Linear Matrix Inequalities in Control

Carsten Scherer and Siep Weiland

Delft Center for Systems and Control
Delft University of Technology
The Netherlands

Department of Electrical Engineering
Eindhoven University of Technology
The Netherlands

Contents

List of Figures

Preface

In recent years linear matrix inequalities (LMI's) have emerged as a powerful tool to approach control problems that appear hard if not impossible to solve in an analytic fashion. Although the history of linear matrix inequalities goes back to the fourties with a major emphasis of their role in control in the sixties through the work of Kalman, Yakubovich, Popov and Willems, only during the last decades powerful numerical interior point techniques have been developed to solve LMI's in a practically efficient manner (Nesterov, Nemirovskii 1994). Currently, several commercial and non-commercial software packages are available that allow a simple coding of general LMI problems and provide efficient tools to solve typical control problems in a numerically efficient manner.

Boosted by the availability of fast LMI solvers, research in robust control has experienced a significant paradigm shift. Instead of arriving at an analytical solution of an optimal control problem and implementing such a solution in software so as to synthesize optimal controllers, the intention is to reformulate a given control problem to verifying whether a specific linear matrix inequality is solvable or, alternatively, to optimizing functionals over linear matrix inequality constraints. This book aims at providing a state of the art treatment of the theory, the usage and the applications of linear matrix inequalities in control. The main emphasis of this book is to reveal the basic principles and background for formulating desired properties of a control system in the form of linear matrix inequalities, and to demonstrate the techniques to reduce the corresponding controller synthesis problem to an LMI problem. The power of this approach is illustrated by several fundamental robustness and performance problems in analysis and design of linear control systems.

This book has been written for a graduate course on the subject of LMI's in systems and control. Within the graduate program of the Dutch Institute of Systems and Control (DISC), this course is intended to provide up-to-date information on the topic for students involved in either the practical or theoretical aspects of control system design. DISC courses have the format of two class hours once per week during a period of eight weeks. Within DISC, the first course on LMI's in control has been given in 1997, when the first draft of this book has been distributed as lecture notes to the students. The lecture notes of this course have been evaluating to the present book, thanks to the many suggestions, criticism and help of many students that followed this course.

Readers of this book are supposed to have an academic background in linear algebra, basic calculus, and possibly in system and control theory.

Chapter 1

Convex optimization and linear matrix inequalities

1.1 Introduction

Optimization questions and decision making processes are abundant in daily life and invariably involve the selection of the best decision from a number of options or a set of candidate decisions. Many examples of this theme can be found in technical sciences such as electrical, mechanical and chemical engineering, in architecture and in economics, but also in the social sciences, in biological and ecological processes and organizational questions. For example, production processes in industry are becoming more and more market driven and require an ever increasing flexibility of product changes and product specifications due to customer demands in quality, price and specification. Products need to be manufactured within strict product specifications, with large variations of input quality, against competitive prices, with minimal waste of resources, energy and valuable production time, with a minimal time-to-market and, of course, with maximal economical profit. Important economical benefits can therefore only be realized by making proper decisions in the operating conditions of production processes. Due to increasing requirements on the safety and flexibility of production processes, there is a constant need for further optimization, for increased efficiency and a better control of processes.

Casting an optimization problem in mathematics involves the specification of the candidate decisions and, most importantly, the formalization of the concept of *best* or *optimal decision*. Let the universum of all possible decisions in an optimization problem be denoted by a set $\mathcal{X}$, and suppose that the set of *feasible (or candidate) decisions* is a subset $\mathcal{S}$ of $\mathcal{X}$. Then one approach to quantify the performance of a feasible decision $x \in \mathcal{S}$ is to express its value in terms of a single real quantity $f(x)$ where f is some real valued function $f : \mathcal{S} \to \mathbb{R}$ called the *objective function* or *cost function*. The value of decision $x \in \mathcal{S}$ is then given by $f(x)$. Depending on the interpretation of the objective function, we may wish to minimize or maximize f over all feasible candidates in $\mathcal{S}$. An optimal decision is then simply an element of $\mathcal{S}$ that minimizes or maximizes f over all feasible alternatives.

The optimization problem to *minimize* the objective function f over a set of feasible decisions $\mathcal{S}$ involves various specific questions:

(a) What is the least possible cost? That is, determine the *optimal value*

$$V_{\text{opt}} := \inf_{x \in \mathcal{S}} f(x) = \inf\{f(x) \mid x \in \mathcal{S}\}.$$

By convention, the optimal value $V_{\text{opt}} = +\infty$ if $\mathcal{S}$ is empty, while the problem is said to be *unbounded* if $V_{\text{opt}} = -\infty$.

(b) How to determine an *almost optimal solution*, i.e., for arbitrary $\varepsilon > 0$, how to determine $x_\varepsilon \in \mathcal{S}$ such that

$$V_{\text{opt}} \leq f(x_\varepsilon) \leq V_{\text{opt}} + \varepsilon.$$

(c) Does there exist an *optimal solution* $x_{\text{opt}} \in \mathcal{S}$ with $f(x_{\text{opt}}) = V_{\text{opt}}$? If so, we say that the *minimum is attained* and we write $f(x_0) = \min_{x \in \mathcal{S}} f(x)$. The set of all optimal solutions is denoted by $\operatorname{argmin}_{x \in \mathcal{S}} f(x)$.

(d) If an optimal solution x_{opt} exists, is it also unique?

We will address each of these questions in the sequel.

1.2 Facts from convex analysis

In view of the optimization problems just formulated, we are interested in finding conditions for optimal solutions to exist. It is therefore natural to resort to a branch of analysis which provides such conditions: convex analysis. The results and definitions in this subsection are mainly basic, but they have very important implications and applications as we will see later.

We start with summarizing some definitions and elementary properties from linear algebra and functional analysis. We assume the reader to be familiar with the basic concepts of vector spaces, norms and normed linear spaces.

Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are two normed linear spaces. A function f which maps $\mathcal{X}$ to $\mathcal{Y}$ is said to be *continuous at* $x_0 \in \mathcal{X}$ if, for every $\varepsilon > 0$, there exists a $\delta = \delta(x_0, \varepsilon)$ such that

$$\|f(x) - f(x_0)\| < \varepsilon \quad \text{whenever} \quad \|x - x_0\| < \delta. \tag{1.2.1}$$

The function f is called *continuous* if it is continuous at all $x_0 \in \mathcal{X}$. Finally, f is said to be *uniformly continuous* if, for every $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon)$, not depending on x_0, such that (1.2.1) holds. Obviously, continuity depends on the definition of the norm in the normed spaces $\mathcal{X}$ and $\mathcal{Y}$. We remark that a function $f : \mathcal{X} \to \mathcal{Y}$ is continuous at $x_0 \in \mathcal{X}$ if and only if for every sequence $\{x_n\}_{n=1}^{\infty}$, $x_n \in \mathcal{X}$, which converges to x_0 as $n \to \infty$, there holds that $f(x_n) \to f(x_0)$.

Now let $\mathcal{S}$ be a subset of the normed linear space $\mathcal{X}$. Then $\mathcal{S}$ is called *compact* if for every sequence $\{x_n\}_{n=1}^{\infty}$ in $\mathcal{S}$ there exists a subsequence $\{x_{n_m}\}_{m=1}^{\infty}$ which converges to an element $x_0 \in \mathcal{S}$. Compact sets in finite dimensional vector spaces are easily characterized. Indeed, if $\mathcal{X}$ is finite dimensional then a subset $\mathcal{S}$ of $\mathcal{X}$ is compact if and only if $\mathcal{S}$ is closed and bounded[1].

The well-known Weierstrass theorem provides a useful tool to determine whether an optimization problem admits a solution. It provides an answer to the third question raised in the previous subsection for special sets $\mathcal{S}$ and special performance functions f.

Proposition 1.1 (Weierstrass) *If $f : \mathcal{S} \to \mathbb{R}$ is a continuous function defined on a compact subset $\mathcal{S}$ of a normed linear space $\mathcal{X}$, then there exists $x_{\min}, x_{\max} \in \mathcal{S}$ such that*

$$f(x_{\min}) = \inf_{x\in\mathcal{S}} f(x) \leq f(x) \leq \sup_{x\in\mathcal{S}} f(x) = f(x_{\max})$$

for all $x \in \mathcal{S}$.

Proof. Define $V_{\min} := \inf_{x\in\mathcal{S}} f(x)$. Then there exists a sequence $\{x_n\}_{n=1}^{\infty}$ in $\mathcal{S}$ such that $f(x_n) \to V_{\min}$ as $n \to \infty$. As $\mathcal{S}$ is compact, there must exist a subsequence $\{x_{n_m}\}_{m=1}^{\infty}$ of $\{x_n\}$ which converges to an element, say $x_{\min}$, which lies in $\mathcal{S}$. Obviously, $f(x_{n_m}) \to V_{\min}$ and the continuity of f implies that $f(x_{n_m}) \to f(x_{\min})$ as $n_m \to \infty$. We claim that $V_{\min} = f(x_{\min})$. By definition of $V_{\min}$, we have $V_{\min} \leq f(x_{\min})$. Now suppose that the latter inequality is strict, i.e., suppose that $V_{\min} < f(x_{\min})$. Then $0 < f(x_{\min}) - V_{\min} = \lim_{n_m\to\infty} f(x_{n_m}) - \lim_{n_m\to\infty} f(x_{n_m}) = 0$, which yields a contradiction. The proof of the existence of a maximizing element is similar. ∎

> Following his father's wishes, Karl Theodor Wilhelm Weierstrass (1815-1897) studied law, finance and economics at the university of Bonn. His primary interest, however, was in mathematics which led to a serious conflict with his father. He started his career as a teacher of mathematics. After various positions and invitations, he accepted a chair at the 'Industry Institute' in Berlin in 1855. Weierstrass contributed to the foundations of analytic functions, elliptic functions, Abelian functions, converging infinite products, and the calculus of variations. Hurwitz and Frobenius were among his students.

Note that Proposition 1.1 does not give a constructive method to find the extremal solutions $x_{\min}$ and $x_{\max}$. It only guarantees the existence of these elements for continuous functions defined on compact sets. For many optimization problems these conditions (continuity and compactness) turn out to be overly restrictive. We will therefore resort to convex sets.

Definition 1.2 (Convex sets) A set $\mathcal{S}$ in a linear vector space is said to be *convex* if

$$\{x_1, x_2 \in \mathcal{S}\} \implies \{x := \alpha x_1 + (1-\alpha)x_2 \in \mathcal{S} \text{ for all } \alpha \in (0, 1)\}.$$

[1] A set $\mathcal{S}$ is *bounded* if there exists a number B such that for all $x \in \mathcal{S}$, $\|x\| \leq B$; it is *closed* if $x_n \to x$ implies that $x \in \mathcal{S}$.

In geometric terms, this states that for any two points of a convex set also the line segment connecting these two points belongs to the set. In general, the *empty set* and *singletons* (sets that consist of one point only) are considered to be convex. The point $\alpha x_1 + (1-\alpha)x_2$ with $\alpha \in (0, 1)$ is called a *convex combination* of the two points x_1 and x_2. More generally, convex combinations are defined for any finite set of points as follows.

Definition 1.3 (Convex combinations) Let $\mathcal{S}$ be a subset of a vector space. The point

$$x := \sum_{i=1}^{n} \alpha_i x_i$$

is called a *convex combination* of $x_1, \ldots, x_n \in \mathcal{S}$ if $\alpha_i \geq 0$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} \alpha_i = 1$.

It is easy to see that the set of all convex combinations of n points $x_1, \ldots, x_n$ in $\mathcal{S}$ is itself convex, i.e.,

$$\mathcal{C} := \{x \mid x \text{ is a convex combination of } x_1, \ldots, x_n\}$$

is convex.

We next define the notion of interior points and closure points of sets. Let $\mathcal{S}$ be a subset of a normed space $\mathcal{X}$. The point $x \in \mathcal{S}$ is called an *interior point* of $\mathcal{S}$ if there exists an $\varepsilon > 0$ such that all points $y \in \mathcal{X}$ with $\|x - y\| < \varepsilon$ also belong the $\mathcal{S}$. The *interior* of $\mathcal{S}$ is the collection of all interior points of $\mathcal{S}$. $\mathcal{S}$ is said to be *open* if it is equal to its interior. The point $x \in \mathcal{X}$ is called a *closure point* of $\mathcal{S}$ if, for all $\varepsilon > 0$, there exists a point $y \in \mathcal{S}$ with $\|x - y\| < \varepsilon$. The *closure* of $\mathcal{S}$ is the collection of all closure points of $\mathcal{S}$. $\mathcal{S}$ is said to be *closed* if it is equal to its closure.

We summarize some elementary properties pertaining to convex sets in the following proposition.

Proposition 1.4 *Let $\mathcal{S}$ and $\mathcal{T}$ be convex sets in a normed vector space $\mathcal{X}$. Then*

(a) the set $\alpha\mathcal{S} := \{x \mid x = \alpha s, s \in \mathcal{S}\}$ is convex for any scalar α.

(b) the sum $\mathcal{S} + \mathcal{T} := \{x \mid x = s + t, s \in \mathcal{S}, t \in \mathcal{T}\}$ is convex.

(c) for any $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$, $(\alpha_1 + \alpha_2)\mathcal{S} = \alpha_1\mathcal{S} + \alpha_2\mathcal{S}$.

(d) the closure and the interior of $\mathcal{S}$ (and $\mathcal{T}$) are convex.

(e) for any linear transformation $T : \mathcal{X} \to \mathcal{X}$, the image $T\mathcal{S} := \{x \mid x = Ts, s \in \mathcal{S}\}$ and the inverse image $T^{-1}\mathcal{S} := \{x \mid Tx \in \mathcal{S}\}$ are convex.

(f) the intersection $\mathcal{S} \cap \mathcal{T} := \{x \mid x \in \mathcal{S} \text{ and } x \in \mathcal{T}\}$ is convex.

The distributive property in the third item is non trivial and depends on the convexity of $\mathcal{S}$. The last property actually holds for the intersection of an *arbitrary collection* of convex sets, i.e, if $\mathcal{S}_\alpha$, with $\alpha \in A$, A an arbitrary index set, is a family of convex sets then the intersection $\cap_{\alpha \in A}\mathcal{S}_\alpha$ is also convex. This property turns out to be very useful in constructing the smallest convex set that contains a given set. It is defined as follows.

As an example, with $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$, the *hyperplane* $\{x \in \mathbb{R}^n \mid a^\top x = b\}$ and the *half-space* $\{x \in \mathbb{R}^n \mid a^\top x \leq b\}$ are convex. A *polyhedron* is the intersection of finitely many hyperplanes and half-spaces and is convex by the last item of Proposition 1.4. A *polytope* is a compact polyhedron.

Definition 1.5 (Convex hull) The *convex hull* $\operatorname{conv} \mathcal{S}$ of any subset $\mathcal{S} \subset \mathcal{X}$ is the intersection of all convex sets containing $\mathcal{S}$. If $\mathcal{S}$ consists of a finite number of elements, then these elements are referred to as the *vertices* of $\operatorname{conv} \mathcal{S}$.

It is easily seen that the convex hull of a finite set is a polytope. The converse is also true: any polytope is the convex hull of a finite set. Since convexity is a property that is closed under intersection, the following proposition is immediate.

Proposition 1.6 (Convex hulls) *For any subset $\mathcal{S}$ of a linear vector space $\mathcal{X}$, the convex hull* $\operatorname{conv}(\mathcal{S})$ *is convex and consists precisely of all convex combinations of the elements of $\mathcal{S}$.*

At a few occasions we will need the concept of a cone. A subset $\mathcal{S}$ of a vector space $\mathcal{X}$ is called a *cone* if $\alpha x \in \mathcal{S}$ for all $x \in \mathcal{S}$ and $\alpha > 0$. A *convex cone* is a cone which is a convex set. Like in Proposition 1.4, if $\mathcal{S}$ and $\mathcal{T}$ are convex cones then the sets $\alpha\mathcal{S}$, $\mathcal{S} + \mathcal{T}$, $\mathcal{S} \cap \mathcal{T}$, $T\mathcal{S}$ and $T^{-1}\mathcal{S}$ are convex cones for all scalars α and all linear transformations T. Likewise, the intersection of an arbitrary collection of convex cones is a convex cone again. Important examples of convex cones are defined in terms of inequalities as follows. If $\mathcal{X}$ is a Hilbert space equipped with the inner product $\langle \cdot, \cdot \rangle$, then

$$\mathcal{S} := \{x \in \mathcal{X} \mid \langle x, k_i \rangle \leq 0, i \in A\}$$

is a (closed) convex cone for any index set A and any collection of elements $k_i \in \mathcal{X}$. Thus, solution sets of systems of linear inequalities define convex cones.

Definition 1.7 (Convex functions) A function $f : \mathcal{S} \to \mathbb{R}$ is called *convex* if $\mathcal{S}$ is a non-empty convex set and if for all $x_1, x_2 \in \mathcal{S}$ and $\alpha \in (0, 1)$ there holds that

$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha) f(x_2). \tag{1.2.2}$$

f is called *strictly convex* if the inequality (1.2.2) is strict for all $x_1, x_2 \in \mathcal{S}$, $x_1 \neq x_2$ and all $\alpha \in (0, 1)$.

Note that the domain of a convex function is *by definition* a convex set. Simple examples of convex functions are $f(x) = x^2$ on $\mathbb{R}$, $f(x) = \sin x$ on $[\pi, 2\pi]$ and $f(x) = -\log x$ on $x > 0$. A function $f : \mathcal{S} \to \mathbb{R}$ is called *concave* if $-f$ is convex. Many operations on convex functions naturally preserve convexity. For example, if f_1 and f_2 are convex functions with domain $\mathcal{S}$ then the function $f_1 + f_2 : x \mapsto f_1(x) + f_2(x)$ is convex. More generally, $\alpha_1 f_1 + \alpha_2 f_2$ and the composite function $g(f_1)$ are convex for any non-negative numbers α_1, α_2 and any non-decreasing convex function $g : \mathbb{R} \to \mathbb{R}$.

Instead of minimizing the function $f : \mathcal{S} \to \mathbb{R}$ we can set our aims a little lower and be satisfied with considering all possible $x \in \mathcal{S}$ that give a guaranteed upper bound of f. For this, we introduce, for any number $\gamma \in \mathbb{R}$, the *sublevel sets* associated with f as follows

$$\mathcal{S}_\gamma := \{x \in \mathcal{S} \mid f(x) \leq \gamma\}.$$

Obviously, $\mathcal{S}_\gamma = \emptyset$ if $\gamma < \inf_{x\in\mathcal{S}} f(x)$ and $\mathcal{S}_\gamma$ coincides with the set of global minimizers of f if $\gamma = \inf_{x\in\mathcal{S}} f(x)$. Note also that $\mathcal{S}_{\gamma'} \subseteq \mathcal{S}_{\gamma''}$ whenever $\gamma' \leq \gamma''$; that is sublevel sets are non-decreasing (in a set theoretic sense) when viewed as function of γ. As one can guess, convex functions and convex sublevel sets are closely related to each other:

Proposition 1.8 *If $f : \mathcal{S} \to \mathbb{R}$ is convex then the sublevel set $\mathcal{S}_\gamma$ is convex for all $\gamma \in \mathbb{R}$.*

Proof. Suppose f is convex. Let $\gamma \in \mathbb{R}$ and consider $\mathcal{S}_\gamma$. If $\mathcal{S}_\gamma$ is empty then the statement is trivial. Suppose therefore that $\mathcal{S}_\gamma \neq \emptyset$ and let $x_1, x_2 \in \mathcal{S}_\gamma$, $\alpha \in [0, 1]$. Then, $f(x_1) \leq \gamma$, $f(x_2) \leq \gamma$ and the convexity of $\mathcal{S}$ implies that $\alpha x_1 + (1-\alpha)x_2 \in \mathcal{S}$. Convexity of f now yields that

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2) \leq \alpha\gamma + (1-\alpha)\gamma = \gamma$$

i.e., $\alpha x_1 + (1-\alpha)x_2 \in \mathcal{S}_\gamma$. ∎

Sublevel sets are commonly used in specifying desired behavior of multi-objective control problems. As an example, suppose that $\mathcal{S}$ denotes a class of (closed-loop) transfer functions and let, for $k = 1, \ldots, K$, $f_k : \mathcal{S} \to \mathbb{R}$ be the kth objective function on $\mathcal{S}$. For a multi-index $\gamma = (\gamma_1, \ldots, \gamma_K)$ with $\gamma_k \in \mathbb{R}$, the kth sublevel set $\mathcal{S}^k_{\gamma_k} := \{x \in \mathcal{S} \mid f_k(x) \leq \gamma_k\}$ then expresses the kth design objective. The multi-objective specification amounts to characterizing

$$\mathcal{S}_\gamma := \mathcal{S}^1_{\gamma_1} \cap \ldots \cap \mathcal{S}^K_{\gamma_K}$$

and the design question will be to decide for which multi-index γ this set is non-empty. Proposition 1.8 together with Proposition 1.4 promises the convexity of $\mathcal{S}_\gamma$ whenever f_k is convex for all k. See Exercise 9.

We emphasize that it is *not* true that convexity of the sublevel sets $\mathcal{S}_\gamma$, $\gamma \in \mathbb{R}$ implies convexity of f. (See Exercise 2 and Exercise 6 in this Chapter). However, the class of functions for which all sublevel sets are convex is that important that it deserves its own name.

Definition 1.9 (Quasi-convex functions) A function $f : \mathcal{S} \to \mathbb{R}$ is *quasi-convex* if its sublevel sets $\mathcal{S}_\gamma$ are convex for all $\gamma \in \mathbb{R}$.

It is easy to verify that f is quasi-convex if and only if

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \max[f(x_1), f(x_2)]$$

for all $\alpha \in (0, 1)$ and for all $x_1, x_2 \in \mathcal{S}$. In particular, every convex function is quasi-convex.

We conclude this section with the introduction of affine sets and affine functions. A subset $\mathcal{S}$ of a linear vector space is called an *affine set* if the point $x := \alpha x_1 + (1-\alpha)x_2$ belongs to $\mathcal{S}$ for every $x_1 \in \mathcal{S}$, $x_2 \in \mathcal{S}$ and $\alpha \in \mathbb{R}$. The geometric idea is that for any two points of an affine set, also the line through these two points belongs to the set. From Definition 1.2 it is evident that every affine set is convex. The empty set and singletons are generally considered to be affine. Each non-empty affine set $\mathcal{S}$ in a finite dimensional vector space $\mathcal{X}$ can be written as

$$\mathcal{S} = \{x \in \mathcal{X} \mid x = x_0 + s,\ s \in \mathcal{S}_0\}$$

where $x_0 \in \mathcal{X}$ is a vector and $\mathcal{S}_0$ is a linear subspace of $\mathcal{X}$. That is, affine sets are *translates* of linear subspaces. For any such representation, the linear subspace $\mathcal{S}_0 \subseteq \mathcal{X}$ is uniquely defined. However, the vector x_0 is not.

Definition 1.10 (Affine functions) A function $f : \mathcal{S} \to \mathcal{T}$ is *affine* if

$$f(\alpha x_1 + (1-\alpha)x_2) = \alpha f(x_1) + (1-\alpha) f(x_2)$$

for all $x_1, x_2 \in \mathcal{S}$ and $\alpha \in \mathbb{R}$.

If $\mathcal{S}$ and $\mathcal{T}$ are finite dimensional then any affine function $f : \mathcal{S} \to \mathcal{T}$ can be represented as $f(x) = f_0 + T(x)$ where $f_0 \in \mathcal{T}$ and $T : \mathcal{S} \to \mathcal{T}$ is a linear map. Indeed, setting $f_0 = f(0)$ and $T(x) = f(x) - f_0$ establishes this representation. In particular, $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine if and only if there exists $x_0 \in \mathbb{R}^n$ such that $f(x) = f(x_0) + T(x - x_0)$ where T is a *matrix* of dimension $m \times n$. Note that all affine functions are convex and concave.

1.3 Convex optimization

After the previous section with definitions and elementary properties of convex sets and convex functions, we hope that this section will convince the most skeptical reader why convexity of sets and functions is such a desirable property for optimization.

1.3.1 Local and global minima

Anyone who gained experience with numerical optimization methods got familiar with the pitfalls of local minima and local maxima. One crucial reason for studying convex functions is related to the *absence of local minima*.

