak is present in the closed loops for both the model and the FRF. The reduced controller has a large sensitivity peak in $S_{\mathbf{zz}}$ at 3.89Hz of 10.2dB. The optimized controller has a much better response at this frequency: 4.95dB. We will see in the next section that this peak causes the bad time-domain performance of the reduced controller.

9.5 Controller implementation

In this section we evaluate the performances of the reduced and optimized controllers on the basis of experiments.
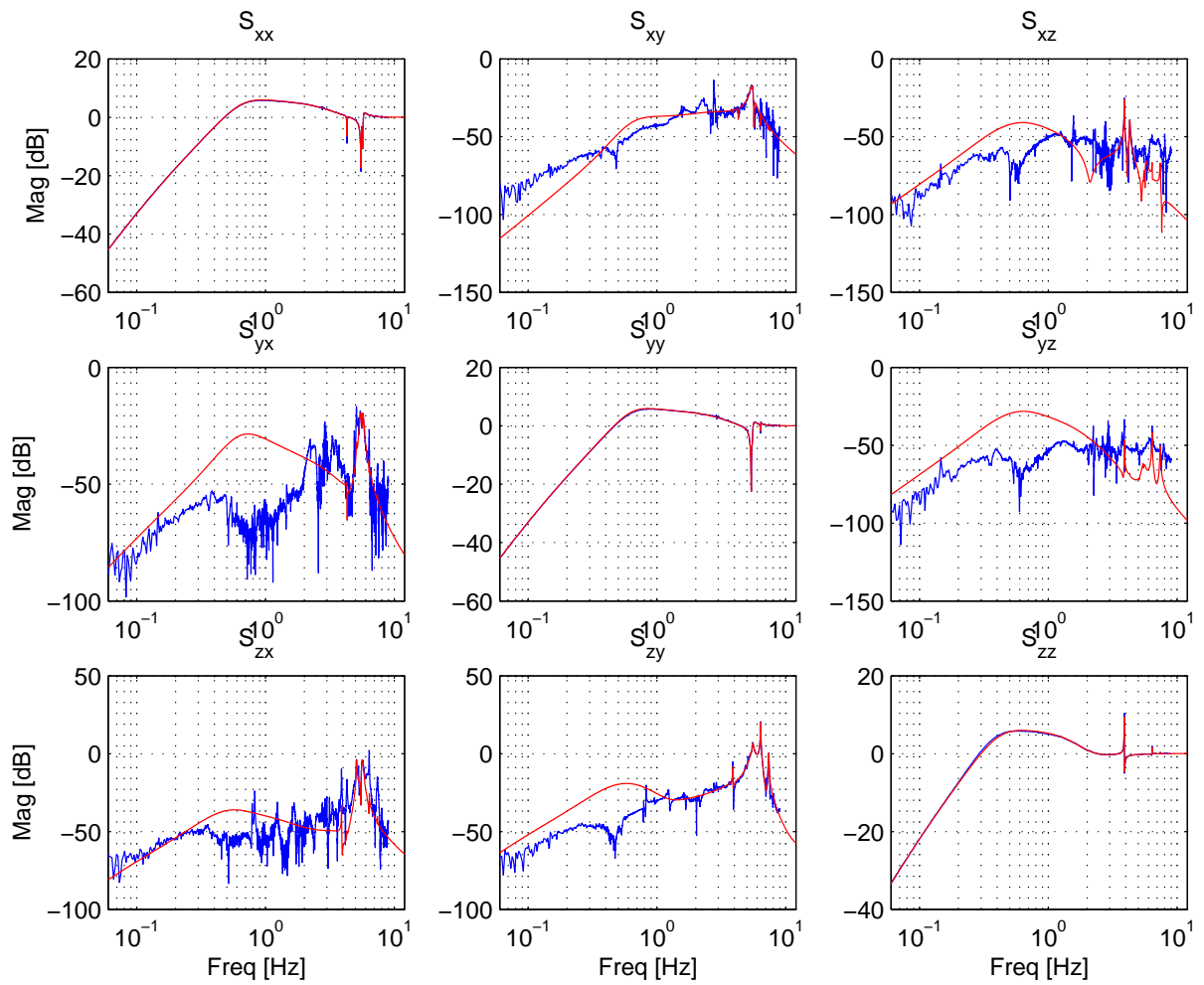
Figure 9.19: Closed-loop sensitivity of the loop with the reduced diagonal controller and the model (blue) and the FRF (red).
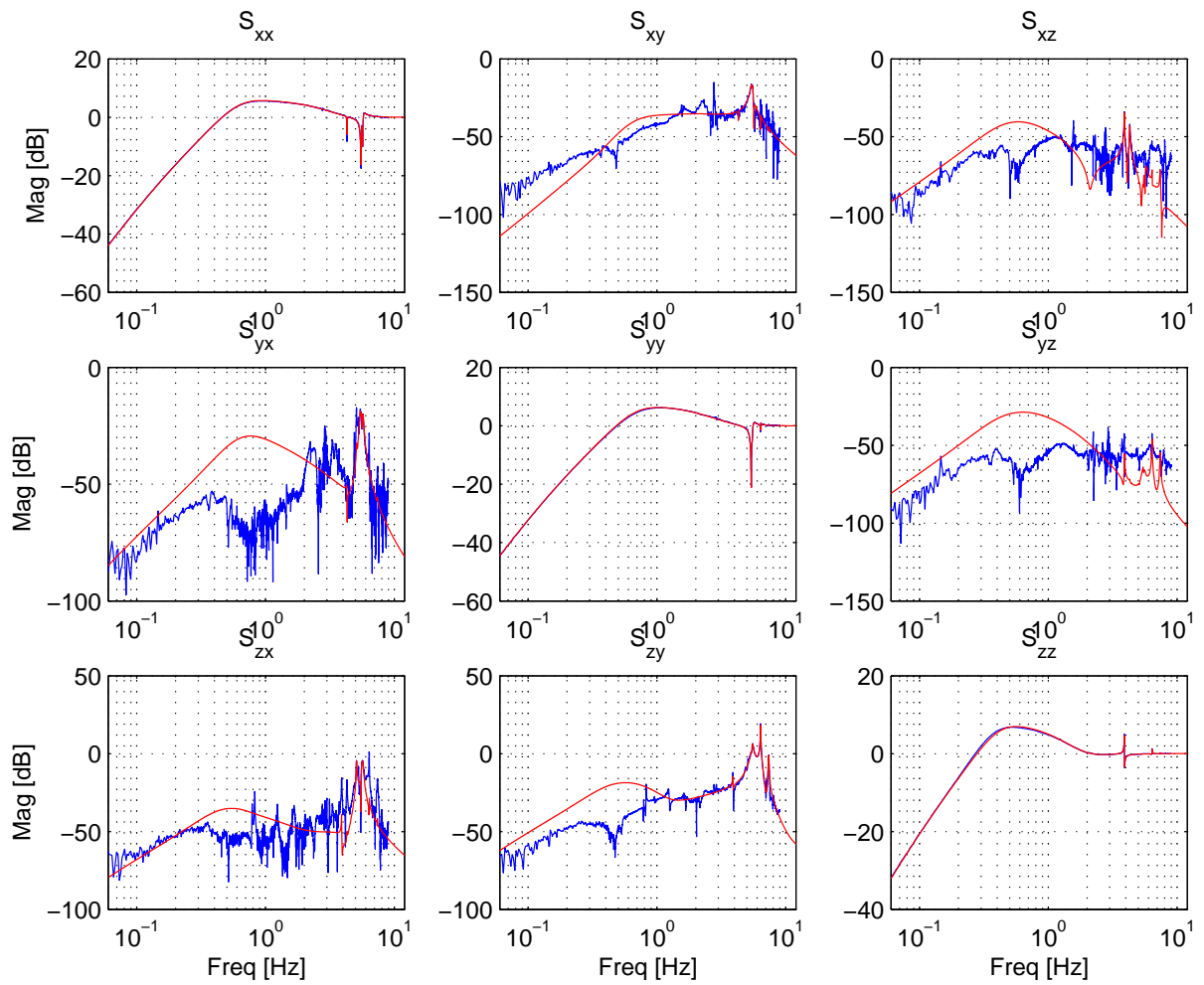
Figure 9.20: Closed-loop sensitivity of the loop with the optimized diagonal controller and the model (blue) and the FRF (red).

9.5.1 Standstill experiments

Measurements have been obtained of the closed loop systems during 'standstill', i.e. when the reference signals are constant in time. This is done at two positions of the wafer chuck: the central position and a position at the left upper corner, as illustrated in Figure 9.21. This latter position is at the edge of the required operating range. The error signal e
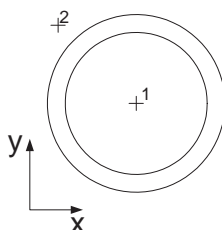
Figure 9.21: Positions on the wafer chuck where experiments have been performed.

(see Figure 9.4) has been measured for 547 sec. For a representative time interval within this period the standstill errors of the three translations **x**, **y** and **z** are shown in Figures 9.22-9.24.
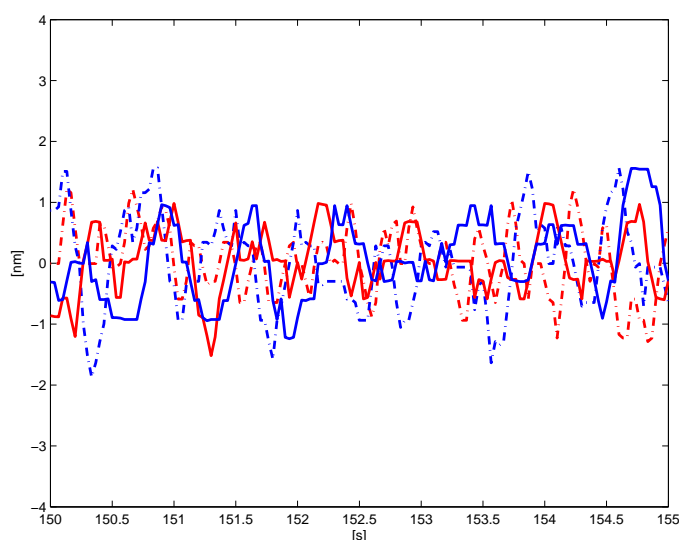
Figure 9.22: Time response of the error in the **x** direction for $t \in [150, 155]$ sec for reduced controller (red) and optimal controller (blue), at center (solid) and the left upper corner (dash-dotted) of the wafer chuck.

The errors in the **x**- and **y**-directions are small. The resolution of the interferometers is even visible in the Figures 9.22 and 9.23. Figure 9.24 reveals that the **z**-errors are much larger for the reduced controller at the corner position. The maximum and the standard deviation of the **z**-error is given in Table 9.10. The errors of both controllers at the central position are satisfactory for the lithographic process. For the corner position the difference between the optimal and reduced controllers is more pronounced.

The error of the closed loop with the reduced controller is much larger than of the optimal controller. Based on Figure 9.24 we have estimated the period of the frequency of
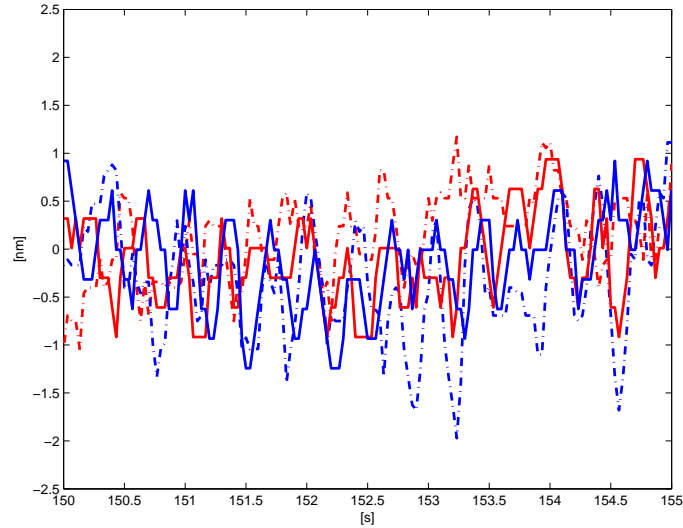
Figure 9.23: Time response of the error in the **y** direction for $t \in [150, 155]$ sec for reduced controller (red) and optimal controller (blue), at center (solid) and the left upper corner (dash-dotted) of the wafer chuck.
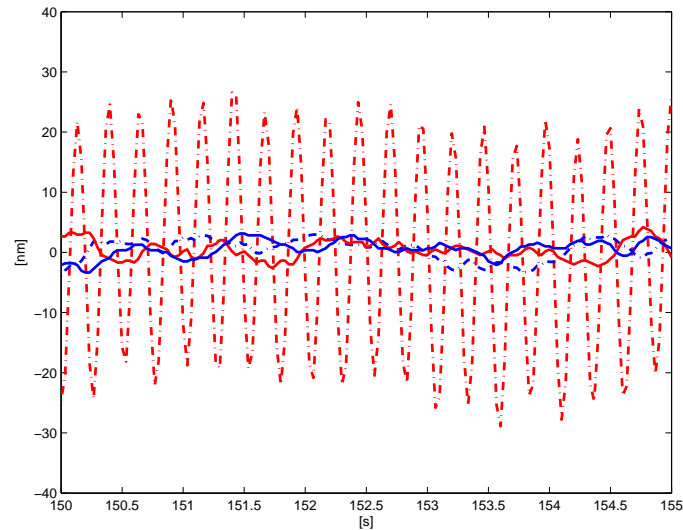
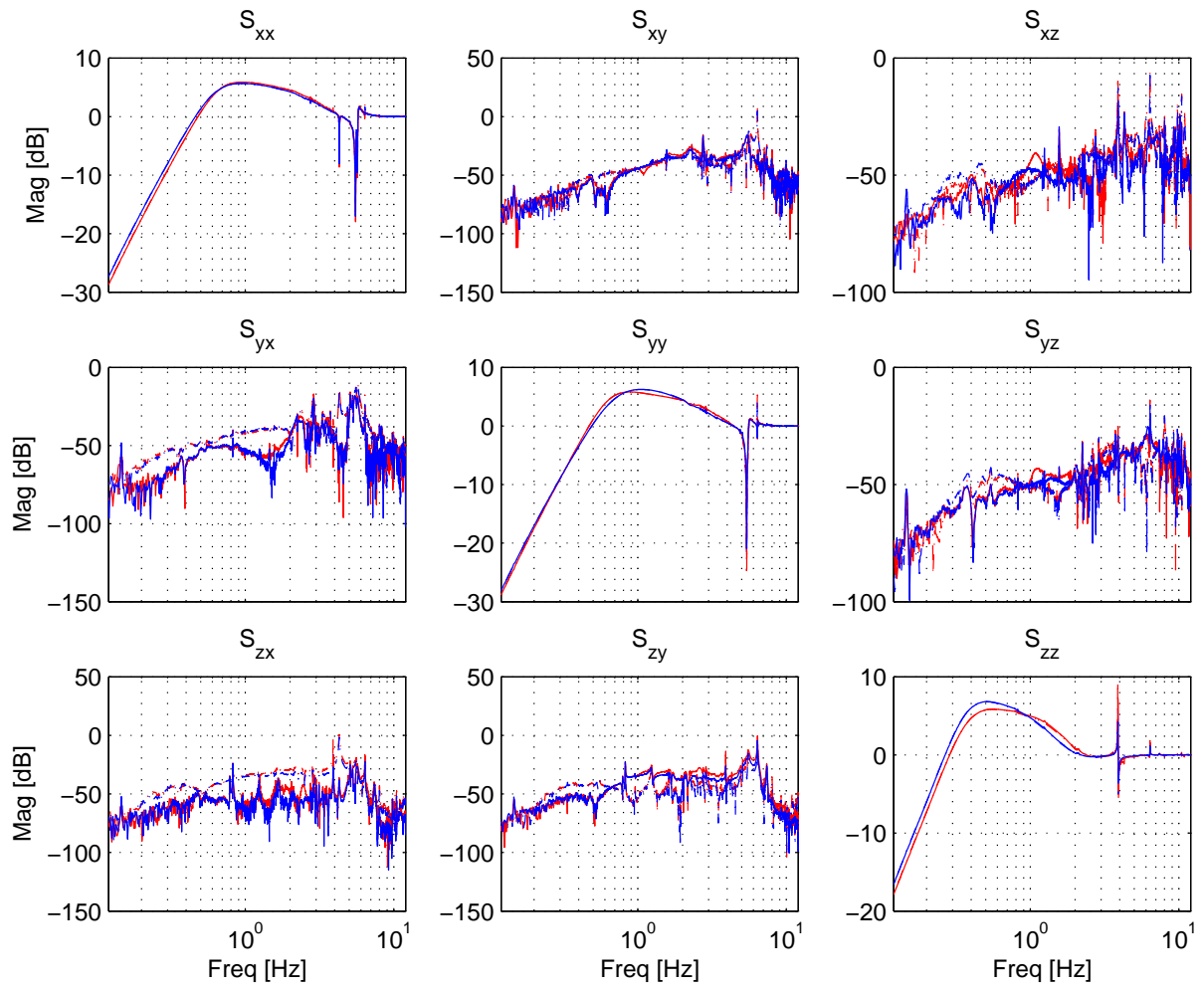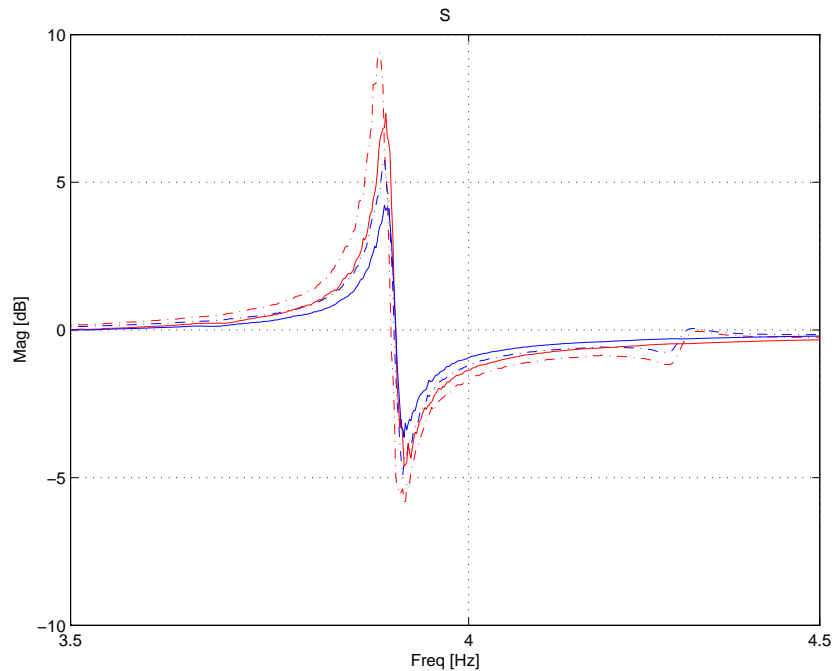Figure 9.24: Time response of the error in the **z** direction for $t \in [150, 155]$ sec for reduced controller (red) and optimal controller (blue), at center (solid) and the left upper corner (dash-dotted) of the wafer chuck.

Table 9.10: Maximum of absolute values $\|e\|_{L_\infty}$ and standard deviation of error signal e in **z**-direction.

direction	Center position		Corner position	
	$\|e\|_{L_\infty}$	standard deviation	$\|e\|_{L_\infty}$	standard deviation
Reduced	7.20	1.79	30.1	15.8
Optimized	7.20	1.91	8.24	1.98

dominant sinusoid in the error signal to be 3.93Hz, for the experiment with the reduced controller. This is very close to the frequency 3.89Hz of the sensitivity peak of the reduced controller, as depicted in Figure 9.19 for the center position. The response is better than was expected based on this figure. Probably, the smaller error must be attributed to a slight change in dynamics of the system, in the time between the identification, performed on February 25th 2005 and the closed-loop experiments performed on April 1st 2005. This is confirmed by Figure 9.28, which shows the frequency response of the sensitivity measured on April 1st 2005, with the reduced controller in the loop.

The contribution of the frequency 3.89Hz to the cumulative spectral density at the corner position is clearly visible in **z**-direction, as shown in Figure 9.25. In the directions **x** and **y** we see large contributions to the variances at 1.55Hz and 2.84Hz respectively.
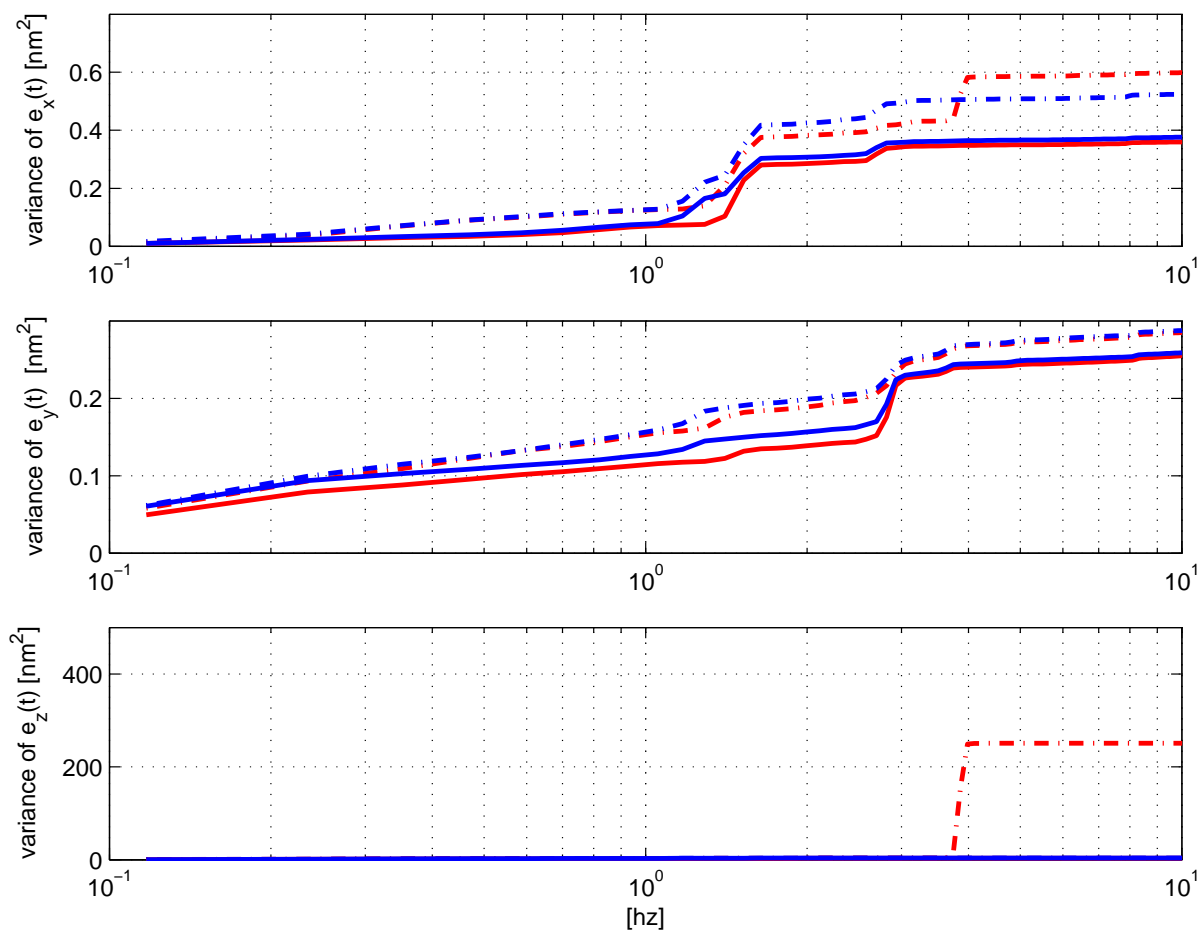
Figure 9.25: Cumulative spectral densities of the errors in the three translations **x**, **y** and **z**, at center (solid) and the left upper corner (dash-dotted) of the wafer chuck for reduced controller (red) and optimal controller (blue).