Definition 1.11 (Local and global optimality) Let $\mathcal{S}$ be a subset of a normed space $\mathcal{X}$. An element $x_0 \in \mathcal{S}$ is said to be a *local optimal solution* of $f : \mathcal{S} \to \mathbb{R}$ if there exists $\varepsilon > 0$ such that

$$f(x_0) \leq f(x) \tag{1.3.1}$$

for all $x \in \mathcal{S}$ with $\|x - x_0\| < \varepsilon$. It is called a *global optimal solution* if (1.3.1) holds for all $x \in \mathcal{S}$.

In words, $x_0 \in \mathcal{S}$ is a local optimal solution if there exists a neighborhood of x_0 such that $f(x_0) \leq f(x)$ for all feasible points nearby x_0. According to this definition, a global optimal solution is also locally optimal. Here is a simple and nice result which provides one of our main interests in convex functions.

Proposition 1.12 *Suppose that $f : \mathcal{S} \to \mathbb{R}$ is convex. Every local optimal solution of f is a global optimal solution. Moreover, if f is strictly convex, then the global optimal solution is unique.*

Proof. Let f be convex and suppose that $x_0 \in \mathcal{S}$ is a local optimal solution of f. Then for all $x \in \mathcal{S}$ and $\alpha \in (0, 1)$ sufficiently small,

$$f(x_0) \leq f(x_0 + \alpha(x - x_0)) = f((1-\alpha)x_0 + \alpha x) \leq (1-\alpha)f(x_0) + \alpha f(x). \tag{1.3.2}$$

This implies that

$$0 \leq \alpha(f(x) - f(x_0)) \tag{1.3.3}$$

or $f(x_0) \leq f(x)$. Hence, x_0 is a global optimal solution of f. If f is strictly convex, then the second inequality in (1.3.2) is strict so that (1.3.3) becomes strict for all $x \in \mathcal{S}$. Hence, x_0 must be unique.
■

It is very important to emphasize that Proposition 1.12 does not make any statement about the *existence* of optimal solutions $x_0 \in \mathcal{S}$ that minimize f. It merely says that all locally optimal solutions are globally optimal. For convex functions it therefore suffices to compute *locally optimal solutions* to actually determine its globally optimal solution.

Remark 1.13 Proposition 1.12 does not hold for quasi-convex functions.

1.3.2 Uniform bounds

The second reason to investigate convex functions comes from the fact that uniform upperbounds of convex functions can be verified on subsets of their domain. Here are the details: let $\mathcal{S}_0$ be a set and suppose that $f : \mathcal{S} \to \mathbb{R}$ is a function with domain

$$\mathcal{S} = \mathrm{conv}(\mathcal{S}_0).$$

As we have seen in Proposition 1.6, $\mathcal{S}$ is convex and we have the following property which is both simple and powerful.

Proposition 1.14 *Let $f : \mathcal{S} \to \mathbb{R}$ be a convex function where $\mathcal{S} = \mathrm{conv}(\mathcal{S}_0)$. Then $f(x) \leq \gamma$ for all $x \in \mathcal{S}$ if and only if $f(x) \leq \gamma$ for all $x \in \mathcal{S}_0$*

Proof. The 'only if' part is trivial. To see the 'if' part, Proposition 1.6 implies that every $x \in \mathcal{S}$ can be written as a convex combination $x = \sum_{i=1}^{n} \alpha_i x_i$ where $n > 0$, $\alpha_i \geq 0$, $x_i \in \mathcal{S}_0$, $i = 1, \ldots, n$ and

$\sum_{i=1}^{n} \alpha_i = 1$. Using convexity of f and non-negativity of the α_i's, we infer

$$f(x) = f(\sum_{i=1}^{n} \alpha_i x_i) \leq \sum_{i=1}^{n} \alpha_i f(x_i) \leq \sum_{i=1}^{n} \alpha_i \gamma = \gamma,$$

which yields the result. ■

Proposition 1.14 states that the uniform bound $f(x) \leq \gamma$ on $\mathcal{S}$ can equivalently be verified on the set $\mathcal{S}_0$. This is of great practical relevance especially when $\mathcal{S}_0$ contains only a *finite number of elements*, i.e., when $\mathcal{S}$ is a polytope. It then requires a *finite number of tests* to conclude whether or not $f(x) \leq \gamma$ for all $x \in \mathcal{S}$. In addition, since

$$\gamma_0 := \sup_{x \in \mathcal{S}} f(x) = \max_{x \in \mathcal{S}_0} f(x)$$

the supremum of f is attained and can be determined by considering $\mathcal{S}_0$ only.

1.3.3 Duality and convex programs

In many optimization problems, the universum of all possible decisions is a real valued finite dimensional vector space $\mathcal{X} = \mathbb{R}^n$ and the space of feasible decisions typically consists of $x \in \mathcal{X}$ which satisfy a finite number of *inequalities* and *equations* of the form

$$\begin{aligned} g_i(x) &\leq 0, \qquad i = 1, \ldots, k \\ h_i(x) &= 0, \qquad i = 1, \ldots, l. \end{aligned}$$

Indeed, saturation constraints, safety margins, physically meaningful variables, and a large number of constitutive and balance equations can be written in this way. The space of feasible decisions $\mathcal{S} \subset \mathcal{X}$ is then expressed as

$$\mathcal{S} := \{x \in \mathcal{X} \mid g(x) \leq 0, \quad h(x) = 0\} \tag{1.3.4}$$

where $g : \mathcal{X} \to \mathbb{R}^k$ and $h : \mathcal{X} \to \mathbb{R}^l$ are the vector valued functions which consist of the component functions g_i and h_i and where the inequality $g(x) \leq 0$ is interpreted component-wise. Hence, the feasibility set $\mathcal{S}$ is defined in terms of k inequality constraints and l equality constraints. With sets of this form, we consider the optimization problem to find the optimal value

$$P_{\text{opt}} := \inf_{x \in \mathcal{S}} f(x),$$

and possibly optimal solutions $x_{\text{opt}} \in \mathcal{S}$ such that $f(x_{\text{opt}}) = P_{\text{opt}}$. Here, $f : \mathcal{X} \to \mathbb{R}$ is a given objective function. In this section, we will refer to this constrained optimization problem as a *primal optimization problem* and to P_{opt} as the *primal optimal value*. To make the problem non-trivial, we will assume that $P_{\text{opt}} > -\infty$ and that $\mathcal{S}$ is non-empty.

Remark 1.15 If $\mathcal{X}$, f and g are convex and h is affine, then it is easily seen that $\mathcal{S}$ is convex, in which case this problem is commonly referred to as a *convex program*. This is probably the only tractable instance of this problem and its study certainly belongs to the most sophisticated area of nonlinear optimization theory. The Karush-Kuhn-Tucker Theorem, presented below, is the key to understanding convex programs. The special instance where f, g and h are all affine functions makes the problem to determine $P_{\rm opt}$ a *linear programming problem*. If f is quadratic (i.e., f is of the form $f(x) = (x - x_0)^\top Q(x - x_0)$ for some $Q = Q^\top \in \mathbb{R}^{n\times n}$ and fixed $x_0 \in \mathbb{R}^n$) and g and h are affine, this is a *quadratic programming problem*.

Obviously, for any $x_0 \in \mathcal{S}$ we have that $P_{\rm opt} \leq f(x_0)$, i.e., an *upperbound* of $P_{\rm opt}$ is obtained from any feasible point $x_0 \in \mathcal{S}$. On the other hand, if $x \in \mathcal{X}$ satisfies $g(x) \leq 0$ and $h(x) = 0$, then for arbitrary vectors $y \geq 0$ (meaning that each component y_i of y is non-negative) and z we have that

$$L(x, y, z) := f(x) + \langle y, g(x)\rangle + \langle z, h(x)\rangle \leq f(x).$$

Here, $L(\cdot,\cdot,\cdot)$ is called a *Lagrangian*, which is a function of $n + k + l$ variables, and $\langle\cdot,\cdot\rangle$ denotes the standard inner product in $\mathbb{R}^n$, that is, $\langle x_1, x_2\rangle := x_2^\top x_1$. It is immediate that for all $y \in \mathbb{R}^k$, $y \geq 0$ and $z \in \mathbb{R}^l$ we have that

$$\ell(y, z) := \inf_{x\in\mathcal{X}} L(x, y, z) \leq \inf_{x\in\mathcal{S}} f(x) = P_{\rm opt}.$$

The function $\ell(\cdot,\cdot)$ is the *Lagrange dual cost*. A pair of vectors (y, z) with $y \geq 0$ is said to be *feasible for the dual problem* if $\ell(y, z) > -\infty$. Suppose that there exists at least one such feasible pair (y, z). Since ℓ is independent of x, we may conclude that

$$D_{\rm opt} := \sup_{y\geq 0,\, z} \ell(y, z) = \sup_{y\geq 0,\, z} \inf_{x\in\mathcal{X}} L(x, y, z) \leq P_{\rm opt}$$

provides a *lower bound* of $P_{\rm opt}$. Since ℓ is the pointwise infimum of the concave functions $L(x,\cdot,\cdot)$, $\ell(y, z)$ is a *concave* function (no matter whether the primal problem is convex program) and the *dual optimization problem* to determine $D_{\rm opt}$ is therefore a concave optimization problem. The main reason to consider this problem is that the constraints in the dual problem are much simpler to deal with than the ones in the primal problem.

Of course, the question arises when $D_{\rm opt} = P_{\rm opt}$. To answer this question, suppose that $\mathcal{X}$, f and g are convex and h is affine. As noted before, this implies that $\mathcal{S}$ is convex. We will say that $\mathcal{S}$ satisfies the *constraint qualification* if there exists a point x_0 in the interior of $\mathcal{X}$ with $g(x_0) \leq 0$, $h(x_0) = 0$ such that $g_j(x_0) < 0$ for all component functions g_j that are not affine[2]. In particular, $\mathcal{S}$ satisfies the constraint qualification if g is affine. We have the following central result.

Theorem 1.16 (Karush-Kuhn-Tucker) *Suppose that $\mathcal{X}$, f and g are convex and h is affine. Assume that $P_{opt} > -\infty$ and $\mathcal{S}$ defined in* (1.3.4) *satisfies the constraint qualification. Then*

$$D_{opt} = P_{opt}.$$

[2] Some authors call $\mathcal{S}$ *superconsistent* and the point x_0 a *Slater point* in that case.

and there exist vectors $y_{opt} \in \mathbb{R}^k$, $y_{opt} \geq 0$ and $z_{opt} \in \mathbb{R}^l$, such that $D_{opt} = \ell(y_{opt}, z_{opt})$, i.e., the dual optimization problem admits an optimal solution. Moreover, x_{opt} is an optimal solution of the primal optimization problem and (y_{opt}, z_{opt}) is an optimal solution of the dual optimization problem, if and only if

(a) $g(x_{opt}) \leq 0$, $h(x_{opt}) = 0$,

(b) $y_{opt} \geq 0$ *and* x_{opt} *minimizes* $L(x, y_{opt}, z_{opt})$ *over all* $x \in \mathcal{X}$ *and*

(c) $\langle y_{opt}, g(x_{opt})\rangle = 0$.

The result of Theorem 1.16 is very general and provides a strong tool in convex optimization. This, because the dual optimization problem is, in general, simpler and, under the stated assumptions, is guaranteed to be solvable. The optimal solutions $(y_{\text{opt}}, z_{\text{opt}})$ of the dual optimization problem are generally called *Kuhn-Tucker points*. The conditions 1, 2 and 3 in Theorem 1.16 are called the *primal feasibility*, the *dual feasibility* and the *alignment* (or *complementary slackness*) condition.

Theorem 1.16 provides a conceptual solution of the primal optimization problem as follows. First construct the dual optimization problem to maximize $\ell(y, z)$. Second, calculate a Kuhn-Tucker point $(y_{\text{opt}}, z_{\text{opt}})$ which defines an optimal solution to the dual problem (existence is guaranteed). Third, determine (if any) the set of optimal solutions $\mathcal{P}'_{\text{opt}}$ which minimize $L(x, y_{\text{opt}}, z_{\text{opt}})$ over all $x \in \mathcal{X}$. Fourth, let $\mathcal{P}_{\text{opt}}$ be the set of points $x_{\text{opt}} \in \mathcal{P}'_{\text{opt}}$ such that $g(x_{\text{opt}}) \leq 0$, $h(x_{\text{opt}}) = 0$ and $\langle y_{\text{opt}}, g(x_{\text{opt}})\rangle = 0$. Then $\mathcal{P}_{\text{opt}}$ is the set of optimal solutions of the primal optimization problem. We emphasize that optimal solutions to the dual problem are guaranteed to exist, optimal solutions of the primal problem may not exist.

Remark 1.17 In order that the triple $(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}})$ defined in Theorem 1.16 exists, it is necessary and sufficient that $(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}})$ be a *saddle point* of the Lagrangian L in the sense that

$$L(x_{\text{opt}}, y, z) \leq L(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}}) \leq L(x, y_{\text{opt}}, z_{\text{opt}})$$

for all x, $y \geq 0$ and z. In that case,

$$\begin{aligned} P_{\text{opt}} &= L(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}}) = \\ &= \inf_x \sup_{y\geq 0,\, z} L(x, y, z) = \sup_{y\geq 0,\, z} \inf_x L(x, y, z) = \\ &= D_{\text{opt}}. \end{aligned}$$

That is, the optimal value of the primal and dual optimization problem coincide with the *saddle point value* $L(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}})$. Under the given conditions, Theorem 1.16 therefore states that x_{opt} is an optimal solution of the primal optimization problem if and only if there exist $(y_{\text{opt}}, z_{\text{opt}})$ such that $y_{\text{opt}} \geq 0$ and $(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}})$ is a saddle point of L.

Remark 1.18 A few generalizations of Theorem 1.16 are worth mentioning. The inequality constraints $g(x) \leq 0$ in (1.3.4) can be replaced by the more general constraint $g(x) \in \mathcal{K}$, where $\mathcal{K} \subset \mathcal{H}$

is a closed convex cone in a Hilbert space $\mathcal{H}$. In the definition of the Lagrangian L, the vectors y and z define *linear functionals* $\langle y, \cdot\rangle$ and $\langle z, \cdot\rangle$ on the Hilbert spaces $\mathbb{R}^k$ and $\mathbb{R}^l$, respectively. For more general Hilbert spaces $(\mathcal{H}, \langle\cdot,\cdot\rangle)$, the constraint $y \geq 0$ needs to be replaced by the requirement that $\langle y, g(x)\rangle \leq 0$ for all $x \in \mathcal{X}$ with $g(x) \in \mathcal{K}$. This is equivalent of saying that the linear functional $\langle y, \cdot\rangle$ is non-positive on $\mathcal{K}$.

Figure 1.1: Joseph-Louis Lagrange

Joseph-Louis Lagrange (1736-1813) studied at the College of Turin and he became interested in mathematics when he read a copy of Halley's work on the use of algebra in optics. Although he decided to devote himself to mathematics, he did not have the benefit of studying under supervision of a leading mathematician. Before writing his first paper, he sent his results to Euler, who at that time was working in Berlin. Lagrange worked on the calculus of variations and regularly corresponded on this topic with Euler. Among many contributions in mathematics, Lagrange worked on the calculus of differential equations and applications in fluid mechanics where he first introduced the Lagrangian function.

1.3.4 Subgradients

Our fourth reason of interest in convex functions comes from the geometric idea that through any point on the graph of a convex function we can draw a line such that the entire graph lies above or on the line. For functions $f : \mathcal{S} \to \mathbb{R}$ with $\mathcal{S} \subseteq \mathbb{R}$, this idea is pretty intuitive from a geometric point of view. The general result is stated in the next Proposition and its proof is a surprisingly simple application of Theorem 1.16.

Proposition 1.19 *Let $\mathcal{S} \subseteq \mathbb{R}^n$. If $f : \mathcal{S} \to \mathbb{R}$ is convex then for all x_0 in the interior of $\mathcal{S}$ there exists a vector $g = g(x_0) \in \mathbb{R}^n$, such that*

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle \tag{1.3.5}$$

for all $x \in \mathcal{S}$.

Proof. The set $\mathcal{S}' := \{x \in \mathcal{S} \mid x - x_0 = 0\}$ has the form (1.3.4) and we note that the primal optimal value $P_{\text{opt}} := \inf_{x \in \mathcal{S}'} f(x) - f(x_0) = 0$. Define the Lagrangian $L(x, z) := f(x) - f(x_0) + \langle z, x - x_0 \rangle$ and the corresponding dual optimization $D_{\text{opt}} := \sup_{z \in \mathcal{S}} \inf_{x \in \mathbb{R}^n} L(x, z)$. Then $D_{\text{opt}} \leq P_{\text{opt}} = 0$ and since $\mathcal{S}'$ trivially satisfies the constraint qualification, we infer from Theorem 1.16 that there exists $z_{\text{opt}} \in \mathbb{R}^n$ such that

$$D_{\text{opt}} = 0 = \inf_{x \in \mathcal{S}} f(x) - f(x_0) + \langle z_{\text{opt}}, x - x_0 \rangle.$$

Consequently, $f(x) - f(x_0) + \langle z_{\text{opt}}, x - x_0 \rangle \geq 0$ for all $x \in \mathcal{S}$ which yields (1.3.5) by setting $g := -z_{\text{opt}}$. ■

A vector g satisfying (1.3.5) is called a *subgradient* of f at the point x_0, and the affine function defined by the right-hand side of (1.3.5) is called a *support functional for f at x_0*. Inequality (1.3.5) is generally referred to as the *subgradient inequality*. We emphasize that the subgradient of a convex function f at a point is in general non-unique. Indeed, the real-valued function $f(x) = |x|$ is convex on $\mathbb{R}$ and has any real number $g \in [-1, 1]$ as its subgradient at $x = 0$. The set of all subgradients of f at x_0 is the *subdifferential of f at x_0* and is denoted by $\partial f(x_0)$ (or $\partial_x f(x_0)$ if the independent variable need to be displayed explicitly). From the subgradient inequality (1.3.5) it is immediate that $\partial(\alpha f)(x) = \alpha \partial f(x)$ for all x and $\alpha > 0$. Also, $x_0 \in \mathcal{S}$ is a global optimal solution of f if and only if $0 \in \partial f(x_0)$. We remark that for a convex function f, $\partial f(x)$ is a closed convex set for any x in the interior of its domain. As a more striking property, let f_1 and f_2 be convex functions with domains $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, then

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x)$$

for all x in the interior of $\mathcal{S}_1 \cap \mathcal{S}_2$.

Remark 1.20 Proposition 1.19 gives a necessary condition for convexity of a function f. It can be shown that if the gradient

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \dots & \frac{\partial f}{\partial x_n} \end{pmatrix}$$

exists and is continuous at $x \in \mathcal{S}$ then $\partial f(x) = \nabla f(x)$. Conversely, if f has a unique subgradient at x, then the gradient of f exists at x and $\partial f(x) = \nabla f(x)$.

The geometric interpretation of Proposition 1.19 is that the graphs of the affine functions $x \mapsto f(x_0) + \langle g, x - x_0 \rangle$, with $g \in \partial f(x_0)$, range over the collection of all hyperplanes which are tangent to the graph of f at the point $(x_0, f(x_0)$. That is, the graph of f lies on or above these hyperplanes each of which contains the point $(x_0, f(x_0))$. If we consider the right hand side of (1.3.5), then trivially $\langle g, x - x_0 \rangle > 0$ implies that $f(x) > f(x_0)$. Thus all points in the half-space

$\{x \in \mathcal{S} \mid \langle g, x - x_0 \rangle > 0\}$ lead to larger values of f than $f(x_0)$. In particular, in searching for the global minimum of f we can disregard this entire half-space.

This last observation is at the basis of the *ellipsoid algorithm*: a simple, numerically robust and straightforward iterative algorithm for the computation of optimal values.

Algorithm 1.21 (Ellipsoid algorithm) Aim: determine the optimal value of a convex function $f : \mathcal{S} \to \mathbb{R}$.

Input: A convex function $f : \mathcal{S} \to \mathbb{R}$ with $\mathcal{S} \subset \mathbb{R}^n$. An ellipsoid

$$\mathcal{E}_0 := \{x \in \mathbb{R}^n \mid (x - x_0)^\top P_0^{-1}(x - x_0) \leq 1\}.$$

centered at $x_0 \in \mathbb{R}^n$ and oriented by a positive definite matrix $P_0 = P_0^\top$ such that it contains an optimal solution of the problem to minimize f. Let $\varepsilon > 0$ be an accuracy level. Let $k = 0$.

Step 1: Compute a subgradient $g_k \in \partial f(x_k)$ and set

$$L_k := \max_{\ell \leq k} \left(f(x_\ell) - \sqrt{g_\ell^\top P_\ell g_\ell} \right)$$
$$U_k := \min_{\ell \leq k} f(x_\ell)$$

If $g_k = 0$ or $U_k - L_k < \varepsilon$, then set $x^* = x_k$ and stop. Otherwise proceed to Step 2.

Step 2: Put $\mathcal{H}_k := \mathcal{E}_k \cap \{x \in \mathbb{R}^n \mid \langle g_k, x - x_k \rangle \leq 0\}$.

Step 3: Set

$$x_{k+1} := x_k - \frac{P_k g_k}{(n+1)\sqrt{g_k^\top P_k g_k}}$$
$$P_{k+1} := \frac{n^2}{n^2 - 1} \left(P_k - \frac{2}{(n+1) g_k^\top P_k g_k} P_k g_k g_k^\top P_k \right)$$

and define the ellipsoid

$$\mathcal{E}_{k+1} := \{x \in \mathbb{R}^n \mid (x - x_{k+1})^\top P_{k+1}^{-1}(x - x_{k+1}) \leq 1\}$$

with center x_{k+1} and orientation P_{k+1}.

Step 4: Set k to $k + 1$ and return to Step 1.

Output: The point x^* with the property that $|f(x^*) - \inf_{x \in \mathcal{S}} f(x)| \leq \varepsilon$.

The algorithm therefore determines the optimal value of f with arbitrary accuracy. We emphasize that the point x^* is generally *not an optimal or almost optimal solution* unless $g_k = 0$ upon termination of the algorithm. Only in that case x^* is an optimal solution. Hence, the algorithm does not necessarily calculate a solution, but only the optimal value $V_{\text{opt}} = \inf_{x \in \mathcal{S}} f(x)$.

The idea behind the algorithm is as follows. The algorithm is initialized by a 'non-automated' choice of x_0 and P_0 such that that there exists an optimal solution x_{opt} in the ellipsoid $\mathcal{E}_0$. If $\mathcal{S}$ is bounded then the safest choice would be such that $\mathcal{S} \subseteq \mathcal{E}_0$. The subgradients $g_k \in \partial f(x_k)$ divide $\mathbb{R}^n$ in the two half-spaces

$$\{x \mid \langle g_k, x - x_k \rangle < 0\} \quad \text{and} \quad \{x \mid \langle g_k, x - x_k \rangle > 0\}$$

while the *cutting plane* $\{x \mid \langle g_k, x - x_k \rangle = 0\}$ passes through the center of the ellipsoid $\mathcal{E}_k$ for each k. Since $f(x) > f(x_k)$ whenever $\langle g_k, x - x_k \rangle > 0$, the optimal solution x_{opt} is guaranteed to be located in $\mathcal{H}_k$. The ellipsoid defined in Step 3 contains $\mathcal{H}_k$ and is the smallest volume ellipsoid with this property. Iterating over k, the algorithm produces a sequence of ellipsoids $\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots$ whose volumes decrease according to

$$\text{vol}(\mathcal{E}_{k+1}) = \det(P_{k+1}) \leq e^{-\frac{1}{2n}} \det(P_k) = e^{-\frac{1}{2n}} \text{vol}(\mathcal{E}_k)$$

and where each ellipsoid is guaranteed to contain x_{opt}. The sequence of centers $x_0, x_1, x_2, \ldots$ of the ellipsoids generate a sequence of function evaluations $f(x_k)$ which *converges to the optimal value* $f(x_{\text{opt}})$. Convergence of the algorithm is in 'polynomial time' due to the fact that the volume of the ellipsoids decreases geometrically. Since $x_{\text{opt}} \in \mathcal{E}_k$ for all k, we have

$$\begin{aligned} f(x_k) \geq f(x_{\text{opt}}) \geq f(x_k) + \langle g_k, x_{\text{opt}} - x_k \rangle \geq \\ \geq f(x_k) + \inf_{\xi \in \mathcal{E}_k} \langle g_k, \xi - x_k \rangle = f(x_k) - \sqrt{g_k^\top P_k g_k} \end{aligned}$$

so that $L_k \leq f(x_{\text{opt}}) \leq U_k$ define an upper and lower bound on the optimal value.