The measured input sensitivity as depicted in Figure 9.26 illustrates that at the corner position the **z**-direction is very sensitive to disturbances around frequency 3.9Hz. The difference is more clearly visible in Figure 9.27, which shows the frequency response of the **z**-direction only, for frequencies between 3.6Hz and 4.2Hz. It is shown in Table 9.11, that the amplitude of the sensitivity of the peak around that frequency is 9.40dB (at 3.89Hz), as opposed to 5.80dB (at 3.89Hz) of the optimal controller. This illustrates once more that

Figure 9.26: Measured input sensitivity, reduced controller (red) and optimal controller (blue), at center (solid) and the left upper corner (dash-dotted) of the wafer chuck.

Table 9.11: Peak value of the measured sensitivity between 3.6Hz and 4.2Hz for the **z**-direction.

Position	Controller	Peak [dB]	Frequency [Hz]
Center	Reduced	7.33	3.89
Center	Optimized	4.22	3.89
Corner	Reduced	9.40	3.88
Corner	Optimized	5.80	3.89

Figure 9.27: Measured input sensitivity, reduced controller (red) and optimal controller (blue), at center (solid) and the left upper corner (dash-dotted) of the wafer chuck.

the large sensitivity of the reduced controller at this frequency results in unsatisfactory closed-loop behavior.

Based on Figure 9.19 we would have expected a larger cumulative error at the central position at 3.9Hz. As already explained, this must be attributed to a change in dynamics.

9.5.2 Conclusions on experiments

The sensitivity peak around 3.89Hz of the reduced controller causes large errors during standstill. By fixed-order optimization we were able to improve the controller's sensitivity at this frequency. This controller has better time-domain performance at the corner position of the system.

The performance can be further reduced by controller re-design. One method is to incorporate approximations of the spectra of the errors in the weighting filters, see e.g. [199]. This is not pursued, since the objective of the experiments is to obtain a fair comparison of the controllers, based on the same generalized plant.

9.6 Discussion

In this chapter we have presented the application of the structured controller synthesis algorithms to very recently developed high-tech servo-system. The results illustrate that structured controller synthesis for industrial servo-systems is feasible with the algorithms presented in this thesis, although there is still room for improvement in terms of computation time and convergence for the MIMO controller designs. This concludes our contributions to the last research question ③ in our problem definition given in Section 2.8.
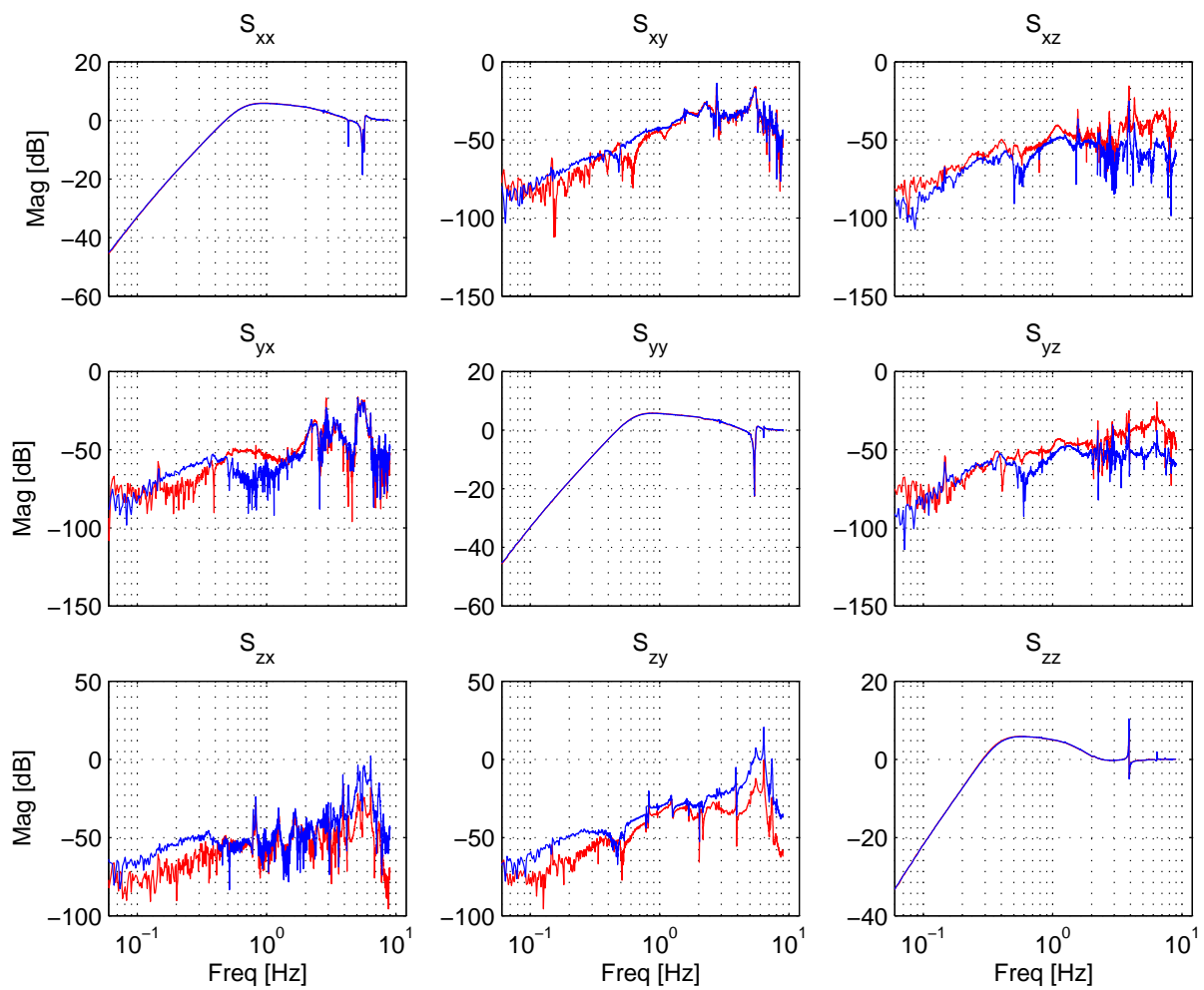
Figure 9.28: Measured input sensitivity on April 1st 2005 (red) and sensitivity of FRF obtained at February 25th 2005 (blue), both with the reduced controller and at center position of the wafer chuck.
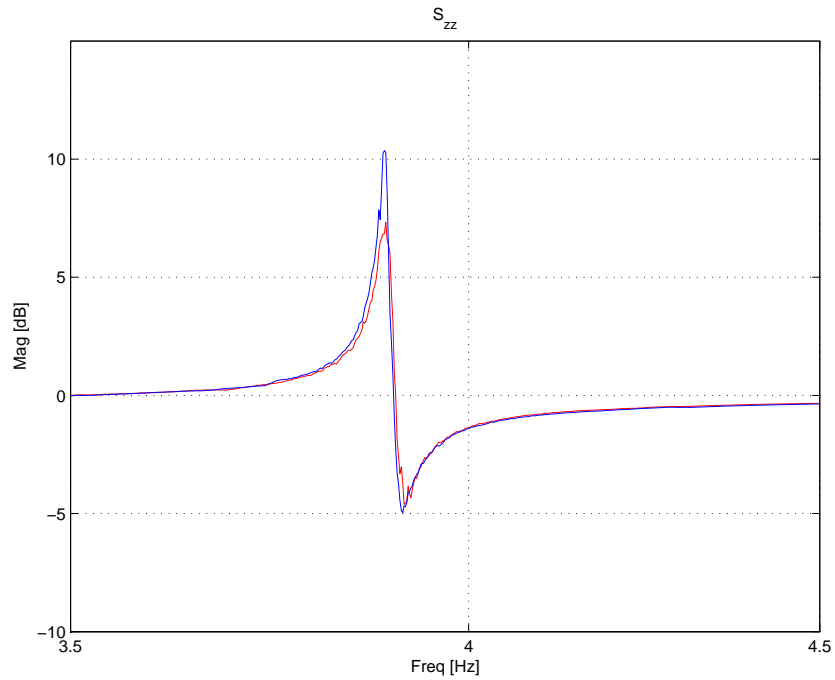
Figure 9.29: Measured input sensitivity at April 1st 2005 (red) and sensitivity of FRF obtained at February 25th 2005 (blue), both with the reduced controller and at center position of the wafer chuck

The controller in this chapter illustrates the benefit of closed-loop optimization to improve the performance of controllers obtained by reduction of full order $\mathcal{H}_\infty$-suboptimal controllers. Referring to research question ③ on page 40 we conclude that the algorithms presented in this thesis can synthesize practically relevant controllers. Furthermore this chapter has revealed that in practice the McMillan degree of the optimized controllers can be much smaller than that of a full-order controller without severe degradation loss. Indeed, from Table 9.6 we conclude that the performance degradation of the SISO Interior Point optimized controllers is only 3%,if compared to the performance of the full-order controller. Since such small performance degradation will often be in the order of the modelling error, we conclude that optimized controllers with significantly lower orders than the plant are a feasible alternative to full-order control, in the sense that the actual closed-loop performance on the real system will be comparable.

Chapter 10

Conclusions and Recommendations

In this thesis we have made several important steps towards the solution of the structured $\mathcal{H}_\infty$-optimal controllers synthesis problem. New algorithms have been developed and local and global certificates have been derived. Furthermore, we have shown that with the developed algorithms we can design low order controllers for industrial servo-systems with a closed-loop performance that is very close to that of a full-order controller. In Section 10.1 we summarize the main contributions of this thesis, give some remarks on relations between the chapters and outlooks on the implications of the results to other control problems. In Section 10.2 we conclude this thesis with some recommendations for future research.

10.1   Conclusions of this thesis

The main contributions of this thesis can be summarized as follows:

① The Interior Point algorithm as presented in Chapter 7 has local convergence guarantees. With the proposed computation of the Newton steps through the solution of Sylvester equations, the computation time can be significantly decreased for high-order plants. The convergence of the algorithm is improved if the reduced parametrization presented in Chapter 8 is applied.

② Sum-Of-Squares relaxations families for two hard problems in control theory have been constructed: the structured controller synthesis problem (or more general a polynomial SDP) and the robust analysis problem (or more precisely a robust SDP). Local first and second order optimality conditions for polynomial SDPs have been derived as well.

③ We have shown that the IP algorithm is suited for MIMO and SISO fixed-order controller design in industrial practice. This is illustrated with the controller designs for the active suspension system and the wafer stage in Chapters 7 and 9, respectively. The resulting closed-loop performance of the low-order controller that have been computed by the IP algorithm is very close to the full-order performance. The performance in real-time experiments is much better than of the controllers obtained by controller reduction.

These conclusion will be worked out in some more detail in the following sections.

10.1.1   Optimality certificates

Global certificates

Global optimality certificates through Sum-Of-Squares relaxation have been developed in Chapters 4 and 6. These strongly rely on LMI duality, and can themselves be considered as generalizations of LMI duality to polynomial SDPs and robust polynomial SDPs. Based on these results, relaxation families for two hard problems in control theory have been constructed: the structured controller synthesis problem and the robust analysis problems (as presented in Chapter 6).

By converting the fixed-order control problem into a robust analysis problem using partial dualization, we have constructed a relaxation scheme that is applicable to a practical fixed-order controller synthesis problem, such as for the active suspension system. It has been explained how a controller can be extracted under certain conditions. Due to the strength of the certainty induced by global certificates and extraction of corresponding solutions, it is expected that these techniques will play a major role in future developments on structured controller synthesis and robust control. This will be stimulated if progress is made to faster computers, better LMI solvers and further exploitation of the control-theoretic characteristics.

Local certificates

Computationally less demanding local optimality certificates have been derived in Section 7.3 for the BMI formulation of the $\mathcal{H}_\infty$-optimal control problem. This analysis is also based on LMI duality. With these conditions one can asses the local optimality of controllers using relatively small-sized LMI problems. Furthermore, the conditions can directly be applied to other polynomial SDPs. The reduced parametrization presented in Chapter 8 enables to assess strict local optimality.

10.1.2   Convergent algorithms

In Chapter 7 an Interior Point method has been presented for structured controller synthesis. This algorithm has local convergence guarantees. The control-theoretic characteristics of the underlying problem have been exploited to obtain a Newton step computation based on Sylvester equations. This improves the computation speed and combines the classical methods based on Lyapunov equations with non-convex Interior Point optimization.

The method has successfully been applied to design a structured order controller for a 27th order generalized plant of an active suspension system.

10.1.3   Controller design for industrial servo-systems

Applicability of the presented algorithms to industrial servo-systems has been illustrated with the wafer stage controller design in Chapter 9 and with the active suspension system in Chapter 7. It has been illustrated that the IP algorithm is able to design fixed-order SISO and MIMO controllers for generalized plants up to about McMillan degree 40. The experience with these practical control problems have shown that the accuracy of the solutions are sensitive to the numerical conditioning and size of the problem data. This sensitivity often hampers the convergence of the algorithms. Effort has been made in this thesis to alleviate the effects of the round-off errors and other numerical artifacts by

- balancing the state-space realizations of the system and the controller

- using Newton-step based algorithms like the Interior-Point algorithm in Chapter 7; this makes the algorithm less sensitive to scaling of the decision variables

- avoiding vectorization as discussed in Section 7.2; the Sylvester equations are computationally much cheaper than the Kronecker formulation and produce in our experience more accurate results

- using a reduced parametrization of controllers in state-space, as has been presented in Chapter 8; by an example it has been illustrated in that chapter that this may improve the convergence

- avoiding uncontrollable modes in the generalized plant with poles close to the imaginary axis, as discussed in Chapter 9

The control designs have also illustrated that for systems with high McMillan degree several algorithms are computationally too expensive. For these systems a Simplex algorithm has been successfully applied to a plant of McMillan degree 89 in Chapter 9. The price to be paid for the smaller computation time is the weak convergence guarantees, if compared to more sophisticated methods.

10.2   Recommendations for further research

Choice of monomial bases

The relaxation gap of the lower bounds described in Chapters 4 and 6 depends on the choice of the monomial bases. The computational complexity of the SOS relaxations could be reduced, if one can a priori determine which monomials are important to include in the SOS bases. A more sophisticated procedure to select monomials based on their expected impact on the quality of the lower bounds would, therefore, be an important step towards more efficient SOS relaxations.

Exploit control-theoretic characteristics further

In Section 7.2.4 it has been explained how the *control-theoretic characteristics* (as defined in Section 2.8) of the problem are exploited to compute the trust region step using a set of Sylvester equations. The number of Sylvester equations depends on the number of generalized disturbance variables m_1 and on the number of controller variables. We think that this is not the end of the story, i.e. it might be possible to reduce the computational complexity of the trust-region step even more by exploiting the control-theoretic characteristics. This may result in even a better merging of the Interior Point method with classical methods based on Lyapunov equations.

Complexity

It has been discussed in Section 2.6 that the fixed-order $\mathcal{H}_\infty$-optimal control problem, without a priori bounds on the controller parameters, is still open. This is a fascinating question that should be solved.

Robust structured controller design for mechanical servo-systems

The algorithms for structured controller synthesis in this thesis are applicable to nominal $\mathcal{H}_\infty$-optimal controller design. It is often important to be able to design controllers that guarantee robust closed-loop performance against dynamic or parametric uncertainty. Although some ideas on the solution of this synthesis problem can be extracted from the results in this thesis, we have not developed an algorithm to solve this problem. Algorithms for this have been addressed literature, but to the best of our knowledge there does not exits an algorithm to efficiently perform a robust, structured controller synthesis for high-order models as the waferstage model in Chapter 9.

Direct controller synthesis on frequency response data

To design fixed-order controllers with the algorithms discussed in this thesis, a state-space model of the system is required. Such a model can be identified based on an experimentally obtained frequency response function, as for instance described in Section 9.2. To speed up the controller design process, it would be convenient if a controller could directly be designed based on the FRF.

Relaxing structural constraints, adding dynamics to the controller

If the desired performance cannot be achieved with a controller of a given structure, some of the structural constraints can be relaxed. For instance, one might introduce extra dynamics to the controller. It would be valuable, if the controller optimization does not fully need to be redone, but might be warm-started with the results of the previous optimization. The results of the previous optimization might indeed contain structural information, which may be valuable to determine a good initial guess of the location of the poles and zeros of the extra dynamics.

Appendix A

Auxiliary technical results

A.1 Direct feedthrough controller reconstruction

Equation (2.7) can be verified as follows. The interconnection with $\tilde{\mathbf{K}}$ is shown Figure 2.3 and is repeated in Figure A.1 for the readers convenience. Its state-space matrix quadruple is $(A_{\tilde{K}}, B_{\tilde{K}}, C_{\tilde{K}}, D_{\tilde{K}})$. The state, input and output variables of $\mathbf{K}$ and $\tilde{\mathbf{K}}$ are related as follows:

$$
\begin{aligned}
x_K &= A_{\tilde{K}} x_K + B_{\tilde{K}} \tilde{y} \\
\tilde{u} &= C_{\tilde{K}} x_K + D_{\tilde{K}} \tilde{y} \\
y &= \tilde{y} + D_{22} \tilde{u} \\
u &= \tilde{u} \\
x_K &= x_{\tilde{K}}
\end{aligned}
$$

Hence

$$
\begin{aligned}
\tilde{y} &= y - D_{22} \tilde{u} \\
u &= C_{\tilde{K}} x_K + D_{\tilde{K}} \tilde{y} = C_{\tilde{K}} x_K + D_{\tilde{K}} y - D_{\tilde{K}} D_{22} u.
\end{aligned}
$$

This implies $(I + D_{\tilde{K}} D_{22}) u = C_{\tilde{K}} x_K + D_{\tilde{K}} y$. For nonsingular $Q := I + D_{\tilde{K}} D_{22}$ we obtain

$$
u = Q^{-1} C_{\tilde{K}} x_K + Q^{-1} D_{\tilde{K}} y.
$$
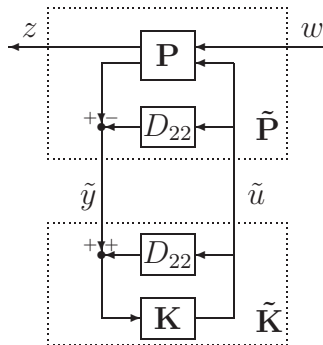
Figure A.1: Loop transformation to eliminate D_{22}

Furthermore

$$x_K = A_{\tilde{K}} x_K + B_{\tilde{K}} \tilde{y} = A_{\tilde{K}} x_K + B_{\tilde{K}} y - B_{\tilde{K}} D_{22} u =$$
$$= A_{\tilde{K}} x_K + B_{\tilde{K}} y - B_{\tilde{K}} D_{22} (Q^{-1} C_{\tilde{K}} x_K + Q^{-1} D_{\tilde{K}} y) =$$
$$= (A_{\tilde{K}} - B_{\tilde{K}} D_{22} Q^{-1} C_{\tilde{K}}) x_K + (B_{\tilde{K}} - B_{\tilde{K}} D_{22} Q^{-1} D_{\tilde{K}}) y$$

Hence K is given by

$$\left(\begin{array}{c|c} A_K & B_K \\ \hline C_K & D_K \end{array} \right) = \left(\begin{array}{c|c} A_{\tilde{K}} - B_{\tilde{K}} D_{22} Q^{-1} C_{\tilde{K}} & B_{\tilde{K}} - B_{\tilde{K}} D_{22} Q^{-1} D_{\tilde{K}} \\ \hline Q^{-1} C_{\tilde{K}} & Q^{-1} D_{\tilde{K}} \end{array} \right) \tag{A.1}$$

which is (2.7).

A.2 Results for Chapter 4

A.2.1 Elementary identities involving Trace$_r$ and the Kronecker product

The in (4.12) defined extension of the trace-operator Trace$_r$ satisfies the following property: for all A and B of appropriate size

$$\mathrm{Trace}_r((I_r \otimes B)A) = \mathrm{Trace}_r(A(I_r \otimes B)). \tag{A.2}$$

The proof of this property is elementary. Indeed, let us partition A in blocks $A_{ij} \in \mathbb{R}^{m \times n}$, $i, j = 1, \ldots, r$ and let $P := \mathrm{Trace}_r((I_r \otimes B)A) \in \mathbb{R}^{r \times r}$ then the element P_{ij} of P at the ith row and j^{th} column is given by

$$P_{ij} = (\mathrm{Trace}_r((I_r \otimes B)A))_{ij} = \mathrm{Trace}(BA_{ij}) = \mathrm{Trace}(A_{ij}B).$$

On the other hand if we define $Q := \mathrm{Trace}_r(A(I_r \otimes B)) \in \mathbb{R}^{r \times r}$, then the element Q_{ij} of Q at the ith row and j^{th} column is given by

$$Q_{ij} = \mathrm{Trace}(A_{ij}B),$$

which implies $P = Q$.

Let $A \in \mathbb{R}^{q \times r}$, $B \in \mathbb{R}^{r \times s}$, $C \in \mathbb{R}^{s \times t}$, $D \in \mathbb{R}^{t \times u}$ and $E \in \mathbb{R}^{u \times v}$ be arbitrary matrices. The Kronecker product is associative:

$$(A \otimes C) \otimes E = A \otimes (C \otimes E).$$

Another important property is the following product rule (see e.g. [30] of [122])

$$(AB) \otimes (CD) = (A \otimes C)(B \otimes D).$$

Using these properties it is not hard to derive the following identities for symmetric matrices $M \in \mathcal{S}^{r \times r}$, $N \in \mathcal{S}^{s \times s}$ and $P \in \mathcal{S}^{rs \times rs}$:

$$
\begin{aligned}
(I_s \otimes M)(N \otimes I_r) &= (N \otimes I_r)(I_s \otimes M) \\
M &= \mathrm{Trace}_r(M) \\
\mathrm{Trace}_r(P)M &= \mathrm{Trace}_r(P(M \otimes I_s)) \\
\mathrm{Trace}_r(P(I_s \otimes N)) &= \mathrm{Trace}_r((I_s \otimes N)P) \\
\mathrm{Trace}_r(M \otimes N) &= \left(\begin{array}{ccc} M_{11}\mathrm{Trace}(N) & \cdots & M_{1r}\mathrm{Trace}(N)) \\ \vdots & \ddots & \vdots \\ M_{r1}\mathrm{Trace}(N) & \cdots & M_{rr}\mathrm{Trace}(N) \end{array} \right) = M\mathrm{Trace}(N)
\end{aligned}
$$

A.2.2 Complete positivity of the Trace operator

A linear map $\Phi : \mathbb{R}^{n \times n} \mapsto \mathbb{R}^{m \times m}$ is [41]:

- positive if Φ maps positive semidefinite matrices into positive semidefinite matrices,

- p-positive if $\Phi \otimes \mathrm{Id}_p$ is positive on $\mathbb{R}^{np \times np}$, where $\mathrm{Id}_p : \mathbb{R}^{p \times p}$ is the identity map and

- completely positive if Φ is p-positive for every $p \in \mathbb{N}$.

Complete positivity of Trace $: \mathbb{R}^{n \times n} \mapsto \mathbb{R}$ follows from a simple extension of Theorem 1 of Choi [41] to real-valued mappings:

Theorem A.1 *Let $\Phi : \mathbb{R}^{n \times n} \mapsto \mathbb{R}^{m \times m}$. Then Φ is completely positive if and only if Φ is of the form $\Phi(A) = \sum_i V_i^T A V_i$ for all A in $\mathbb{R}^{n \times n}$ where V_i are $n \times m$ matrices.*

Proof. Simple extension of the proof of Theorem 1 in [41].

∎

This implies that Trace $: \mathbb{R}^{n \times n} \mapsto \mathbb{R}$ is completely positive for all $n \in \mathbb{N}$ since it admits an expression

$$\mathrm{Trace}(A) = \sum_{i=1}^{n} e_i^T A e_i \quad \text{for all} \quad A \in \mathbb{R}^{n \times n}.$$

A.2.3 Proof of equivalent constraint qualification

We will show that (4.18) can be equivalently formulated as follows: there exist an SOS matrix-valued polynomial $P(x)$ and an SOS polynomial $p(x)$ such that

$$\{x \in \mathbb{R}^{n_x} \mid \langle P(x), G(x) \rangle - p(x) \leq 0\} \quad \text{is compact.} \tag{A.3}$$

Indeed, (4.18) implies the existence of a $M > 0$ and SOS polynomials $\Psi(x)$ and $\psi(x)$ such that

$$\langle \Psi(x), G(x) \rangle - \psi(x) = \|x\|^2 - M \text{ for all } x \in \mathbb{R}^m.$$

Since $\{x \in \mathbb{R}^{n_x} \mid \|x\|^2 - M \leq 0\}$ is obviously compact, (A.3) follows. On the other hand, let $P(x)$ and $p(x)$ be a matrix-valued and a scalar-valued SOS polynomial, respectively, such that (A.3) holds true and let us set $h(x) := \langle P(x), G(x) \rangle - p(x)$. Obviously, (A.3) implies that $\{x \in \mathbb{R}^{n_x} \mid h(x) \leq 0\}$ is compact, and Assumption 4.1 implies that it is nonempty as well. This implies there exists a large enough M such $M - \|x\|^2$ is positive on this set. Now we apply Schmüdgen's theorem [177] to conclude that there exist SOS polynomials $s_0(x)$ and $s_1(x)$ with $M - \|x\|^2 + h(x)s_0(x) = s_1(x)$. Finally (4.18) follows for $\Psi(x) = s_0(x)P(x)$ and $\psi(x) = s_1(x) + p(x)s_0(x)$, which are obviously SOS polynomials.

A.3 Results for Chapter 5

A.3.1 Derivation of the dual of (5.7)

We derive the Lagrange dual of the optimization problem

$$\begin{array}{ll}
\text{infimize} & t \\
\text{subject to} & X \succ 0,\ \mathcal{B}_{\mathrm{S}}(t, X, K_0) \succ 0,\ t > 0
\end{array}$$

for fixed K_0, where

$$\mathcal{B}_{\mathrm{S}}(t, X, K_0) := \begin{pmatrix} -A_{\mathrm{cl}}(K_0)^T X - X A_{\mathrm{cl}}(K_0) & -X B_{\mathrm{cl}}(K_0) \\ -B_{\mathrm{cl}}(K_0)^T X & tI \end{pmatrix} - \begin{pmatrix} C_{\mathrm{cl}}(K_0)^T \\ D_{\mathrm{cl}}(K_0)^T \end{pmatrix} \begin{pmatrix} C_{\mathrm{cl}}(K_0)^T \\ D_{\mathrm{cl}}(K_0)^T \end{pmatrix}^T.$$

The Lagrangian is

$$L(S, Z, \lambda, t, X) = t - \mathrm{Trace}(SX) - \mathrm{Trace}(Z\mathcal{B}_{\mathrm{S}}(t, X, K_0)) - \lambda t =$$

$$= (1 - \lambda - \mathrm{Trace}(Z_{22}))t + \mathrm{Trace}\left(\begin{bmatrix} C_{\mathrm{cl}}(K_0) & D_{\mathrm{cl}}(K_0) \end{bmatrix} Z \begin{bmatrix} C_{\mathrm{cl}}(K_0) & D_{\mathrm{cl}}(K_0) \end{bmatrix}^T\right) +$$

$$+ \mathrm{Trace}\left(\left(-S + A_{\mathrm{cl}}(K_0)Z_{11} + Z_{11}A_{\mathrm{cl}}(K_0)^T + B_{\mathrm{cl}}(K_0)Z_{12}^T + Z_{12}B_{\mathrm{cl}}(K_0)^T\right) X\right),$$

where we used the partition

$$Z = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix}.$$

Hence, the dual problem reads as

$$\begin{aligned} \text{supremize} \quad & \mathrm{Trace}\left(\begin{bmatrix} C_{\mathrm{cl}}(K_0) & D_{\mathrm{cl}}(K_0) \end{bmatrix} Z \begin{bmatrix} C_{\mathrm{cl}}(K_0) & D_{\mathrm{cl}}(K_0) \end{bmatrix}^T\right) \\ \text{subject to} \quad & A_{\mathrm{cl}}(K_0)Z_{11} + Z_{11}A_{\mathrm{cl}}(K_0)^T + B_{\mathrm{cl}}(K_0)Z_{12}^T + Z_{12}B_{\mathrm{cl}}(K_0)^T = S \\ & 1 - \lambda - \mathrm{Trace}(Z_{22}) = 0, \ Z \succeq 0, \ S \succeq 0 \text{ and } \lambda \geq 0 \end{aligned}$$

Eliminating the dual variables S and λ yields the desired formulation of the dual problem:

$$\begin{aligned} \text{supremize} \quad & \mathrm{Trace}\left(\begin{bmatrix} C_{\mathrm{cl}}(K_0) & D_{\mathrm{cl}}(K_0) \end{bmatrix} Z \begin{bmatrix} C_{\mathrm{cl}}(K_0) & D_{\mathrm{cl}}(K_0) \end{bmatrix}^T\right) \\ \text{subject to} \quad & A_{\mathrm{cl}}(K_0)Z_{11} + Z_{11}A_{\mathrm{cl}}(K_0)^T + B_{\mathrm{cl}}(K_0)Z_{12}^T + Z_{12}B_{\mathrm{cl}}(K_0)^T \succeq 0 \\ & \mathrm{Trace}(Z_{22}) \leq 1, \qquad Z \succeq 0 \end{aligned}$$

A.3.2 Proof of strict feasibility of the dual problem

Let us prove that (5.10) is strictly feasible for all $K_0 \in \mathcal{K}$. For an arbitrary $K_0 \in \mathcal{K}$, we need to show that there exists some W with

$$\begin{aligned} A(K_0)W_{11} + W_{11}A(K_0)^T + B(K_0)W_{12}^T + W_{12}B(K_0)^T &\succ 0, \\ \mathrm{Trace}(W_{22}) < 1, \qquad W &\succ 0. \end{aligned} \tag{A.4}$$

Since $(A(K_0), B(K_0))$ is controllable, one can construct an anti-stabilizing state-feedback gain, i.e., a matrix L such that $A(K_0) + B(K_0)L$ has all its eigenvalues in the open right-half plane. Hence there exists some $P \succ 0$ with

$$(A(K_0) + B(K_0)L)P + P(A(K_0) + B(K_0)L)^T \succ 0 \tag{A.5}$$

and rP also satisfies (A.5) for any $r > 0$. Then W defined by the blocks $W_{11} = rP$, $W_{12}^T = rLP$ and

$$W_{22} = W_{12}^T W_{11}^{-1} W_{12} + rI = r(L^T PL + I)$$

in the partition (5.9) satisfies (A.4) and $W \succ 0$ for arbitrary $r > 0$. The constructed W does the job if we choose in addition $r > 0$ sufficiently small to achieve $\mathrm{Trace}(W_{22}) = r\mathrm{Trace}(L^T PL + I) < 1$.

A.4 Results for Chapter 7

A.4.1 Derivatives of the barrier function

Since the term $-\log\det(\mathcal{B}(x))$ is involved in the barrier, we first give the first- and second-order derivatives of $\mathcal{B}(x)$ and $-\log\det(\mathcal{B}(x))$.

Derivatives of $\mathcal{B}(x)$

The derivative of $\mathcal{B}(x)$ as given in (7.15), which we repeat for the convenience of the reader:

$$\partial\mathcal{B}(x)d_x := \mathcal{L}_\gamma(x)(d_\gamma) + \mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K), \tag{A.6}$$

where $d_x = (d_\gamma, d_X, d_K)$ and $\mathcal{L}_\gamma(x)(d_\gamma)$, $\mathcal{L}_X(x)(d_X)$ and $\mathcal{L}_K(x)(d_K)$ are defined in (7.16), (7.17) and (7.18), respectively. For the second derivative of $\mathcal{B}(x)$, we need the derivatives of the $\mathcal{L}_\gamma(x)(d_\gamma)$, $\mathcal{L}_X(x)(d_X)$ and $\mathcal{L}_K(x)(d_K)$ with respect to x:

$$\partial_x\mathcal{L}_\gamma(x)(d_{\gamma 1})(d_{x2}) \;=\; 0 \tag{A.7}$$

$$\partial_x\mathcal{L}_X(x)(d_{X1})(d_{x2}) \;=\; -\mathrm{sy}\left(\left(\begin{array}{c} I_{n+n_c} \\ 0 \\ 0 \end{array}\right) d_{X1} \left(\begin{array}{c} B_2 d_{K2} C_2{}^T \\ B_2 d_{K2} D_{21}{}^T \\ 0 \end{array}\right)^T\right) \tag{A.8}$$

$$\partial_x\mathcal{L}_K(x)(d_{K1})(d_{x2}) \;=\; -\mathrm{sy}\left(\left(\begin{array}{c} d_{X2}B_2 \\ 0 \\ 0 \end{array}\right) d_{K1} \left(\begin{array}{c} C_2^T \\ D_{21}^T \\ 0 \end{array}\right)^T\right), \tag{A.9}$$

where $d_{x1} = (d_{\gamma 1}, d_{X1}, d_{K1})$, $d_{x2} = (d_{\gamma 2}, d_{X2}, d_{K2})$ and $\mathrm{sy}(A) = A + A^T$. Since $\partial_x\mathcal{L}_\gamma(x)(d_\gamma)d_{x2} = 0$, the second derivative of $\mathcal{B}(x)$ is

$$\partial^2\mathcal{B}(x)(d_{x1},d_{x2}) := \partial_x\mathcal{L}_X(x)(d_{X1})(d_{x2}) + \partial_x\mathcal{L}_K(x)(d_{K1})(d_{x2}). \tag{A.10}$$

Derivatives of $-\log\det(\mathcal{B}(x))$

The matrix derivative rules as for instance presented in [5] imply for arbitrary $Y \succ 0$

$$\partial(-\log(\det(Y)))(d_Y) \;=\; -\mathrm{Trace}(Y^{-1}d_Y) \tag{A.11}$$

$$\partial(-\mathrm{Trace}(Y^{-1}))(d_Y) \;=\; \mathrm{Trace}(Y^{-1}d_Y Y^{-1}). \tag{A.12}$$

By combining (A.6), (A.11) with the chain rule of derivation, we obtain the following first-order derivative of $-\log(\det(\mathcal{B}(x)))$:

$$\partial\left(-\log\left(\det\left(\mathcal{B}(x)\right)\right)\right)(d_x) = -\mathrm{Trace}\left(\mathcal{B}(x)^{-1}[\mathcal{L}_\gamma(x)(d_\gamma) + \mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K)]\right) =$$
$$= -\mathrm{Trace}\left(\mathcal{B}(x)^{-1}\partial\mathcal{B}(x)(d_x)\right).$$

In terms of the dual mappings $\mathcal{L}_\gamma^*$, $\mathcal{L}_X^*$ and $\mathcal{L}_K^*$, this derivative is equivalently expressed as

$$\partial\left(-\log\left(\det\left(\mathcal{B}(x)\right)\right)\right)(d_x) = -\mathcal{L}_\gamma^*(x)\left(\mathcal{B}(x)^{-1}\right)d_\gamma -$$
$$- \mathrm{Trace}(\mathcal{L}_X^*(x)\left(\mathcal{B}(x)^{-1}\right)d_X) - \mathrm{Trace}(\mathcal{L}_K^*(x)\left(\mathcal{B}(x)^{-1}\right)d_K). \tag{A.13}$$

Using (A.12) we compute the second-order derivative of $-\log\left(\det\left(\mathcal{B}(x)\right)\right)$:

$$\partial^2\left(-\log\left(\det\left(\mathcal{B}(x)\right)\right)\right)(d_{x1},d_{x2}) = \partial_x\left[\text{Trace}(\mathcal{B}(x)^{-1})\partial\mathcal{B}(x)(d_{x1})\right](d_{x2}) =$$
$$= \text{Trace}\left(\mathcal{B}(x)^{-1}\partial\mathcal{B}(x)(d_{x1})\mathcal{B}(x)^{-1}\partial\mathcal{B}(x)(d_{x2})\right) - \text{Trace}\left(\mathcal{B}(x)^{-1}\partial^2\mathcal{B}(x)(d_{x1},d_{x2})\right). \tag{A.14}$$

The first term of the last expression of (A.14), i.e. $\text{Trace}\left(\mathcal{B}(x)^{-1}\partial\mathcal{B}(x)(d_{x1})\mathcal{B}(x)^{-1}\partial\mathcal{B}(x)(d_{x2})\right)$, can be written in terms of the dual mappings as follows

$$\text{Trace}\left(\mathcal{B}(x)^{-1}\partial\mathcal{B}(x)(d_{x1})\mathcal{B}(x)^{-1}\partial\mathcal{B}(x)(d_{x2})\right) =$$
$$= \mathcal{L}_\gamma^*\left(\mathcal{B}(x)^{-1}\partial\mathcal{B}(x)(d_{x1})\mathcal{B}(x)^{-1}\right)d_{\gamma2} + \text{Trace}\left(\mathcal{L}_X^*\left(\mathcal{B}(x)^{-1}\partial\mathcal{B}(x)(d_{x1})\mathcal{B}(x)^{-1}\right)d_{X2}\right) +$$
$$+ \text{Trace}\left(\mathcal{L}_K^*\left(\mathcal{B}(x)^{-1}\partial\mathcal{B}(x)(d_{x1})\mathcal{B}(x)^{-1}\right)d_{K2}\right) \tag{A.15}$$

The mappings $\mathcal{L}_{KX}$ and $\mathcal{L}_{XK}$ defined in (7.24) and (7.25) satisfy the following identities

$$\text{Trace}(\mathcal{L}_{KX}(d_{X1})d_{K2}) = -\text{Trace}\left(\mathcal{B}(x)^{-1}\partial_x\mathcal{L}_X(x)(d_{X1})(d_{x2})\right)$$
$$\text{Trace}(\mathcal{L}_{XK}(d_{K1})d_{X2}) = -\text{Trace}\left(\mathcal{B}(x)^{-1}\partial_x\mathcal{L}_K(x)(d_{K1})(d_{x2})\right),$$

where $d_{x2} = (d_{\gamma2}, d_{X2}, d_{K2})$. This follows from (7.24), (7.25) and (A.8)-(A.9). Using these identities and (A.10), we can express second term $-\text{Trace}\left(\mathcal{B}(x)^{-1}\partial^2\mathcal{B}(x)(d_{x1},d_{x2})\right)$ of the last expression of (A.14) as

$$- \text{Trace}\left(\mathcal{B}(x)^{-1}\partial^2\mathcal{B}(x)(d_{x1},d_{x2})\right) =$$
$$= -\text{Trace}\left(\mathcal{B}(x)^{-1}\partial_x\mathcal{L}_X(x)(d_{X1})(d_{x2})\right) - \text{Trace}\left(\mathcal{B}(x)^{-1}\partial_x\mathcal{L}_K(x)(d_{K1})(d_{x2})\right) =$$
$$= \text{Trace}(\mathcal{L}_{KX}(d_{X1})d_{K2}) + \text{Trace}(\mathcal{L}_{XK}(d_{K1})d_{X2}) \tag{A.16}$$

This completes the computation of the derivatives of $-\log\det(\mathcal{B}(x))$.

Derivatives of remaining terms in barrier

The first and second order derivative of $-\log(\det(X))$ follow directly from (A.11) and (A.12)

$$\partial[-\log(\det(X))](d_X) = -\text{Trace}(X^{-1}d_X) \tag{A.17}$$
$$\partial^2[-\log(\det(X))](d_{X1},d_{X2}) = \text{Trace}(X^{-1}d_{X1}X^{-1}d_{X2}). \tag{A.18}$$

Similarly we obtain for the derivatives of $-\log(\det(\rho_X I - X))$ the following expressions:

$$\partial[-\log(\det(\rho_X I - X))](d_X) = \text{Trace}((\rho_X I - X)^{-1}d_X) \tag{A.19}$$
$$\partial^2[-\log(\det(\rho_X I - X))](d_{X1},d_{X2}) = \text{Trace}\left((\rho_X I - X)^{-1}d_{X1}(\rho_X I - X)^{-1}d_{X2}\right). \tag{A.20}$$

Finally observe that

$$\partial\left(-\log(\rho_K - \|K\|_F^2)\right)(d_K) = \frac{2}{\rho_K - \|K\|_F^2}\text{Trace}(K^T d_K) \tag{A.21}$$

and

$$\partial^2\left(-\log(\rho_K - \|K\|_F^2)\right)(d_{K1},d_{K2}) = \frac{2}{\rho_K - \|K\|_F^2}\text{Trace}(d_{K1}^T d_{K2}) +$$
$$+ \frac{4\text{Trace}(K^T d_{K1})\text{Trace}(K^T d_{K2})}{(\rho_K - \|K\|_F^2)^2} \tag{A.22}$$

Partial derivatives of the barrier

Equations (A.13), (A.17), (A.19) and (A.21) imply the following first order partial derivatives of ϕ:

$$\partial_\gamma \phi(\gamma, X, K) = -\mathcal{L}_\gamma^*(x)\left(\mathcal{B}(x)^{-1}\right) + \frac{1}{\mu} \tag{A.23}$$

$$\partial_X \phi(\gamma, X, K) = -\mathcal{L}_X^*(x)\left(\mathcal{B}(x)^{-1}\right) - X^{-1} + (\rho_X I - X)^{-1} \tag{A.24}$$

$$\partial_K \phi(\gamma, X, K) = -\mathcal{L}_K^*(x)\left(\mathcal{B}(x)^{-1}\right) + \frac{2}{\rho_K - \|K\|_F^2} K^T \tag{A.25}$$

Similarly, the following second-order partial derivatives of the barrier ϕ can be derived from (A.14)-(A.16), (A.18), (A.20) and (A.22):

$$\partial_\gamma^2 \phi(\gamma, X, K)(d_\gamma, d_X, d_K) = \mathcal{L}_\gamma^*(x)\left(\mathcal{B}(x)^{-1}\mathcal{L}_\gamma(x)(d_\gamma)\mathcal{B}(x)^{-1}\right) \tag{A.26}$$

$$\partial_\gamma \partial_X \phi(\gamma, X, K)(d_\gamma, d_X, d_K) = \mathcal{L}_\gamma^*(x)\left(\mathcal{B}(x)^{-1}\mathcal{L}_X(x)(d_X)\mathcal{B}(x)^{-1}\right) \tag{A.27}$$

$$\partial_\gamma \partial_K \phi(\gamma, X, K)(d_\gamma, d_X, d_K) = \mathcal{L}_\gamma^*(x)\left(\mathcal{B}(x)^{-1}\mathcal{L}_K(x)(d_K)\mathcal{B}(x)^{-1}\right) \tag{A.28}$$

$$\partial_X^2 \phi(\gamma, X, K)(d_\gamma, d_X, d_K) = \mathcal{L}_X^*(x)\left(\mathcal{B}(x)^{-1}\mathcal{L}_X(x)(d_X)\mathcal{B}(x)^{-1}\right) + X^{-1}d_X X^{-1} +$$
$$+ (\rho_X I - X)^{-1}d_X(\rho_X I - X)^{-1} \tag{A.29}$$

$$\partial_X \partial_K \phi(\gamma, X, K)(d_\gamma, d_X, d_K) = \mathcal{L}_X^*(x)\left(\mathcal{B}(x)^{-1}\mathcal{L}_K(x)(d_K)\mathcal{B}(x)^{-1}\right) + \mathcal{L}_{XK}(x)(d_K) \tag{A.30}$$

$$\partial_K^2 \phi(\gamma, X, K)(d_\gamma, d_X, d_K) = \mathcal{L}_K^*(x)\left(\mathcal{B}(x)^{-1}\mathcal{L}_K(x)(d_K)\mathcal{B}(x)^{-1}\right) +$$
$$+ \frac{2}{\rho_K - \|K\|_F^2}(K + d_K)^T + \frac{4\text{Trace}(K^T d_K)}{(\rho_K - \|K\|_F^2)^2} K^T. \tag{A.31}$$

For reasons of clarity we finally remark that the derivatives in (A.26)-(A.31) have been presented as mappings of the form $\partial^2 \phi(d_x, \cdot)$.