The algorithm is easy to implement, is very robust from a numerical point of view and implies low memory requirements in its performance. However, convergence may be rather slow which may be a disadvantage for large optimization problems.

1.4 Linear matrix inequalities

1.4.1 What are they?

A *linear matrix inequality* is an expression of the form

$$F(x) := F_0 + x_1 F_1 + \ldots + x_n F_n \prec 0 \tag{1.4.1}$$

where

- $x = (x_1, \ldots, x_n)$ is a vector of n real numbers called the *decision variables*.
- $F_0, \ldots, F_n$ are real symmetric matrices, i.e., $F_j = F_j^\top$, for $j = 0, \ldots, n$.
- the inequality $\prec 0$ in (1.4.1) means 'negative definite'. That is, $u^\top F(x)u < 0$ for all non-zero real vectors u. Because all eigenvalues of a real symmetric matrix are real, (1.4.1) is equivalent to saying that all eigenvalues $\lambda(F(x))$ are negative. Equivalently, the maximal eigenvalue $\lambda_{\max}(F(x)) < 0$.

It is convenient to introduce some notation for symmetric and Hermitian matrices. A matrix A is *Hermitian* if it is square and $A = A^* = \bar{A}^\top$ where the bar denotes taking the complex conjugate of each entry in A. If A is real then this amounts to saying that $A = A^\top$ and we call A *symmetric*. The sets of all $m \times m$ Hermitian and symmetric matrices will be denoted by $\mathbb{H}^m$ and $\mathbb{S}^m$, respectively, and we will omit the superscript m if the dimension is not relevant for the context.

Definition 1.22 (Linear Matrix Inequality) A *linear matrix inequality* (LMI) is an inequality

$$F(x) \prec 0 \tag{1.4.2}$$

where F is an *affine function* mapping a finite dimensional vector space $\mathcal{X}$ to either $\mathbb{H}$ or $\mathbb{S}$.

Remark 1.23 Recall from Definition 1.10 that an affine mapping $F : \mathcal{X} \to \mathbb{S}$ necessarily takes the form $F(x) = F_0 + T(x)$ where $F_0 \in \mathbb{S}$ (i.e., F_0 is real symmetric) and $T : \mathcal{X} \to \mathbb{S}$ is a linear transformation. Thus if $\mathcal{X}$ is finite dimensional, say of dimension n, and $\{e_1, \ldots, e_n\}$ constitutes a basis for $\mathcal{X}$, then every $x \in \mathbb{X}$ can be represented as $x = \sum_{j=1}^n x_j e_j$ and we can write

$$T(x) = T\left(\sum_{j=1}^n x_j e_j\right) \sum_{j=1}^n x_j F_j$$

where $F_j = T(e_j) \in \mathbb{S}$. Hence we obtain (1.4.1) as a special case.

Remark 1.24 In most control applications, LMI's arise as functions of *matrix variables* rather than scalar valued decision variables. This means that we consider inequalities of the form (1.4.2) where $\mathcal{X} = \mathbb{R}^{m_1 \times m_2}$ is the set of real matrices of dimension $m_1 \times m_2$. A simple example with $m_1 = m_2 = m$ is the Lyapunov inequality $F(X) = A^\top X + XA + Q \prec 0$ where $A, Q \in \mathbb{R}^{m \times m}$ are assumed to be given and X is the unknown matrix variable of dimension $m \times m$. Note that this defines an LMI only if $Q \in \mathbb{S}^m$. We can view this LMI as a special case of (1.4.1) by defining an arbitrary basis $e_1, \ldots, e_n$ of $\mathcal{X}$ and expanding $X \in \mathcal{X}$ as $X = \sum_{j=1}^n x_j e_j$. Then

$$F(X) = F\left(\sum_{j=1}^n x_j e_j\right) = F_0 + \sum_{j=1}^n x_j F(e_j) = F_0 + \sum_{j=1}^n x_j F_j$$

which is of the form (1.4.1). The coefficients x_j in the expansion of X define the decision variables. The number of (independent) decision variables n corresponds to the dimension of $\mathcal{X}$. The number n

is at most m^2 (or $m_1 \times m_2$ for non-square matrix variabes) and will depend on the structure imposed on the matrix variable X. For example, if the matrix variable X is required to be symmetric, $\mathcal{X} = \mathbb{S}^m$ which has a basis of $n = m(m+1)/2$ matrix-valued elements. If X is required to be diagonal then $n = m$.

Remark 1.25 A *non-strict LMI* is a linear matrix inequality where $\prec$ in (1.4.1) and (1.4.2) is replaced by $\preceq$. The matrix inequalities $F(x) \succ 0$, and $F(x) \prec G(x)$ with F and G affine functions are obtained as special cases of Definition 1.22 as they can be rewritten as the linear matrix inequalities $-F(x) \prec 0$ and $F(x) - G(x) \prec 0$.

1.4.2 Why are they interesting?

The linear matrix inequality (1.4.2) defines a *convex constraint* on x. That is, the set

$$\mathcal{S} := \{x \mid F(x) \prec 0\}$$

of solutions of the LMI $F(x) \prec 0$ is convex. Indeed, if $x_1, x_2 \in \mathcal{S}$ and $\alpha \in (0, 1)$ then

$$F(\alpha x_1 + (1-\alpha)x_2) = \alpha F(x_1) + (1-\alpha)F(x_2) \prec 0$$

where we used that F is affine and where the inequality follows from the fact that $\alpha > 0$ and $(1-\alpha) > 0$.

Although the convex constraint $F(x) \prec 0$ on x may seem rather special, it turns out that many convex sets can be represented in this way and that these sets have more attractive properties than general convex sets. In this subsection we discuss some seemingly trivial properties of linear matrix inequalities which turn out to be of eminent help to reduce multiple constraints on an unknown variable to an equivalent constraint involving a single linear matrix inequality.

Definition 1.26 (System of LMI's) A *system of linear matrix inequalities* is a finite set of linear matrix inequalities

$$F_1(x) \prec 0, \ldots, F_k(x) \prec 0. \tag{1.4.3}$$

From Proposition 1.4 we infer that the intersection of the feasible sets of each of the inequalities (1.4.3) is convex. In other words, the set of all x that satisfy (1.4.3) is convex. The question now arises whether or not this set can be represented as the feasibility set of another LMI. The answer is yes. Indeed, $F_1(x) \prec 0, \ldots, F_k(x) \prec 0$ if and only if

$$F(x) := \begin{pmatrix} F_1(x) & 0 & \ldots & 0 \\ 0 & F_2(x) & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & F_k(x) \end{pmatrix} \prec 0.$$

The last inequality indeed makes sense as $F(x)$ is symmetric (or Hermitian) for any x. Further, since the set of eigenvalues of $F(x)$ is simply the union of the eigenvalues of $F_1(x), \ldots, F_k(x)$, any x that satisfies $F(x) \prec 0$ also satisfies the system of LMI's (1.4.3) and vice versa. Conclude that multiple LMI constraints can always be converted to a single LMI constraint.

A second important property amounts to incorporating *affine constraints* in linear matrix inequalities. By this, we mean that *combined constraints* (in the unknown x) of the form

$$\begin{cases} F(x) \prec 0 \\ Ax = a \end{cases} \qquad \text{or} \qquad \begin{cases} F(x) \prec 0 \\ x = Bu + b \text{ for some } u \end{cases}$$

where the affine function $F : \mathbb{R}^n \to \mathbb{S}$, matrices A and B and vectors a and b are given, can be *lumped* in one linear matrix inequality $\widetilde{F}(\widetilde{x}) \prec 0$. More generally, the combined equations

$$\begin{cases} F(x) \prec 0 \\ x \in \mathcal{M} \end{cases} \tag{1.4.4}$$

where $\mathcal{M}$ is an *affine set* in $\mathbb{R}^n$ can be rewritten in the form of one single linear matrix inequality $\hat{F}(\hat{x}) \prec 0$ so as to *eliminate* the affine constraint. To do this, recall that affine sets $\mathcal{M}$ can be written as

$$\mathcal{M} = \{x \mid x = x_0 + m,\ m \in \mathcal{M}_0\}$$

with $x_0 \in \mathbb{R}^n$ and $\mathcal{M}_0$ a linear subspace of $\mathbb{R}^n$. Suppose that $\tilde{n} = \dim(\mathcal{M}_0)$ and let $e_1, \ldots, e_{\tilde{n}} \in \mathbb{R}^n$ be a basis of $\mathcal{M}_0$. Let $F(x) = F_0 + T(x)$ be decomposed as in Remark 1.23. Then (1.4.4) can be rewritten as

$$\begin{aligned} 0 \succ F(x) = F_0 + T(x_0 + \sum_{j=1}^{\tilde{n}} x_j e_j) = \\ = F_0 + T(x_0) + \sum_{j=1}^{\tilde{n}} x_j T(e_j) = \widetilde{F}_0 + x_1 \widetilde{F}_1 + \ldots + x_{\tilde{n}} \widetilde{F}_{\tilde{n}} =: \widetilde{F}(\widetilde{x}) \end{aligned}$$

where $\widetilde{F}_0 = F_0 + T(x_0)$, $\widetilde{F}_j = T(e_j)$ and $\widetilde{x} = \mathrm{col}(x_1, \ldots, x_{\tilde{n}})$ are the coefficients of $x - x_0$ in the basis of $\mathcal{M}_0$. This implies that $x \in \mathbb{R}^n$ satisfies (1.4.4) if and only if $\widetilde{F}(\widetilde{x}) \prec 0$. Note that the dimension $\tilde{n}$ of $\widetilde{x}$ is at most equal to the dimension n of x.

A third property of LMI's is obtained from a simple algebraic observation. If M is a square matrix and T is non-singular, then the product T^*MT is called a *congruence transformation* of M. For Hermitian matrices M such a transformation does not change the number of positive and negative eigenvalues of M. Indeed, if vectors u and v are related according to $u = Tv$ with T non-singular, then $u^*Mu < 0$ for all nonzero u is equivalent to saying that $v^*T^*MTv < 0$ for all nonzero v. Hence $M \prec 0$ if and only if $T^*MT \prec 0$. Apply this insight to a partitioned Hermitian matrix

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

with M_{11} square and non-singular by computing the congruence transformation T^*MT with the non-singular matrix $T = \begin{pmatrix} I & 0 \\ -M_{11}^{-1}M_{12} & I \end{pmatrix}$. This yields that

$$M \prec 0 \iff \begin{pmatrix} M_{11} & 0 \\ 0 & S \end{pmatrix} \prec 0 \iff \begin{cases} M_{11} \prec 0 \\ S \prec 0 \end{cases}$$

where

$$S := M_{22} - M_{21}M_{11}^{-1}M_{12}$$

is the *Schur complement* of M_{11} in M. A similar result is obtained with $T = \begin{pmatrix} I & -M_{22}^{-1}M_{21} \\ 0 & I \end{pmatrix}$. This observation can be exploited to derive a powerful result to linearize some *non-linear* inequalities to *linear inequalities*:

Proposition 1.27 (Schur complement) *Let $F : \mathcal{X} \to \mathbb{H}$ be an affine function which is partitioned according to*

$$F(x) = \begin{pmatrix} F_{11}(x) & F_{12}(x) \\ F_{21}(x) & F_{22}(x) \end{pmatrix}$$

where $F_{11}(x)$ is square. Then $F(x) \prec 0$ if and only if

$$\begin{cases} F_{11}(x) \prec 0 \\ F_{22}(x) - F_{21}(x)\,[F_{11}(x)]^{-1}\,F_{12}(x) \prec 0. \end{cases} \tag{1.4.5}$$

if and only if

$$\begin{cases} F_{22}(x) \prec 0 \\ F_{11}(x) - F_{12}(x)\,[F_{22}(x)]^{-1}\,F_{21}(x) \prec 0 \end{cases} . \tag{1.4.6}$$

The second inequalities in (1.4.5) and (1.4.6) are *non-linear* constraints in x. Using this result, it follows that non-linear matrix inequalities of the form (1.4.5) or (1.4.6) can be converted to linear matrix inequalities. In particular, non-linear inequalities of the form (1.4.5) or (1.4.6) define convex constraints on the variable x in the sense that the solution set of these inequalities is convex.

1.4.3 What are they good for?

As we will see, many optimization problems in control, identification and signal processing can be formulated (or reformulated) using linear matrix inequalities. Clearly, it only makes sense to cast these problems in an LMI setting if these inequalities can be solved in an efficient and reliable way. Since the linear matrix inequality $F(x) \prec 0$ defines a *convex constraint* on the variable x, optimization problems involving the minimization (or maximization) of a performance function $f : \mathcal{S} \to \mathbb{R}$ with $\mathcal{S} := \{x \mid F(x) \prec 0\}$ belong to the class of *convex optimization problems*. Casting this in the setting of the previous section, it may be apparent that the full power of convex optimization theory can be employed if the performance function f is known to be convex.

Suppose that $F : \mathcal{X} \to \mathbb{S}$ is affine. There are two generic problems related to the study of linear matrix inequalities:

(a) **Feasibility:** The question whether or not there exist elements $x \in \mathcal{X}$ such that $F(x) \prec 0$ is called a *feasibility problem*. The LMI $F(x) \prec 0$ is called *feasible* if such x exists, otherwise it is said to be *infeasible*.

(b) **Optimization:** Let an *objective function* $f : \mathcal{S} \to \mathbb{R}$ where $\mathcal{S} = \{x \mid F(x) \prec 0\}$. The problem to determine

$$V_{\text{opt}} = \inf_{x \in \mathcal{S}} f(x)$$

is called an *optimization problem with an LMI constraint*. This problem involves the determination of V_{opt}, the calculation of an *almost optimal solution* x (i.e., for arbitrary $\varepsilon > 0$ the calculation of an $x \in \mathcal{S}$ such that $V_{\text{opt}} \leq f(x) \leq V_{\text{opt}} + \varepsilon$), or the calculation of a optimal solutions x_{opt} (elements $x_{\text{opt}} \in \mathcal{S}$ such that $V_{\text{opt}} = f(x_{\text{opt}})$).

Let us give some simple examples to motivate the study of these problems.

Example 1: stability

Consider the problem to determine exponential stability of the linear autonomous system

$$\dot{x} = Ax \tag{1.4.7}$$

where $A \in \mathbb{R}^{n\times n}$. By this, we mean the problem to decide whether or not there exists positive constants K and α such that for any initial condition x_0 the solution $x(t)$ of (1.4.7) with $x(t_0) = x_0$ satisfies the bound

$$\|x(t)\| \leq \|x(t_0)\| M e^{-\alpha(t-t_0)}, \quad \text{for all } t \geq t_0. \tag{1.4.8}$$

Lyapunov taught us that the system (1.4.7) is exponentially stable if and only if there exists $X = X^\top$ such that $X \succ 0$ and $A^\top X + XA \prec 0$. Indeed, in that case the function $V(x) := x^\top X x$ qualifies as a Lyapunov function in that it is positive for all non-zero x and strictly decaying along solutions x of (1.4.7). In Chapter 3 we show that (1.4.8) holds with $K^2 = \lambda_{\max}(X)/\lambda_{\min}$ and $\alpha > 0$ such that $A^\top X + XA + \alpha X \prec 0$. Thus, exponential stability of the system (1.4.7) is equivalent to the feasibility of the LMI

$$\begin{pmatrix} -X & 0 \\ 0 & A^\top X + XA \end{pmatrix} \prec 0.$$

Example 2: μ-analysis

Experts in μ-analysis (but other people as well!) regularly face the problem to determine a diagonal matrix D such that $\|DMD^{-1}\| < 1$ where M is some given matrix. Since

$$\begin{aligned}\|DMD^{-1}\| < 1 &\iff D^{-\top}M^\top D^\top DMD^{-1} \prec I\\ &\iff M^\top D^\top DM \prec D^\top D\\ &\iff M^\top XM - X \prec 0\end{aligned}$$

where $X := D^\top D \succ 0$, we see that the existence of such a matrix is an LMI feasibility problem where $\mathcal{X}$ needs to be taken as the set of diagonal matrices.

Example 3: singular value minimization

Let $F : \mathcal{X} \to \mathbb{S}$ be an affine function and let $\sigma_{\max}(\cdot)$ denote the maximal singular value of a real symmetric matrix. Consider the problem to minimize $f(x) := \sigma_{\max}(F(x))$ over x. Clearly,

$$\begin{aligned}f(x) < \gamma \iff \lambda_{\max}\left(F^\top(x)F(x)\right) < \gamma^2 \iff \frac{1}{\gamma}F^\top(x)F(x) - \gamma I \prec 0\\ \iff \begin{pmatrix}\gamma I & F(x)\\ F^\top(x) & \gamma I\end{pmatrix} \succ 0\end{aligned}$$

where the latter inequality is obtained by taking a Schur complement. If we define

$$y := \begin{pmatrix}\gamma\\ x\end{pmatrix}, \quad G(y) := -\begin{pmatrix}\gamma I & F(x)\\ F^\top(x) & \gamma I\end{pmatrix}, \quad g(y) := \gamma$$

then G is an *affine function* of y and the problem to minimize f over x is equivalent to the problem to minimize g over all y such that $G(y) \prec 0$. Hence, this is an optimization problem with an LMI constraint and a *linear objective function* g.

Example 4: simultaneous stabilization

Consider k linear time-invariant systems of the form

$$\dot{x} = A_i x + B_i u$$

where $A_i \in \mathbb{R}^{n\times n}$ and $B_i \in \mathbb{R}^{n\times m}$, $i = 1, \ldots, k$. The question of *simultaneous stabilization* amounts to finding a state feedback law $u = Fx$ with $F \in \mathbb{R}^{m\times n}$ such that each of the k autonomous systems $\dot{x} = (A_i + B_i F)x$, $i = 1, \ldots, k$ is asymptotically stable. Using Example 1 above, this problem is solved when we can find matrices F and X_i, $i = 1, \ldots, k$, such that for all of these i's

$$\begin{cases}X_i \succ 0\\ (A_i + B_i F)^\top X_i + X_i(A_i + B_i F) \prec 0\end{cases}. \tag{1.4.9}$$

Since both X_i and F are unknown, this is *not* a system of LMI's in the variables X_i and F. One way out of this inconvenience is to require that $X_1 = \ldots = X_k =: X$. If we introduce new variables $Y = X^{-1}$ and $K = FY$ then (1.4.9) reads

$$\begin{cases} Y \succ 0 \\ A_i Y + Y A_i^\top + B_i K + K^\top B_i^\top \prec 0 \end{cases}$$

for $i = 1, \ldots, k$. The latter is a system of LMI's in the variables Y and K. The joint stabilization problem therefore has a solution $F = KY^{-1}$ if the latter system of LMI's is feasible. We will see in Chapter 3 that the quadratic function $V(x) := x^\top X x$ serves as a joint Lyapunov function for the k autonomous systems.

Example 5: evaluation of quadratic cost

Consider the linear autonomous system

$$\dot{x} = Ax \tag{1.4.10}$$

together with an arbitrary (but fixed) initial value $x(0) = x_0$ and the criterion function $J := \int_0^\infty x^\top(t) Q x(t)\, dt$ where $Q = Q^\top \succeq 0$. Assume that the system is asymptotically stable. Then all solutions x of (1.4.10) are square integrable so that $J < \infty$. Now consider the non-strict linear matrix inequality $A^\top X + XA + Q \preceq 0$. For any solution $X = X^\top$ of this LMI we can differentiate $x^\top(t) X x(t)$ along solutions x of (1.4.10) to get

$$\frac{d}{dt}[x^\top(t) X x(t)] = x^\top(t)[A^\top X + XA]x(t) \leq -x^\top(t) Q x(t).$$

If we assume that $X \succ 0$ then integrating the latter inequality from $t = 0$ till ∞ yields the upper bound

$$J = \int_0^\infty x^\top(t) Q x(t)\, dt \leq x_0^\top X x_0.$$

where we used that $\lim_{t\to\infty} x(t) = 0$. Moreover, the smallest upperbound of J is obtained by minimizing the function $f(X) := x_0^\top X x_0$ over all $X = X^\top$ which satisfy $X \succ 0$ and $A^\top X + XA + Q \preceq 0$. Again, this is an optimization problem with an LMI constraint.

Example 6: a Leontief economy

A manufacturer is able to produce n different products from m different resources. Assume that the selling price of product j is p_j and that it takes the manufacturer a_{ij} units of resource i to produce one unit of product j. Let x_j denote the amount of product j that is to be produced and let a_i denote the amount of available units of resource i, $i = 1, \ldots, m$. A smart manager advised the manufacturer to maximize his profit

$$p(x_1, \ldots, x_n) := p_1 x_1 + p_2 x_2 + \ldots + p_n x_n,$$

but the manager can do this only subject to the production constraints

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &\le a_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &\le a_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &\le a_m \end{aligned}$$

and $x_j \ge 0$, $j = 1, \ldots, n$. Note that the manufacturer faces an optimization problem subject to a system of non-strict linear matrix inequalities.

Wassily Leontief was born in 1906 in St. Petersburg and is winner of the 1973 Nobel Prize of Economics. Among many things, he used input-output analysis to study the characteristics of trade flow between the U.S. and other countries.

1.4.4 How are they solved?

The problems defined in the previous subsection can be solved with efficient numerical tools. In this section we discuss the basic ideas behind the 'LMI-solvers'.

Ellipsoid method

We first give a solution which is based on the Ellipsoid Algorithm 1.21 discussed in Section 1.3. As mentioned before, this algorithm is simple, numerically robust and easy to implement but may be slow for larger optimization problems.

We will apply this algorithm to the feasibility problem defined in subsection 1.4.3. Let $F : \mathbb{R}^n\mathbb{S}$ be affine. Recall that $F(x) \prec 0$ if and only if $\lambda_{\max}(F(x)) < 0$. Define, for $x \in \mathbb{R}^n$, the function $f(x) := \lambda_{\max}(F(x))$ and consider the optimal value $V_{\text{opt}} := \inf_{x \in \mathbb{R}^n} f(x)$. Clearly, the LMI $F(x) \prec 0$ is feasible if and only if $V_{\text{opt}} < 0$. It is infeasible if and only if $V_{\text{opt}} \ge 0$.

There are a few observations to make to apply Proposition 1.19 and the ellipsoid algorithm for this optimization problem. The first one is to establish that f is a convex function. Indeed, $\mathcal{S}$ is convex and for all $0 < \alpha < 1$ and $x_1, x_2 \in \mathcal{S}$ we have that

$$\begin{aligned} f(\alpha x_1 + (1-\alpha)x_2) &= \lambda_{\max}(F(\alpha x_1 + (1-\alpha)x_2)) \\ &= \lambda_{\max}(\alpha F(x_1) + (1-\alpha)F(x_2)) \\ &\le \alpha\lambda_{\max}(F(x_1)) + (1-\alpha)\lambda_{\max}(F(x_2)) \\ &= \alpha f(x_1) + (1-\alpha) f(x_2) \end{aligned}$$

which shows that f is convex. Second, to apply Step 1 of the algorithm, for any x_0 we need to determine a subgradient g of f at the point x_0. To do this, we will use the fact that

$$f(x) = \lambda_{\max}(F(x)) = \max_{u^\top u=1} u^\top F(x)u.$$

This means that for an arbitrary $x_0 \in \mathcal{S}$ we can determine a vector $u_0 \in \mathbb{R}^N$, depending on x_0, with $u_0^\top u_0 = 1$ such that $\lambda_{\max}(F(x_0)) = u_0^\top F(x_0)u_0$. But then

$$\begin{aligned} f(x) - f(x_0) &= \max_{u^\top u=1} u^\top F(x)u - u_0^\top F(x_0)u_0 \\ &\geq u_0^\top F(x)u_0 - u_0^\top F(x_0)u_0 \\ &= u_0^\top \left(F(x) - F(x_0)\right) u_0. \end{aligned}$$

The last expression is an affine functional which vanishes at x_0. This means that the right-hand side of this expression must be of the form $\langle g, x - x_0\rangle$ for some vector $g \in \mathbb{R}^n$. To obtain g, we can write

$$\begin{aligned} u_0^\top F(x)u_0 &= \underbrace{u_0^\top F_0 u_0}_{g_0} + \sum_{j=1}^{n} x_j \underbrace{u_0^\top F_j u_0}_{g_j} \\ &= g_0 + \langle g, x\rangle, \end{aligned}$$

where g_j are the components of g. In particular, we obtain that $f(x) - f(x_0) \geq \langle g, x - x_0\rangle$. The remaining steps of the ellipsoid algorithm can be applied in a straightforward way.