A.4.2 Derivation of Sylvester equations for the Newton step

The Newton step is the solution to

$$\partial^2 \phi(\gamma, X, K)(d_\gamma, d_X, d_K) = -\partial \phi(\gamma, X, K),$$

where '=' means equality of the linear mappings on the left and right-hand side. In the IP algorithm we compute trust region steps, which are the solutions to

$$\partial^2 \phi(\gamma, X, K)(d_\gamma, d_X, d_K) + \lambda q(d_\gamma, d_X, d_K) = -\partial \phi(\gamma, X, K), \tag{A.32}$$

for some trust-region parameter $\lambda \geq 0$, where $q(\cdot)(\cdot) : \left(\mathbb{R} \times \mathcal{S}^{n+n_c} \times \mathbb{R}^{(n_c+m_2) \times (n_c+p_2)}\right) \times \left(\mathbb{R} \times \mathcal{S}^{n+n_c} \times \mathbb{R}^{(n_c+m_2) \times (n_c+p_2)}\right) \to \mathbb{R}$ is defined by

$$q(d_{\gamma 1}, d_{X1}, d_{K1})(d_{\gamma 2}, d_{X2}, d_{K2}) = d_{\gamma 1} d_{\gamma 2} + \text{Trace}(d_{X1} d_{X2}) + \text{Trace}(d_{K1}^T d_{K2}).$$

Sylvester equations

Equation (A.32) can be subdivided into the following three equations:

$$\partial_\gamma^2 \phi(\gamma, X, K)(d_\gamma) + \partial_\gamma \partial_X \phi(\gamma, X, K)(d_X) + \partial_\gamma \partial_K \phi(\gamma, X, K)(d_K) + \lambda q_\gamma(d_\gamma) = -\partial_\gamma \phi(\gamma, X, K)$$

$$\partial_X \partial_\gamma \phi(\gamma, X, K)(d_\gamma) + \partial_X^2 \phi(\gamma, X, K)(d_X) + \partial_X \partial_K \phi(\gamma, X, K)(d_K) + \lambda q_X(d_X) = -\partial_X \phi(\gamma, X, K)$$

$$\partial_K \partial_\gamma \phi(\gamma, X, K)(d_\gamma) + \partial_K \partial_X \phi(\gamma, X, K)(d_X) + \partial_K^2 \phi(\gamma, X, K)(d_K) + \lambda q_K(d_K) = -\partial_K \phi(\gamma, X, K)$$

where $q_\gamma(\gamma_1)(\gamma_2) = \gamma_1\gamma_2$, $q_X(X_1)(X_2) = \text{Trace}(X_1X_2)$ and $q_K(K_1)(K_2) = \text{Trace}(K_1^T K_2)$. By substituting in these equations the partial derivatives given in (A.23)-(A.25) and (A.26)-(A.31), we obtain the following set of Sylvester equations which are (7.26), (7.27) and (7.28) respectively:

$$\mathcal{L}_\gamma^*(x)\left(\mathcal{B}(x)^{-1}\left[\mathcal{L}_\gamma(x)(d_\gamma) + \mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K)\right]\mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1}\right) + \frac{1}{\mu} + \lambda d_\gamma = 0$$

$$\mathcal{L}_X^*(x)\left(\mathcal{B}(x)^{-1}\left[\mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K) + \mathcal{L}_\gamma(x)(d_\gamma)\right]\mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1}\right) + \mathcal{L}_{XK}(x)(d_K) -$$
$$- X^{-1} + X^{-1}d_X X^{-1} + (\rho_X I - X)^{-1} + (\rho_X I - X)^{-1}d_X(\rho_X I_{n_X} - X)^{-1} + \lambda d_X = 0,$$

and

$$\mathcal{L}_K^*(x)\left(\mathcal{B}(x)^{-1}\left[\mathcal{L}_X(x)(d_X) + \mathcal{L}_K(x)(d_K) + \mathcal{L}_\gamma(x)(d_\gamma)\right]\mathcal{B}(x)^{-1} - \mathcal{B}(x)^{-1}\right) + \mathcal{L}_{KX}(x)(d_X) +$$
$$+ \frac{2}{\rho_K - \|K\|_F^2}(K + d_K)^T + \frac{4\text{Trace}(K^T d_K)}{(\rho_K - \|K\|_F^2)^2}K^T + \lambda d_K = 0.$$

A.4.3 Adjoint mappings of $\mathcal{L}_\gamma$, $\mathcal{L}_X$ and $\mathcal{L}_K$

The domains of the mappings $\mathcal{L}_\gamma$, $\mathcal{L}_X$ and $\mathcal{L}_K$ are $\mathbb{R}$, $\mathcal{S}^{n+n_c}$ and $\mathbb{R}^{(n_c+m_2)\times(n_c+m_2)}$, which are linear finite-dimensional vector spaces. The ranges of these functions are $\mathcal{S}^{n+n_c+m_1+p_1}$, which is also a linear vector space. To be more general, consider an arbitrary finite-dimensional linear operator $\mathcal{L} : \mathcal{U} \to \mathcal{V}$, where $\mathcal{U}$ and $\mathcal{V}$ are vector spaces with finite dimensions of m and n, respectively. In terms of arbitrary bases $\{u_1, \ldots u_m\}$ of $\mathcal{U}$ and $\{v_1, \ldots v_n\}$ of $\mathcal{V}$, $\mathcal{L}$ has a matrix representation. This implies that there exists a matrix $A \in \mathbb{R}^{m \times n}$ such that

$$\mathcal{L}\left(\sum_{i=1}^m \alpha_i u_i\right) = \sum_{i=1}^n \beta_i v_i, \qquad \beta = A\alpha, \qquad \text{for all } \alpha \in \mathbb{R}^m.$$

The *adjoint* $\mathcal{L}^* : \mathcal{V} \to \mathcal{U}$ is the linear mapping represented by A^T, i.e. the mapping defined by

$$\mathcal{L}^*\left(\sum_{i=1}^n \beta_i v_i\right) = \sum_{i=1}^m \alpha_i u_i, \ \alpha = A^T \beta, \ \text{for all } \beta \in \mathbb{R}^n.$$

The *adjoint* appears to be independent of the choice of the bases $\{u_1, \ldots u_m\}$ and $\{v_1, \ldots v_n\}$ [124]. The dual of $\mathcal{L}_\gamma$ satisfies [124] for all $d_\gamma \in \mathbb{R}$, $d_W \in \mathcal{S}^{n+n_c+m_1+p_1}$

$$\text{Trace}(d_W(\mathcal{L}_\gamma(d_\gamma))) = \text{Trace}\left(d_W \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_{m_1} & 0 \\ 0 & 0 & I_{p_1} \end{pmatrix} d_\gamma\right) = \text{Trace}\left(d_\gamma \mathcal{L}_\gamma^*(d_W)\right),$$

which implies that $\mathcal{L}_\gamma^*$ is as given in (7.22). Similarly, $\mathcal{L}_{d_X}^*$ satisfies for all $d_X \in \mathcal{S}^{n+n_c}$ and $d_U \in \mathcal{S}^{n+n_c+m_1+p_1}$

234

$$\text{Trace}\,(d_U(\mathcal{L}_X(d_X))) = 2\text{Trace}\left(-d_U\begin{pmatrix} I_{n+n_c} \\ 0 \\ 0 \end{pmatrix} d_X \begin{pmatrix} A_{\text{cl}}(K)^T \\ B_{\text{cl}}(K)^T \\ 0 \end{pmatrix}^T\right) =$$

$$= 2\text{Trace}\left(-\begin{pmatrix} A_{\text{cl}}(K)^T \\ B_{\text{cl}}(K)^T \\ 0 \end{pmatrix}^T d_U \begin{pmatrix} I_{n+n_c} \\ 0 \\ 0 \end{pmatrix} d_X\right) = \text{Trace}\,\left(\mathcal{L}^*_{d_X}(d_U)d_X\right)$$

and for all $d_K \in \mathbb{R}^{(m_2+n_c)\times(p_2+n_c)}$ and $d_V \in \mathcal{S}^{n+n_c+m_1+p_1}$, $\mathcal{L}^*_{d_K}$ satisfies

$$\text{Trace}\,(d_V(\mathcal{L}_K(d_K))) = \text{Trace}\left(-d_V \begin{pmatrix} XB_2 \\ 0 \\ D_{12} \end{pmatrix} d_K \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix}^T - \begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix} d_K{}^T \begin{pmatrix} XB_2 \\ 0 \\ D_{12} \end{pmatrix}^T\right) =$$

$$= 2\text{Trace}\left(-\begin{pmatrix} C_2^T \\ D_{21}^T \\ 0 \end{pmatrix}^T d_V \begin{pmatrix} XB_2 \\ 0 \\ D_{12} \end{pmatrix} d_K\right) = \text{Trace}(\mathcal{L}^*_{d_K}(d_V)d_K).$$

Using these properties, (7.23) and (7.24) follow immediately.

A.4.4 Unique solution of trust region problem

Lemma A.2 *Let H be a symmetric matrix. For every $\rho > 0$ the problem*

$$\begin{array}{ll} \text{supremize} & t \\ \text{subject to} & \begin{pmatrix} H+\lambda I & g \\ g^T & -2t-\lambda\rho^2 \end{pmatrix} \succeq 0, \quad \lambda \geq 0, \quad \lambda \in \mathbb{R} \end{array} \tag{A.33}$$

has a unique optimal solution.

Proof. First observe that the problem is feasible, since we can always choose λ large enough and t small enough to render the constraints in (A.33) satisfied. The optimal value is also bounded above by zero, since the constraints in (A.33) imply $\lambda \geq 0$ and $t \leq -\frac{1}{2}\lambda\rho^2 \leq 0$.

Let us denote the (finite) optimal value of (A.33) by p_{opt}. The optimal value of (A.33) does obviously not change, if we add the constraints $2p_{\text{opt}} \leq t \leq 0$. This condition combined with the constraint in (A.33) imply $-4p_{\text{opt}} - \lambda\rho^2 \geq -2t - \lambda\rho^2 \geq 0$ such that $\lambda \leq \frac{-4p_{\text{opt}}}{\rho^2}$. The feasible set is therefore compact and Weierstrass' Approximation Theorem implies that the supremum of (A.33) is attained. Furthermore the maximum is unique, for suppose $(t_{\text{opt}}, \lambda_1)$ and $(t_{\text{opt}}, \lambda_2)$ are both optimal for (A.33) with $0 \leq \lambda_1 < \lambda_2$. This implies that

$$\begin{pmatrix} H+\lambda I & g \\ g^T & -2t_{\text{opt}}-\lambda\rho^2 \end{pmatrix} \succeq 0$$

for all $\lambda \in [\lambda_1, \lambda_2]$. Since

$$\lim_{\lambda \to \infty} \frac{1}{\lambda}\begin{pmatrix} H+\lambda I & g \\ g^T & -2t_{\text{opt}}-\lambda\rho^2 \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & -\rho^2 \end{pmatrix}$$

is nonsingular, the polynomial

$$f(\lambda) := \det\left(\begin{pmatrix} H + \lambda I & g \\ g^T & -2t_{\text{opt}} - \lambda\rho^2 \end{pmatrix}\right)$$

is nonzero for very large λ and therefore has a finite number of roots. This implies that there exists a $\lambda_3 \in [\lambda_1, \lambda_2]$ with $f(\lambda_3) \neq 0$, such that

$$\begin{pmatrix} H + \lambda_3 I & g \\ g^T & -2t_{\text{opt}} - \lambda_3\rho^2 \end{pmatrix} \succ 0.$$

Since this inequality holds strict, there exists a $t_3 > t_{\text{opt}}$ such that

$$\begin{pmatrix} H + \lambda_3 I & g \\ g^T & -2t_3 - \lambda_3\rho^2 \end{pmatrix} \succ 0.$$

This contradicts optimality of t_{opt}.

∎

A.4.5 Derivatives of constraints

For $x = (\gamma, K, X)$ and a direction $d_x = (d_\gamma, d_K, d_X)$, the directional derivatives of $G = \text{diag}(G_1, G_2, G_3, g_4)$ with

$$G_1(x) = -\mathcal{B}(\gamma, X, K), \; G_2(x) = -X, \; G_3(x) = X - \rho_X I, \; g_4(x) = \|K\|_{\text{F}}^2 - \rho_K,$$

are as follows:

$$\partial g_1(x)d_x = -\partial\mathcal{B}(x)d_x = -\mathcal{L}_\gamma(x)(d_\gamma) - \mathcal{L}_X(x)(d_X) - \mathcal{L}_K(x)(d_K) =$$
$$= \begin{pmatrix} d_X A_{\text{cl}} + A_{\text{cl}}^T d_X + X B_2 d_K C_2 + (X B_2 d_K C_2)^T & d_X B_{\text{cl}} + X B_2 d_K D_{21} & (D_{12} d_K C_2)^T \\ (d_X B_{\text{cl}} + X B_2 d_K D_{21})^T & -d_\gamma & (D_{12} d_K D_{21})^T \\ D_{12} d_K C_2 & D_{12} d_K D_{21} & -d_\gamma \end{pmatrix}$$

$$\partial G_2(x)d_x = -d_X$$
$$\partial G_3(x)d_x = d_X$$
$$\partial G_4(x)d_x = 2\text{Trace}(K^T d_K)$$

For directions $d_{x1} = (d_{\gamma 1}, d_{K1}, d_{X1})$ and $d_{x2} = (d_{\gamma 2}, d_{K2}, d_{X2})$, the second derivative of G_1 (as also given in (A.10)) is:

$$\partial^2 G_1(x)(d_{x1}, d_{x2}) = -\partial^2\mathcal{B}(\gamma, X, K)(d_{x1}, d_{x2}) =$$
$$= \begin{pmatrix} \text{sy}\,(d_{X1} B_2 d_{K2} C_2 + d_{X2} B_2 d_{K1} C_2) & d_{X1} B_2 d_{K2} D_{21} + d_{X2} B_2 d_{K1} D_{21} & 0 \\ (d_{X1} B_2 d_{K2} D_{21} + d_{X2} B_2 d_{K1} D_{21})^T & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $\text{sy}(A) = A + A^T$. Note that, due to the bilinear structure, $\partial^2 G_1(x)$ is constant over x. The second derivatives of G_2 and G_3 are identically zero and finally

$$\partial^2 G_4(x)(d_{x1}, d_{x2}) = 2\text{Trace}(d_{K1}^T d_{K2}).$$

A.4.6 Newton step in Y

LMI for stability analysis

First, let us consider for $A \in \mathbb{R}^{n \times n}$ the barrier function for the stability analysis LMIs $A^T Y + YA \prec 0$ and $Y \succ 0$:

$$\phi(Y, A) := -\log(\det(-(A^T Y + YA))) - \log(\det(Y)). \tag{A.34}$$

We assume $A \in \mathbb{R}^{n \times n}$ is stable, such that there exists an $Y \in \mathcal{S}^n$ with $Y \succ 0$ and

$$A^T Y + YA \prec 0. \tag{A.35}$$

Furthermore we assume that A has the following block-diagonal structure

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_N \end{pmatrix},$$

where $A_k \in \mathbb{R}^{m_k \times m_k}, k = 1, 2, \ldots, N$ are square matrices.

Define

$$F_k := \begin{pmatrix} 0_{m_k \times m_1} & \cdots & 0_{m_i \times m_{k-1}} & I_{m_k} & 0_{m_k \times m_{k+1}} & \cdots & 0_{m_k \times m_N} \end{pmatrix}^T, \quad k = 1, \ldots, N \tag{A.36}$$

Consider $P : \mathcal{S}^n \mapsto \mathcal{S}^n$ defined by:

$$P(A) = \sum_{k=1}^N F_k F_k^T A F_k F_k^T,. \tag{A.37}$$

Since $P(P(A)) = P(A)$ for all $A \in \mathcal{S}^n$, P is an idempotent mapping and its definition is motivated by the following property: if Y is block-partitioned as A, i.e.

$$Y = \begin{pmatrix} Y_1 & Y_{12} & \cdots & Y_{1N} \\ Y_{12}^T & Y_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & Y_{(N-1)N} \\ Y_{1N}^T & \cdots & Y_{(N-1)N}^T & Y_{NN} \end{pmatrix},$$

then

$$P(Y) = \begin{pmatrix} Y_1 & 0 & \cdots & 0 \\ 0 & Y_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & Y_{NN} \end{pmatrix}.$$

We want to prove that for all A with the block diagonal structure and $Y \succ 0$ such that $A^T Y + YA \prec 0$ the following holds true

$$\phi(Y, A) \geq \phi(P(Y), A). \tag{A.38}$$

237

Let us first assume that A has two blocks ($k = 2$), i.e. $A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$, where A_1 and A_2 are square matrices. Partition Y accordingly into

$$Y = \begin{pmatrix} Y_1 & Y_{12} \\ Y_{12}^T & Y_2 \end{pmatrix}$$

then

$$\det(-(A^T Y + YA)) = \det\left(-\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}^T \begin{pmatrix} Y_1 & Y_{12} \\ Y_{12}^T & Y_2 \end{pmatrix} - \begin{pmatrix} Y_1 & Y_{12} \\ Y_{12}^T & Y_2 \end{pmatrix} \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}\right) =$$

$$= \det\left(-\begin{pmatrix} A_1^T Y_1 & A_1^T Y_{12} \\ A_2^T Y_{12}^T & A_2^T Y_2 \end{pmatrix} - \begin{pmatrix} Y_1 A_1 & Y_{12} A_2 \\ Y_{12}^T A_1 & Y_2 A_2 \end{pmatrix}\right) =$$

$$= \det\left(-\begin{pmatrix} A_1^T Y_1 + Y_1 A_1 & A_1^T Y_{12} + Y_{12} A_2 \\ A_2^T Y_{12}^T + Y_{12}^T A_1 & A_2^T Y_2 + Y_2 A_2 \end{pmatrix}\right)$$

To simplify notation we define

$$M := \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} := -\begin{pmatrix} A_1^T Y_1 + Y_1 A_1 & A_1^T Y_{12} + Y_{12} A_2 \\ A_2^T Y_{12}^T + Y_{12}^T A_1 & A_2^T Y_2 + Y_2 A_2 \end{pmatrix}.$$

Note that (A.35) implies that M is a symmetric positive definite matrix, which implies

$$M_{21} M_{11}^{-1} M_{12} \succeq 0. \tag{A.39}$$

Writing $\det(-(A^T Y + YA))$ in terms of M leads to

$$\det(-(A^T Y + YA)) = \det\left(\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}\right) =$$

$$= \det(M_{11}) \det(M_{22} - M_{21} M_{11}^{-1} M_{12}) \leq \det(M_{11}) \det(M_{22}) \tag{A.40}$$

where the inequality follows from (A.39) and the monotonicity of the determinant with respect to the Löwner partial ordering, i.e. $A \succeq B \succeq 0 \Rightarrow \det(A) \geq \det(B)$ [94]. Since the right-hand-side of the inequality in (A.40) is

$$\det(M_1) \det(M_2) = \det\left(-\begin{pmatrix} A_1^T Y_1 + Y_1 A_1 & 0 \\ 0 & A_2^T Y_2 + Y_2 A_2 \end{pmatrix}\right) =$$

$$= \det(-(A^T P(Y) + P(Y)A)),$$

we conclude that

$$\det(-(A^T Y + YA)) \leq \det(-(A^T P(Y) + P(Y)A))$$

and by the monotonicity of the logarithm this implies that

$$-\log(\det(-(A^T Y + YA)) \geq -\log(\det(-(A^T P(Y) + P(Y)A)). \tag{A.41}$$

Next, consider the second term of the barrier in (A.34), i.e.

$$-\log(\det(Y)) = -\log\left(\det\left(\begin{pmatrix} Y_1 & Y_{12} \\ Y_{12}^T & Y_2 \end{pmatrix}\right)\right) =$$

$$= -\log\left(\det(Y_1) \det(Y_2 - Y_{21} Y_1^{-1} Y_{12})\right) \geq \det(Y_1) \det(Y_2) = -\log(\det(P(Y)))$$

$$, \tag{A.42}$$

238

where the inequality follows from the fact that $Y_{21}Y_1^{-1}Y_{12} \succeq 0$ and (again) the monotonicity of the determinant with respect to the Löwner partial ordering. Combining (A.41) and (A.42) yields

$$-\log(\det(-(A^T Y + Y A))) - \log(\det(Y)) \geq$$
$$\geq -\log(\det(-(A^T P(Y) + P(Y)A))) - \log(\det(P(Y)))$$

which implies (A.38). This shows that (A.38) holds for $N = 2$.

Now suppose that $N > 2$ and consider

$$\tilde{A}_2 = \begin{pmatrix} A_2 & 0 & \cdots & 0 \\ 0 & A_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_N \end{pmatrix}$$

as a single block matrix. The A-matrix can then compactly be written as the two-block matrix

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & \tilde{A}_2 \end{pmatrix}.$$

Let us similarly partition Y as follows:

$$Y = \begin{pmatrix} Y_1 & \tilde{Y}_{12} \\ \tilde{Y}_{12}^T & \tilde{Y}_2 \end{pmatrix}.$$

Since (A.38) holds for two blocks we infer that

$$\phi\left(\begin{pmatrix} Y_1 & \tilde{Y}_{12} \\ \tilde{Y}_{12}^T & \tilde{Y}_2 \end{pmatrix}, A\right) \geq \phi\left(\begin{pmatrix} Y_1 & 0 \\ 0 & \tilde{Y}_2 \end{pmatrix}, A\right). \tag{A.43}$$

For the right-hand side of this equation the following holds true

$$\phi\left(\begin{pmatrix} Y_1 & 0 \\ 0 & \tilde{Y}_2 \end{pmatrix}, A\right) =$$
$$= -\log(\det(-(A_1^T Y_1 + Y_1 A_1)) \det(-(\tilde{A}_2^T \tilde{Y}_2 + \tilde{Y}_2 \tilde{A}_2))) - \log(\det(Y_1) \det(\tilde{Y}_2)) =$$
$$= -\log(\det(-(A_1^T Y_1 + Y_1 A_1))) - \log(\det(Y_1)) - \log(\det(-(\tilde{A}_2^T \tilde{Y}_2 + \tilde{Y}_2 \tilde{A}_2))) - \log(\det(\tilde{Y}_2)) =$$
$$= \phi(Y_1, A_1) + \phi(\tilde{Y}_2, \tilde{A}_2),$$

such that we can proceed by partitioning $\tilde{A}_2$ in two smaller blocks. Doing this for all N blocks finally implies (A.38).

BMI for synthesis

Now let us consider the BMI for controller synthesis

$$\phi(Y, A, B_2, C_2) := -\log(\det(-((A + B_2 K C_2)^T Y + Y(A + B_2 K C_2)))) - \log(\det(Y)). \tag{A.44}$$

If we transform A, B, C, and Y with a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ into

$$\tilde{A} = T^{-1}AT, \quad \tilde{B}_2 = T^{-1}B_2, \quad \tilde{C}_2 = C_2 T, \quad \tilde{Y} = T^T Y T,$$

then

$$T^T \left((A + B_2KC_2)^TY + Y(A + B_2KC_2) \right) T = (\tilde{A} + \tilde{B}_2K\tilde{C}_2)^T\tilde{Y} + \tilde{Y}(\tilde{A} + \tilde{B}_2K\tilde{C}_2).$$

The constraints in terms of the new variables are obviously feasible, i.e.

$$(\tilde{A} + \tilde{B}_2K\tilde{C}_2)^T\tilde{Y} + \tilde{Y}(\tilde{A} + \tilde{B}_2K\tilde{C}_2) \prec 0 \tag{A.45}$$

and

$$\tilde{Y} \succ 0 \tag{A.46}$$

If $P(\cdot)$ is the projection in (A.37) corresponding to the block structure of $\tilde{A} + \tilde{B}_2K\tilde{C}_2$, then, by a similar argument as above, it can be shown that (A.45) and (A.46) also hold for the projected $P(\tilde{Y})$, i.e. that

$$(\tilde{A} + \tilde{B}_2K\tilde{C}_2)^T P(\tilde{Y}) + P(\tilde{Y})(\tilde{A} + \tilde{B}_2K\tilde{C}_2) \prec 0 \tag{A.47}$$

and

$$P(\tilde{Y}) \succ 0. \tag{A.48}$$

Log-barrier optimization of stability constraint

We are now in the position to derive the Newton step equations for the log-barrier function for the constraints in (7.45) (or equivalently in (A.45) and (A.48)), that is given by

$$\phi(Y, K) = -\log(\det(-(A + BKC)^TY - Y(A + BKC))) - \log(\det(Y)), \tag{A.49}$$

where we write in this section B and C instead of B_2 and C_2 for ease of notation. Using the notation

$$L := -(A + BKC)^TY - Y(A + BKC) \tag{A.50}$$

and

$$d_L(d_Y, d_K) := -(A + BKC)^T d_Y - d_Y(A + BKC) - (Bd_KC)^TY - YBd_KC, \tag{A.51}$$

we can write the barrier $\phi(Y, K)$ and its derivatives as

$$\phi(Y, K) = -\log(\det(L)) - \log(\det(Y))$$
$$\partial_Y\phi(Y, K)(d_Y) = -\text{Trace}(L^{-1}d_L(d_Y, 0) + Y^{-1}d_Y) =$$
$$= \text{Trace}\left(\left[L^{-1}A_{\text{cl}}^T + A_{\text{cl}}L^{-1} - Y^{-1} \right] d_Y \right)$$
$$\partial_Y\phi(Y, K)(d_K) = \text{Trace}\left(B^TYL^{-1}C^Td_K^T + CL^{-1}YBd_K \right)$$
$$\partial_Y^2\phi(Y, K)(d_{Y1}, d_{Y2}) = \text{Trace}(L^{-1}d_L(d_{Y1}, 0)L^{-1}d_L(d_{Y2}, 0) + Y^{-1}d_{Y1}Y^{-1}d_{Y2})$$
$$\partial_Y\partial_K\phi(Y, K)(d_{Y1}, d_{K2}) = \text{Trace}\left(L^{-1}d_L(d_{Y1}, 0)L^{-1}d_L(0, d_{K2}) \right) +$$
$$+ \text{Trace}\left(Bd_{K2}CL^{-1}d_{Y1} + d_{Y1}L^{-1}(Bd_{K1}C)^T \right)$$
$$\partial_K^2\phi(Y, K)(d_{K1}, d_{K2}) = \text{Trace}\left(L^{-1}d_L(0, d_{K1})L^{-1}d_L(0, d_{K2}) \right),$$

where $\partial_Y^2\phi(Y, K)(d_{Y1}, d_{Y2})$ satisfies the following identity

$$\partial_Y^2\phi(Y, K)(d_{Y1}, d_{Y2}) =$$
$$= \text{Trace}\left(\left[-L^{-1}d_L(d_{Y1}, 0)L^{-1}A_{\text{cl}}^T - A_{\text{cl}}L^{-1}d_L(d_{Y1}, 0)L^{-1} + Y^{-1}d_{Y1}Y^{-1} \right] d_{Y2} \right).$$

Similarly, for $\partial_Y \partial_K \phi(Y, K)(d_{Y1}, d_{K2})$ the relation

$$\partial_Y \partial_K \phi(Y, K)(d_{Y1}, d_{K2}) = \text{Trace}\left(L^{-1}\left[YBd_KC + (Bd_KC)^TY\right]L^{-1}A_{\text{cl}}^Td_Y\right) +$$
$$+ \text{Trace}\left(Bd_KCL^{-1}d_Y\right) + \text{Trace}\left(A_{\text{cl}}L^{-1}\left[YBd_KC + (Bd_KC)^TY\right]L^{-1}d_Y\right)$$

holds true. The Newton step d_Y is the solution to

$$\partial^2\phi(Y, K)(d_Y, d_K) = -\partial\phi(Y, K)),$$

which with the partial derivatives derived above is equivalent to

$$-L^{-1}d_L(d_Y, 0)L^{-1}A_{\text{cl}}^T - A_{\text{cl}}L^{-1}d_L(d_Y, 0)L^{-1} + Y^{-1}d_YY^{-1} = Q_Y(d_K), \qquad \text{(A.52)}$$
$$-B^TYL^{-1}d_L(0, d_K)L^{-1}C^T - CL^{-1}d_L(0, d_K)L^{-1}BY = Q_K(d_Y), \qquad \text{(A.53)}$$

where

$$Q_Y(d_K) = -L^{-1}A_{\text{cl}}^T - A_{\text{cl}}L^{-1} + Y^{-1} -$$
$$- L^{-1}d_L(0, d_K)L^{-1}A_{\text{cl}}^T - A_{\text{cl}}L^{-1}d_L(0, d_K)L^{-1} + Bd_KCL^{-1} + L^{-1}(Bd_KC)^T$$

and

$$Q_K(d_Y) = -2CL^{-1}YB + 2CL^{-1}d_L(d_Y, 0)L^{-1}YB - CL^{-1}d_YB.$$

Let us first solve (A.52) for d_Y with $d_K = E_i$ and some fixed $i \in \{0, 1, \ldots, m_{\text{c}}\}$, where $\{E_1, \ldots, E_{m_{\text{c}}}\}$ is a basis for $\mathbb{R}^{(n+n_{\text{c}}+m_2) \times (n+n_{\text{c}}+p_2)}$ and $E_0 = 0$. (Note that a similar approach has been used in Section 7.2.4.)

If Y and $A_{\text{cl}} := A + BKC$ have block-diagonal structure, this equation can be solved efficiently as follows. Suppose that

$$A_{\text{cl}} = \begin{pmatrix} A_{\text{cl}1} & 0 & \cdots & 0 \\ 0 & A_{\text{cl}2} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & A_{\text{cl}N} \end{pmatrix}$$

with $A_{\text{cl}k} \in \mathbb{R}^{m_k \times m_k}$, $k = 1, \ldots, N$ and

$$Y = \begin{pmatrix} Y_1 & 0 & \cdots & 0 \\ 0 & Y_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & Y_N \end{pmatrix},$$

with $Y_k \in \mathcal{S}^{m_k}$, $k = 1, \ldots, N$, respectively. This implies that L has similar block-diagonal structure

$$L = \begin{pmatrix} L_1 & 0 & \cdots & 0 \\ 0 & L_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & L_N \end{pmatrix},$$

with $L_k \in \mathcal{S}^{m_k}$, $k = 1, \ldots, N$, Observe that the block-diagonal structure of A_{cl}, Y and L imply that

$$
\begin{aligned}
F_k^T A_{\mathrm{cl}} &= A_{\mathrm{cl}k} F_k^T \\
A_{\mathrm{cl}} F_k &= F_k A_{\mathrm{cl}k} \\
F_k^T Y &= Y_k F_k^T \\
Y F_k &= F_k Y_k \\
F_k^T L &= L_k F_k^T \\
L F_k &= F_k L_k,
\end{aligned}
$$

where F_k, $k = 1, \ldots, N$ are as defined in (A.36). Hence, for arbitrary $k, l = 1, \ldots, N$ we infer

$$
(d_L(dY, 0))_{kl} := F_k^T d_L(dY, 0) F_l = -F_k^T A_{\mathrm{cl}}^T d_Y F_l - F_k^T d_Y A_{\mathrm{cl}} F_l = -A_{\mathrm{cl}k}^T d_{Ykl} - d_{Ykl} A_{\mathrm{cl}l},
$$

where $d_{Ykl} := F_k^T d_Y F_l$. Furthermore, if we multiply (A.52) from left and right by F_k^T and F_l, respectively, we obtain

$$
\begin{aligned}
F_k^T(-L^{-1} d_L(dY, 0) L^{-1} A_{\mathrm{cl}}^T - A_{\mathrm{cl}} L^{-1} d_L(dY, 0) L^{-1} + Y^{-1} d_Y Y^{-1}) F_l = \\
= -L_k^{-1}(d_L(dY, 0))_{kl} L_l^{-1} A_{\mathrm{cl}l} - A_{\mathrm{cl}k} L_k^{-1}(d_L(dY, 0))_{kl} L_l^{-1} + Y_k^{-1} d_{Ykl} Y_l^{-1} = F_k^T Q_Y(d_K) F_l.
\end{aligned}
\tag{A.54}
$$

The only unknown in (A.54) is d_{Ykl} and this equation can hence be solved for d_{Ykl}, independently of the other variables in d_Y. Since $k, l \in \{1, \ldots, N\}$ were arbitrary, this shows that (A.52) can be solved block-wise.

Procedure for the Newton step computation

Finally, we present the procedure to compute the Newton directions (d_Y, d_K) in the corrector step of the Interior Point algorithm, if we use the barrier as in (A.49). First we transform the closed-loop matrix A_{cl} to block diagonal form. Then, as in Section 7.2.4, we write $d_K = \sum_{i=1}^{m_c} k_i E_i$, where $\{E_1, \ldots, E_{m_c}\}$ is a basis for $\mathbb{R}^{(n+n_c+m_2) \times (n+n_c+p_2)}$ and $E_0 = 0$, and we compute the solution to (A.52) for $d_K = E_i$ for each $i \in \{0, \ldots, m_c\}$, using the block-wise solution procedure described in the previous section. Let $Z_i, i = 1, \ldots, n_p$ and Z_0 be the solutions of (A.52) for $d_K = E_i, i = 1, \ldots, n_p$ and $d_K = 0$ respectively. We express the solution to (A.52) as

$$
d_Y(d_{k1}, \ldots, d_{km_c}) = Z_0 + \sum_{i=1}^{m_c} Z_i d_{ki}.
$$

This can be substituted in (A.53), and this equation can be solved for d_{ki}, $i = 1, \ldots, m_c$ by scalarization. The procedure is more precisely explained in the sequel.

Suppose that at the lth main iteration and k^{th} corrector step of Algorithm 7.1.5 we have Y and K as decision variables. To find the Newton step of d_Y and d_K we perform the following steps

1. find a nonsingular $T \in \mathbb{R}^{(n+n_c) \times (n+n_c)}$ with $\det(T) = 1$, such that $\tilde{A}_{\mathrm{cl}} = T^{-1} A_{\mathrm{cl}} T$ is in real-Jordan form [94]

2. find the projection $P(\cdot)$ as in (A.37)

3. replace Y by $P(T^T Y T)$, B_{cl} by $T^{-1} B_{\text{cl}}$ and C_{cl} by $C_{\text{cl}} T$. Note that the value of the barrier function is not increased by this replacement.

4. introduce a basis $\{E_1, \ldots, E_{m_c}\}$ for d_K and parameterize d_K as $d_K(k_i) = \sum_{i=1}^{m_c} k_i E_i$. Set $E_0 = 0$.

5. solve for each $i = 0, 1, \ldots, m_c$ (A.52) with right-hand side $Q_Y(E_i)$. Do this block-wise using (A.54) for all $k, l \in \{1, \ldots, N\}$. Denote the solution of (A.52) by Z_i

6. find the coefficients k_i by solving the Sylvester equation for $d_K(k_i)$ by scalarization.

7. compute the Newton directions $d_K = \sum_{i=1}^{m_c} k_i E_i$ and $d_Y = Z_0 + \sum_{i=1}^{m_c} k_i Z_i$.

A.5   Results for Chapter 8

A.5.1   Proof that $d_{T11} = 0$

Let e_j be the jth basis vector in $\mathbb{R}^{n_c}$, $j = 1, \ldots, n_c$ and A_K be as in (8.47). We need to show that
$$-d_T A_K + A_K d_T = e_{n_c} d_a^T \text{ and } d_T e_{n_c} = 0$$
has the unique solution $d_T = 0$ and $d_a = 0$. With $a^T = \begin{pmatrix} -a_{n_c} & \cdots & -a_2 & -a_1 \end{pmatrix} \in \mathbb{R}^{n_c}$ and
$$N = \begin{pmatrix} 0_{(n_c-1)\times 1} & I_{n_c-1} \\ 0 & 0_{1\times(n_c-1)} \end{pmatrix}, \tag{A.55}$$
we write A_K as $A_K = N + e_{n_c} a^T$. The result to be shown is formulated in the following lemma.

Lemma A.3 *Let $A_K = N + e_{n_c} a^T$ for arbitrary $a \in \mathbb{R}^{n_c}$, N as in (A.55) and e_{n_c} be the n_c^{th} basis vector in $\mathbb{R}^{n_c}$. Then the set of equations*
$$-d_T A_K + A_K d_T = e_{n_c} d_a^T \text{ and } d_T e_{n_c} = 0 \tag{A.56}$$
has the unique solution $d_T = 0$ and $d_a = 0$.

Proof. Observe that (A.56) implies that
$$-d_T N + N d_T = e_{n_c}(d_a^T - a^T d_T) + d_T e_{n_c} a^T = e_{n_c}(d_a^T - a^T d_T).$$

If we define $d_g := (d_a^T - a^T d_T)^T \in \mathbb{R}^{n_c}$, this equation simplifies to
$$-d_T N + N d_T = e_{n_c} d_g^T.$$

Hence, it suffices to show that the system of equations
$$-d_T N + N d_T = e_{n_c} d_g^T \tag{A.57}$$
$$d_T e_{n_c} = 0 \tag{A.58}$$

has a unique solution $d_T = 0$ and $d_g = 0$.

First, we will show that d_T is upper triangular. This will be done by an induction argument. Let us define

$$E_j := \left(\begin{array}{cc} I_{n_c-j} & 0_{(n_c-j)\times j} \end{array} \right) \in \mathbb{R}^{(n_c-j)\times n_c}$$

for $j = 1, \ldots, n_c - 1$, where $0_{(n_c-j)\times j}$ denotes the zero matrix in $\mathbb{R}^{(n_c-j)\times j}$. Observe that

$$E_j N^{j-1} = \left(\begin{array}{ccc} 0_{(n_c-j)\times(j-1)} & I_{n_c-j} & 0_{(n_c-j)\times 1} \end{array} \right)$$

such that

$$E_k N^{k-1} e_{n_c} = 0 \quad \text{for all} \quad k \in \{1, \ldots, n_c - 1\}. \tag{A.59}$$

Multiplying (A.57) from the left by E_1 and from the right by e_1 implies that

$$-E_1 d_T N e_1 + E_1 N d_T e_1 = E_1 e_{n_c} d_g{}^T e_1 \tag{A.60}$$

Since $N e_1 = 0$ and $E_1 e_{n_c} = 0$, we conclude that

$$E_1 N d_T e_1 = 0. \tag{A.61}$$

As the induction hypothesis, let us assume that

$$E_j N^j d_T e_j = 0 \tag{A.62}$$

for some $j \in \{1, \ldots n_c - 2\}$. By multiplying (A.57) from the left by $E_{j+1} N^j$ and from the right by e_{j+1} we conclude that

$$-E_{j+1} N^j d_T N e_{j+1} + E_{j+1} N^{j+1} d_T e_{j+1} = E_{j+1} N^j e_{n_c} d_g{}^T e_{j+1} = 0, \tag{A.63}$$

where the utmost right equality follows from (A.59).

Furthermore,

$$E_{j+1} N^j d_T N e_{j+1} = E_{j+1} N^j d_T e_j = 0,$$

because it is a subset of the rows of $E_j N^j d_T e_j$, which is zero by (A.62). From (A.61) we conclude that the induction hypothesis (A.62) is true for $j = 1$, such that by induction we conclude that

$$E_k N^k d_T e_k = 0 \quad \text{for all} \quad k \in \{1, \ldots, n_c - 1\}, \tag{A.64}$$

which shows that d_T is upper triangular.

As a second step, let us multiply (A.57) from the left by E_1 and from the right by e_{n_c} to infer that

$$-E_1 d_T N e_{n_c} + E_1 N d_T e_{n_c} = 0. \tag{A.65}$$

Since $d_T e_{n_c}$, we conclude $E_1 d_T e_{n_c-1} = 0$. Since $E_{n_c-1} N^{n_c-1} d_T e_{n_c-1} = 0$ by (A.64), we actually infer $d_T e_{n_c-1} = 0$. As induction hypothesis we assume $d_T e_j = 0$ for some $j \in \{2, \ldots n_c - 1\}$. Then multiplying (A.57) from the left by E_1 and from the right by e_j implies

$$-E_1 d_T N e_j + E_1 N d_T e_j = 0. \tag{A.66}$$

Since $E_1 N d_T e_j = 0$ holds true because of the induction hypothesis, we conclude $E_1 d_T e_{j-1} = 0$. Together with $E_{j-1} N^{j-1} d_T e_{j-1} = 0$ from (A.64) this implies $d_T e_{j-1} = 0$. Hence we conclude $d_T e_j = 0$, $j = 1, \ldots, n_c - 1$ and thus $d_T = 0$. This implies $d_g = 0$, which completes the proof.

∎

A.5.2 Proof of $\mathbf{Im}(\partial g(x_1, I)) = \mathbf{Im}(\partial g(x_0, y_0))$

Let $x_0 = (\gamma_0, X_0, K_0)$, $x_1 = (\gamma_1, X_1, K_1)$ and $T_0 \in \mathbb{R}^{n_c \times n_c}$ be arbitrary such that $\det(T_0) = 1$ and $g(x_0, T_0) = x_1$. The function g given in (8.46) and is repeated here for the reader's convenience:

$$g(x, T) := \left(\gamma, \mathrm{svec}\left(X^{\mathrm{trans}}(X, T)\right), \mathrm{vec}\left(K^{\mathrm{trans}}(K, T)\right)\right),$$

where

$$X^{\mathrm{trans}}(X, T) := \begin{pmatrix} I_n & 0 \\ 0 & T \end{pmatrix}^T X \begin{pmatrix} I_n & 0 \\ 0 & T \end{pmatrix}$$

and

$$K^{\mathrm{trans}}(K, T) := \begin{pmatrix} T & 0 \\ 0 & I_{m_2} \end{pmatrix}^{-1} K \begin{pmatrix} T & 0 \\ 0 & I_{p_2} \end{pmatrix}.$$

We need to show that $\mathrm{Im}(\partial g(x_1, I_{n_c})) = \mathrm{Im}(\partial g(x_0, T_0))$. Let us partition K_0 and K_1 into its state-space matrices

$$K_0 =: \begin{pmatrix} A_{K_0} & B_{K_0} \\ C_{K_0} & D_{K_0} \end{pmatrix}$$

and

$$K_1 =: \begin{pmatrix} A_{K_1} & B_{K_1} \\ C_{K_1} & D_{K_1} \end{pmatrix},$$

where $A_{K_0}, A_{K_1} \in \mathbb{R}^{n_c \times n_c}$. The identity $g(x_0, T_0) = x_1$ implies that

$$X^{\mathrm{trans}}(X_0, T_0) \;=\; X_1 \tag{A.67}$$

$$K^{\mathrm{trans}}(K_0, T_0) \;=\; \begin{pmatrix} A_K^{\mathrm{trans}}(K_0, T_0) & B_K^{\mathrm{trans}}(K_0, T_0) \\ C_K^{\mathrm{trans}}(K_0, T_0) & D_K^{\mathrm{trans}}(K_0, T_0) \end{pmatrix} = K_1 \tag{A.68}$$

where A_K^{trans}, B_K^{trans}, C_K^{trans} and D_K^{trans} are as defined in (8.4). Recall from (8.12) that the first-order derivative of X^{trans} at arbitrary $X_* \in \mathcal{S}^{n+n_c}$ and $T_* \in \mathbb{R}^{n_c \times n_c}$, $\det(T_*) \neq 0$ reads as

$$\partial_T X^{\mathrm{trans}}(X_*, T_*)(d_T) = d_T^T X_* T_* + T_*^T X_* d_T.$$

For arbitrary $T_* \in \mathbb{R}^{n_c \times n_c}$ with $\det(T_*) \neq 0$, the derivatives of A_K^{trans} and B_K^{trans} with respect to T are given in (8.13) and (8.14) and are repeated here:

$$\partial_T A_K^{\mathrm{trans}}(K_*, T_*)(d_T) \;=\; -T_*^{-1} d_T T_*^{-1} A_{K_*} T_* + T_*^{-1} A_{K_*} d_T$$

$$\partial_T B_K^{\mathrm{trans}}(K_*, T_*)(d_T) \;=\; -T_*^{-1} d_T T_*^{-1} B_{K_*},$$

where and

$$K_* := \begin{pmatrix} A_{K_*} & B_{K_*} \\ C_{K_*} & D_{K_*} \end{pmatrix} \in \mathbb{R}^{(n_c+m_2) \times (n_c+p_2)}.$$

Furthermore, the derivatives of C_K^{trans} and D_K^{trans} are as follows

$$\partial_T C_K^{\mathrm{trans}}(K_*, T_*)(d_T) \;=\; C_{K_*} d_T$$

$$\partial_T D_K^{\mathrm{trans}}(K_*, T_*)(d_T) \;=\; 0.$$

Using these derivatives (A.67) and (A.68), we observe that

$$\partial_T A_K^{\mathrm{trans}}(K_0, T_0)(d_T) \;=\; -T_0^{-1} d_T T_0^{-1} A_{K_0} T_0 + T_0^{-1} A_{K_0} d_T =$$
$$= -(T_0^{-1} d_T) A_{K_1} + A_{K_1} T_0^{-1} d_T = \partial_T A_K^{\mathrm{trans}}(K_1, I)(T_0^{-1} d_T)$$

245

and

$$
\begin{aligned}
\partial_T B_K^{\text{trans}}(K_0, T_0)(d_T) &= -T_0^{-1} d_T T_0^{-1} B_{K_0} = -(T_0^{-1} d_T) B_{K_1} = \partial_T B_K^{\text{trans}}(K_1, I)(T_0^{-1} d_T) \\
\partial_T C_K^{\text{trans}}(K_0, T_0)(d_T) &= C_{K_0} d_T = C_{K_1}(T_0^{-1} d_T) = \partial_T C_K^{\text{trans}}(K_1, I)(T_0^{-1} d_T) \\
\partial_T D_K^{\text{trans}}(K_0, T_0)(d_T) &= 0 = \partial_T D_K^{\text{trans}}(K_1, I)(T_0^{-1} d_T)
\end{aligned}
$$

and finally

$$
\begin{aligned}
\partial_T X^{\text{trans}}(X_0, T_0)(d_T) = d_T{}^T X_0 T_0 + T_0^T X_0 d_T &= \\
&= d_T{}^T T_0^{-T} X_1 + X_1 T_0^{-1} d_T = \partial_T X^{\text{trans}}(X_1, I)(T_0^{-1} d_T).
\end{aligned}
$$

We conclude $x^{\text{trans}}(K_0, T_0)(d_T) = x^{\text{trans}}(K_1, I)(T_0^{-1} d_T)$ for all d_T, such that indeed

$$
\text{Im}(\partial g(x_1, I)) = \text{Im}(\partial g(x_0, T_0)).
$$

Appendix B

Fixed-order controllers for the active suspension system

The discrete-time transfer functions of the controllers are:

$$K_{\mathbf{Balreal}}(z) \;\; := \;\; 0.01\left(\frac{z+1}{z}\right)\frac{1.402z^5 - 2.769z^4 + 3.126z^3 - 1.885z^2 + 0.1928z + 0.05143}{z^5 - 3.161z^4 + 4.657z^3 - 4.29z^2 + 2.355z - 0.5605}$$

$$K_{\mathbf{CC}}(z) \;\; := \;\; 0.01\left(\frac{z+1}{z}\right)\frac{2.402z^5 - 7.781z^4 + 10.11z^3 - 6.653z^2 + 2.254z - 0.3242}{z^5 - 3.795z^4 + 5.69z^3 - 4.205z^2 + 1.524z - 0.2135}$$

$$K_{\mathbf{CLIP}}(z) \;\; := \;\; 0.01\left(\frac{z+1}{z}\right)\frac{1.331z^5 - 2.357z^4 + 2.413z^3 - 1.427z^2 + 0.1925z + 0.01808}{z^5 - 2.991z^4 + 4.071z^3 - 3.48z^2 + 1.841z - 0.4377}$$

The phase of their frequency responses is shown in Figure B.1.
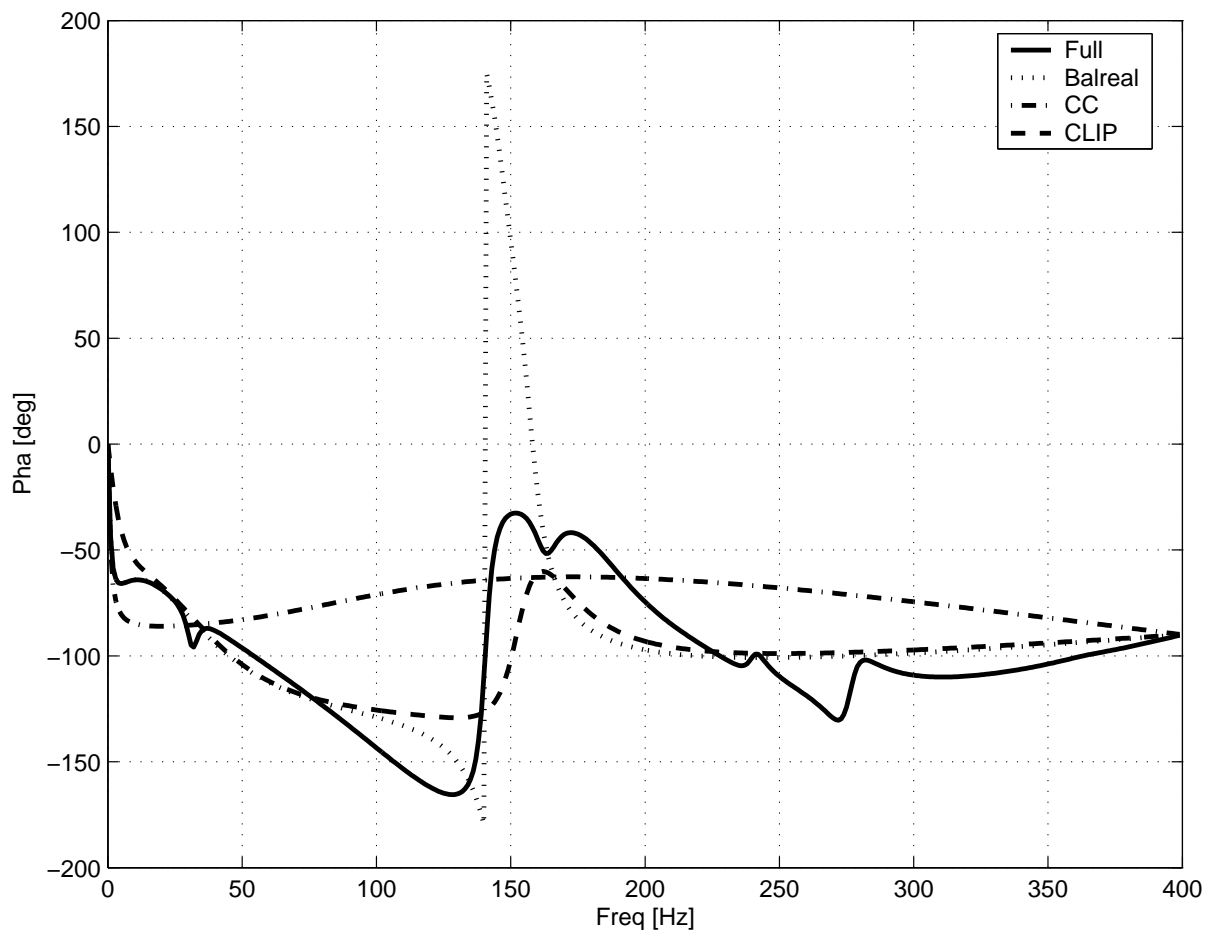
Figure B.1: Controllers: phase of the frequency response