Interior point methods

A major breakthrough in convex optimization lies in the introduction of interior-point methods. These methods were developed in a series of papers [11] and became of true interest in the context of LMI problems in the work of Yurii Nesterov and Arkadii Nemirovskii [20].

The main idea is as follows. Let F be an affine function and let $\mathcal{S} := \{x \mid F(x) \prec 0\}$ be the domain of a convex function $f : \mathcal{S} \to \mathbb{R}$ which we wish to minimize. That is, we consider the convex optimization problem

$$V_{\text{opt}} = \inf_{x\in\mathcal{S}} f(x).$$

To solve this problem (that is, to determine optimal or almost optimal solutions), it is first necessary to introduce a *barrier function*. This is a smooth function ϕ which is required to

(a) be *strictly convex* on the interior of $\mathcal{S}$ and

(b) approach $+\infty$ along each sequence of points $\{x_n\}_{n=1}^{\infty}$ in the interior of $\mathcal{S}$ that converges to a boundary point of $\mathcal{S}$.

Given such a barrier function ϕ, the constraint optimization problem to minimize $f(x)$ over all $x \in \mathcal{S}$ is replaced by the *unconstrained optimization problem* to minimize the functional

$$f_t(x) := t f(x) + \phi(x) \tag{1.4.11}$$

where $t > 0$ is a so called *penalty parameter*. Note that f_t is strictly convex on $\mathbb{R}^n$. The main idea is to determine a mapping $t \mapsto x(t)$ that associates with any $t > 0$ a minimizer $x(t)$ of f_t. Subsequently, we consider the behavior of this mapping as the penalty parameter t varies. In almost all interior point methods, the latter unconstrained optimization problem is solved with the classical Newton-Raphson iteration technique to approximate the minimum of f_t. Under mild assumptions and for a suitably defined sequence of penalty parameters t_n with $t_n \to \infty$ as $n \to \infty$, the sequence $x(t_n)$ with $n \in \mathbb{Z}_+$ will converge to a point x_{opt} which is a solution of the original convex optimization problem. That is, the limit $x_{\text{opt}} := \lim_{t\to\infty} x(t)$ exists and $V_{\text{opt}} = f(x_{\text{opt}})$.

A small modification of this theme is obtained by replacing the original constraint optimization problem by the unconstrained optimization problem to minimize

$$g_t(x) := \phi_0(t - f(x)) + \phi(x) \tag{1.4.12}$$

where $t > t_0 := V_{\text{opt}}$ and ϕ_0 is a barrier function for the non-negative real half-axis. Again, the idea is to determine, for every $t > 0$ a minimizer $x(t)$ of g_t (typically using the classical Newton-Raphson algorithm) and to consider the 'path' $t \mapsto x(t)$ as function of the penalty parameter t. The curve $t \mapsto x(t)$ with $t > t_0$ is called the *path of centers* for the optimization problem. Under suitable conditions the solutions $x(t)$ are analytic and have a limit as $t \downarrow t_0$, say x_{opt}. The point $x_{\text{opt}} := \lim_{t\downarrow t_0} x(t)$ is optimal in the sense that $V_{\text{opt}} = f(x_{\text{opt}})$ since for $t > t_0$, $x(t)$ is feasible and satisfies $f(x(t)) < t$.

Interior point methods can be applied to either of the two LMI problems as defined in the previous section. If we consider the *feasibility problem* associated with the LMI $F(x) \prec 0$ then (f does not play a role and) one candidate barrier function is the logarithmic function

$$\phi(x) := \begin{cases} \log\det -F(x)^{-1} & \text{if } x \in \mathcal{S} \\ \infty & \text{otherwise} \end{cases}.$$

Under the assumption that the feasible set $\mathcal{S}$ is bounded and non-empty, it follows that ϕ is strictly convex and hence it defines a barrier function for the feasibility set $\mathcal{S}$. By invoking Proposition 1.12, we know that there exists a unique x_{opt} such that $\phi(x_{\text{opt}})$ is the global minimum of ϕ. The point x_{opt} obviously belongs to $\mathcal{S}$ and is called the *analytic center* of the feasibility set $\mathcal{S}$. It is usually obtained in a very efficient way from the classical Newton iteration

$$x_{k+1} = x_k - \left(\phi''(x_k)\right)^{-1} \phi'(x_k). \tag{1.4.13}$$

Here ϕ' and ϕ'' denote the gradient and the Hessian of ϕ, respectively.

The convergence of this algorithm can be analyzed as follows. Since ϕ is strongly convex and sufficiently smooth, there exist numbers L and M such that for all vectors u with norm $\|u\| = 1$

there holds

$$\begin{aligned} &u^\top \phi''(x)u \geq M \\ &\|\phi''(x)u - \phi''(y)u\| \leq L\|x-y\|. \end{aligned}$$

In that case,

$$\|\phi'(x_{k+1})\|^2 \leq \frac{L}{2M^2}\|\phi'(x_k)\|^2$$

so that whenever the initial value x_0 is such that $\frac{L}{2M^2}\|\phi'(x_0)\| < 1$ the method is guaranteed to converge *quadratically*.

The idea will be to implement this algorithm in such a way that quadratic convergence can be guaranteed for the largest possible set of initial values x_0. For this reason the iteration (1.4.13) is modified as follows

$$x_{k+1} = x_k - \alpha_k(\lambda(x_k))\left(\phi''(x_k)\right)^{-1}\phi'(x_k)$$

where

$$\alpha_k(\lambda) := \begin{cases} 1 & \text{if } \lambda < 2-\sqrt{3} \\ \frac{1}{1+\lambda} & \text{if } \lambda \geq 2-\sqrt{3} \end{cases}.$$

and $\lambda(x) := \sqrt{\phi'(x)^\top \phi''(x)\phi'(x)}$ is the so called *Newton decrement* associated with ϕ. It is this damping factor that guarantees that x_k will converge to the analytic center x_{opt}, the unique minimizer of ϕ. It is important to note that the step-size is variable in magnitude. The algorithm guarantees that x_k is always feasible in the sense that $x_k \in \mathcal{S}$ and that x_k converges globally to a minimizer x_{opt} of ϕ. It can be shown that $\phi(x_k) - \phi(x_{\text{opt}}) \leq \varepsilon$ whenever

$$k \geq c_1 + c_2 \log\log(1/\varepsilon) + c_3\left(\phi(x_0) - \phi(x_{\text{opt}})\right)$$

where c_1, c_2 and c_3 are constants. The first and second terms on the right-hand side do not dependent on the optimization criterion and the specific LMI constraint. The second term can almost be neglected for small values of ε.

The *optimization problem* to minimize $f(x)$ subject to the LMI $F(x) \prec 0$ can be viewed as a feasibility problem for the LMI

$$\widetilde{F}_t(x) := \begin{pmatrix} f(x)-t & 0 \\ 0 & F(x) \end{pmatrix} \prec 0.$$

where $t > t_0 := \inf_{x\in\mathcal{S}} f(x)$ is a penalty parameter. Using the same barrier function for this linear matrix inequality yields the unconstrained optimization problem to minimize

$$g_t(x) := \log\det -\hat{F}_t(x)^{-1} = \underbrace{\log\frac{1}{t-f(x)}}_{\phi_0(t-f(x))} + \underbrace{\log\det -F(x)^{-1}}_{\phi(x)}$$

which is of the form (1.4.12). Due to the strict convexity of g_t the minimizer $x(t)$ of g_t is unique for all $t > t_0$. It can be shown that the sequence $x(t)$ is feasible for all $t > t_0$ and approaches the infimum $\inf_{x\in\mathcal{S}} f(x)$ as $t \downarrow t_0$.

1.4.5 What are dual LMI problems?

Let $\mathcal{X}$ be a finite dimensional vector space and define a *linear objective function* $f : \mathcal{X} \to \mathbb{R}$ by setting $f(x) = c^\top x = \langle x, c\rangle$ where $c \in \mathcal{X}$. We consider the optimization problem to determine

$$P_{\rm opt} = \inf_{x \in \mathcal{S}} f(x)$$

where the space of feasible decisions $\mathcal{S}$ is defined as

$$\mathcal{S} = \{x \in \mathcal{X} \mid G(x) \preceq 0,\ H(x) = 0\}.$$

Here, $G : \mathcal{X} \to \mathcal{Y}$ and $H : \mathcal{X} \to \mathcal{Z}$ are affine functions, and $\mathcal{Y}$ and $\mathcal{Z}$ are finite dimensional spaces. To properly interpret the inequality $G(x) \preceq 0$, the space $\mathcal{Y}$ is equal to the set $\mathbb{S}$ of real symmetric matrices. Since G and H are affine, we can write $G(x) = G_0 + G_1(x)$ and $H(x) = H_0 + H_1(x)$ where G_1 and H_1 are linear transformations. This optimization problem is similar to the 'primal problem' which we discussed in Section 1.3.3, except for the fact that the constraints $G(x) \preceq 0$ and $H(x) = 0$ are now matrix valued. Following the terminology of Section 1.3.3, this is an example of a *linear convex program*. The aim of this subsection is to generalize the duality results of Section 1.3.3 to the matrix valued case and to establish a precise formulation of the *dual optimization problem*.

The set $\mathcal{Y} = \mathbb{S}$ of real symmetric matrices becomes a Hilbert space when equipped with the inner product $\langle y_1, y_2\rangle := \mathrm{trace}(y_1 y_2)$. Here, y_1 and y_2 are matrices in $\mathcal{Y}$. Similarly, we equip $\mathcal{Z}$ with an inner product $\langle\cdot,\cdot\rangle$, so as to make $\mathcal{Z}$ a Hilbert space as well.

With reference to Remark 1.18, define the closed convex cone $\mathcal{K} := \{Y \in \mathcal{Y} \mid Y \preceq 0\}$ and note that the inequality $G(x) \preceq 0$ is equivalent to $G(x) \in \mathcal{K}$. Obviously, every $y \in \mathcal{Y}$ defines a linear functional $g(\cdot) := \langle y, \cdot\rangle$ on $\mathcal{Y}$. Conversely, every linear functional $g : \mathcal{Y} \to \mathbb{R}$ uniquely defines an element $y \in \mathcal{Y}$ such that $g(\cdot) = \langle y, \cdot\rangle$. In particular, for all $k \in \mathcal{K}$ the linear functional $g(k) = \langle y, k\rangle$ is non-positive if and only if $y \succeq 0$ (in the sense that the smallest eigenvalue of y is non-negative). Hence, the Lagrangian $L : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}$ defined by

$$L(x, y, z) := \langle x, c\rangle + \langle y, G(x)\rangle + \langle z, H(x)\rangle$$

satisfies

$$\ell(y, z) := \inf_{x \in \mathcal{X}} L(x, y, z) \leq \inf_{x \in \mathcal{S}} \langle x, c\rangle = P_{\rm opt}.$$

for all $y \in \mathcal{Y}$, $z \in \mathcal{Z}$, $y \succeq 0$ and the dual optimization problem amounts to determining

$$D_{\rm opt} := \sup_{y \succeq 0,\, z} \ell(y, z) = \sup_{y \succeq 0,\, z} \inf_{x \in \mathcal{X}} L(x, y, z).$$

We then obtain the following generalization of Theorem 1.16.

Theorem 1.28 *Under the conditions given in this subsection, if there exists* $(x_0, y_0, z_0) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ *such that* $G(x_0) \prec 0$, $H(x_0) = 0$, $y_0 \succ 0$ *and* $c + G_1^*(y_0) + H_1^*(z_0) = 0$ *then both the primal and the dual optimization problem admit optimal solutions and*

$$P_{opt} = \min_{G(x) \preceq 0,\ H(x)=0} \langle x, c\rangle = \max_{y \succeq 0,\ c + G_1^*(y) + H_1^*(z) = 0} \langle G_0, y\rangle + \langle H_0, z\rangle = D_{opt}.$$

Moreover, the triple $(x_{opt}, y_{opt}, z_{opt})$ *is optimal for both the primal and the dual problem if and only if*

(a) $G(x_{opt}) \preceq 0$, $H(x_{opt}) = 0$,

(b) $y_{opt} \succeq 0$, $c + G_1^*(y_{opt}) + H_1^*(z_{opt}) = 0$ *and*

(c) $\langle y_{opt}, G(x_{opt})\rangle = 0$.

Proof. Under the given feasibility conditions, the dual optimal value

$$\begin{aligned} D_{\text{opt}} &= \max_{y \succeq 0,\, z} \inf_x \langle c, x\rangle + \langle y, G_0\rangle + \langle z, H_0\rangle + \langle G_1^*(y), x\rangle + \langle H_1^*(z), x\rangle \\ &= \max_{y \succeq 0,\, z} \inf_x \langle c + G_1^*(y) + H_1^*(z), x\rangle + \langle G_0, y\rangle + \langle H_0, z\rangle \\ &= \max_{y \succeq 0,\, c + G_1^*(y) + H_1^*(z) = 0} \langle G_0, y\rangle + \langle H_0, z\rangle \end{aligned}$$

where the last equality follows by dualization of the dual problem. Since the dual problem satisfies the constraint qualification by assumption, the primal problem admits an optimal solution. Let $(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}})$ be optimal for both the primal and the dual problem. Then items 1 and 2 are immediate and item 1 implies that $\langle x_{\text{opt}}, c\rangle \leq L(x, y_{\text{opt}}, z_{\text{opt}})$ for all x. With $x = x_{\text{opt}}$ this yields $\langle y_{\text{opt}}, G(x\text{opt})\rangle \geq 0$. On the other hand, items 1 and 2 imply $\langle y_{\text{opt}}, G(x\text{opt})\rangle \leq 0$ and we may conclude item 3. Conversely, if items 1, 2 and 3 hold, then it is easily verified that L satisfies the saddle-point property

$$L(x_{\text{opt}}, y, z) \leq L(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}}) \leq L(x, y_{\text{opt}}, z_{\text{opt}}), \qquad \text{for all } x, y \geq 0, z$$

The first inequality shows that $(y_{\text{opt}}, z_{\text{opt}})$ is an optimal solution for the dual problem. Likewise, the second inequality shows that x_{opt} is optimal for the primal problem. ∎

1.4.6 When were they invented?

Contrary to what many authors nowadays seem to suggest, the study of linear matrix inequalities in the context of dynamical systems and control goes back a long way in history and probably starts with the fundamental work of Aleksandr Mikhailovich Lyapunov on the stability of motion. Lyapunov was a school friend of Markov (yes, the one of the Markov parameters) and later a student of Chebyshev. Around 1890, Lyapunov made a systematic study of the local expansion and contraction properties of motions of dynamical systems around an attractor. He worked out the idea that an invariant set of a differential equation is stable in the sense that it attracts all solutions if one can find a function that is bounded from below and decreases along all solutions outside the invariant set.

Aleksandr Mikhailovich Lyapunov was born on May 25, 1857 and published in 1892 his work 'The General Problem of the Stability of Motion' in which he analyzed the question of stability of

Figure 1.2: Aleksandr Mikhailovich Lyapunov

equilibrium motions of mechanical systems. This work served as his doctoral dissertation and was defended on September 1892 in Moscow University. Put into modern jargon, he studied stability of differential equations of the form

$$\dot{x} = A(x)$$

where $A : \mathbb{R}^n \to \mathbb{R}^n$ is some analytic function and x is a vector of positions and velocities of material taking values in a finite dimensional state space $\mathcal{X} = \mathbb{R}^n$. As Theorem I in Chapter 1, section 16 it contains the statement[3] that

> *if the differential equation of the disturbed motion is such that it is possible to find a definite function V of which the derivative V' is a function of fixed sign which is opposite to that of V, or reduces identically to zero, the undisturbed motion is stable.*

The intuitive idea behind this result is that the so called *Lyapunov function V* can be viewed as a generalized 'energy function' (in the context of mechanical systems the kinetic and potential energies always served as typical Lyapunov functions). A system is then stable if it is 'dissipative' in the sense that the Lyapunov function decreases. Actually, this intuitive idea turns out to be extremely fruitful in understanding the role of linear matrix inequalities in many problems related to analysis and synthesis of systems. This is why we devote the next chapter to dissipative dynamical systems. We will consider stability issues in much more detail in a later chapter.

[3]Translation by A.T. Fuller as published in the special issue of the International Journal of Control in March 1992 and in [16].

1.5 Further reading

Foundations on convex sets and convex functions have been developed around 1900, mainly by the work of Minkowski in [19]. Detailed modern treatments on the theory of convex functions and convex set analysis can be found in [1, 22, 24, 41]. The theory of convex programs and its relation with Lagrange multipliers and saddle points originates with the work by Kuhn-Tucker in [14]. The constraint qualification assumption is due to Slater. For details on the theory of subgradients we refer to [1, 25]. A standard work on optimization methods with applications in systems theory is the classical book by Leunberger [15]. Interior point methods were developed in a series of papers [11] and have led to major breakthroughs in LMI-solvers in [20]. For general sources on linear algebra we refer to the classics [4] and [8].

1.6 Exercises

Exercise 1
Which of the following statements are true.

(a) $\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \succ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

(b) $A \succ B$ implies that $\lambda_{\max}(A) > \lambda_{\max}(B)$.

(c) $\lambda_{\max}(A + B) \leq \lambda_{\max}(A) + \lambda_{\max}(B)$.

(d) Write $A \in \mathbb{H}$ has $A = X + iY$ with X and Y real. Then $A \succeq 0$ if and only if $\begin{pmatrix} X & Y \\ -Y & X \end{pmatrix} \succeq 0$.

(e) Let $T \in \mathbb{C}^{n\times m}$ and $A \in \mathbb{H}^n$. If $A \succeq 0$ then $T^*AT \succeq 0$.

(f) Let $T \in \mathbb{C}^{n\times m}$ and $A \in \mathbb{H}^n$. If $T^*AT \succeq 0$ then $A \succeq 0$.

Exercise 2
Give an example of a non-convex function $f : \mathcal{S} \to \mathbb{R}$ whose sublevel sets $\mathcal{S}_\alpha$ are convex for all $\alpha \in \mathbb{R}$.

Exercise 3
Let $f : \mathcal{S} \to \mathbb{R}$ be a convex function.

(a) Show the so called *Jensen's inequality* which states that for a convex combination $x = \sum_{i=1}^n \alpha_i x_i$ of $x_1, \ldots x_n \in \mathcal{S}$ there holds that

$$f(\sum_{i=1}^n \alpha_i x_i) \leq \sum_{i=1}^n \alpha_i f(x_i).$$

Hint: A proof by induction on n may be the easiest.

(b) Show that $\mathrm{conv}(\mathcal{S})$ is equal to the set of all convex combinations of $\mathcal{S}$

Exercise 4
Let $\mathcal{S}$ be a subset of a finite dimensional vector space. The *affine hull* $\mathrm{aff}(\mathcal{S})$ of $\mathcal{S}$ is the intersection of all affine sets containing $\mathcal{S}$ (cf. Definition 1.5). An *affine combination* of $x_1, \ldots, x_n \in \mathcal{S}$ is a point

$$x := \sum_{i=1}^{n} \alpha_i x_i$$

where $\sum_{i=1}^{n} \alpha_i = 1$ (cf. Definition 1.3). Show that for any set $\mathcal{S}$ in $\mathbb{R}^n$ the affine hull $\mathrm{aff}(\mathcal{S})$ is affine and consists of all affine combinations of the elements of $\mathcal{S}$. (cf. Proposition 1.6).

Exercise 5
Let $\mathcal{S}$ and $\mathcal{T}$ be finite dimensional vector spaces and let $f : \mathcal{S} \to \mathcal{T}$ be an affine function. Show that

(a) f^{-1} is affine whenever the inverse function exists.

(b) the *graph* $\{(x, y) \mid x \in \mathcal{S}, y = f(x)\}$ of f is affine.

(c) $f(\mathcal{S}')$ is affine (as a subset of $\mathcal{T}$) for every affine subset $\mathcal{S}' \subseteq \mathcal{S}$.

Exercise 6
It should not be surprising that the notion of a convex set and a convex function are related. Let $\mathcal{S} \subseteq \mathbb{R}^n$. Show that a function $f : \mathcal{S} \to \mathbb{R}$ is convex if and only if its *epigraph* $\{(x, y) \mid x \in \mathcal{S}, y \in \mathbb{R}, f(x) \leq y\}$ is a convex set.

Exercise 7
Perform a feasibility test to verify the asymptotic stability of the system $\dot{x} = Ax$, where

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -3 & -4 \end{pmatrix}.$$

Exercise 8
Show that

(a) the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by the quadratic form $f(x) = x^\top Rx + 2s^\top x + q$ is convex if and only if the $n \times n$ matrix $R = R^\top \succeq 0$.

(b) the intersection of the sets $\mathcal{S}_j := \{x \in \mathbb{R}^n \mid x^\top R_j x + 2s_j^\top x + q_j \leq 0\}$ where $j = 1, \ldots, k$ and $R_j \succeq 0$ is convex.

How does $\mathcal{S}_j$ look like if $R_j = 0$? And how if $s_j = 0$?

Exercise 9
Let $\mathcal{S}$ denote the vector space of (closed-loop) single-input and single-output rational transfer functions. Let $f_{\min} : \mathbb{R}_+ \to \mathbb{R}$ and $f_{\max} : \mathbb{R}_+ \to \mathbb{R}$ be two functions such that $f_{\min}(\cdot) \leq f_{\max}(\cdot)$. Consider the following rather typical time and frequency domain specifications:

(a) Frequency response shaping: For $\gamma \geq 0$,

$$\mathcal{S}_\gamma = \{ S \in \mathcal{S} \mid \sup_{\omega \geq 0} | f_{\max}(\omega) - S(i\omega)| \leq \gamma \}.$$

(b) Bandwidth shaping: For $\gamma \geq 0$,

$$\mathcal{S}_\gamma = \{ S \in \mathcal{S} \mid f_{\min}(\omega) \leq |S(i\omega)| \leq f_{\max}(\omega) \text{ for all } 0 \leq \omega \leq \frac{1}{\gamma} \}.$$

(c) Step response shaping:

$$\mathcal{S}_\gamma = \{ S \in \mathcal{S} \mid f_{\min}(t) \leq s(t) \leq f_{\max}(t) \text{ for all } t \geq 0 \}.$$

Here, $s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S(i\omega)}{i\omega} e^{i\omega t}\, d\omega$ is the step response of the system $S \in \mathcal{S}$.

(d) Over- and undershoot: For $\gamma > 0$,

$$\mathcal{S}_\gamma = \{ S \in \mathcal{S} \mid \sup_{t \geq 0} \max(s(t) - f_{\max}(t),\ f_{\min}(t) - s(t), 0) \leq \gamma \}$$

with s the step response as defined in item 3.

Verify for each of these specifications whether or not $\mathcal{S}_\gamma$ defines a convex subset of $\mathcal{S}$.

Exercise 10
Consider the systems $\dot{x} = A_i x + B_i u$ where $i = 1, \ldots, 4$ and

$$A_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} -1 & 2 \\ 1 & 2 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$
$$A_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 0 & 2 \\ 1 & -1 \end{pmatrix}, \quad B_4 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Find a state feedback law $u = Fx$ such that each of the 4 autonomous systems $\dot{x} = (A_i + B_i F)x$, $i = 1, \ldots, k$ is asymptotically stable. (See example 4 in subsection (1.4.3)).

Exercise 11
In this exercise we investigate the stability of the linear time-varying system

$$\dot{x} = A(t)x \tag{1.6.1}$$

where for all $t \in \mathbb{R}_+$ the matrix $A(t)$ is a convex combination of the triple

$$A_1 := \begin{pmatrix} -1 & 1 \\ -1 & -0.2 \end{pmatrix}, \quad A_2 := \begin{pmatrix} -1 & 1 \\ -2 & -0.7 \end{pmatrix}, \quad A_3 := \begin{pmatrix} -2 & 1 \\ -1.2 & 0.4 \end{pmatrix}.$$

That is,

$$A(t) \in \mathrm{conv}(A_1, A_2, A_3)$$

for all values of $t \in \mathbb{R}_+$. This is a *polytopic model*. It is an interesting fact that the time-varying system (1.6.1) is asymptotically stable in the sense that for any initial condition $x_0 \in \mathbb{R}^n$, the solution $x(\cdot)$ of (1.6.1) satisfies $\lim_{t\to\infty} x(t) = 0$ whenever there exists a matrix $X = X^\top \succ 0$ such that

$$\begin{aligned} A_1^\top X + X A_1 &\prec 0 \\ A_2^\top X + X A_2 &\prec 0 \\ A_3^\top X + X A_3 &\prec 0. \end{aligned}$$

(We will come to this fact later!) If such an X exists then (1.6.1) is asymptotically stable *irrespective of how fast the time variations of $A(t)$ take place*! Reformulate the question of asymptotic stability of (1.6.1) as a feasibility problem and find, if possible, a feasible solution X to this problem.

Exercise 12
Consider the dynamical system

$$\dot{x} = Ax + Bu$$

where x is an n-dimensional state and u is a scalar-valued input which is supposed to belong to $\mathcal{U} = \{u : \mathbb{R} \to \mathbb{R} \mid -1 \leq u(t) \leq 1 \text{ for all } t \geq 0\}$. Define the *null controllable subspace* of this system as the set

$$\mathcal{C} := \{x_0 \in \mathbb{R}^n \mid \exists T \geq 0 \text{ and } u \in \mathcal{U} \text{ such that } x(T) = 0\}$$

i.e., the set of initial states that can be steered to the origin of the state space in finite time with constrained inputs. Show that $\mathcal{C}$ is a convex set.

Exercise 13
Let $F : \mathcal{X} \to \mathbb{H}$ be affine and suppose that the LMI $F(x) \prec 0$ is feasible. Prove that there exists $\varepsilon > 0$ such that the LMI $F(x) + \varepsilon I \prec 0$ is also feasible. Does this statement hold for I replaced with any Hermitian matrix?

Chapter 2

Dissipative dynamical systems

2.1 Introduction

The notion of dissipativity is a most important concept in systems theory both for theoretical considerations as well as from a practical point of view. Especially in the physical sciences, dissipativity is closely related to the notion of energy. Roughly speaking, a dissipative system is characterized by the property that at any time the amount of energy which the system can conceivably supply to its environment can not exceed the amount of energy that has been supplied to it. Stated otherwise, when time evolves, a dissipative system absorbs a fraction of its supplied energy and transforms it for example into heat, an increase of entropy, mass, electro-magnetic radiation, or other kinds of energy 'losses'. In many applications, the question whether a system is dissipative or not can be answered from physical considerations on the way the system interacts with its environment. For example, by observing that the system is an interconnection of dissipative components, or by considering systems in which a loss of energy is inherent to the behavior of the system due to friction, optical dispersion, evaporation losses, etc.

In this chapter we will formalize the notion of a dissipative dynamical system for a very general class of systems. It will be shown that linear matrix inequalities occur in a very natural way in the study of linear dissipative systems. Perhaps the most appealing setting for studying LMI's in system and control theory is within the framework of dissipative dynamical systems. It will be shown that solutions of LMI's have a natural interpretation as *storage functions* associated with a dissipative system. This interpretation will play a key role in understanding the importance of LMI's in questions related to stability, performance, robustness, and a large variety of controller design problems.

2.2 Dissipative dynamical systems

2.2.1 Definitions and examples

Consider a continuous time, time-invariant dynamical system Σ described by the equations

$$\dot{x} = f(x, w) \tag{2.2.1a}$$

$$z = g(x, w) \tag{2.2.1b}$$

Here, x is the state which takes its values in a *state space* X, w is the input taking its values in an *input space* W and z denotes the output of the system which assumes its values in the *output space* Z. Throughout this section, the precise representation of the system will not be relevant. What we need, though, is that for any initial condition $x_0 \in X$ and for any input w belonging to an input class $\mathcal{W}$, there exists uniquely defined signals $x : \mathbb{R}_+ \to X$ and $z : \mathbb{R}_+ \to Z$ which satisfy (2.2.1) subject to $x(0) = x_0$. Here, $\mathbb{R}_+ = [0, \infty)$ is the time set. In addition, the output z is assumed to depend on w in a *causal* way; that is, if $w_1 \in \mathcal{W}$ and $w_2 \in \mathcal{W}$ are two input signals that are identical on $[0, T]$ then the outputs z_1 and z_2 of (2.2.1) corresponding to the inputs w_1 and w_2 and the same (but arbitrary) initial condition $x(0) = x_0$ are also identical on $[0, T]$. The system (2.2.1) therefore generates outputs from inputs and initial conditions while future values of the inputs do not have an effect on the past outputs. Let

$$\mathrm{s} : W \times Z \to \mathbb{R}$$

be a mapping and assume that for all $t_0, t_1 \in \mathbb{R}$ and for all input-output pairs (w, z) satisfying (2.2.1) the composite function $\mathrm{s}(w(t), z(t))$ is locally absolutely integrable, i.e., $\int_{t_0}^{t_1} |\mathrm{s}(w(t), z(t))| dt < \infty$. The mapping s will be referred to as the *supply function*.

Definition 2.1 (Dissipativity) The system Σ with supply function s is said to be *dissipative* if there exists a function $V : X \to \mathbb{R}$ such that

$$V(x(t_0)) + \int_{t_0}^{t_1} \mathrm{s}(w(t), z(t)) dt \;\geq\; V(x(t_1)) \tag{2.2.2}$$

for all $t_0 \leq t_1$ and all signals (w, x, z) which satisfy (2.2.1). The pair (Σ, s) is said to be *conservative* if equality holds in (2.2.2) for all $t_0 \leq t_1$ and all (w, x, z) satisfying (2.2.1).

Interpretation 2.2 The supply function (or *supply rate*) s should be interpreted as the supply *delivered to the system*. This means that $\mathrm{s}(w(\cdot), z(\cdot))$ represents the rate at which supply flows into the system if the system generates the input-output pair $(w(\cdot), z(\cdot))$. In other words, in the time interval $[0, T]$ work has been done *on* the system whenever $\int_0^T \mathrm{s}(w(t), z(t)) dt$ is positive, while work is done *by* the system if this integral is negative. The function V is called a *storage function* and generalizes the notion of an energy function for a dissipative system. With this interpretation, inequality (2.2.2) formalizes the idea that a dissipative system is characterized by the property that the change of internal storage $V(x(t_1)) - V(x(t_0))$ in any time interval $[t_0, t_1]$ will never exceed the amount of supply that flows into the system. This means that part of what is supplied to the system is stored, while the remaining part is dissipated. Inequality (2.2.2) is known as the *dissipation inequality*.

We stress that, contrary to the definition in the classical papers [45,46], we do *not* require the storage function V in (2.2.2) to be non-negative. This difference is an important one and stems mainly from applications in mechanical and thermodynamical systems where energy or entropy functions need not necessarily be bounded from below. See Example 2.3.

If the composite function $\overline{V}(t) := V(x(t))$ is differentiable, then (2.2.2) is equivalent to

$$\frac{d}{dt}\overline{V}(t) = V_x(x(t))\,f(x(t), w(t)) \leq \mathrm{s}(w(t), z(t))$$

for all t and all solutions (w, x, z) of (2.2.1). Here, V_x denotes the gradient of V. This observation makes dissipativity of a dynamical system a *local property* in the sense that (Σ, s) is dissipative if and only if

$$V_x(x)\,f(x, w) \leq \mathrm{s}(w, g(x, w)) \tag{2.2.3}$$

holds for all points $x \in \mathbb{R}^n$ and $w \in \mathbb{R}^m$. In words, (2.2.3) states that the rate of change of storage along trajectories of the system will never exceed the rate of supply. We will refer to (2.2.3) as the *differential dissipation inequality*.

The classical motivation for the study of dissipativity comes from circuit theory. In the analysis of electrical networks the product of voltages and currents at the external branches of a network, i.e. the power, is an obvious supply function. Similarly, the product of forces and velocities of masses is a candidate supply function in mechanical systems. For those familiar with the theory of bond-graphs we remark that every bond-graph can be viewed as a representation of a dissipative dynamical system where inputs and outputs are taken to be effort and flow variables and the supply function is the product of these two variables. A bond-graph is therefore a special case of a dissipative system.

Example 2.3 Consider a thermodynamic system at uniform temperature T on which mechanical work is being done at rate W and which is being heated at rate Q. Let (T, Q, W) be the external variables of such a system and assume that –either by physical or chemical principles or through experimentation– the mathematical model of the thermodynamic system has been decided upon and is given by the time invariant system (2.2.1). The first and second law of thermodynamics may then be formulated in the sense of Definition 2.1 by saying that the system Σ is *conservative* with respect to the supply function $\mathrm{s}_1 := (W + Q)$ and *dissipative* with respect to the supply function $\mathrm{s}_2 := -Q/T$. Indeed, the first law of thermodynamics states that for all system trajectories (T, Q, W) and all time instants $t_0 \leq t_1$,

$$E(x(t_0)) + \int_{t_0}^{t_1} Q(t) + W(t)\,dt = E(x(t_1))$$

(conservation of thermodynamic energy). The second law of thermodynamics states that all system trajectories satisfy

$$S(x(t_0)) + \int_{t_0}^{t_1} -\frac{Q(t)}{T(t)}\,dt \geq S(x(t_1))$$

for all $t_0 \leq t_1$. Here, E is called the *internal energy* and S the *entropy*. The first law promises that the change of internal energy is equal to the heat absorbed by the system and the mechanical work which is done on the system. The second law states that the entropy decreases at a higher rate than the quotient of absorbed heat and temperature. It follows that thermodynamic systems are dissipative with respect to two supply functions. Nernst's third law of thermodynamics –the entropy of any object of zero temperature is zero– is only a matter of scaling of the entropy function S and does not further constrain the trajectories of the system.

Example 2.4 Other examples of supply functions $\mathrm{s} : W \times Z \to \mathbb{R}$ are the quadratic forms

$$\mathrm{s}(w,z) = w^\top z, \qquad \mathrm{s}(w,z) = \|w\|^2 - \|z\|^2,$$
$$\mathrm{s}(w,z) = \|w\|^2 + \|z\|^2, \qquad \mathrm{s}(w,z) = \|w\|^2$$

which arise in network theory, bondgraph theory, scattering theory, H_∞ theory, game theory and LQ-optimal control and H_2-optimal control theory. We will come across these examples in more detail below.

There are a few refinements to Definition 2.1 which are worth mentioning. Definition 2.1 can be generalized to *time-varying* systems by letting the supply rate s explicitly depend on time. Many authors have proposed a definition of dissipativity for discrete time systems, but since we can not think of any physical example of such a system, there seems little practical point in doing this. Another refinement consists of the idea that a system may be dissipative with respect to more than one supply function. See Example 2.3 for an example. Also, a notion of *robust dissipativity* may be developed in which the system description (2.2.1) is not assumed to be perfectly known, but uncertain to some well defined extend. An uncertain system is then called robustly dissipative if (2.2.2) holds for all $t_0 \leq t_1$ and all trajectories (w, x, z) that can conceivably be generated by the uncertain system. See Section **??** in Chapter **??** for more details. The notion of *strict dissipativity* is a refinement of Definition 2.1 which will prove useful in the sequel. It is defined as follows.

Definition 2.5 (Strict dissipativity) The system Σ with supply rate s is said to be *strictly dissipative* if there exists a storage function $V : X \to \mathbb{R}$ and an $\varepsilon > 0$ such that

$$V(x(t_0)) + \int_{t_0}^{t_1} \mathrm{s}(w(t), z(t))dt - \varepsilon^2 \int_{t_0}^{t_1} \|w(t)\|^2 dt \ \geq \ V(x(t_1)) \tag{2.2.4}$$

for all $t_0 \leq t_1$ and all trajectories (w, x, z) which satisfy (2.2.1).

A strictly dissipative system satisfies (2.2.2) with strict inequality, which justifies its name. As a final comment we mention the notion of *cyclo dissipativity* which has been introduced in (**??**). For $T > 0$, the function $w : \mathbb{R} \to W$ is said to be T*-periodic* if for all $t \in \mathbb{R}$ we have that $w(t) = w(t + T)$. A system Σ with supply function s is called *cyclo dissipative* if for all $T > 0$ there holds

$$\int_0^T \mathrm{s}(w(t), z(t))dt \ \geq \ 0$$

for all T*-periodic trajectories* $(w(\cdot), z(\cdot))$ which satisfy (2.2.1). Cyclo dissipativity is therefore a system property defined in terms of T-periodic trajectories only. The importance of this notion lies in the fact that it avoids reference to the internal state space structure of the system and requires a condition on signals in the external (input-output) behavior of the system only. It is easily seen that a dissipative system is cyclo dissipative whenever the state x is observable from (w, z), that is, whenever x is uniquely defined by any (w, z) which satisfies (2.2.1). Conversely, under some mild minimality and connectability conditions on the state X, a cyclo dissipative system is also dissipative. We refer to [44] for details.

2.2.2 A classification of storage functions

Suppose that (Σ, s) is dissipative and let $x^* \in X$ be a fixed reference point in the state space of Σ. Instead of considering the set of all possible storage functions associated with (Σ, s), we will restrict attention to the set of *normalized storage functions* defined by

$$\mathcal{V}(x^*) := \{V : X \to \mathbb{R} \mid V(x^*) = 0 \text{ and (2.2.2) holds}\}.$$

Hence, x^* is a reference point of *neutral storage*. Clearly, if V is a storage function satisfying (2.2.2), then we can obtain a normalized storage function by defining $\widetilde{V}(x) := V(x) - V(x^*)$ by noting that $\widetilde{V} \in \mathcal{V}(x^*)$.

Two mappings $V_{\mathrm{av}} : X \to \mathbb{R} \cup \{+\infty\}$ and $V_{\mathrm{req}} : X \to \mathbb{R} \cup \{-\infty\}$ will play a crucial role in the sequel. They are defined by

$$V_{\mathrm{av}}(x_0) := \sup \left\{ -\int_0^{t_1} \mathrm{s}(w(t), z(t))\, dt \;\mid\; t_1 \geq 0;\; (w, x, z) \text{ satisfies (2.2.1) with } x(0) = x_0 \text{ and } x(t_1) = x^* \right\} \tag{2.2.5a}$$

$$V_{\mathrm{req}}(x_0) := \inf \left\{ \int_{t_{-1}}^{0} \mathrm{s}(w(t), z(t))\, dt \;\mid\; t_{-1} \leq 0;\; (w, x, z) \text{ satisfies (2.2.1) with } x(0) = x_0 \text{ and } x(t_{-1}) = x^* \right\} \tag{2.2.5b}$$

Here, $V_{\mathrm{av}}(x)$ denotes the maximal amount of internal storage that may be recovered from the system over all state trajectories starting in x and eventually ending in x^*. Similarly, $V_{\mathrm{req}}(x)$ reflects the minimal supply which needs to be delivered to the system in order to steer the state to x via any trajectory originating in x^*. We refer to V_{av} and V_{req} as the *available storage* and the *required supply*, (measured with respect to x^*). In (2.2.5) it is silently assumed that for $x_0 \in X$ there exists an input $w \in \mathcal{W}$ which steers the state from x^* at some time instant $t_{-1} < 0$ to x_0 at time $t = 0$ and back to x^* at time $t_1 > 0$. We call x_0 *connectable with* x^* if this property holds. If such a loop can be run in finite time for any $x_0 \in X$, then we say that *every state is connectable with* x^*. The following characterization is the main result of this section.

Proposition 2.6 *Let the system Σ be represented by* (2.2.1) *and let* s *be a supply function. Suppose that every state is connectable with x^* for some $x^* \in X$. Then the following statements are equivalent*

(a) (Σ, s) *is dissipative.*

(b) $-\infty < V_{av}(x) < \infty$ *for all* $x \in X$.

(c) $-\infty < V_{req}(x) < \infty$ *for all* $x \in X$.

Moreover, if one of these equivalent statements hold, then

(a) $V_{av}, V_{req} \in \mathcal{V}(x^*)$.

(b) $\{V \in \mathcal{V}(x^*)\} \Rightarrow \{$*for all* $x \in X$ *there holds* $V_{av}(x) \leq V(x) \leq V_{req}(x)\}$.

(c) $\mathcal{V}(x^*)$ *is a convex set. In particular,* $V_\alpha := \alpha V_{av} + (1-\alpha)V_{req} \in \mathcal{V}(x^*)$ *for all* $\alpha \in (0, 1)$.

Interpretation 2.7 Proposition 2.6 confirms the intuitive idea that a dissipative system can neither supply nor store an infinite amount of energy during any experiment that starts or ends in a state of neutral storage. Proposition 2.6 shows that a system is dissipative if and only if the available storage and the required supply are real (finite) valued functions. Moreover, both the available storage and the required supply are possible storage functions of a dissipative system, these functions are normalized and define *extremal storage functions* in $\mathcal{V}(x^*)$ in the sense that $V_{\rm av}$ is the smallest and $V_{\rm req}$ is the largest element in $\mathcal{V}(x^*)$. In particular, for any state of a dissipative system, the available storage can not exceed its required supply. In addition, convex combinations of the available storage and the required supply are candidate storage functions.

Proof. Let (Σ, s) be dissipative, and let V be a storage function. Since $\widetilde{V}(x) := V(x) - V(x^*) \in \mathcal{V}(x^*)$ it follows that $\mathcal{V}(x^*) \neq \emptyset$ so that we may equally assume that $V \in \mathcal{V}(x^*)$. Let $x_0 \in X$, $t_{-1} \leq 0 \leq t_1$ and (w, x, z) satisfy (2.2.1) with $x(t_{-1}) = x(t_1) = x^*$ and $x(0) = x_0$. Since Σ is x^*-connectable such trajectories exist. From (2.2.2) we then infer that

$$-\infty < -\int_0^{t_1} \mathrm{s}(t)dt \leq \int_{t_{-1}}^0 \mathrm{s}(t)dt < +\infty.$$

First take in this inequality the supremum over all $t_1 \geq 0$ and $(w, x, z)|_{[0,t_1]}$ which satisfy (2.2.1) with $x(0) = x_0$ and $x(t_1) = x^*$. This yields that $-\infty < V_{\rm av}(x_0) < \infty$. Second, by taking the infimum over all $t_{-1} \leq 0$ and $(w, x, z)|_{[t_{-1},0]}$ with $x(t_{-1}) = x^*$ and $x(0) = x_0$ we infer that $-\infty < V_{\rm req}(x_0) < \infty$. Since x_0 is arbitrary, we obtain 2 and 3. To prove the converse implication, it suffices to show that $V_{\rm av}$ and $V_{\rm req}$ define storage functions. To see this, let $t_0 \leq t_1 \leq t_2$ and (w, x, z) satisfy (2.2.1) with $x(t_2) = x^*$. Then

$$V_{\rm av}(x(t_0)) \geq -\int_{t_0}^{t_1} \mathrm{s}(w(t), z(t))dt - \int_{t_1}^{t_2} \mathrm{s}(w(t), z(t))dt.$$

Since the second term in the right hand side of this inequality holds for arbitrary $t_2 \geq t_1$ and arbitrary $(w, x, z)|_{[t_1,t_2]}$ (with $x(t_1)$ fixed and $x(t_2) = x^*$), we can take the supremum over all such trajectories to conclude that

$$V_{\rm av}(x(t_0)) \geq -\int_{t_0}^{t_1} \mathrm{s}(w(t), z(t))dt + V_{\rm av}(x(t_1))$$

which shows that V_{av} satisfies (2.2.2). In a similar manner it is seen that V_{req} satisfies (2.2.2).

We next prove the remaining claims.

1. We already proved that V_{av} and V_{req} are storage functions. It thus remains to show that $V_{\mathrm{av}}(x^*) = V_{\mathrm{req}}(x^*) = 0$. Obviously, $V_{\mathrm{av}}(x^*) \geq 0$ and $V_{\mathrm{req}}(x^*) \leq 0$ (take $t_1 = t_{-1} = 0$ in (2.2.5)). Suppose that the latter inequalities are strict. Then, since the system is x^*-connectable, there exists $t_{-1} \leq 0 \leq t_1$ and a state trajectory x with $x(t_{-1}) = x(0) = x(t_1) = x^*$ such that $-\int_0^{t_1} \mathrm{s}(t)dt > 0$ and $\int_{t_{-1}}^0 \mathrm{s}(t)dt < 0$. But this contradicts the dissipation inequality (2.2.2) as both $\int_0^{t_1} \mathrm{s}(t)dt \geq 0$ and $\int_{t_{-1}}^0 \mathrm{s}(t)dt \geq 0$. Thus, $V_{\mathrm{av}}(x^*) = V_{\mathrm{req}}(x^*) = 0$.

2. If $V \in \mathcal{V}(x^*)$ then

$$-\int_0^{t_1} \mathrm{s}(w(t), z(t))dt \leq V(x_0) \leq \int_{t_{-1}}^0 \mathrm{s}(w(t), z(t))dt$$

for all $t_{-1} \leq 0 \leq t_1$ and (w, x, z) satisfying (2.2.1) with $x(t_{-1}) = x^* = x(t_1)$ and $x(0) = x_0$. Now take the supremum and infimum over all such trajectories to obtain that $V_{\mathrm{av}}(x_0) \leq V(x_0) \leq V_{\mathrm{req}}(x_0)$.

3. Follows trivially from the dissipation inequality (2.2.2). ■

2.3 Linear dissipative systems with quadratic supply rates

In the previous section we analyzed the notion of dissipativity at a fairly high level of generality. In this section we will apply the above theory to linear input-output systems Σ described by

$$\begin{pmatrix} \dot{x} \\ z \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix} \tag{2.3.1}$$

with state space $X = \mathbb{R}^n$, input space $W = \mathbb{R}^m$ and output space $Z = \mathbb{R}^p$. Let $x^* = 0$ be the point of neutral storage and consider supply functions that are general *quadratic functions* $\mathrm{s} : W \times Z \to \mathbb{R}$ defined by

$$\mathrm{s}(w, z) = \begin{pmatrix} w \\ z \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} w \\ z \end{pmatrix} = w^\top Q w + w^\top S z + z^\top S^\top w + z^\top R z. \tag{2.3.2}$$

Here, the matrix

$$P := \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix}$$

is a real symmetric matrix (that is, $P \in \mathbb{S}^{m+p}$) which is partitioned conform with w and z. No *a priori* definiteness assumptions are made on P.

Substituting the output equation $z = Cx + Dw$ in (2.3.2) shows that (2.3.2) can equivalently be viewed as a quadratic function in the variables x and w. Indeed,

$$\mathrm{s}(w,z) = \mathrm{s}(w, Cx + Dw) = \begin{pmatrix} x \\ w \end{pmatrix}^\top \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix}$$

This quadratic form in x and w will be used frequently in the sequel.

2.3.1 Main results

The following theorem is the main result of this chapter. It provides necessary and sufficient conditions for the pair (Σ, s) to be dissipative. In addition, it provides a complete parametrization of all normalized storage functions, together with a useful frequency domain characterization of dissipativity.

Theorem 2.8 *Suppose that the system* Σ *described by* (2.3.1) *is controllable and let the supply function* s *be defined by* (2.3.2). *Then the following statements are equivalent.*

(a) (Σ, s) *is dissipative.*

(b) (Σ, s) *admits a quadratic storage function* $V(x) := x^\top K x$ *with* $K = K^\top$.

(c) There exists $K = K^\top$ *such that*

$$F(K) := \begin{pmatrix} A^\top K + KA & KB \\ B^\top K & 0 \end{pmatrix} - \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} \preceq 0. \tag{2.3.3}$$

(d) There exists $K_- = K_-^\top$ *such that* $V_{av}(x) = x^\top K_- x$.

(e) There exists $K_+ = K_+^\top$ *such that* $V_{req}(x) = x^\top K_+ x$.