Appendix C

Frequency responses of controller design for the wafer stage

C.1 MIMO controllers
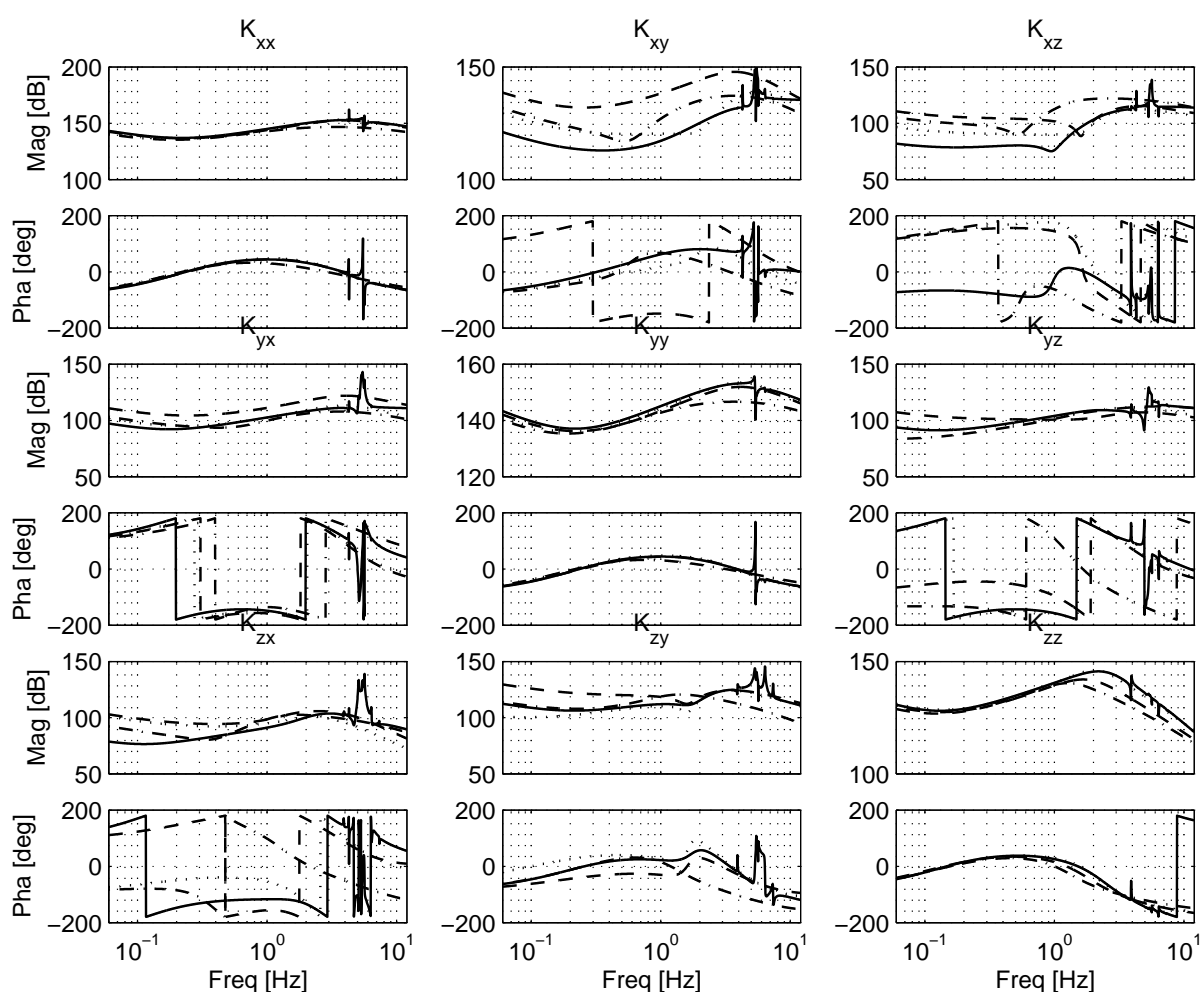
Figure C.1: Frequency response of the MIMO controllers: full order solid, reduced dotted, IP optimized dashed and Simplex optimized dash-dotted.

C.2   Frequency responses of SG, KS and KSG for full order MIMO and diagonal SISO controllers
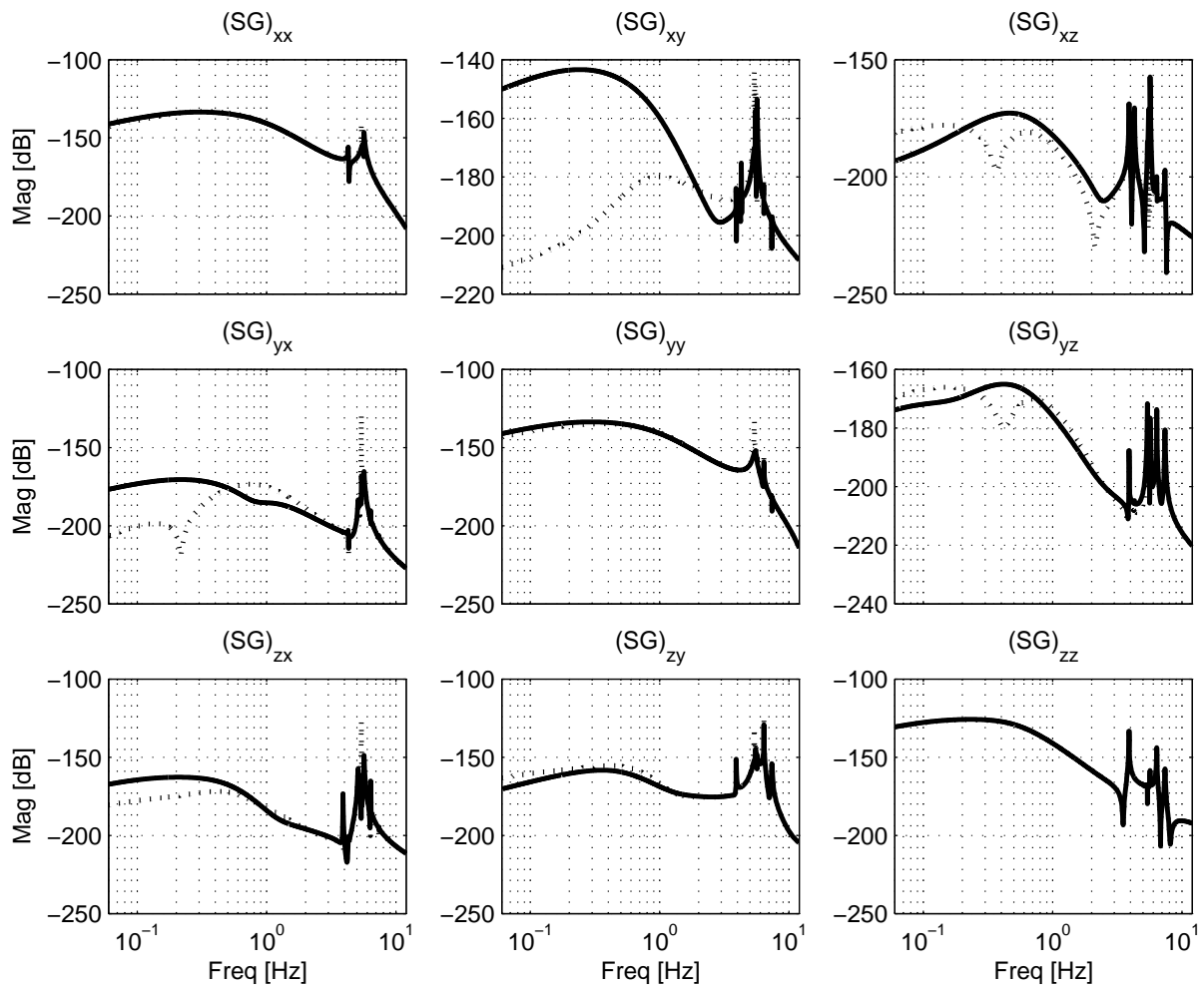
Figure C.2: Frequency response of the process sensitivity of the full order controllers. Solid: full MIMO controller, Dotted: diagonally augmented SISO controllers
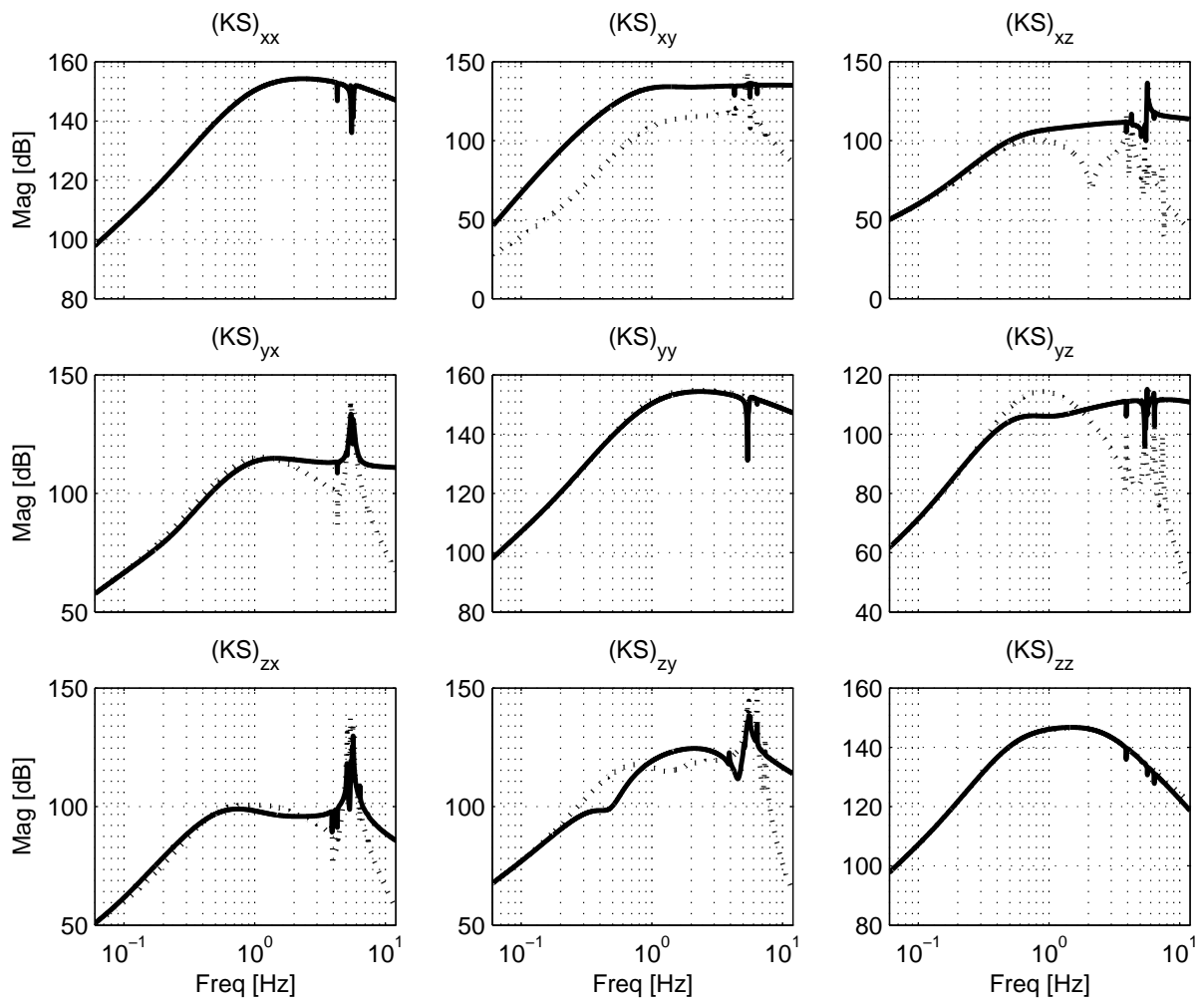
Figure C.3: Frequency response of the controller sensitivity of the full order controllers. Solid: full MIMO controller, Dotted: diagonally augmented SISO controllers
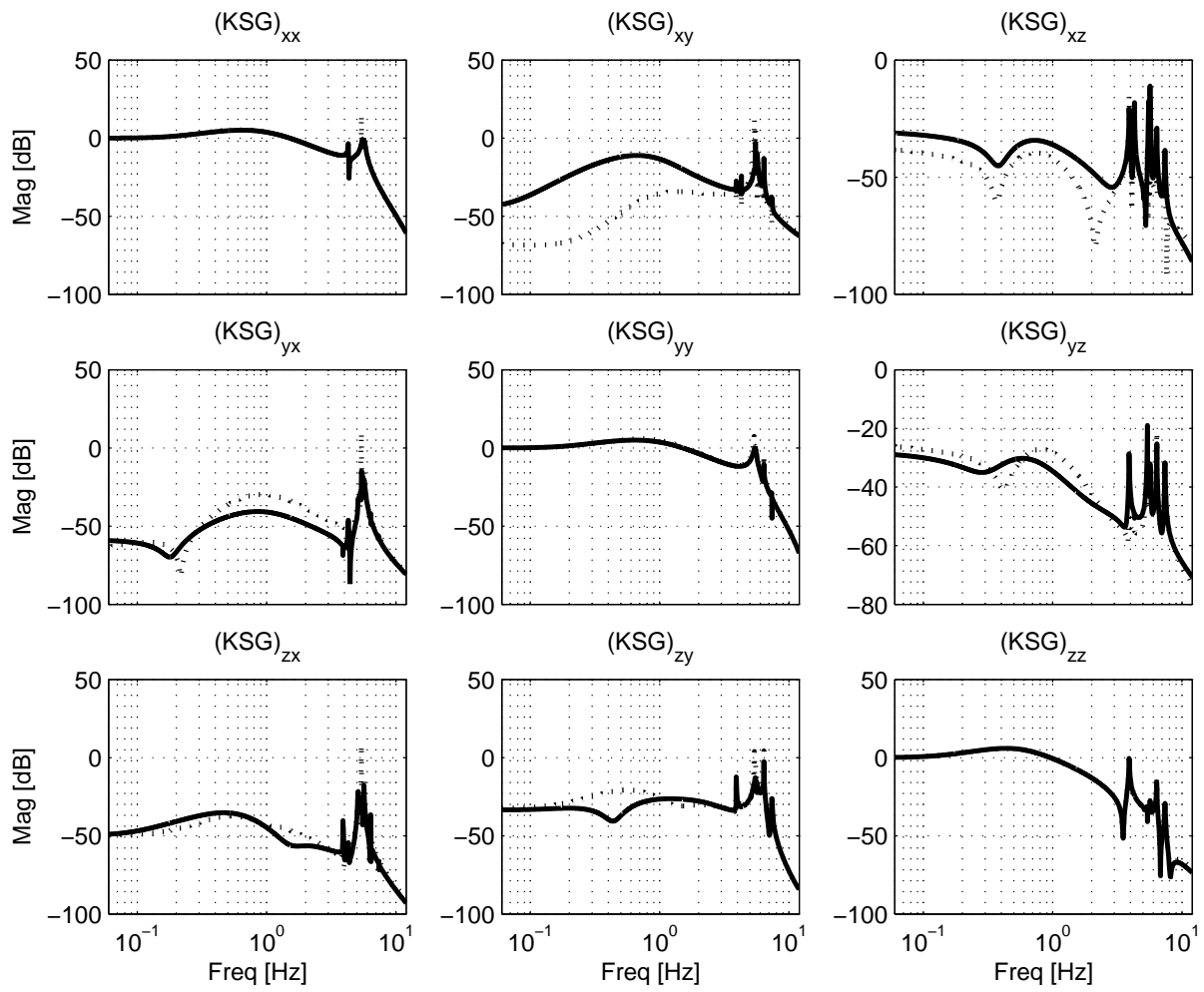
Figure C.4: Frequency response of the controller sensitivity of the full order controllers. Solid: full MIMO controller, Dotted: diagonally augmented SISO controllers

Appendix D

Modifications to Simplex algorithm

In this appendix we present a description of our modifications to the Simplex algorithm as explained in Section 3.4, that speed up the convergence.

The Simplex algorithm minimizes the function $f_p(x)$ as given in (3.21). This function is evaluated $3m + 1$ times at each iteration of Algorithm 3.6. These function evaluations are the computationally most demanding steps. It is therefore important to reduce the computation time of them as much as possible.

After a few iterations the step-sizes tend to be very small. This implies that the computational cost of the function evaluation can possible be reduced using the results from previous evaluations.

Suppose the algorithm evaluates f_p at x_{eval}. If $f_p(x_{\text{eval}})$ is larger than the best value p_{current} obtained thus far, it is not necessary in the algorithm to compute $f_p(x_{\text{eval}})$ exactly: we only need to know that $f_p(x_{\text{eval}}) > p_{\text{current}}$ to decline x_{eval} as the better controller. In practice, the number of declinations exceeds by large the number of times a new minimum point is found as the algorithm proceeds. This motivates to make the declinations more efficient as follows.

Suppose that for some $K \in \mathbb{R}^{(n_c+m_2)\times(n_c+p_2)}$, the closed-loop matrix $A_{\text{cl}}(K)$ is strictly stable and $\mathcal{H}_\infty$-norm of the closed loop $G_{\text{cl}}(P, K)$ is attained at a certain frequency ω. Since $\sigma_{\max}(G_{\text{cl}}(P, K)(i\omega))$ depends continuously on the element in the matrix K, the $\mathcal{H}_\infty$-norm of the closed-loop $G_{\text{cl}}(P, K_{\text{eval}})$ will often be close to $\|G_{\text{cl}}(P, K)\|_\infty$, if $\|K - K_{\text{eval}}\|$ is small. It will therefore often be the case that $\sigma_{\max}(G_{\text{cl}}(P, K_{\text{eval}})(i\omega)) > p_{\text{current}}$. This motivates to evaluate $\sigma_{\max}(G_{\text{cl}}(P, K_{\text{eval}})(i\omega))$ and compare it with γ. If $\gamma < \sigma_{\max}(G_{\text{cl}}(P, K_{\text{eval}})(i\omega))$, then K_{eval} can be declined as a possible better controller without computation of its $\mathcal{H}_\infty$-norm. This can save significant computation time, since the computation of $\sigma_{\max}(G_{\text{cl}}(P, K)(\omega))$ is in general cheaper than the computation of the $\mathcal{H}_\infty$-norm of the closed loop, i.e. $\|G_{\text{cl}}(P, K)\|_\infty$.

Appendix E

Generalized Plant of the two-mass system

A state-space representation of the two-mass system is as follows:

$$
A = \begin{pmatrix}
-8.216 \cdot 10^{-5} & 0 & 0 & 0 & -0.9064 & 0 & 0 \\
0 & -4502 & -3355 & 0 & 0 & 0 & 0 \\
0 & 3355 & -1249 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -8.216 \cdot 10^{-5} & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & -8.216 \cdot 10^{-5} & 0 & 1 \\
0 & 0 & 0 & -1500 & 1500 & -0.5001 & 0.5 \\
0 & 0 & 0 & 1500 & -1500 & 0.5 & -0.5001
\end{pmatrix}
$$

$$
\left(\begin{array}{c|c} B_1 & B_2 \end{array} \right) = \left(\begin{array}{cc|c}
-0.9064 & 0 & 0 \\
0 & 0 & -482.8 \\
0 & 0 & 131.6 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 132 & 1 \\
0 & 0 & 0
\end{array} \right)
$$

$$
\left(\frac{C_1}{C_2} \right) = \left(\begin{array}{ccccccc}
0.9064 & 0 & 0 & 0 & -0.5 & 0 & 0 \\
0 & 482.8 & 131.6 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & -1 & 0 & 0
\end{array} \right)
$$

$$
\left(\begin{array}{c|c} D_{11} & D_{12} \\ \hline D_{21} & D_{22} \end{array} \right) = \left(\begin{array}{cc|c}
-0.5 & 0 & 0 \\
0 & 0 & 37.89 \\
\hline
-1 & 0 & 0
\end{array} \right)
$$

Bibliography

[1] U.M Al-Saggaf and G.F. Franklin. Model reduction via balanced realization: An extension and frequency weighting techniques. *IEEE Transactions on Automatic Control*, 33(7):687–692, 1998.

[2] B.D.O. Anderson and J.B. Moore. *Linear Optimal Control*. Englewood Cliffs, New Jersey, 1971.

[3] P. Apkarian and D. Noll. Controller Design via Nonsmooth Multidirectional Search. *SIAM J. Contr. Optim.*, 44(6):1923–1949, 2005.

[4] P. Apkarian, D. Noll, and H.D. Tuan. Fixed-order $\mathcal{H}_\infty$ control design via a partially augmented Lagrangian method. *International Journal of Robust and Nonlinear Control*, 13:1137–1148, 2003.

[5] M. Athans and F.C. Schweppe. Gradient Matrices and Matrix Calculations. Technical report, MIT Lincoln Laboratory, 1965.

[6] V. Balakrishan and L. Vandenberghe. Semidefinite Programming Duality and Linear Time-Invariant Systems. *IEEE Transactions on Automatic Control*, 48(1):30–41, 2003.

[7] G.J. Balas, J.C. Doyle, K. Glover, A. Packard, and R. Smith. *μ-analysis and synthesis toolbox*. The MathWorks Inc. and MYSYNC Inc., 1995.

[8] D. Banjerdpongchai and J.P. How. Convergence analysis of a parametric robust $\mathcal{H}_2$ controller synthesis algorithm. In *Proc. 36th IEEE Conf. Decision and Control*, volume 2, pages 1020 – 1025, San Diego, CA, USA, 1997.

[9] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM-MPS Series in Optimizaton. SIAM Publications, Philadelphia, 2001.

[10] A. Ben-Tal and A. Nemirovski. On tractable approximations of uncertain Linear Matrix Inequalities affected by interval uncertainties. *SIAM J. Optim.*, 12(3):811–833, 2002.

[11] A. Ben-Tal, A. Nemirovski, and C. Roos. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM J. Optim.*, 13(2):535–560, 2002.

[12] P Benner, R. Byers, V. Mehrmann, and H. Xu. A unified deflating subspace approach for classes of polynomial and rational matrix equations. Technical report, Fakultät für Mathematik, University Chemnitz, 2000.