(f) For all $\omega \in \mathbb{R}$ *with* $\det(i\omega I - A) \neq 0$, *the transfer function* $T(s) := C(Is - A)^{-1}B + D$ *satisfies*

$$\begin{pmatrix} I \\ T(i\omega) \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ T(i\omega) \end{pmatrix} \succeq 0 \tag{2.3.4}$$

Moreover, if one of the above equivalent statements holds, then $V(x) := x^\top K x$ *is a quadratic storage function in* $\mathcal{V}(0)$ *if and only if* $F(K) \preceq 0$.

Proof. (1⇔4,5). If (Σ, s) is dissipative then by Proposition 2.6 V_{av} and V_{req} are storage functions. We claim that both $V_{\mathrm{av}}(x)$ and V_{req} are quadratic functions of x. This follows from [**?**] upon noting that both V_{av} and V_{req} are defined as optimal values corresponding to a linear quadratic optimization problem. Hence, if $x^* = 0$, $V_{\mathrm{av}}(x)$ is of the form $x^\top K_- x$ and $V_{\mathrm{req}}(x)$ takes the form $x^\top K_+ x$ for some matrices $K_- = K_-^\top$ and $K_+ = K_+^\top$. The converse implication is obvious from Proposition 2.6.

(1⇔2). Using the previous argument, item 1 implies item 4. But item 4 implies item 2 by Proposition 2.6. Hence, 1⇒ 2. The reverse implication is trivial.

(2⇒3). If $V(x) = x^\top Kx$ with $K \in \mathbb{S}^n$ a storage function then the differential dissipation inequality (2.2.3) can be rewritten as

$$2x^\top K(Ax + Bw) \leq \mathrm{s}(w, Cx + Dw)$$

for all $x \in X$ and all $w \in W$. This is equivalent to the algebraic condition

$$\begin{pmatrix} x \\ w \end{pmatrix}^\top \underbrace{\left\{ \begin{pmatrix} A^\top K + KA & KB \\ B^\top K & 0 \end{pmatrix} - \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} \right\}}_{F(K)} \begin{pmatrix} x \\ w \end{pmatrix} dt \ \leq\ 0 \qquad (2.3.5)$$

for all w and x. Controllability of the system implies that through every x we can pass a state trajectory. Hence, (2.3.5) reduces to the condition that $K \in \mathbb{S}^n$ satisfies $F(K) \preceq 0$.

(3⇒2). If there exist $K \in \mathbb{S}$ such that $F(K) \preceq 0$ then (2.3.5) holds from which it follows that $V(x) = x^\top Kx$ is a storage function satisfying the dissipation inequality.

The equivalence (1⇔6) is an application of Lemma 2.12, which we present below. An alternative proof for the implication (1⇒6) can be given as follows. Let $\omega > 0$ be such that $\det(i\omega I - A) \neq 0$ and consider the (complex) input $w(t) = \exp(i\omega t)w_0$ with $w_0 \in \mathbb{R}^m$. Define $x(t) := \exp(i\omega t)(i\omega I - A)^{-1}Bw_0$ and $z(t) := Cx(t) + Dw(t)$. Then $z(t) = \exp(i\omega t)T(i\omega)w_0$ and the (complex valued) triple (w, x, z) is a τ-periodic harmonic solution of (2.3.1) with $\tau = 2\pi/\omega$. Moreover,

$$\mathrm{s}(w(t), z(t)) = w_0^* \begin{pmatrix} I \\ T(i\omega) \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ T(i\omega) \end{pmatrix} w_0$$

which is *constant* for all $t \in \mathbb{R}$. Now suppose that (Σ, s) is dissipative. Then for all $k \in \mathbb{Z}$, $x(t_0) = x(t_0 + k\tau)$ and hence $V(x(t_0)) = V(x(t_0 + k\tau))$. For $t_1 = t_0 + k\tau$, the dissipation inequality (2.2.2) thus reads

$$\begin{aligned} \int_{t_0}^{t_1} \mathrm{s}(w(t), z(t))dt &= \int_{t_0}^{t_1} w_0^* \begin{pmatrix} I \\ T(i\omega) \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ T(i\omega) \end{pmatrix} w_0 \, dt \\ &= k\tau w_0^* \begin{pmatrix} I \\ T(i\omega)I \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ T(i\omega) \end{pmatrix} w_0 \ \geq\ 0 \end{aligned}$$

Since $k\tau > 0$ and w_0 is arbitrary, this yields 6. ■

We recognize in (2.3.3) a non-strict linear matrix inequality. The matrix $F(K)$ is usually called the *dissipation matrix*. Observe that in the above theorem the set of *quadratic storage functions* in $\mathcal{V}(0)$ is completely characterized by the linear matrix inequality $F(K) \preceq 0$. In other words, the set of normalized quadratic storage functions associated with (Σ, s) coincides with the feasibility set of the system of LMI $F(K) \preceq 0$. In particular, the available storage and the required supply are quadratic storage functions and hence K_- and K_+ also satisfy $F(K_-) \preceq 0$ and $F(K_+) \preceq 0$. Using Proposition 2.6, it moreover follows that any solution $K \in \mathbb{S}$ of $F(K) \preceq 0$ has the property that

$$K_- \preceq K \preceq K_+.$$

In other words, among the set of symmetric solutions K of the LMI $F(K) \preceq 0$ there exists a smallest and a largest element. The inequality (2.3.4) is called the *frequency domain inequality*. The equivalence between statements 1 and the frequency domain characterization in statement 6 has a long history in system theory. The result goes back to Popov (1962), V.A. Yakubovich (1962) and R. Kalman (1963) and is an application of the 'Kalman-Yakubovich-Popov Lemma', which we present and discuss in subsection 2.3.3 below.

For conservative systems with quadratic supply functions a similar characterization can be given. The precise formulation is evident from Theorem 2.8 and is left to the reader. Strictly dissipative systems are characterized as follows.

Theorem 2.9 *Suppose that the system Σ is described by* (2.3.1) *and let* s *be a supply function defined by* (2.3.2). *Then the following statements are equivalent.*

(a) (Σ, s) *is strictly dissipative.*

(b) There exists $K = K^\top$ such that

$$F(K) := \begin{pmatrix} A^\top K + KA & KB \\ B^\top K & 0 \end{pmatrix} - \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} \prec 0 \qquad (2.3.6)$$

(c) For all $\omega \in \mathbb{R}$ with $\det(i\omega I - A) \neq 0$, *the transfer function $T(s) := C(Is - A)^{-1}B + D$ satisfies*

$$\begin{pmatrix} I \\ T(i\omega) \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ T(i\omega) \end{pmatrix} \succ 0 \qquad (2.3.7)$$

Moreover, if one of the above equivalent statements holds, then $V(x) := x^\top K x$ is a quadratic storage function satisfying (2.2.4) *for some $\varepsilon > 0$ if and only if $F(K) \prec 0$.*

Proof. (1⇒3). By definition, item 1 implies that for some $\varepsilon > 0$ the pair (Σ, s') is dissipative with $\mathrm{s}'(w, z) := \mathrm{s}(w, z) - \varepsilon^2 \|w\|^2$. If Σ is controllable, Theorem 2.8 yields that for all $\omega \in \mathbb{R}$ with $\det(i\omega I - A) \neq 0$,

$$\begin{pmatrix} I \\ T(i\omega) \end{pmatrix}^* \begin{pmatrix} Q - \varepsilon^2 I & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ T(i\omega) \end{pmatrix} = \begin{pmatrix} I \\ T(i\omega) \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ T(i\omega) \end{pmatrix} - \varepsilon^2 I \succeq 0 \qquad (2.3.8)$$

The strict inequality (2.3.7) then follows. If Σ is not controllable, we use a perturbation argument to arrive at the inequality (2.3.7). Indeed, for $\delta > 0$ let B_δ be such that $B_\delta \to B$ as $\delta \to 0$ and (A, B_δ) controllable. Obviously, such B_δ exist. Define $T_\delta(s) := C(Is - A)^{-1}B_\delta + D$ and let (w, x_δ, z_δ) satisfy $\dot{x}_\delta = Ax_\delta + B_\delta w$, $z_\delta = Cx_\delta + Dw$. It follows that $B_\delta \to B$ implies that for every $w(t)$

which is bounded on the interval $[t_0, t_1]$, $x_\delta(t) \to x(t)$ and $z_\delta(t) \to z(t)$ pointwise in t as $\delta \to 0$. Here, (w, x, z) satisfy (2.3.1) and (2.2.4). Since V and s are continuous functions, the dissipation inequality (2.2.4) also holds for the perturbed system trajectories (w, x_δ, z_δ). This means that the perturbed system is strictly dissipative which, by Theorem 2.8, implies that (2.3.8) holds with $T(i\omega)$ replaced by $T_\delta(i\omega) := C(Ii\omega - A)^{-1}B_\delta + D$. Since for every $\omega \in \mathbb{R}$, $\det(Ii\omega - A) \neq 0$, we have that $T(i\omega) = \lim_{\delta\to 0} T_\delta(i\omega)$, $T(i\omega)$ satisfies (2.3.8) which, in turn, yields (2.3.7).

(3⇒2) is a consequence of Lemma 2.12, presented below.

(2⇒1). If $K = K^\top$ satisfies $F(K) \prec 0$, then there exists $\varepsilon > 0$ such that

$$F'(K) := F(K) + \begin{pmatrix} 0 & 0 \\ 0 & \varepsilon^2 I \end{pmatrix} = F(K) + \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^\top \begin{pmatrix} \varepsilon^2 I & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} \preceq 0$$

By Theorem 2.8, this implies that (Σ, s') is dissipative with $\mathrm{s}'(w, z) = \mathrm{s}(w, z) - \varepsilon^2\|w\|^2$. Inequality (2.2.4) therefore holds and we conclude that (Σ, s) is strictly dissipative. ∎

Remark 2.10 Contrary to Theorem 2.8, the system Σ is not assumed to be controllable in Theorem 2.9.

The dissipation matrix $F(K)$ in (2.3.3) and (2.3.6) can be written in various equivalent and sometimes more convenient forms. Indeed:

$$\begin{aligned} F(K) &= \begin{pmatrix} I & 0 \\ A & B \end{pmatrix}^\top \begin{pmatrix} 0 & K \\ K & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} - \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} = \\ &= \left(\begin{array}{cc} I & 0 \\ A & B \\ \hline 0 & I \\ C & D \end{array}\right)^\top \left(\begin{array}{cc|cc} 0 & K & 0 & 0 \\ K & 0 & 0 & 0 \\ \hline 0 & 0 & -Q & -S \\ 0 & 0 & -S^\top & -R \end{array}\right) \left(\begin{array}{cc} I & 0 \\ A & B \\ \hline 0 & I \\ C & D \end{array}\right) \\ &= \begin{pmatrix} A^\top K + KA - C^\top RC & KB - (SC)^\top - C^\top RD \\ B^\top K - SC - D^\top RC & -Q - SD - (SD)^\top - D^\top RD \end{pmatrix}. \end{aligned}$$

If we set $T := Q + SD + (SD)^\top + D^\top RD \prec 0$ and use a Schur complement, then the LMI $F(K) \prec 0$ is equivalent to $T \succ 0$ and

$$A^\top K + KA + C^\top RC + \left(KB + (SC)^\top + C^\top RD\right) T^{-1} \left(B^\top K + SC + D^\top RC\right) \prec 0.$$

The latter is a quadratic inequality in the unknown K.

2.3.2 Dissipation functions and spectral factors

If the system Σ is dissipative with respect to the supply function s then for any storage function V, the inequality (2.2.3) implies that

$$d(x, w) := \mathrm{s}(w, g(x, w)) - V_x(x) f(x, w) \tag{2.3.9}$$

is non-negative for all x and w. Conversely, if there exists a non-negative function d and a differentiable $V : X \to \mathbb{R}$ for which (2.3.9) holds, then the pair (Σ, s) is dissipative. The function d quantifies the amount of supply that is dissipated in the system when it finds itself in state x while the input w is exerted. We will call $d : \boldsymbol{X} \times W \to \mathbb{R}$ a *dissipation function* for (Σ, s) if (2.3.9) is satisfied for a differentiable storage function $V : X \to \mathbb{R}$.

For linear systems with quadratic supply functions, (2.3.9) reads

$$\begin{aligned} d(x, w) &= \begin{pmatrix} x \\ w \end{pmatrix}^\top \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix} - \begin{pmatrix} x \\ w \end{pmatrix}^\top \begin{pmatrix} A^\top K + KA & KB \\ B^\top K & 0 \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix} \\ &= - \begin{pmatrix} x \\ w \end{pmatrix}^\top F(K) \begin{pmatrix} x \\ w \end{pmatrix}. \end{aligned}$$

If $K = K^\top$ is such that $F(K) \preceq 0$ (or $F(K) \prec 0$) then the dissipation matrix can be factorized as

$$-F(K) = \begin{pmatrix} M_K & N_K \end{pmatrix}^\top \begin{pmatrix} M_K & N_K \end{pmatrix}. \tag{2.3.10}$$

where $\begin{pmatrix} M_K & N_K \end{pmatrix}$ is a real matrix with $n + m$ columns and at least $r_K := \operatorname{rank}(F(K))$ rows. For any such factorization, the function

$$d(x, w) := (M_K x + N_K w)^\top (M_K x + N_K w) = \|M_K x + N_K w\|^2$$

is therefore a dissipation function. If we extend the system equations (2.3.1) with the output equation $v = M_K x + N_K w$, then the output v incorporates the dissipated supply at each time instant we infer from (2.3.9) that the extended system

$$\begin{pmatrix} \dot{x} \\ z \\ v \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \\ M_K & N_K \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix} \tag{2.3.11}$$

becomes *conservative* with respect to the quadratic supply function $\mathrm{s}'(w, z, v) := \mathrm{s}(w, z) - v^\top v$.

This observation leads to an interesting connection between dissipation functions and spectral factorizations of rational functions. A complex valued rational function $\Phi : \mathbb{C} \to \mathbb{H}$ is called a *spectral density* if $\Phi(s) = \Phi^*(s)$ and Φ is analytic on the imaginary axis. A rational function V is a *spectral factor* of Φ if $\Phi(s) = V^*(s)V(s)$ for all but finitely many $s \in \mathbb{C}$.

Theorem 2.11 *Consider the spectral density*

$$\Phi(s) = \begin{pmatrix} I \\ T(s) \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ T(s) \end{pmatrix}$$

where $T(s) = C(Is - A)^{-1}B + D$, A has no eigenvalues on the imaginary axis and where (A, B) is controllable. Then there exists a spectral factor of Φ if and only if there exists $K = K^\top$ such that $F(K) \preceq 0$. In that case, $V(s) := M_K(Is - A)^{-1}B + N_K$ is a spectral factor of Φ where M_K and N_K are defined by the factorization (2.3.10).

Proof. If $F(K) \preceq 0$ then $F(K)$ can be factorized as (2.3.10) and for any such factorization the system (2.3.11) is conservative with respect to the supply function $s'(w, z, v) := s(w, z) - v^\top v$. Applying Theorem 2.8 for conservative systems, this means that

$$\begin{pmatrix} I \\ T(i\omega) \\ V(i\omega) \end{pmatrix}^* \begin{pmatrix} Q & S & 0 \\ S^\top & R & 0 \\ 0 & 0 & -I \end{pmatrix} \begin{pmatrix} I \\ T(i\omega) \\ V(i\omega) \end{pmatrix} = \Phi(i\omega) - V^*(i\omega)V(i\omega) = 0$$

for all $\omega \in \mathbb{R}$. But a rational function that vanishes identically on the imaginary axis, vanishes for all $s \in \mathbb{C}$. Hence, we infer that $\Phi(s) = V^*(s)V(s)$ for all but finitely many $s \in \mathbb{C}$. Conversely, if no $K = K^\top$ exists with $F(K) \preceq 0$, it follows from Theorem 2.8 that $\Phi(i\omega) \not\preceq 0$ and hence Φ admits no factorization on $\mathbb{C}^0$. ■

2.3.3 The Kalman-Yakubovich-Popov lemma

As mentioned in the proofs of Theorem 2.8 and Theorem 2.9, the Kalman-Yakubovich-Popov lemma is at the basis of the relation between frequency dependent matrix inequalities and an algebraic feasibility property of a linear matrix inequality. We will use this important result at various instances. The Lemma originates from a stability criterion of nonlinear feedback systems given by Popov in 1962 ([**?**]). Yakubovich and Kalman introduced the lemma by showing that the frequency condition of Popov is equivalent to the existence of a Lyapunov function.

We present a very general statement of the lemma which is free of any hypothesis on the system parameters. The proof which we present here is an elementary proof based on the dualization result that we stated in Theorem 1.16 of Chapter 1.

Lemma 2.12 (Kalman-Yakubovich-Popov) *For any triple of matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $M \in \mathbb{S}^{(n+m)\times(n+m)} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$, the following statements are equivalent:*

(a) There exists a symmetric $K = K^\top$ such that

$$\begin{pmatrix} I & 0 \\ A & B \end{pmatrix}^\top \begin{pmatrix} 0 & K \\ K & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} + M \prec 0 \tag{2.3.12}$$

(b) $M_{22} \prec 0$ and for all $\omega \in \mathbb{R}$ and complex vectors $\mathrm{col}(x, w) \neq 0$

$$\begin{pmatrix} A - i\omega I & B \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix} = 0 \quad \textit{implies} \quad \begin{pmatrix} x \\ w \end{pmatrix}^* M \begin{pmatrix} x \\ w \end{pmatrix} < 0. \tag{2.3.13}$$

If (A, B) is controllable, the corresponding equivalence also holds for non-strict inequalities.

If

$$M = -\begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}$$

then statement 2 is equivalent to the condition that for all $\omega \in \mathbb{R}$, with $\det(i\omega I - A) \neq 0$,

$$\begin{pmatrix} I \\ C(i\omega I - A)^{-1}B + D \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ C(i\omega I - A)^{-1}B + D \end{pmatrix} \prec 0.$$

Lemma 2.12 therefore reduces to the equivalence between the linear matrix inequality and the frequency domain inequality in Theorem 2.8 and Theorem 2.9. The Kalman-Yakubovich-Popov lemma therefore completes the proofs of these theorems. We proceed with an elegant proof of Lemma 2.12.

Proof. (1⇒2). Let $w \in \mathbb{C}^m$ and $(i\omega I - A)x = Bw$, $\omega \in \mathbb{R}$. The implication then follows from

$$\begin{pmatrix} x \\ w \end{pmatrix}^* \begin{pmatrix} I & 0 \\ A & B \end{pmatrix}^\top \begin{pmatrix} 0 & K \\ K & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix} + \begin{pmatrix} x \\ w \end{pmatrix} M \begin{pmatrix} x \\ w \end{pmatrix} = \begin{pmatrix} x \\ w \end{pmatrix}^* M \begin{pmatrix} x \\ w \end{pmatrix}.$$

(1⇐2). Suppose that (2.3.12) has no solution $K \in \mathbb{S}$. This means that the optimal value

$$P_{\text{opt}} := \inf \left\{ \gamma \mid \exists K = K^\top \text{ such that } G(K, \gamma) := \begin{pmatrix} A^\top K + KA & KB \\ B^\top K & 0 \end{pmatrix} + M - \gamma I \leq 0 \right\}$$

is non-negative. Note that this is a convex optimization problem with a linear objective function. As in subsection 1.4.5, let $\langle Y, X \rangle := \text{trace}(YX)$ be the natural inner product of the Hilbert space $\mathbb{S}^{n+m}$. Infer from Theorem 1.28 of Chapter 1 that

$$\begin{aligned}
D_{\text{opt}} &:= \max_{Y \succeq 0} \inf_{K \in \mathbb{S}, \gamma \in \mathbb{R}} \gamma + \langle Y, G(K, \gamma) \rangle = \\
&= \max_{Y \succeq 0} \inf_{K \in \mathbb{S}, \gamma \in \mathbb{R}} \left\{ \langle Y, M \rangle + \langle Y, \begin{pmatrix} A & B \end{pmatrix}^\top K \begin{pmatrix} I & 0 \end{pmatrix} + \begin{pmatrix} I & 0 \end{pmatrix}^\top K \begin{pmatrix} A & B \end{pmatrix} - \gamma I \rangle \right\} = \\
&= \max \left\{ \langle Y, M \rangle \mid Y \geq 0, \ \begin{pmatrix} A & B \end{pmatrix} Y \begin{pmatrix} I & 0 \end{pmatrix}^\top + \begin{pmatrix} I & 0 \end{pmatrix} Y \begin{pmatrix} A & B \end{pmatrix}^\top = 0 \right\} = \\
&= P_{\text{opt}}
\end{aligned}$$

is also non-negative. Moreover, by Theorem 1.16, there exists $Y = Y^\top$ such that

$$\langle Y, M \rangle = \text{trace}(YM) \geq 0, \qquad \begin{pmatrix} A & B \end{pmatrix} Y \begin{pmatrix} I & 0 \end{pmatrix}^\top + \begin{pmatrix} I & 0 \end{pmatrix} Y \begin{pmatrix} A & B \end{pmatrix}^\top = 0, \qquad Y \succeq 0. \tag{2.3.14}$$

Partition Y as the $(n + m) \times (n + m)$ matrix

$$Y = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix}.$$

We claim that $Y \succeq 0$ implies that $Y_{21} = Y_{12}^\top = FY_{11}$ for some $F \in \mathbb{R}^{m \times n}$. To see this, define $F|_{\text{im}\, Y_{11}}$ as $Fx := u$ where $u = Y_{21}v$ and v is such that $Y_{11}v = x$ and let F vanish on any complement of $\text{im}\, Y_{11}$ in $\mathbb{R}^n$. Then F is well defined unless there exist v_1, v_2, with $Y_{11}v_1 = Y_{11}v_2$ and $Y_{21}v_1 \neq Y_{21}v_2$. That is, unless $(v_1 - v_2) \in \ker Y_{11}$ implies $(v_1 - v_2) \notin \ker Y_{21}$. Hence, F is well defined if (and also only if) $\ker Y_{11} \subseteq \ker Y_{21}$. To see this inclusion, let $x \in \ker Y_{11}$, $u \in \mathbb{R}^m$. Then $Y \geq 0$ implies that for all $\alpha \in \mathbb{R}$, $\alpha x^\top Y_{12}u + u^{tr} Y_{21} \alpha x + u^\top Y_{22}u \geq 0$. Since α is arbitrary, this means that $x \in \ker Y_{21}$ so that $\ker Y_{11} \subseteq \ker Y_{21}$. Conclude that $Y_{21} = FY_{11}$. The second expression in (2.3.14) now reads $(A + BF)Y_{11} + Y_{11}^\top(A + BF)^\top = 0$, which shows that $\operatorname{im} Y_{11}$ is $(A + BF)$-invariant and $\ker Y_{11}$ is $(A + BF)^\top$-invariant. Hence there exist a basis $x_j \neq 0$ of $\operatorname{im} Y_{11}$ for which $(A + BF)x_j = \lambda_j x_j$ and $x_j = Y_{11}z_j$, $0 \neq z_j \in \mathbb{R}^n$. But then

$$0 = z_j^\top \Big[(A + BF)Y_{11} + Y_{11}^\top(A + BF)\Big] z_j = (\lambda_j + \bar{\lambda}_j)\underbrace{z_j^\top Y_{11} z_j}_{>0}$$

which yields that $\lambda_j \in \mathbb{C}^0$. Since

$$Y = \begin{pmatrix} I & 0 \\ F & I \end{pmatrix} \begin{pmatrix} Y_{11} & 0 \\ 0 & Y_{22} - FY_{11}F^\top \end{pmatrix} \begin{pmatrix} I & 0 \\ F & I \end{pmatrix}^\top$$

we may factorize $Y_{11} = UU^\top$, and $Y_{22} - FY_{11}F^\top = VV^\top$ with U and V full rank matrices. Then

$$\begin{aligned} 0 \leq \operatorname{trace}(YM) &= \operatorname{trace}(\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} I & 0 \\ F & I \end{pmatrix}^\top M \begin{pmatrix} I & 0 \\ F & I \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}^\top) \\ &= \operatorname{trace}(UM_{11}U^\top) + \operatorname{trace}(VMV^\top) \end{aligned}$$

■

2.3.4 The positive real lemma

Consider the system (2.3.1) together with the quadratic supply function $\mathrm{s}(w, z) = z^\top w + w^\top z$. Then the following result is worth mentioning as a special case of Theorem 2.8.