[13] P. Benner, Quintana-Ortí E.S., and Quintana-Ortí G. State-space truncation methods for parallel model reduction of large-scale systems. *Parallel Comput.*, 29(11-12):1701–1722, 2003.

[14] P. Benner, V. Mehrmann, V. Sima, S. van Huffel, and A. Varga. Slicot - a Subroutine Libary in Systems and Control Theory. *Applied and Computational Control, Signals and Circuits*, 1:499–539, 1999.

[15] E. Beran and K. Grigoriadis. Computational issues in alternating projection algorithms for fixed-order control design. In *Proc. Amer. Contr. Conf.*, pages 81–85, Albuquerque, New Mexico, USA, 1997.

[16] D.S Bernstein and W.M Haddad. LQG control with an H_∞ performance bound: A Riccati equation approach. *IEEE Transactions on Automatic Control*, 34:293–305, 1989.

[17] D.P. Bertsekas. *Convex analysis and optimization.* Athena Scientific, Belmont, Massachusetts, 2003.

[18] P.A. Bliman. On robust semidefinite programming. In *Sixteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, Leuven, Belgium, 2004.

[19] P.A. Bliman. Stabilisation of LPV systems. In D. Henrion and A.Garulli, editors, *Positive polynomials in Control*, Springer Lecture Notes in Control and Information Sciences. Springer, 2005.

[20] V.D. Blondel and J.N. Tsitsiklis. NP-hardness of some linear control design problems. *SIAM J. Contr. Optim.*, 35(6):2118–2127, 1997.

[21] V.D. Blondel and J.N. Tsitsiklis. A survey of computational complexity results in Systems and Control. *Automatica*, 36, 2000.

[22] P.M.M. Bongers. *Modeling and Identification of Flexible Wind Turbines, a Factorial Approach to Robust Control Design.* PhD thesis, Delft University of Technology, 1994.

[23] P.M.M. Bongers and O.H. Bosgra. Low order robust $\mathcal{H}_\infty$ controller synthesis. In *Proc. 29th IEEE Conference on Decision and Control*, volume 1, pages 194 – 199, 1990.

[24] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems.* New York, 2000.

[25] O.H. Bosgra and H. Kwakernaak. *Design Methods for Control Systems.* Lecture Notes, Dutch Insititue of Systems and Control, Delft, The Netherlands, 2000.

[26] S.P. Boyd and G.H. Barratt. *Linear Controller Design - Limits of Performance.* Prentice-Hall, Englewood Cliffs, New Jersey, 1991.

[27] S.P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishan. *Linear Matrix Inequalities in System and Control Theory.* SIAM Studies in Applied Mathematics 15. SIAM, Philadelphia, 1994.

[28] S.P. Boyd and L. Vandenberghe. Semidefinite programming relaxations of non-convex problems in control and combinatorial optimization. In A. Paulraj, V. Roychowdhuri, and C. Schaper, editors, *Communications, Computation, Control and Signal Processing: a Tribute to Thomas Kailath*, pages 279–288. Kluwer, 1997.

[29] S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[30] J.W. Brewer. Kronecker products and matrix calculus in system theory. *IEEE Transactions on Circuits and Systems*, 25(9):772–781, 1978.

[31] N.A. Bruinsma and M. Steinbuch. A fast algorithm to compute the $\mathcal{H}_\infty$-norm of a transfer function matrix. *Systems and Control Letters*, 14:287–293, 1990.

[32] J.V. Burke, D. Henrion, A.S. Lewis, and M.L. Overton. HIFOO - a Matlab package for fixed-order controller design and $\mathcal{H}_\infty$ optimization. In *submitted to Rocond 2006*, 2006.

[33] J.V. Burke, A.S. Lewis, and M.L. Overton. A nonsmooth, nonconvex optimization approach to robust stabilization by static output feedback and low-order controllers. In *Rocond*, Milan, Italy, 2003.

[34] J.V. Burke, A.S. Lewis, and M.L. Overton. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM J. Optim.*, 15(3):751–779, 2005.

[35] G. Calafiore, F. Dabbene, and R. Tempo. Randomized algorithms for reduced order $\mathcal{H}_\infty$ controller design. In *Proc. American Control Conf.*, Chicago, IL, USA, 2000.

[36] M. Capiński and P.E. Kopp. *Measure, Integral and Probability*. Springer-Verlag, London, 2003.

[37] G. Chesi, A. Garulli, A. Tesi, and A. Vicino. An LMI-based approach for characterizing the solution set of polynomial systems. In *Proc. 39th IEEE Conf. Decision and Control*, Sydney, Australia, 2000.

[38] G. Chesi, A. Garulli, A. Tesi, and A. Vicino. Homogeneous Lyapunov functions for systems with structured uncertainties. *Automatica*, 39(6):1027–1035, 2003.

[39] G. Chesi, A. Garulli, A. Tesi, and A. Vicino. Robust stability of polytopic systems via polynomially parameter-dependent lyapunov functions. In *42nd IEEE Conference on Decision and Control*, pages 4670–4675, Maui, Hawaii, 2003.

[40] M. Chilali and P. Gahinet. H_∞ design with pole placement constraints: An LMI approach. *IEEE Transactions on Automatic Control*, 41(3):358–367, 1996.

[41] M.D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra Appl.*, 10:285–290, 1975.

[42] Y. Chou, T. Wu, and J. Leu. On strong stabilization and $\mathcal{H}_\infty$ strong stabilization problems. In *Proc. 42nd IEEE Conf. Decision and Control*, Maui, Hawaii, USA, 2003.

[43] A.G. Collins Jr., W.M. Haddad, and S.S. Ying. Reduced-order compensation using the Hyland-Bernstein optimal projection equations. *Journal of Guidance, Control and Dynamics*, 19:407–417, 1996.

[44] E.G. Collins Jr., W.M. Haddad, and S.S. Ying. Reduced-order dynamic compensation using the Hyland-Bernstein optimal projection equations. In *Proc. American Control Conf.*, volume 1, pages 539–543, Seattle, WA, USA, 1995.

[45] D. de Roover. *Motion Control of a Wafer Stage. A Design Approach for Speeding Up IC Production.* Ph.D. thesis, Delft University of Technology, 1997.

[46] M. Dettori. *LMI techniques for control with applications to a Compact Disc player mechanism.* Ph.D. thesis, Delft University of Technology, 2001.

[47] P. Dorato. *Analytic Feedback Design, An Interpolation Approach.* Brooks Cole, 2000.

[48] J. Doyle, B. Francis, and A. Tannenbaum. *Feedback Control Theory.* Macmillan, New York, 1990.

[49] J. Doyle and K. Glover. State-space formulae for all stabilizing controllers that satisfy an $\mathcal{H}_\infty$ norm bound and relations to risk sensitivity. *Syst. Contr. Letters*, 11:167–172, 1988.

[50] J. Doyle, K. Glover, P. Khargonekar, and B. Francis. State-space solutions to standard H_∞ and H_2 control problems. *IEEE Transactions on Automatic Control*, 34:831–847, 1989.

[51] J. Doyle, A. Packard, and K. Zhou. Review of LFT's, LMI's, and μ. In *Proc. 30th IEEE Conf. Decision and Control*, pages 1227–1232, Brighton, UK, 1991.

[52] G.E. Dullerud and F. Paganini. *A course in robust control theory: a convex approach*, volume 36 of *Texts in Applied Mathematics.* Springer-Verlag, New York, 1999.

[53] T.F. Edgar, D.M. Himmelblau, and L.S. Lasdon. *Optimization of Chemical Processes.* McGraw-Hill, New York, 1988.

[54] L. El Ghaoui and V. Balakrishnan. Synthesis of fixed-structure controllers via numerical optimization. In *Proc. 33rd IEEE Conf. Decision and Control*, pages 2678–2683, Lake Buena Vista, FL, USA, 1994.

[55] L. El Ghaoui and S.I. Niculescu, editors. *Advances in Linear Matrix Inequality Methods in Control.* SIAM, Philadelphia, 2000.

[56] L. El Ghaoui, F. Oustry, and M. AitRami. A cone complementarity linearization algorithm for static output-feedback and related problems. *IEEE Transactions on Automatic Control*, 42:1171–1176, 1997.

[57] D. F. Enns. Model reduction with balanced realization: An error bound and a frequency weighted generalization. In *Proc. 23rd IEEE Conf. Decision and Control*, pages 127–132, 1984.

[58] A.C. Farag and H.C. Werner. Fixed-order control of active suspension: a hybrid approach. In *IFAC World Congress*, Prague, Chech Republic, 2005.

[59] M. Fazel, H. Hindi, and S.P. Boyd. Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In *Proc. American Control Conf.*, pages 2156–2162, Denver, Colorado, USA, 2003.

[60] A. Forsgren. Optimality conditions for nonconvex semidefinite programming. *Math. Program.*, 88:105–128, 2000.

[61] G.F. Franklin, J.D. Powell, and A. Emami-Naeini. *Feedback Control of Dynamic Systems*. Addison-Wesley, New York, 3rd edition, 1994.

[62] B. Friedland. *Control System Design: An Introduction to State-Space Methods*. McGraw-Hill, Singapore, 1986.

[63] M. Fu and Z. Luo. Computational complexity of a problem arising in fixed order output feedback design. *Syst. Contr. Letters*, 30:209–215, 1997.

[64] P. Gahinet and P. Apkarian. A linear matrix inequality approach to $\mathcal{H}_\infty$ control. *Int. J. of Robust and Nonlinear Control*, 4:421–448, 1994.

[65] P. Gahinet, A. Nemirovskii, A.J. Laub, and M. Chilali. The LMI control toolbox. In *Proc. 33rd IEEE Conf. Decision and Control*, pages 2038–2041, Lake Buena Vista, FL, USA, 1994.

[66] J.D. Gardiner, A.J. Laub, J.J. Amato, and C.B. Moler. Solution of the sylvester matrix equation $AXB + CXD = E$. *ACM Transactions Mathematical Software*, 18(2):223–231, 1992.

[67] K. Gatermann and P.A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *Journal of Pure and Appl. Algebra, to appear*, 2004.

[68] J.C. Geromel, C.C. de Souza, and R.E. Skelton. Static output feedback controllers: Stability and convexity. *IEEE Transactions on Automatic Control*, 43(1):120–126, 1998.

[69] K. Glover. All optimal hankel-norm approximations of linear multivariable systems and their $\mathcal{L}_1$-error bounds. *Int. J. Control*, 39:1115–1193, 1984.

[70] P.J. Goddard and K. Glover. Controller approximation: Approaches for preserving $\mathcal{H}_\infty$ performance. *IEEE transactions on Automatic Control*, 43:858–871, 1998.

[71] K.C. Goh, L. Turan, M.G. Safonov, G.P. Papavassiloupoulos, and J.H. Ly. Biaffine matrix inequality properties and computational methods. In *Proc. American Control Conf.*, pages 850–855, Baltimore, Maryland, USA, 1994.

[72] G.H. Golub and C.F. Van Loan. *Matrix Computations*. John Hopkins, Baltimore, 1989.

[73] E. Grassi, K.S. Tsakalis, S. Dash, S.V. Gaikwad, W. MacArthur, and G. Stein. Integrated system identification and PID controller tuning by frequency loop-shaping. *IEEE Transactions on Control Systems Technology*, 9(2):285–294, 2001.

[74] M. Green and D.J.N. Limebeer. *Linear Robust Control*. Prentice Hall, Englewood Cliffs, New Jersey, 1996.

[75] K.M. Grigoriadis and R.E. Skelton. Fixed-order control design for LMI control problems using alternating projection methods. In *Proc. 33rd IEEE Conf. Decision and Control*, pages 2003–2008, Lake Buena Vista, FL, USA, 1994.

[76] M. Groot Wassink, M.M.J. van de Wal, C.W. Scherer, and O.H. Bosgra. LPV control for a wafer stage: beyond the theoretical solution. *Control Engineering Practice*, 13(2):231–246, 2004.

[77] G. Gu. Model reduction with relative/multiplicative error bounds and relations to controller reduction. *IEEE Transactions on Automatic Control*, 40(8):1478–1485, 1995.

[78] R. Guidorzi. Canonical structures in identification of multivariable systems. *Automatica*, 11:361–374, 1975.

[79] A. Hassibi, J.P. How, and S.P. Boyd. Low-authority controller design via convex optimization. In *Proc. 37th IEEE Conf. Decision and Control*, volume 1, pages 140–145, Tampa, Florida USA, 1998.

[80] A. Hassibi, J.P. How, and S.P. Boyd. A path-following method for solving BMI problems in control. In *Proc. American Control Conf.*, volume 2, pages 1385–1389, San Diego, California, USA, 1999.

[81] J.W. Helton. Positive noncommutative polynomials are sums of squares. *Annals of Mathematics*, 2002.

[82] D. Henrion. LMI optimization for fixed-order H_∞ controller design. In *Proc. 42nd IEEE Conf. Decision and Control*, Maui, Hawaii, USA, 2003.

[83] D. Henrion and A. Garulli. *Positive polynomials in control*. Lecture Notes in Control and Information Sciences. Springer-Verlag, Berlin, Heidelberg, New York, 2005.

[84] D. Henrion and J.-B. Lasserre. Convergent relaxations of polynomial matrix inequalities and static output feedback. Technical report, LAAS-CNRS, 2004.

[85] D. Henrion and J.-B. Lasserre. Solving nonconvex problems. *IEEE Control Systems Magazine*, 2004.

[86] D. Henrion, M. Sebek, and V. Kucera. Positive polynomials and robust stabilization with fixed-order controllers. *IEEE Transactions on Automatic Control*, 48:1178–1186, 2003.

[87] P.M.J. van den Hof. System identification. Technical report, Delft University of Technology, 1999.

[88] C.W.J. Hol and C.W. Scherer. Computing optimal fixed order $\mathcal{H}_\infty$-synthesis values by matrix sum of squares relaxations. In *Proc. 43rd IEEE Conf. Decision and Control*, Bahamas, 2004.

[89] C.W.J. Hol and C.W. Scherer. Fixed order $\mathcal{H}_\infty$-synthesis: Computing optimal values by robust performance analysis. In *Proc. American Control Conf.*, Boston, MA, USA, 2004.

[90] C.W.J. Hol and C.W. Scherer. Sum of squares relaxations for polynomial semi-definite programming. In *Sixteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, Leuven, Belgium, 2004.

[91] C.W.J. Hol and C.W. Scherer. A sum-of-squares approach to fixed-order $\mathcal{H}_\infty$-synthesis. In D. Henrion and A.Garulli, editors, *Positive polynomials in Control*, Springer Lecture Notes in Control and Information Sciences. Springer, 2005.

[92] C.W.J. Hol and C.W. Scherer. Sum of squares relaxations for robust polynomial semi-definite programs. In *IFAC World Congress*, Prague, 2005.

[93] C.W.J. Hol, C.W. Scherer, E.G. van der Meché, and O.H. Bosgra. A nonlinear SDP approach to fixed-order controller synthesis and comparison with two other methods applied to an active suspension system. *European Journal of Control*, 9(1):13–28, 2003.

[94] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press., 1985.

[95] D.C. Hyland and D.S. Bernstein. The optimal projection equations for fixed-order dynamic compensation. *IEEE Transactions on Automatic Control*, 29:1034–1037, 1984.

[96] T. Iwasaki. The dual iteration for fixed order control. In *Proc. American Control Conf.*, pages 62–66, Albuquerque, New Mexico, USA, 1997.

[97] T. Iwasaki and S. Hara. Well-posedness of feedback systems: insights into exact robustness analysis. In *Proc. 35th IEEE Conf. Decision and Control*, pages 1863–1868, Kobe, Japan, 1996.

[98] T. Iwasaki and G. Shibata. LPV system analysis via quadratic separator for uncertain implicit systems. *IEEE Transactions on Automatic Control*, 46(8):1195–1208, 2001.

[99] T. Iwasaki and R.E. Skelton. All controllers for the general $\mathcal{H}_\infty$ control problem: LMI existence conditions and state space formulas. *Automatica*, 30:1307–1317, 1994.

[100] T. Iwasaki and R.E. Skelton. The XY-centering algorithm for the dual LMI problem: A new approach to fixed-order controller design. *International Journal of Control*, 62:1257–1272, 1995.

[101] T. Jacobi and A. Prestel. Distinguished representations of strictly positive polynomials. *J. Reine Angew. Math.*, 532:223–235, 2001.

[102] J. Jahn. *Introduction to the Theory of Nonlinear Optimization*. Springer-Verlag, Berlin, 1994.

[103] F. Jarre. An interior method for minimizing the maximum eigenvalue of a linear combination of matrices. *SIAM J. Contr. Optim.*, 31(5):1360–1377, 1993.

[104] F. Jarre. Optimal ellipsoidal approximations around the analytic center. *Appl. Math. Optim.*, 30:15–19, 1994.

[105] F. Jarre. An interior point method for nonconvex semidefinite programs. Technical report, Universität Düsseldorf, 2001.

[106] D. Jibetean and E. De Klerk. Global optimization of rational functions: a semidefinite programming approach. Technical report, CWI (Centrum voor Wiskunde en Informatica), 2003.

[107] I. Jonsson and B. Kagström. Recursive blocked algorithms for solving triangular systemsPart I: one-sided and coupled Sylvester-type matrix equations. *ACM Trans. Math. Softw.*, 28(4):392–415, 2002.

[108] I. Jonsson and B. Kagström. Recursive blocked algorithms for solving triangular systemsPart II: two-sided and generalized Sylvester and Lyapunov matrix equations. *ACM Trans. Math. Softw.*, 28(4):416–435, 2002.

[109] U.T. Jönsson. A Lecture on the S-procedure. Technical report, Royal Institute of Technology, 2001.

[110] D. Jungnickel. *Optimierungsmethoden: Eine Einführung.* Springer-Verlag, Berlin, 1999.

[111] S. Kanev. *Robust Fault-Tolerant Control.* PhD thesis, Delft University of Technology, 2004.

[112] M. Kocvara, F. Leibfritz, M. Stingel, and D. Henrion. A nonlinear sdp algorithm for static output feedback problems in COMPleib. 2004.

[113] M. Kocvara and M. Stingel. PENNON, A Code for Convex Nonlinear and Semidenite Programming. *Optimization Methods and Software*, 18:317–333, 2003.

[114] M. Kojima. Sums of squares relaxations of polynomial semidefinite programs. Technical Report Reserach Report B-397, Dept. of Mathematical and Computing Scienses, Tokyo Institute of Technolgy, November 2003.

[115] V. Koltchinski, M. Ariola, and C. T. Abdallah. Statistical controller design for the linear benchmark problem. In *Proceedings IEEE Conference on Decision and Control*, pages 507–509, Phoenix, Arizona, USA, 1999.

[116] E. Kreyszig. Introductory functional analysis with applications, 1978.

[117] I.D. Landau, editor. *European Journal of Control.* Special issue on Restricted Complexity Controllers. The European Union Control Association, 2003.

[118] I.D. Landau, A. Karimi, L. Miskovic, and H. Prochazka. Control of an active suspension system as a benchmark for design and optimisation of restricted complexity controllers. *European Journal of Control*, 9(1):1–12, 2003.

[119] J.-B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal of Optimization*, 11:796–817, 2001.

[120] M. Laurent. Revisiting two theorems of Curto and Fialkow on moment matrices. *Proc. Am. Math. Soc.*, 133:2965–2976, 2005.

[121] F. Leibfritz and E.M.E. Mostafa. An interior point constrained trust region method for a special class of nonlinear semidefinite programming problems. *SIAM J. Optim.*, 12(4):1048–1074, 2002.

[122] C. Van Loan. The ubiquitous kronecker product. *Journal of Computational and Applied Mathematics*, 123(1-2), 2000.

[123] D.G. Luenberger. Canonical forms for linear multivariable systems. *IEEE Transactions on Automatic Control*, AC-12(3):290–293, 1967.

[124] D.G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, New York, 1969.

[125] D.G. Luenberger. *Linear and Nonlinear Programming*. Kluwer Academic Publishers, Massachusetts, 2003.

[126] D. McFarlane, K. Glover, and M. Vidyasagar. Reduced-order controller design using coprime factor model reduction. *IEEE Transactions on Automatic Control*, 35(3):369–373, 1990.

[127] T. McKelvey and A. Helmerson. State-space parametrizations of multivariable liner systems using tridiagonal matrix forms. In *Proc. 35th IEEE Conf. Decision and Control*, pages 3654–3659, Kobe, Japan, 1996.

[128] T. McKelvey and A. Helmerson. System identification using an over-parametrized model class - Improving the optimization algorithm. In *Proc. 36th IEEE Conf. Decision and Control*, pages 2984–2989, San Diego, California, 1997.

[129] E.G. van der Meché and O.H. Bosgra. A convex relaxation approach to multivariable state-space frequency response approximation. In *Proc. American Control Conf.*, volume 6, pages 5274–5279, Denver, USA, 2003.

[130] J.D. Meindl. Beyond Moore's Law: the Interconnect Era. *Computing in Science & Engineering*, 5(1):20–24, 2003.

[131] M. Mesbahi. Fixed-order output feedback synthesis via semidefinite programming. *preprint*, 2004.

[132] M. Mesbahi, M.G. Safonov, and G.P. Papavassilopoulos. Bilinearity and complementarity in robust control. In L. El Ghaoui and S.I. Niculescu, editors, *Advances in Linear Matrix Inequality Methods in Control*, pages 269–292. SIAM, Philadelphia, 2000.

[133] V.F. Montagner and P.L.D. Peres. A new LMI condition for the robust stability of linear time-varying systems. In *Proc. 42nd IEEE Conf. Decision and Control*, volume 6, pages 6133 – 6138, Maui, Hawaii, USA, 2003.

[134] G. Moore. Cramming more components onto integrated circuits. *Proc. IEEE*, 68(1):82–85, 1985.

[135] R. Nagamune. Closed-loop shaping by analytic functions with a bounded degree. In *European Control Conf.*, pages 722–726, Porto, Portugal, 2001.

[136] S.G. Nash and A. Sofer. *Linear and Nonlinear Programming*. McGraw-Hill, New York, 1996.

[137] Y. Nesterov and A. Nemirovski. *Interior point polynomial methods in convex programming*. SIAM Series in Applied Mathematics. SIAM Publications, Philadelphia, 1994.

[138] J. Nie and M. Schweighofer. On the complexity of Putinar's Positivstellensatz. *preprint*, 2005.

[139] G. Obinata and B.D.O. Anderson. *Model Reduction for Control System Design*. Communications and Control Engineering. Springer, Berlin, 2001.

[140] A. Packard and J. Doyle. The Complex Structured Singular Value. *Automatica*, 29:71–109, 1993.

[141] A. Papachristodoulou and S. Prajna. On the construction of Lyapunov functions using the sum of squares decomposition. In *Proc. Conf. Decision and Control*, Las Vegas, NV, USA, 2002.

[142] P. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming Ser. B*, 96(2):293–320, 2003.

[143] P. Parrilo and B. Sturmfels. Minimizing polynomial functions. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 2001.

[144] P.A. Parrilo. *Structure semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

[145] J.A. Pensar and H.T. Toivonen. On the design of fixed-structure $\mathcal{H}_\infty$ optimal controllers. In IEEE, editor, *Proc. 32nd IEEE Conf. Decision and Control*, volume 1, pages 668 – 673, San Antonio, TX, USA, 1993.

[146] H. Pillai and J.C. Willems. Lossless and dissipative distributed systems. *SIAM J. Control Optim.*, 40(5):1406–1430, 2002.

[147] V. Powers and T. Wörman. An algorithm for sums of squares of real polynomials. *Journal of pure and applied algebra*, 127(1):99–104, 1998.