Corollary 2.13 *Suppose that the system Σ described by* (2.3.1) *is controllable and has transfer function T. Let $\mathrm{s}(w, z) = z^\top w + w^\top z$ be a supply function. Then equivalent statements are*

(a) (Σ, s) *is dissipative.*

(b) the LMI

$$\left(\begin{array}{cc} I & 0 \\ A & B \\ \hline 0 & I \\ C & D \end{array}\right)^\top \left(\begin{array}{cc|cc} 0 & K & 0 & 0 \\ K & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -I \\ 0 & 0 & -I & 0 \end{array}\right) \left(\begin{array}{cc} I & 0 \\ A & B \\ \hline 0 & I \\ C & D \end{array}\right) \preceq 0$$

is feasible

(c) For all $\omega \in \mathbb{R}$ with $\det(i\omega I - A) \neq 0$ *one has* $T(i\omega)^* + T(i\omega) \succeq 0$.

Moreover, $V(x) = x^\top Kx$ defines a quadratic storage function if and only if K satisfies the above LMI.

Corollary 2.13 is known as the *positive real lemma* and has played a crucial role in questions related to the stability of control systems and synthesis of passive electrical networks. Transfer functions which satisfy the third statement are generally called *positive real*. Note that for single-input and single-output systems, positive realness of a transfer function is graphically verified by the condition that the Nyquist plot of the system lies entirely in the right-half complex plane.

2.3.5 The bounded real lemma

Consider the quadratic supply function

$$\mathrm{s}(w, z) = \gamma^2 w^\top w - z^\top z \tag{2.3.15}$$

where $\gamma \geq 0$. We obtain the following result as an immediate consequence of Theorem 2.8.

Corollary 2.14 *Suppose that the system* Σ *described by* (2.3.1) *is controllable and has transfer function* T. *Let* $\mathrm{s}(w, z) = \gamma^2 w^\top w - z^\top z$ *be a supply function where* $\gamma \geq 0$. *Then equivalent statements are*

(a) (Σ, s) *is dissipative.*

(b) The LMI

$$\begin{pmatrix} I & 0 \\ A & B \\ \hline 0 & I \\ C & D \end{pmatrix}^\top \left(\begin{array}{cc|cc} 0 & K & 0 & 0 \\ K & 0 & 0 & 0 \\ \hline 0 & 0 & -\gamma^2 I & 0 \\ 0 & 0 & 0 & I \end{array}\right) \begin{pmatrix} I & 0 \\ A & B \\ \hline 0 & I \\ C & D \end{pmatrix} \preceq 0$$

is feasible.

(c) For all $\omega \in \mathbb{R}$ *with* $\det(i\omega I - A) \neq 0$ *one has* $T(i\omega)^* T(i\omega) \preceq \gamma^2 I$.

Moreover, $V(x) = x^\top K x$ *defines a quadratic storage function if and only if* K *satisfies the above LMI.*

Let us analyze the importance of this result. If the transfer function T of a system satisfies item 3 of Corollary 2.14 then for all $\omega \in \mathbb{R}$ for which $i\omega$ is not an eigenvalue of A and all complex vectors $\hat{w}(\omega) \in \mathbb{C}^m$ we have

$$\hat{w}(\omega)^* T(i\omega)^* T(i\omega) \hat{w}(\omega) \leq \gamma^2 \|\hat{w}(\omega)\|^2.$$

Now suppose that $\hat{w}$, viewed as a function of $\omega \in \mathbb{R}$ is square integrable in the sense that

$$\|\hat{w}\|_2^2 := \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{w}(\omega)^* \hat{w}(\omega) d\omega \ < \ \infty$$

Using Parseval's theorem, $\hat{w}$ is the Fourier transform of a function $w : \mathbb{R}_+ \to W$ in that $\hat{w}(\omega) = \int_0^\infty w(t)e^{-i\omega t}dt$ and we have that $\|\hat{w}\|_2 = \|w\|_2$ provided that the latter norm is defined as

$$\|w\|_2^2 = \int_0^\infty w(t)^\top w(t)dt$$

Let $\hat{z}(\omega) = T(i\omega)\hat{w}(\omega)$. Then $\hat{z}$ is the Fourier transform of the output z of (2.3.1) due to the input w and initial condition $x(0) = 0$. Consequently, item 3 is equivalent to saying that

$$\|z\|_2^2 \leq \gamma^2 \|w\|_2$$

for all inputs w for which $\|w\|_2 < \infty$. That is, with initial condition $x(0) = 0$, the 2-norm of the output of (2.3.1) is uniformly bounded by γ^2 times the 2-norm of the input. This crucial observation is at the basis of H_∞ optimal control theory and will be further exploited in the next chapter.

2.4 Interconnected dissipative systems

In this section we consider the question whether the interconnection of a number of dissipative dynamical systems is dissipative. To answer this question, we restrict attention to the case where just two dynamical systems are interconnected. Generalizations of these results to interconnections of more than two dynamical systems follow immediately from these ideas. Consider the two dynamical systems

$$\Sigma_1 : \begin{cases} \dot{x}_1 = f_1(x_1, w_1) \\ z_1 = g_1(x_1, w_1) \end{cases}, \qquad \Sigma_2 : \begin{cases} \dot{x}_2 = f_2(x_2, w_2) \\ z_2 = g_2(x_2, w_2) \end{cases}. \tag{2.4.1}$$

We will first need to formalize what we mean by an interconnection of these systems. For this, assume that the input and output variable w_i and z_i ($i = 1, 2$) of each of these systems is partitioned as

$$w_i = \begin{pmatrix} w_i' \\ w_i'' \end{pmatrix}, \qquad z_i = \begin{pmatrix} z_i' \\ z_i'' \end{pmatrix}.$$

The input space W_i and the output space Z_i can therefore be written as the Cartesian set products $W_i = W_i' \times W_i''$ and $Z_i = Z_i' \times Z_i''$. We will think of the variables (w_i', z_i') as the *external variables* and (w_i'', z_i'') as the *interconnection variables* of the interconnected system. It will be assumed that $\dim Z_1'' = \dim W_2''$ and $\dim W_1'' = \dim Z_2''$. The *interconnection constraint* is then defined by the algebraic relations

$$z_1'' = w_2'', \qquad w_1'' = z_2'' \tag{2.4.2}$$

and the interconnected system is defined by the laws (2.4.1) of Σ_1 and Σ_2 and the interconnection constraint (2.4.2). The idea behind this concept is displayed in Figure 2.1.

It is not evident that the joint equations (2.4.1)-(2.4.2) will have a unique solution for every pair of input variables (w_1', w_2'). This property is referred to as *well posedness* of the interconnection. We decided to disappoint the reader and avoid a thorough discussion on this issue here. For the time being, we will assume that the interconnected system is well defined, takes (w_1', w_2') as its

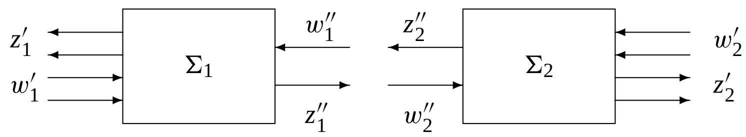


Figure 2.1: System interconnections

input variables and (z_1', z_2') as its outputs. Whenever well defined, the interconnected system will be denoted by $\Sigma = \Sigma_1 \sqcap \Sigma_2$.

We now get to the discuss the question whether or not the interconnection is dissipative. Suppose that (Σ_i, s_i) is dissipative for $i = 1, 2$. Assume that the supply functions s_i admit an additive structure in that there exists functions $\mathrm{s}_i' : W_i' \times Z_i' \to \mathbb{R}$ and $\mathrm{s}_i'' : W_i'' \times Z_i'' \to \mathbb{R}$ such that

$$\mathrm{s}_i(w_i, z_i) = \mathrm{s}_i'(w_i', z_i') + \mathrm{s}_i''(w_i'', z_i'')$$

for all $w_i \in W_i$ and $z_i \in Z_i$. The interconnection of Σ_1 and Σ_2 is the said to be *neutral* if

$$\mathrm{s}_1''(w_1'', z_1'') + \mathrm{s}_2''(w_2'', z_2'') = 0$$

for all $w_1'' \in W_1''$, $z_1 \in Z_1''$, $w_2'' \in W_2''$ and $z_2'' \in Z_2''$. In words, this means that there is no dissipation in the interconnection variables, i.e., the interconnected system is conservative with respect to the supply function $\mathrm{s}'' : W_1'' \times W_2'' \to \mathbb{R}$ defined as $\mathrm{s}''(w_1'', w_2'') := \mathrm{s}_1''(w_1'', w_2'') + \mathrm{s}_2''(w_2'', w_1'')$. Neutrality seems a rather natural requirement on many interconnected systems. The following result states that a neutral interconnection of dissipative systems is dissipative again.

Theorem 2.15 *Let* (Σ_i, s_i), $i = 1, 2$ *be dissipative dynamical systems and suppose that the interconnection* $\Sigma = \Sigma_1 \sqcap \Sigma_2$ *is well defined and neutral. Then* Σ *is dissipative with respect to the supply function* $\mathrm{s} : W_1' \times Z_1' \times W_2' \times Z_2'$ *defined as*

$$\mathrm{s}(w_1', z_1', w_2', z_2') := \mathrm{s}_1'(w_1', z_1') + \mathrm{s}_2'(w_2', z_2').$$

Moreover, if $V_i : X_i \to \mathbb{R}$ *is a storage function of* (Σ_i, s_i) *then* $V : X_1 \times X_2 \to \mathbb{R}$ *defined by* $V(x_1, x_2) := V_1(x_1) + V_2(x_2)$ *is a storage function for* (Σ, s).

Proof. Since (Σ_i, s_i), $i = 1, 2$ is dissipative, there exists $V_1 : X_1 \to \mathbb{R}$, $V_2 : X_2 \to \mathbb{R}$ such that for all $t_0 \leq t_1$,

$$V_1(x_1(t_0)) + \int_{t_0}^{t_1} \mathrm{s}_1'(w_1', z_1') + \mathrm{s}_1''(w_1'', z_1'')dt \geq V_1(x_1(t_1))$$
$$V_2(x_2(t_0)) + \int_{t_0}^{t_1} \mathrm{s}_2'(w_2', z_2') + \mathrm{s}_2''(w_2'', z_2'')dt \geq V_2(x_2(t_1))$$

Adding these two equations and applying the neutrality condition of the interconnection yields that

$$V(x(t_0)) + \int_{t_0}^{t_1} \mathrm{s}_1'(w_1', z_1') + \mathrm{s}_2'(w_2', z_2')dt \geq V(x(t_1))$$

i.e., (Σ, s) is dissipative with V as its storage function. ■

We finally remark that the above ideas can be easily generalized to interconnections of any finite number of dynamical systems.

2.5 Further reading

Many of the ideas on dissipative dynamical systems originate from the work of Willems in [45, 46]. The definitions on dissipative dynamical systems and their characterizations have been reported in [43, 44] and originate from the behavioral approach of dynamical systems. See also similar work in [39]. Extensions to nonlinear systems are discussed in [40]. The proof of the Kalman-Yakubovich-Popov lemma which is presented here is very much based on ideas in [42]. For alternative proofs of this important lemma, see [?, 23].

2.6 Exercises

Exercise 1
Show that for conservative controllable systems the set of normalized storage functions $\mathcal{V}(x^*)$ consist of one element only.

Conclude that storage functions of conservative systems are unique up to normalization!.

Exercise 2
Show that the set of dissipation functions associated with a dissipative system is convex.

Exercise 3
Consider the suspension system Σ of a transport vehicle as depicted in Figure 2.2. The system is modeled by the equations

$$\begin{aligned} m_2\ddot{q}_2 + b_2(\dot{q}_2 - \dot{q}_1) + k_2(q_2 - q_1) - F = 0 \\ m_1\ddot{q}_1 + b_2(\dot{q}_1 - \dot{q}_2) + k_2(q_1 - q_2) + k_1(q_1 - q_0) + F = 0 \end{aligned}$$

where F (resp. $-F$) is a force acting on the chassis mass m_2 (the axle mass m_1). Here, $q_2 - q_1$ is the distance between chassis and axle, and $\ddot{q}_2$ denotes the acceleration of the chassis mass m_2. b_2 is a damping coefficient and k_1 and k_2 are spring coefficients. ($b_1 = 0$). The variable q_0 represents the road profile. A 'real life' set of system parameters is given in Table 2.1.

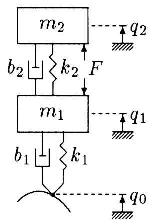


Figure 2.2: Model for suspension system

m_1	m_2	k_1	k_2	b_2
1.5×10^3	1.0×10^4	5.0×10^6	5.0×10^5	50×10^3

Table 2.1: Physical parameters

(a) Derive a state space model of the form 2.3.1 of the system which assumes $w = \operatorname{col}(q_0, F)$ and $z = \operatorname{col}(q_1, \dot{q}_1, q_2, \dot{q}_2)$ as its input and output, respectively.

(b) Define a supply function $s : W \times Z \to \mathbb{R}$ such that (Σ, s) is dissipative. (Base your definition on physical insight).

(c) Characterize the set of all quadratic storage functions of the system as the feasibility set of a linear matrix inequality.

(d) Compute a quadratic storage function $V(x) = x^\top K x$ for this system.

(e) Determine a dissipation function $d : X \times W \to \mathbb{R}$ for this system.

Exercise 4
Consider the transfer functions

(a) $T_1(s) = 1/(s+1)$

(b) $T_2(s) = (s-1)/(s+1)$

(c) $T_3(s) = \begin{pmatrix} (s+2)(s-1)/(s+1)^2 & (s+3)/(s+4) \\ (s-1)/(s+0.5) & (s+1)/(s+2) \end{pmatrix}.$

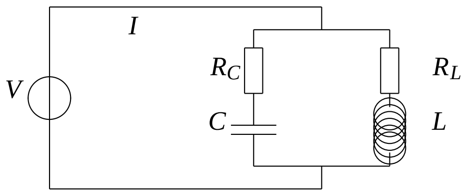
Determine for each of these transfer functions

(a) whether or not they are positive real and

(b) the smallest value of $\gamma > 0$ for which $T^*(i\omega)T(i\omega) \preceq \gamma^2 I$.

Reformulate these problems as a feasibility test involving a suitably defined LMI. (See Corollaries 2.13 and 2.14).

Exercise 5
Consider the following electrical circuit. We will be interested in modeling the relation between



the external voltage V and the current I through the circuit. Assume that the resistors $R_C = 1$ and $R_L = 1$, the capacitor $C = 2$ and the inductance $L = 1$.

(a) Derive a linear, time-invariant system Σ that models the relation between the voltage V and the current I.

(b) Find a state space representation of the form (2.3.1) which represents Σ. Is the choice of input and output variable unique?

(c) Define a supply function $s : W \times Z \to \mathbb{R}$ such that (Σ, s) is dissipative.

(d) Characterize the set of all quadratic storage functions of the system as the feasibility set of a linear matrix inequality.

(e) Compute a quadratic storage function $V(x) = x^\top K x$ of this system.

(f) Does dissipativity of (Σ, s) depend on whether the voltage V or the current I is taken as input of your system?

Exercise 6
Consider a first-order unstable system $P(s) = 1/(-3s + 1)$. It is desirable to design a feedback compensator C, so that the feedback system is dissipative. Assume that the compensator C is a simple gain $C(s) = k$, $k \in \mathbb{R}$. Find the range of gains k that will make the system depicted in Figure 2.3 dissipative with respect to the supply function $s(w, z) = wz$.

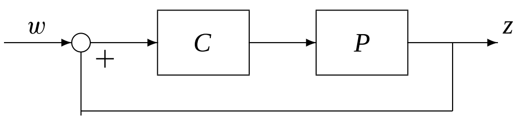


Figure 2.3: Feedback configuration

Exercise 7
Consider the example of an orbit of a mass m moving in a central conservative force field, the force being given by Newton's inverse square law, i.e., at time t, the mass m is at position $z(t) \in \mathbb{R}^3$ and moves under influence of a Newtonian gravitational field of a second mass M, which is assumed to be at the origin of $\mathbb{R}^3$. Newton's inverse square law claims that

$$\frac{d^2z}{dt^2} + \frac{Mg}{\|z\|^3} z = 0. \tag{2.6.1}$$

Let $v = \dot{z}$ denote the velocity vector of m and let $x = \mathrm{col}(z, v)$ be a state vector.

(a) Represent this (autonomous) system in the form $\dot{x} = f(x)$, $z = g(x)$.

(b) Consider the function
$$V(z, v) := \frac{1}{2} m \|v\|^2 - \frac{Mg}{\|z\|}$$
and show that V is a storage function of the system with supply function $\mathrm{s}(z) = 0$.

(c) Prove that the orbit of the mass m is a hyperbola, parabola or ellipse depending on whether $V > 0$, $V = 0$ or $V < 0$ along solutions z of (2.6.1).

Conclude from the last item that Kepler's first law of planetary motions tells us that the solar system is an autonomous dissipative system with a negative storage function.

Chapter 3

Nominal stability and nominal performance

3.1 Lyapunov stability

As mentioned in Chapter 1, Aleksandr Mikhailovich Lyapunov studied contraction and expansion phenomena of the motions of a mechanical system around an equilibrium. Translated in modern jargon, the study of *Lyapunov stability* concerns the asymptotic behavior of the state of an autonomous dynamical system. The main contribution of Lyapunov has been to define the concept of stability, asymptotic stability and instability of such systems and to give a method of verification of these concepts in terms of the existence of functions, called Lyapunov functions. Both his definition and his verification method characterize, in a local way, the stability properties of an autonomous dynamical system. Unfortunately, for the general class of nonlinear systems there are no systematic procedures for finding Lyapunov functions. However, we will see that for linear systems the problem of finding Lyapunov functions can be solved adequately as a feasibility test of a linear matrix inequality.

3.1.1 Nominal stability of nonlinear systems

Let $\mathcal{X}$ be a set and $T \subseteq \mathbb{R}$. A *flow* is a mapping $\phi : T \times T \times \mathcal{X} \to \mathcal{X}$ which satisfies the

(a) consistency property: $\phi(t, t, x) = x$ and the

(b) semi-group property: $\phi(t_1, t_{-1}, x_0) = \phi(t_1, t_0, \phi(t_0, t_{-1}, x_0))$ for all $t_{-1} \leq t_0 \leq t_1$ and $x_0 \in \mathcal{X}$.

The set $\mathcal{X}$ is called the *state space* (or the *phase space*) and we will think of a flow as a *state evolution map*. A flow defines an *unforced* or *autonomous dynamical system* in the sense that the evolution of a flow is completely determined by an initial state and not by any kind of external input. A typical example of a flow is the solution of a differential equation of the form

$$\dot{x}(t) = f(x(t), t) \tag{3.1.1}$$

with finite dimensional state space $\mathcal{X} = \mathbb{R}^n$ and where $f : \mathcal{X} \times T \to \mathcal{X}$ is a function. By a *solution* of (3.1.1) over a (finite) time interval $T \subset \mathbb{R}$ we will mean a function $x : T \to \mathcal{X}$ which is differentiable everywhere and which satisfies (3.1.1) for all $t \in T$. By a solution over the unbounded intervals $\mathbb{R}_+$, $\mathbb{R}_-$ or $\mathbb{R}$ we will mean a solution over all finite intervals T contained in either of these sets. Let us denote by $\phi(t, t_0, x_0)$ the solution of (3.1.1) at time t with initial condition $x(t_0) = x_0$. If for all pairs $(t_0, x_0) \in T \times \mathcal{X}$ one can guarantee existence and uniqueness of a solution $\phi(t, t_0, x_0)$ of (3.1.1) with $t \in T$, then the mapping ϕ satisfies the consistency and semi-group property, and therefore defines a flow. In that case, ϕ is said to be the *flow associated with the differential equation* (3.1.1). This flow is called *time-invariant* if

$$\phi(t + \tau, t_0 + \tau, x_0) = \phi(t, t_0, x_0)$$

for any $\tau \in T$. It is said to be *linear* if $\mathcal{X}$ is a vector space and $\phi(t, t_0, \cdot)$ is a linear mapping for all time instances t and t_0. For time-invariant flows we usually take (without loss of generality) $t_0 = 0$ as initial time. We remark that the condition on existence and uniqueness of solutions of (3.1.1) holds whenever f satisfies a *global Lipschitz condition*. That is, if $\mathcal{X}$ has the structure of a normed vector space with norm $\|\cdot\|$, and if for some $L > 0$, the inequality

$$\|f(x, t) - f(y, t)\| \leq L\|x - y\|$$

holds for all $x, y \in \mathcal{X}$, $t \in T$.

An element x^* is a *fixed point* or an *equilibrium point* of (3.1.1) if $f(x^*, t) = 0$ for all $t \in T$. It is easy to see that x^* is a fixed point if and only if $\phi(t, t_0, x^*) = x^*$ is a solution of (3.1.1) for all t and t_0. In other words, solutions of (3.1.1) remain in x^* once they started there.

There exists a wealth of concepts to define the stability of a flow ϕ. The various notions of Lyapunov stability pertain to a distinguished fixed point x^* of a flow ϕ and express to what extend an other trajectory $\phi(t, t_0, x_0)$, whose initial state x_0 lies in the neighborhood of x^* at time t_0, remains or gets close to $\phi(t, t_0, x^*)$ for all time instances $t \geq t_0$.

Definition 3.1 (Lyapunov stability) Let $\phi : T \times T \times \mathcal{X}$ be a flow and suppose that $T = \mathbb{R}$ and $\mathcal{X}$ is a normed vector space. The fixed point x^* is said to be

(a) *stable* (in the sense of Lyapunov) if given any $\varepsilon > 0$ and $t_0 \in T$, there exists $\delta = \delta(\varepsilon, t_0) > 0$ (not depending on t) such that

$$\|x_0 - x^*\| \leq \delta \implies \|\phi(t, t_0, x_0) - x^*\| \leq \varepsilon \text{ for all } t \geq t_0. \tag{3.1.2}$$

(b) *attractive* if for all $t_0 \in T$ there exists $\delta = \delta(t_0) > 0$ with the property that

$$\|x_0 - x^*\| \leq \delta \implies \lim_{t\to\infty} \|\phi(t, t_0, x_0) - x^*\| = 0. \tag{3.1.3}$$

(c) *exponentially stable* if for all $t_0 \in T$ there exists $\delta = \delta(t_0)$, $\alpha = \alpha(t_0) > 0$ and $\beta = \beta(t_0) > 0$ such that

$$\|x_0 - x^*\| \leq \delta \implies \|\phi(t, t_0, x_0) - x^*\| \leq \beta\|x_0 - x^*\|e^{-\alpha(t-t_0)} \text{ for all } t \geq t_0. \tag{3.1.4}$$

(d) *asymptotically stable* (in the sense of Lyapunov) if it is both stable (in the sense of Lyapunov) and attractive.

(e) *unstable* if it is not stable (in the sense of Lyapunov).

(f) *uniformly stable* (in the sense of Lyapunov) if given any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ (not depending on t_0) such that (3.1.2) holds for all $t_0 \in T$.

(g) *uniformly attractive* if there exists $\delta > 0$ (not depending on t_0) such that (3.1.3) holds for all $t_0 \in T$.

(h) *uniformly exponentially stable* if there exists $\delta > 0$ (not depending on t_0) such that (3.1.3) holds for all $t_0 \in T$.

(i) *uniformly asymptotically stable* (in the sense of Lyapunov) if it is both uniformly stable (in the sense of Lyapunov) and uniformly attractive.

In words, a fixed point is stable if the graphs of all flows that initiate sufficiently close to x^* at time t_0, remain as close as desired to x^* for all time $t \geq t_0$. Stated otherwise, a fixed point is stable if the mapping $\phi(t, t_0, \cdot)$ is continuous at x^*, uniformly in $t \geq t_0$. The *region of attraction* associated with a fixed point x^* is defined to be the set of all initial states $x_0 \in \mathcal{X}$ for which $\phi(t, t_0, x_0) \to x^*$ as $t \to \infty$. If this region does not depend on t_0, it is said to be *uniform*, if it coincides with $\mathcal{X}$ then x^* is *globally attractive*. Similarly, we can define the region of stability, the region of asymptotic stability and the region of exponential stability associated with x^*. Again, these regions are said to be *uniform* if they do not depend on t_0. If these regions cover the entire state space $\mathcal{X}$, then the fixed point is called *globally stable*, *globally asymptotically stable*, or *globally exponentially stable*, respectively.