[148] S. Prajna, A. Papachristodoulou, P. Seiler, and P. A. Parrilo. New Developments in Sum of Squares Optimization and SOSTOOLS. In *Proc. Conf. Decision and Control*, Boston, MA, USA, 2004.

[149] M. Putinar. Positive Polynomials on Compact Semi-Algebraic Sets. *Indiana Univ. Math. J.*, 42:969–984, 1993.

[150] T. Rautert and E.W. Sachs. Computational design of optimal output feedback controllers. *SIAM J. Optim.*, 7(3):837–852, 1997.

[151] M.C.B. Reurings. *Symmetric Matrix Equations*. PhD thesis, Vrije Universiteit Amsterdam, 2003.

[152] B. Reznick. Extremal PSD forms with few terms. *Duke Mathematical Journal*, 45(2):363–374, 1978.

[153] B. Reznick. Some concrete aspects of Hilbert's 17th problem. In C.N. Delzell and J.J. Madden, editors, *Real Algebraic Geometry and Ordered Structures*, volume 253 of *Contemp. Math.*, pages 251–272. American Mathematical Society, 2000.

[154] S.L. Richter and R.A. DeCarlo. Continuation methods: theory and applications. *IEEE Transactions on Automatic Control*, AC-28(6):660–665, 1983.

[155] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[156] C. Roos, T. Terlaky, and J.Ph. Vial. *Theory and algorithms for linear optimization; an interior point approach.* Wiley-Interscience series in discrete mathematics and optimization. Wiley, Chichester, 1997.

[157] C. Roos, T. Terlaky, and J.Ph. Vial. *Interior Point Methods for Linear Optimization.* Springer, New York, 2006.

[158] M. Rotunno and R.A. de Callafon. Fixed order $\mathcal{H}_\infty$ control design for dual stage hard disk drives. In *Proc. 39th IEEE Conf. Decision and Control*, pages 3118–3119. Sydney, Australia, 2000.

[159] M. Rotunno and R.A. de Callafon. A bundle method for solving the fixed order control problem. In *Proc. 41th IEEE Conf. Decision and Control*, volume 3, pages 3156–3161, Las Vegas, Nevada, USA, 2002.

[160] W. Rudin. *Principles of Mathematical Analysis.* McGraw-Hill, Singapore, 3rd edition, 1976.

[161] W. Rudin. *Real and complex analysis.* Mathematics Series. McGraw-Hill, New York, 3rd edition, 1986.

[162] H. G. Sage and M. F. de Mathelin. Canonical $\mathcal{H}_\infty$ state-space parametrization. *Automatica*, 36:1049–1055, 2000.

[163] G. Schelfhout and B. de Moor. A note on closed-loop balanced truncation. *IEEE Transactions on Automatic Control*, 41(10):1498 – 1500, 1996.

[164] C. Scherer. *The Riccati inequality and state-space $\mathcal{H}_\infty$-optimal control.* Ph.d. thesis, Bayerische Julius-Maximilians Universität Würzburg, 1991.

[165] C.W. Scherer. $\mathcal{H}_\infty$-control by state-feedback for plants with zeros on the imaginary axis. *SIAM J. Contr. Optim.*, 30:123–142, 1992.

[166] C.W. Scherer. Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control. In A. Isidori, editor, *Trends in Control, A European Prospective*, pages 173–216. Springer-Verlag, New York, 1995.

[167] C.W. Scherer. Multiobjective $\mathcal{H}_2/\mathcal{H}_\infty$ control. *IEEE Transactions on Automatic Control*, 40:1054–1062, 1995.

[168] C.W. Scherer. Theory of robust control. Lecture notes, Delft University of Technology, 1995.

[169] C.W. Scherer. Design of structured controllers with applications. In *Proc. 39th IEEE Conf. Decision and Control*, Sydney, Australia, 2000.

[170] C.W. Scherer. LPV control and full block multipliers. *Automatica*, 37:361–375, 2001.

[171] C.W. Scherer. Higher-order relaxations for robust LMI problems with verifications for exactness. In *Proc. 42nd IEEE Conf. Decision and Control*, pages 4652–4657, Maui, Hawaii, USA, 2003.

[172] C.W. Scherer. Relaxations for robust linear matrix inequality problems with verifications for exactness. *SIAM J. Matrix Anal. Appl.*, 27(2):365–395, 2004.

[173] C.W. Scherer, P. Gahinet, and M. Chilali. Multi-objective output-feedback control via LMI optimization. *IEEE Transactions on Automatic Control*, 42:896–911, 1997.

[174] C.W. Scherer and C.W.J. Hol. Asymptotically exact relaxations for robust LMI problems based on matrix-valued sum-of-squares. In *Sixteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, Leuven, Belgium, 2004.

[175] C.W. Scherer and C.W.J. Hol. Matrix sum-of-squares relaxations for robust semidefinite-programs. *Math. Program. Series B, to appear*, 2005.

[176] C.W. Scherer and S. Weiland. *Linear Matrix Inequalities in Control*. Lecture Notes, Dutch Institute for Systems and Control, Delft, The Netherlands, 2000.

[177] K. Schmüdgen. The K-moment problem for compact semi-algebraic sets. *Math. Ann.*, 289(2):203–206, 1991.

[178] R. Schrama. *Approximate Identification and Control Design with Application to a Mechanical System*. PhD thesis, Delft University of Technology, 1994.

[179] G.A. Schultz, R.B. Schnabel, and R.H. Byrd. A family of trust region based algorithms for unconstrained minimization with strong global convergence properties. *SIAM Journal on Numerical Analysis*, 22(1):47–67, 1985.

[180] M. Schweighofer. Optimization of polynomials on compact semialgebraic sets. *SIAM J. Optim.*, 15(3):805–825, 2003.

[181] M. Schweighofer. On the complexity of Schmüdgen's Positivstellensatz. *Journal of Complexity*, 20(4):529–542, 2004.

[182] A. Shapiro and K. Sheinberg. Duality and optimality conditions. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming*, pages 67–92. Kluwer, 2000.

[183] S. Skogestad and I. Postlethwaite. *Multivariable Feedback Control, Analysis and Design*. John Wiley & Sons, 1996.

[184] V. Sreeram, B.D.O. Anderson, and A.G. Madievski. New results on frequency weighted balanced reduction technique. In *Proc. American Control Conf.*, Seattle, WA, USA, 1995.

[185] M. Steinbuch and O.H. Bosgra. Necessary conditions for static and fixed order dynamic mixed $\mathcal{H}_2/\mathcal{H}_\infty$ optimal control. In *Proc. Amer. Contr. Conf.*, pages 1137–1143, 1991.

[186] M. Steinbuch and M.L. Norg. Industrial perspective on robust control: Application to storage systems. In *2nd IFAC Symp. Robust Control Design (ROCOND)*, 1997.

[187] M. Steinbuch, G. Schootstra, and O.H. Bosgra. Robust control of a compact disc player. In *Proc. 31st IEEE Conference on Decision and Control*, pages 2596–2600, Tucson AR, USA, 1992.

[188] A.A. Stoorvogel. *The $\mathcal{H}_\infty$ control problem: a state space approach*. Prentice Hall, Hemel Hempstead, UK, 1992.

[189] B. Sturmfels. Polynomial equations and convex polytopes. *American Mathematical Monthly*, 105(10):907922, 1998.

[190] T. Stykel. Numerical solution and perturbation theory for generalized Lyapunov equations. *Linear Alg. Appl.*, 349:155–18, 2002.

[191] O. Toker and H. Ozbay. On the NP-hardness of solving bilinear matrix inequalities and simultaneous stabilization with static output feedback. In *Proc. American Control Conf.*, volume 4, pages 2525 – 2526, Seattle, WA, USA, 1995.

[192] V. Torczon. On the convergence of the multidirectional search algorithm. *SIAM J. Optim.*, 1(1):123–145, 1991.

[193] H.D. Tuan and P. Apkarian. Low nonconvexity-rank bilinear matrix inequalities: Algorithms and applications in robust controller and structure designs. *IEEE Transactions on Automatic Control*, 45:2111–2117, 2000.

[194] L. Vandenberghe, V. Balakrishan, R. Wallin, and A. Hansson. On the implementation of primal-dual interior-point methods for semidefinite programming problems derived from the kyp lemma. In *Proc. 42nd IEEE Conf. Decision and Control*, pages 4685–4663, Maui, Hawaii, USA, 2003.

[195] L. Vandenberghe, V. Balakrishan, R. Wallin, A. Hansson, and T. Roh. Interior-point methods for semidefinite programming problems derived from the KYP lemma. In D. Henrion and A. Garulli, editors, *Positive Polynomials in Control*, volume 312 of *Lecture Notes in Control and Information Sciences*, pages 195–238. Springer Verlag, 2005.

[196] L. Vandenberghe and V. Balakrishnan. Semidefinite programming duality and linear system theory: connections and implications for computation. In *Proc. 38th IEEE Conf. Decision and Control*, pages 989 –994, Phoenix, Arizona, USA, 1999.

[197] A. Varga. Coprime factor reduction of $\mathcal{H}_\infty$ controllers. In *European Control Conference (ECC)*, Cambridge, UK, 2003.

[198] V. Verdult, N. Bergboer, and M. Verhagen. Maximum likelihood identification of multivariable bilinear state-space systems by projected gradient search. In *Proc. 41th IEEE Conf. Decision and Control*, Las Vegas, Nevada USA, 2002.

[199] M.M.J. van de Wal, G.E. van Baars, F. Sperling, and O.H. Bosgra. Multivariable $\mathcal{H}_\infty/\mu$ feedback control design for high-precision wafer stage motion. *Control engineering practice*, 10:739–755, 2002.

[200] R. Wallin, A. Hansson, and L. Vandenberghe. Comparison of Two Structure-Exploiting Optimization Algorithms for Integral Quadratic Constraints. In 4$^{\text{th}}$ *IFAC Symposium on Robust Control Design*, 2003.

[201] G. Wang, V. Sreeram, and W.Q. Liu. Balanced performance preserving controller reduction. *System and Control Letters*, 46:99–110, 2002.

[202] M.S. Whorton, Buschek H., and Calise A.J. Homotopy algorithms for fixed order $\mathcal{H}_2$ and $\mathcal{H}_\infty$ design. *AIAA J. of Guidance, Control and Dynamics*, 19(6):1262–1269, 1996.

[203] J.C. Willems. Least squares stationary optimal control and the algebraic Riccati equation. *IEEE Transactions on Automatic Control*, 16:621–634, 1971.

[204] L.G. van Willigenburg and W.L. De Koning. Optimal reduced-order compensators for time-varying discrete-time systems with deterministic and white parameters. *Automatica*, 35(1):129–138, 1999.

[205] H. Wolkowicz, B. Saigal, and L. Vandenberghe, editors. *Handbook on semidefinite programming: theory, algorithms, applications*. International Series in Operations Research and Management Science. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[206] P.M.R. Wortelboer. *Frequency-Weighted Balanced Reduction of Closed-Loop Mechanical Servo-Systems: Theory and Tools*. PhD thesis, Delft University of Technology, 1994.

[207] P.M.R. Wortelboer and O.H. Bosgra. Generalized frequency weighted balanced reduction. In *Proc. 31st IEEE Conf. Decision and Control*, volume 3, pages 2848 – 2849, Tucson AR, USA, 1992.

[208] P.M.R. Wortelboer, M. Steinbuch, and O.H. Bosgra. Iterative model and controller reduction using closed-loop balancing, with application to a compact disc mechanism. *International Journal of Robust and Nonlinear Control*, 9:123–142, 1999.

[209] S.J. Wright. *Primal-dual interior-point methods*. SIAM, Philadelphia, 1997.

[210] Yang X.Y. Packard A. Becker G. Wu, F. Induced $\mathcal{L}_2$-Norm Control for LPV Systems with Bounded Parameter Variation Rates. *Int. J. of Robust and Nonlinear Control*, 6:983–998, 1996.

[211] V.A. Yakubovich. S-procedure in nonlinear control theory. *Vestnik Leningrad University*, 4:73–93, 1977.

[212] M.J. Zandvliet, C.W. Scherer, C.W.J. Hol, and M.M.J. van de Wal. Multiobjective feedback control applied to a waferstage model. In *Proc. 43rd IEEE Conf. Decision and Control*, Bahamas, 2004.

[213] K. Zhou. Frequency weighted model reduction with $\mathcal{L}_\infty$ error bounds. *Syst. Contr. Letters*, 21:115–125, 1993.

[214] K. Zhou, J. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, Upper Saddle River, New Jersey, 1996.

Samenvatting

Voor vele toepassingsgebieden van mechanische servo-systemen is een zeer hoge positineer-nauwkeurigheid en een snel volggedrag vereist. Een goed voorbeeld hiervan is een wafer stage, die gebruikt wordt bij de productie van Integrated Circuits (IC's). Om aan de eisen te voldoen is het dikwijls noodzakelijk om de regelaar strak te tunen en daarbij rekening te houden met dynamische interactie. Modelgebaseerd $\mathcal{H}_\infty$-optimaal regelaarontwerp is daarvoor een handige en systematische methodiek. De bestaande manier om zo een regelaar te berekenen, levert echter een regelaar op met een zelfde McMillan graad (d.w.z. het aantal toestanden in een minimale toestandsrealisatie) als die van de generalised plant, oftewel het systeem en de wegingsfuncties. Door beperkte capaciteit in de real-time processoren is het vaak onmogelijk deze regelaars op hoge sample-frequentie te implementeren. Andere a priori eisen op de structuur van de regelaar, zoals PID-structuur, diagonaliteit of stabiliteit, zijn ook vaak belangrijk. Het ontwerpen van dit soort regelaars noemen we *gestructeerde regelaarsynthese*. Het onderwerp van dit proefschrift is het ontwikkelen van nieuwe methoden voor $\mathcal{H}_\infty$-optimale gestructeerde regelaarsynthese voor servo-systemen.

Met behulp van polynomen die bestaan uit een som van kwadraten is er in dit proefschrift een familie van relaxaties ontwikkeld, waarmee ondergrenzen uitgerekend kunnen worden voor het $\mathcal{H}_\infty$-optimale gestructureerde regelaarsynthese probleem. Deze ondergrenzen kunnen worden gevonden door een convex probleem met Lineaire Matrix-ongelijkheden (LMIs) op te lossen. Het is gegarandeerd, dat deze ondergrenzen convergeren naar de globaal optimale gestructureerde regelaar-prestatie in termen van de $\mathcal{H}_\infty$-norm, indien we kiezen voor steeds betere relaxaties. Door de snelle groei van het aantal beslissingsvariablen in de LMIs, is de toepassing van deze techniek echter beperkt tot generalised plants van lage McMillan graad.

Door middel van gedeeltelijke dualisatie verkrijgt men een andere formulering van het gestructureerde regelaarontwerpprobleem, waardoor de relaxaties over het algemeen minder beslissingsvariabelen hebben en kunnen worden toegepast op regelaarproblemen met een generalised plant van hogere McMillan graad. Deze relaxaties maken het ook mogelijk globale ondergrenzen uit te rekenen met behulp van technieken voor robuuste analyse, zoals bijvoorbeeld de S-procedure.

De relaxaties op basis van sommen van kwadraten zijn ook toegepast op robuuste polynomiale Semi-Definiete Programma's (SDP's). Ook hiervoor is een familie van relaxaties ontwikkeld, wiens optimale waarden van beneden convergeren naar de optimale waarde van het robuuste SDP.

De hiervoor genoemde ondergrenzen kunnen gebruikt worden als stop-criterium van algoritmes voor gestructureerd regelaarontwerp. Zo is een algoritme ontwikkeld op basis van een inwendige punts-methode. Bij deze methode moet er in elke iteratie een stelsel van Sylvester vergelijkingen opgelost worden. Er is aangetoond hoe dit stelsel efficiënt is op te lossen, door de onderliggende structuur van het probleem uit te buiten. Het algoritme maakt gebruik van een trust-region en een plane search, die zorgen voor een

gegarandeerde convergentie naar een stationair punt. Locale optimaliteit kan in dit punt geverifieerd worden met behulp van de eerste en tweede orde optimaliteitscondities, die zijn afgeleid voor algemene polynomiale semi-definiete programma's. De gegarandeerd goede convergentie van het algoritme is bevestigd met behulp van het regelaarontwerp voor een actief trillingsonderdrukkingssysteem.

De keuze van de toestandsrealisatie van de regelaar geeft vrijheden in de beslissingsvariabelen van vele algoritmes voor regelaaroptimalisatie en in het bijzonder voor het inwendige punts-algoritme. Dit kan trage convergentie tot gevolg hebben. Bovendien is het daardoor vaak onmogelijk om locale optimaliteit van regelaars te verifiëren. Er is daarom een parametrisatie ontwikkeld voor multi-variabele regelaars, die deze problemen verhelpt en waarmee toch alle regelaars (met de vereiste structuur) kunnen worden geparametriseerd.

Tenslotte zijn vescheidene algoritmes gebruikt om regelaars met beperking op de McMillan graad te ontwerpen voor een nieuw prototype van een wafer stage. Hiervoor is een lineair tijdsinvariant model van dit servo-systeem geïdentificeerd. Een diagonale, geoptimaliseerde regelaar is geïmplementeerd en vergeleken met een diagonale gereduceerde regelaar. Deze analyse is zowel gedaan op basis van het model als op basis van experimenten met het systeem. Hieruit blijkt dat de geoptimaliseerde regelaar beter presteerde over het beschouwde deel van het werkgebied.

Abstract

In many application areas of mechanical servo-systems a high positioning accuracy and fast tracking is required. A good example is a wafer stage, that is used for the production of Integrated Circuits (IC's). The high demands on the performance often imply a tightly tuned feedback controller, that takes dynamical interaction into account. Model-based $\mathcal{H}_\infty$-optimal controller synthesis is a well-suited systematic design technique for this purpose. However, the state-of-the-art synthesis approach yields a controller with the same Mc-Millan degree as the generalized plant, i.e. the order of the system model and the weighting filters. Often these controllers can not be implemented in real-time at high sampling-rates, because of the limited computational capacity of the processors. Besides restrictions on the Mc-Millan degree it is often also important that the controller satisfies other structural constraints, such as PID-structure, strong stabilization or zero off-diagonal entries. The design of controllers with an a priori structure is denoted by *structured controller synthesis*. The aim of this thesis is to provide numerical tools for $\mathcal{H}_\infty$-optimal structured controller synthesis.

As the first result of this thesis work, relaxations have been developed that are based on Sum-Of-Squares polynomials. Their optimal values are lower bounds on the globally optimal structured controller synthesis problem. Those bounds can be computed by solving Linear Matrix Inequalities (LMI) problems. It is guaranteed, that the bounds converge to best achievable performance when restricting to structured control as we improve our relaxations. However, the technique is limited to generalized plants of small Mc-Millan degree, due to the fast growth of the number of decision variables in the LMIs.

We obtain a different scheme for lower bound computation, if we first apply partial dualization. The corresponding relaxations have often fewer variables than the ones mentioned above and are therefore applicable to plants with higher McMillan degree. By partial dualization it is also possible to reformulate the structured controller synthesis problem as a robust analysis problem. Known robust analysis techniques, such as those based on the S-procedure, can therefore be used to compute lower bounds on the optimal controller performance.

The Sum-Of-Squares relaxations have also been applied to robust polynomial Semi-Definite Programs (SDPs). Also for this case a sequence of relaxations has been developed, whose optimal values converge from below to the optimal value of the robust SDP.

The aforementioned lower bounds can be used as a stopping criterion for the Interior Point algorithm, that has been developed for controller synthesis. At each iteration of this method a system of Sylvester equations has to be solved. It is shown how this can be done efficiently, by exploiting the control-theoretic characteristics of the problem. The algorithm uses a trust region and a plane search and has a guaranteed convergence to a point satisfying the first-order necessary conditions. Local optimality of this point can be verified using first- and second-order optimality conditions, that have been derived for polynomial semi-definite programs in general and the structured synthesis problem in

273

particular. The guaranteed good local convergence of the algorithm is confirmed by means of a controller design for an active suspension system.

The choice of the state-space realization of the controller gives freedom in the decision variables. This freedom plays a role in several structured controller synthesis algorithms and in the Interior Point method in particular. It may cause slow convergence of the algorithm and it makes it impossible to verify the local optimality conditions. A parametrization for multi-variable controller synthesis has been developed to resolve this problem for the Interior Point algorithm. At the same time, this parametrization can describe all controllers of the required structure.

Finally, several algorithms described in this thesis have been applied to synthesize fixed-order controllers for a new prototype of a wafer stage. To this purpose a linear time-invariant model of the servo-system has been identified. A diagonal optimized controller has been implemented. Its performance has been compared to that of a reduced controller, in terms of the closed-loop $\mathcal{H}_\infty$-norm, as well as by experiments. The optimized controller exhibited better performance over the considered part of the operational domain.

Index

stability theorem, 13

matrix
-valued polynomial, 59
orthogonal, 175
quasi-upper-triangular, 125
Maximum Modulus Principle, 14
McMillan degree, 10
MDS, 52
MIMO, 2
monomial, 61
Multi-Directional-Search, 52
Multiple-Input-Multiple-Ouput, 2

Newton
Polytope, 74
step, 233
NP-hard, 38

observability, 41
optimal value, 35
optimality, 35
optimization
BMI, 37
Interior-Point, 113–157
LMI, 33
matrix inequalities, 32

Padé approximation, 192
phase-one problem, 121
pink noise, 188
plane search, 119
plant, 9
polynomial, 60
matrix-valued, 59
SDP, 59
Sum-Of-Squares, 65
time, 33
proper, 10
Proportional Integral Derivative, PID, 2

realization
balanced, 42
minimal, 10
reduction, 41
controller, 41
relaxations, 71, 83, 99
residualization, 43
Riccati
Equations, 24

Inequality, 15

S-procedure, 36
full-block, 36
relaxations, 89
Schur
complement lemma, 15
decomposition, 175
Semi-Definite Program, SDP
Polynomial, 59
Semi-Definite Programming, SDP, 32
Single Input Single Output, 2
SISO, 2
Slater's constraint qualification, 34
stability, 12
static decoupling, 189
Static Output Feedback,SOF, 12
strong stabilization, 31
Sum-Of-Squares
matrix-valued, 65
relaxations, 88
Sylvester Equations, 124

tangent cone, 134
transfer
function, 9
matrix, 10
transformation
similarity, 24
truncation, 43
trust region, 118
two-strokes principle, 186

uncertainty
polynomial description, 100
polytopic, 101
real, 91
S-procedure, 37

well-posedness, 11

Curriculum Vitae

Camile Wilbert José was born in 1973 in Enschede. He finished his secondary school (Ongedeeld Atheneum) in 1991 at the Thijcollege in Oldenzaal. In 1991-1992 he has been an exchange student to Russia with exchange organization 'Youth for Understanding'. In 1992 he started with his study Mechanical Engineering at the Technical University in Delft. In 1994 he received his first-years degree (propaedeuse) Cum Laude. During his study he participated in the EU-Japan exchange Programme for Engineering students in 1998-1999. This programme consisted of a Japanese Language Course in Tokyo of four months, followed by an internship of eight months in the company NTT Data in Kawasaki. He graduated in 2000 in Mechanical Engineering at the Delft University of Technology with the MSc. thesis entitled "Nonlinear state estimator design for model predictive control, application on a high density polyethylene process" with the grade 9 (out of 10). From 2000-2001 he conducted research at the Wageningen University and Research Center. He started his PhD. at the Delft Center for Systems and Control of the Delft University of Technology in April 2001. In May 2005 he started working at the Department Rolling Metal Strip of Corus Research, Development and Technology in IJmuiden.