We remark that there exist examples of stable fixed points that are not attractive, and examples of attractive fixed points that are not stable. The notion of exponential stability is the strongest in the sense that an exponentially stable fixed point is also asymptotically stable (i.e., a stable attractor). Similarly, it is easily seen that uniform exponential stability implies uniform asymptotic stability.

A set $\mathcal{S} \subset \mathcal{X}$ is called a *positive invariant set* of a flow ϕ if $x_0 \in \mathcal{S}$ implies that there exists a $t_0 \in T$ such that $\phi(t, t_0, x_0) \in \mathcal{S}$ for all $t \geq t_0$. It is called a *negative invariant set* if this condition holds for $t \leq t_0$ and it is said to be an *invariant set* if it is both positive and negative invariant. The idea of a (positive, negative) invariant set is, therefore, that a flow remains in the set once it started there at time t_0. Naturally, a set $\mathcal{S} \subseteq \mathcal{X}$ is said to be an invariant set of the differential equation (3.1.1), if it is an invariant set of its associated flow. Also, any point $x_0 \in \mathcal{X}$ naturally generates the invariant set $\mathcal{S} = \{\phi(t, t_0, x_0) \mid t \geq t_0\}$ consisting of all points that the flow $\phi(t, t_0, x_0)$ hits when time evolves. In particular, for every fixed point x^* of (3.1.1) the singleton $\mathcal{S} = \{x^*\}$ is an invariant set.

The following proposition gives a first reason not to distinguish all of the above notions of stability.

Proposition 3.2 *Let $\phi : T \times T \times \mathcal{X}$ be a flow with $T = \mathbb{R}$ and suppose that x^* is a fixed point. If ϕ is linear then*

(a) x^* is attractive if and only if x^* is globally attractive.

(b) x^* is asymptotically stable if and only if x^* is globally asymptotically stable.

(c) x^* is exponentially stable if and only if x^* is globally exponentially stable.

If ϕ is time-invariant then

(a) x^* is stable if and only if x^* is uniformly stable.

(b) x^* is asymptotically stable if and only if x^* is uniformly asymptotically stable.

(c) x^* is exponentially stable if and only if x^* is uniformly exponentially stable.

Proof. All *if* parts are trivial. To prove the *only if* parts, let ϕ be linear, and suppose that x^* is attractive. Without loss of generality we will assume that $x^* = 0$. Take $x_0 \in \mathcal{X}$ and δ as in Definition 3.1. Then there exists α such that $\|\alpha x_0\| < \delta$ and by linearity of the flow, $\lim_{t\to\infty} \|\phi(t, t_0, x_0)\| = \alpha^{-1} \lim_{t\to\infty} \|\phi(t, t_0, \alpha x_0)\| = 0$, i.e., 0 is a global attractor. The second and third claim for linear flows is now obvious. Next, let ϕ be time-invariant and $x^* = 0$ stable. Then for $\varepsilon > 0$ and $t_0 \in T$ there exists $\delta > 0$ such that $\|x_0\| \leq \delta$ implies $\|\phi(t + \tau, t_0 + \tau, x_0)\| = \|\phi(t, t_0, x_0)\| \leq \varepsilon$ for all $t \geq t_0$ and $\tau \in T$. Set $t' = t + \tau$ and $t_0' = t_0 + \tau$ to infer that $\|x_0\| \leq \delta$ implies $\|\phi(t', t_0', x_0)\| \leq \varepsilon$ for all $t' \geq t_0'$. But these are the trajectories passing through x_0 at time t_0' with t_0' arbitrary. Hence 0 is uniformly stable. The last claims follow with a similar reasoning. ■

Definition 3.3 (Definite functions) Let $\mathcal{S} \subseteq \mathbb{R}^n$ have the point x^* in its interior and let $T \subseteq \mathbb{R}$. A function $V : \mathcal{S} \times T \to \mathbb{R}$ is said to be

(a) *positive definite (with respect to x^*)* if there exists a continuous, strictly increasing function $a : \mathbb{R}_+ \to \mathbb{R}_+$ with $a(0) = 0$ such that $V(x, t) \geq a(\|x - x^*\|)$ for all $(x, t) \in \mathcal{S} \times T$.

(b) *positive semi-definite* if $V(x, t) \geq 0$ for all $(x, t) \in \mathcal{S} \times T$.

(c) *decrescent (with respect to x^*)* if there exists a continuous, strictly increasing function $b : \mathbb{R}_+ \to \mathbb{R}_+$ with $b(0) = 0$ such that $V(x, t) \leq b(\|x - x^*\|)$ for all $(x, t) \in \mathcal{S} \times T$..

(d) *negative definite (around x^*)* or *negative semi-definite* if $-V$ is positive definite or positive semi-definite, respectively.

We introduced positive definite and positive semi-definite *matrices* in Chapter 1. This use of terminology is consistent with Definition 3.3 in the following sense. If X is a real symmetric matrix then for all vectors x we have that

$$\lambda_{\min}(X)\|x\|^2 \leq x^\top X x \leq \lambda_{\max}(X)\|x\|^2.$$

Hence, the function $V(x) := x^\top X x$ is positive definite with respect to the origin if, and only if, the smallest eigenvalue $\lambda_{\min}(X)$ is positive. Hence, X is positive definite as a matrix (denoted $X \succ 0$) if and only if the function $V(x) = x^\top X x$ is positive definite. Similarly, V is negative definite with respect to the origin if and only if $\lambda_{\max}(X) < 0$ which we denoted this by $X \prec 0$.

Consider the system (3.1.1) and suppose that x^* is an equilibrium point. Let $\mathcal{S}$ be a set which has x^* in its interior and suppose that $V : \mathcal{S} \times T \to \mathbb{R}$ has continuous partial derivatives (i.e., V is continuously differentiable). Consider, for $(x_0, t_0) \in \mathcal{S} \times T$, the function $V^* : T \to \mathcal{X}$ defined by the composition

$$V^*(t) := V(\phi(t, t_0, x_0), t).$$

Then this function is differentiable and its derivative reads

$$\dot{V}^*(t) = \partial_x V(\phi(t, t_0, x_0), t) f(\phi(t, t_0, x_0), t) + \partial_t V(\phi(t, t_0, x_0), t).$$

Now introduce the mapping $V' : \mathcal{S} \times T \to \mathbb{R}$ by setting

$$V'(x, t) := \partial_x V(x, t) f(x, t) + \partial_t V(x, t). \tag{3.1.5}$$

V' is called the *derivative of V along trajectories of* (3.1.1) and, by construction, we have that $\dot{V}^*(t) = V'(\phi(t, t_0, x_0), t)$ for all $t \in T$. It is very important to observe that V' not only depends on V but also on the differential equation (3.1.1). It is rather common to write $\dot{V}$ for V' and even more common to confuse $\dot{V}^*$ with V'. Formally, these objects are different as V^* is a function of time, whereas V' is a function of the state and time.

The main stability results for autonomous systems of the form (3.1.1) are summarized in the following result.

Theorem 3.4 (Lyapunov theorem) *Consider the differential equation* (3.1.1) *and let $x^* \in \mathcal{X}$ be an equilibrium point which belongs to the interior of a set $\mathcal{S}$.*

(a) If there exists a positive definite, continuously differentiable function $V : \mathcal{S} \times T \to \mathbb{R}$ with $V(x^*, t) = 0$ and V' negative semi-definite, then x^* is stable. If, in addition, V is decrescent, then x^* is uniformly stable.

(b) If there exists a positive definite decrescent and continuously differentiable function $V : \mathcal{S} \times T \to \mathbb{R}$ with $V(x^*, t) = 0$ and V' negative definite, then x^* is uniformly asymptotically stable.

Proof. 1. Let $t_0 \in T$ and $\epsilon > 0$. Since $V(\cdot, t_0)$ is continuous at x^* and $V(x^*, t_0) = 0$, there exists $\delta > 0$ such that $V(x_0, t_0) \leq a(\epsilon)$ for every $x_0 \in \mathcal{S}$ with $\|x_0 - x^*\| < \delta$. Since V is a positive definite and $\dot{V}$ negative semi-definite, we have that for every $x_0 \in \mathcal{S}$ with $\|x_0 - x^*\| < \delta$ and $t \geq t_0$:

$$a(\|x(t) - x^*\|) \leq V(x(t), t) \leq V(x_0, t_0) \leq a(\epsilon).$$

where we denoted $x(t) = \phi(t, t_0, x_0)$. Since a is strictly increasing, this implies (3.1.2), i.e., x^* is stable. If, in addition, V is decrescent, then $V(x, t) \leq b(\|x - x^*\|)$ for all $(x, t) \in \mathcal{S} \times T$. Apply the previous argument with δ such that $b(\delta) \leq a(\epsilon)$. Then δ is independent of t_0 and $V(x_0, t_0) \leq b(\delta) \leq a(\epsilon)$ for every $(x_0, t_0) \in \mathcal{S} \times T$ such that $\|x_0 - x^*\| < \delta$. Hence, (3.1.2) holds for all t_0.

2. By item 1, x^* is uniformly stable. It thus suffices to show that x^* is uniformly attractive. Let $\delta > 0$ be such that all x_0 with $\|x_0 - x^*\| < \delta$ belong to $\mathcal{S}$. Since x^* is an interior point of $\mathcal{S}$ such δ obviously exists. Let x_0 satisfy $\|x_0 - x^*\| < \delta$ and let $t_0 \in T$. Under the given hypothesis, there exist continuous, strictly increasing functions a, b and c such that $a(\|x - x^*\|) \leq V(x, t) \leq b(\|x - x^*\|)$ and $\dot{V}(x, t) \geq -c(\|x - x^*\|)$ for all $(x, t) \in \mathcal{S} \times T$. Let $\epsilon > 0$, $\gamma > 0$ such that $b(\gamma) < a(\epsilon)$, and $t_1 > t_0 + b(\delta)/c(\gamma)$. We claim that there exists $\tau \in [t_0, t_1]$ such that $x(\tau) := \phi(\tau, t_0, x_0)$ satisfies $\|x(\tau) - x^*\| \leq \gamma$. Indeed, if no such τ exists, integration of both sides of the inequality $\dot{V}(x(t), t) \leq -c(\|x(t) - x^*\|)$ yields that

$$\begin{aligned} V(x(t_1), t_1) &\leq V(x_0, t_0) - \int_{t_0}^{t_1} c(\|x(t) - x^*\|) \\ &< b(\|x_0 - x^*\|) - (t_1 - t_0)c(\gamma) \\ &< b(\delta) - \frac{b(\delta)}{c(\gamma)} c(\gamma) = 0, \end{aligned}$$

which contradicts the assumption that $V(x(t_1), t_1) \geq 0$. Consequently, it follows from the hypothesis that for all $t \geq \tau$:

$$a(\|x(t) - x^*\|) \leq V(x(t), t) \leq V(x(\tau), \tau) \leq b(\|x(\tau) - x^*\|) \leq b(\gamma) \leq a(\epsilon).$$

Since a is strictly increasing, this yields that $\|x(t) - x^*\| \leq \epsilon$ for all $t \geq \tau$. As ϵ is arbitrary, this proves (3.1.3) for all t_0. ∎

The functions V that satisfy either of the properties of Theorem 3.4 are generally referred to as *Lyapunov functions*. The main implication of Theorem 3.4 is that stability of equilibrium points of differential equations of the form (3.1.1) can be verified by searching for suitable Lyapunov functions.

3.1.2 Stability and dissipativity

An intuitive way to think about the result of Theorem 3.4 is to view V as a storage function of the autonomous dynamical system described by the differential equation (3.1.1). At a distinguished equilibrium state x^*, the storage vanishes but is positive everywhere else. For example, a damped pendulum described by the nonlinear differential equations

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = -\sin(x_1) - dx_2$$

has all points $(k\pi, 0)$ with $k \in \mathbb{Z}$ as its fixed points (corresponding to the vertical positions of the pendulum). The function $V(x_1, x_2) := \frac{1}{2}x_2^2 + (1 - \cos(x_1))$, representing the sum of kinetic and potential mechanical energy in the system, vanishes at the fixed points $(k\pi, 0)$ with k even and is a Lyapunov function where $V'(x_1, x_2) := \tilde{V}(x_1, x_2) = -dx_2^2$ is negative definite provided that the damping coefficient $d > 0$. Hence, the fixed points $(k\pi, 0)$ with k even are stable equilibria. In fact, these points are uniformly asymptotically stable.

Storage functions, introduced in Chapter 2, and Lyapunov functions are closely related as can been seen by comparing (3.1.5) with the differential dissipation inequality (2.2.3) from Chapter 2. Indeed, if $w(t) = w^*$ is taken to a constant input in the input-state-output system (2.2.1) then we obtain the *autonomous time-invariant system*

$$\begin{aligned} \dot{x} &= f(x, w^*) \\ z &= g(x, w^*). \end{aligned}$$

If x^* is an equilibrium point of this system and if this system is known to be dissipative with respect to a supply function that satisfies

$$\mathrm{s}(w^*, g(x, w^*)) \ \leq \ 0$$

for all x in a neighborhood of x^*, then any (differentiable) storage function V satisfies the differential dissipation inequality (2.2.3) and is monotone non-increasing along solutions in a neighborhood of x^*. If the storage function is moreover non-negative in this neighborhood, then it follows from Theorem 3.4 that x^* is a stable equilibrium at x^*. In that case, the storage function V is nothing else than a Lyapunov function defined in a neighborhood of x^*. In particular, the dissipativity of the system implies the stability of the system provided the storage function is non-negative.

Unfortunately, for this general class of nonlinear differential equations there are no systematic procedures for actually finding such functions. We will see next that more explicit results can be obtained for the analysis of stability of solutions of linear differential equations.

3.1.3 Nominal stability of linear systems

Let us consider the *linear autonomous system*

$$\dot{x} = Ax \tag{3.1.6}$$

where $A : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map obtained as the *linearization* of $f : \mathcal{X} \to \mathcal{X}$ around an equilibrium point $x^* \in \mathcal{X}$ of (3.1.1). Precisely, for $x^* \in \mathcal{X}$ we write

$$f(x) = f(x^*) + \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x^*)[x - x^*] + \ldots$$

where we assume that f is at least once differentiable. The linearization of f around x^* is defined by the system (3.1.6) with A defined by the real $n \times n$ matrix

$$A := \partial_x f(x^*).$$

All elements in $\ker A$ are equilibrium points, but we will investigate the stability of (3.1.6) at the equilibrium $x^* = 0$. The positive definite quadratic function $V : \mathcal{X} \to \mathbb{R}$ defined by

$$V(x) = x^\top X x$$

serves as a quadratic Lyapunov function. Indeed, V is continuous at $x^* = 0$, is assumes a strong local minimum at $x = 0$ (actually this is a strong global minimum of V), while the derivative of $V(x)$ in the direction of the vector field Ax is given by

$$V_x Ax = x^\top [A^\top X + XA]x$$

which should be negative to guarantee that the origin is an asymptotic stable equilibrium point of (3.1.6). We thus obtain the following result:

Proposition 3.5 *Let the system* (3.1.6) *be a linearization of* (3.1.1) *at the equilibrium* x^*. *The following statements are equivalent.*

(a) The origin is an asymptotically stable equilibrium for (3.1.6).

(b) The origin is a globally asymptotically stable equilibrium for (3.1.6).

(c) All eigenvalues $\lambda(A)$ *of* A *have strictly negative real part.*

(d) The linear matrix inequalities

$$A^\top X + XA \prec 0, \qquad X \succ 0$$

are feasible.

Moreover, if one of these statements hold, then the equilibrium x^* *of the flow* (3.1.1) *is asymptotically stable.*

As the most important implication of Proposition 3.5, asymptotic stability of the equilibrium x^* of the nonlinear system (3.1.1) can be concluded from the asymptotic stability of its linearization at x^*.

3.2 Generalized stability regions for LTI systems

As we have seen, the autonomous linear system

$$\dot{x} = Ax$$

is asymptotically stable if and only if all eigenvalues of A lie in $\mathbb{C}^-$, the open left half complex plane. For many applications in control and engineering we may be interested in more general stability regions. Let us define a *stability region* as a subset $\mathbb{C}_{\rm stab} \subseteq \mathbb{C}$ with the following two properties

$$\begin{cases} \text{Property 1:} & \lambda \in \mathbb{C}_{\rm stab} \implies \bar{\lambda} \in \mathbb{C}_{\rm stab} \\ \text{Property 2:} & \mathbb{C}_{\rm stab} \text{ is convex .} \end{cases}$$

Typical examples of common stability sets include

$\mathbb{C}_{\rm stab\,1} = \mathbb{C}^-$	open left half complex plane
$\mathbb{C}_{\rm stab\,2} = \mathbb{C}$	no stability requirement
$\mathbb{C}_{\rm stab\,3} = \{s \in \mathbb{C} \mid \mathrm{Re}(s) < -\alpha\}$	guaranteed damping
$\mathbb{C}_{\rm stab\,4} = \{s \in \mathbb{C} \mid \lvert s\rvert < r\}$	circle centered at origin
$\mathbb{C}_{\rm stab\,5} = \{s \in \mathbb{C} \mid \alpha_1 < \mathrm{Re}(s) < \alpha_2\}$	vertical strip
$\mathbb{C}_{\rm stab\,6} = \{s \in \mathbb{C} \mid \mathrm{Re}(s)\tan(\theta) < -\lvert \mathrm{Im}(s)\rvert\}$	conic stability region.

Here, $\theta \in (0, \pi/2)$ and $r, \alpha, \alpha_1, \alpha_2$ are real numbers. We consider the question whether we can derive a feasibility test to verify whether the eigen-modes of the system $\dot{x} = Ax$ belong to either of these sets. This can indeed be done in the case of the given examples. To see this, let us introduce the notion of an LMI-region as follows:

Definition 3.6 For a real symmetric matrix $P \in \mathbb{S}^{2m\times 2m}$, the set of complex numbers

$$L_P := \left\{ s \in \mathbb{C} \mid \begin{pmatrix} I \\ sI \end{pmatrix}^* P \begin{pmatrix} I \\ sI \end{pmatrix} \prec 0 \right\}$$

is called an *LMI region*.

If P is partitioned according to $P = \left(\begin{smallmatrix} Q & S \\ S^\top & R \end{smallmatrix}\right)$, then an LMI region is defined by those points $s \in \mathbb{C}$ for which

$$Q + sS + \bar{s}S^\top + \bar{s}Rs < 0.$$

All of the above examples fit in this definition. Indeed, by setting

$$P_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad P_2 = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \quad P_3 = \begin{pmatrix} 2\alpha & 1 \\ 1 & 0 \end{pmatrix}$$

$$P_4 = \begin{pmatrix} -r^2 & 0 \\ 0 & 1 \end{pmatrix}, \quad P_5 = \begin{pmatrix} 2\alpha_1 & 0 & -1 & 0 \\ 0 & -2\alpha_2 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad P_6 = \begin{pmatrix} 0 & 0 & \sin(\theta) & \cos(\theta) \\ 0 & 0 & -\cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) & 0 & 0 \\ \cos(\theta) & \sin(\theta) & 0 & 0 \end{pmatrix}$$

we obtain that $\mathbb{C}_{{\rm stab}\,i} = L_{P_i}$. More specifically, LMI regions include regions bounded by circles, ellipses, strips, parabolas and hyperbolas. Since any finite intersection of LMI regions is again an LMI region one can virtually approximate any convex region in the complex plane.

To present the main result of this section, we will need to introduce the notation for *Kronecker products*. Given two matrices $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{k\times \ell}$, the Kronecker product of A and B is the $mk \times n\ell$ matrix

$$A \otimes B = \begin{pmatrix} A_{11}B & \dots & A_{1n}B \\ \vdots & & \vdots \\ A_{m1}B & \dots & A_{mn}B \end{pmatrix}$$

Some properties pertaining to the Kronecker product are as follows

- $1 \otimes A = A = A \otimes 1$.
- $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$.
- $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.
- $(A \otimes B)^* = A^* \otimes B^*$.
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.
- in general $A \otimes B \neq B \otimes A$.

These properties are easily verified and we will not prove them here. Stability regions described by LMI regions lead to the following interesting generalization of the Lyapunov inequality.

Theorem 3.7 *All eigenvalues of $A \in \mathbb{R}^{n\times n}$ are contained in the LMI region*

$$\left\{ s \in \mathbb{C} \mid \begin{pmatrix} I \\ sI \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ sI \end{pmatrix} \prec 0 \right\}$$

if and only if there exists $X \succ 0$ such that

$$\begin{pmatrix} I \\ A \otimes I \end{pmatrix}^* \begin{pmatrix} X \otimes Q & X \otimes S \\ X \otimes S^\top & X \otimes R \end{pmatrix} \begin{pmatrix} I \\ A \otimes I \end{pmatrix} \prec 0$$

Note that the latter is an LMI in X and that the Lyapunov theorem (Theorem 3.4) corresponds to taking $Q = 0$, $S = I$ and $R = I$. Among the many interesting special cases of LMI regions that are covered by Theorem 3.7, we mention the stability set $\mathbb{C}_{\text{stab }4}$ with $r = 1$ used for the characterization of stability of the *discrete time* system $x(t + 1) = Ax(t)$. This system is stable if and only if the eigenvalues of A are inside the unit circle. Equivalently, $\lambda(A) \in L_P$ with $P = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, which by Theorem 3.7 is equivalent to saying that there exist $X \succ 0$ such that

$$\begin{pmatrix} I \\ A \end{pmatrix}^* \begin{pmatrix} -X & 0 \\ 0 & X \end{pmatrix} \begin{pmatrix} I \\ A \end{pmatrix} = A^* X A - X \prec 0.$$

3.3 Nominal performance and LMI's

In this section we will use the results on dissipative systems of Chapter 2 to characterize a number of relevant performance criteria for dynamical systems. In view of forthcoming chapters we consider the system

$$\begin{cases} \dot{x} &= Ax + Bw \\ z &= Cx + Dw \end{cases} \tag{3.3.1}$$

where $x(t) \in \mathcal{X} = \mathbb{R}^n$ is the state, $w(t) \in \mathcal{W} = \mathbb{R}^m$ the input and $z(t) \in \mathcal{Z} = \mathbb{R}^p$ the output. Let $T(s) = C(Is - A)^{-1}B + D$ denote the corresponding transfer function and assume throughout this section that the system is asymptotically stable (i.e., the eigenvalues of A are in the open left-half complex plane). We will view w as an input variable (a 'disturbance') whose effect on the output z (an 'error indicator') we wish to minimize. There are various ways to quantify the effect of w on z. For example, for a given input w, and for suitable signal norms, the quotient $\|z\|/\|w\|$ indicates the relative gain which the input w has on the output z. More generally, the *worst case gain* of the system is the quantity

$$\|T\| := \sup_{0<\|w\|<\infty} \frac{\|z\|}{\|w\|}$$

which, of course, depends on the chosen signal norms. Other indicators for nominal performance could be the energy in the impulse response of the system, the (asymptotic) variance of the output when the system is fed with inputs with a prescribed stochastic nature, percentage overshoot in step responses, etc.

3.3.1 Quadratic nominal performance

We start this section by reconsidering Theorem 2.9 from Chapter 2. The following proposition is obtained by making sign changes in Theorem 2.9 and applying the Kalman-Yabubovich-Popov lemma. We will see that it has important implications for the characterization of performance criteria.

Proposition 3.8 *Consider the system* (3.3.1) *with transfer function T. Suppose that A has its eigenvalues in $\mathbb{C}^-$ and let $x(0) = 0$. The following statements are equivalent.*

(a) there exists $\varepsilon > 0$ such that for all $w \in \mathcal{L}_2$

$$\int_0^\infty \begin{pmatrix} w \\ z \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} w \\ z \end{pmatrix} dt \le -\varepsilon^2 \int_0^\infty w^\top(t)w(t)dt \tag{3.3.2}$$

(b) for all $\omega \in \mathbb{R} \cup \{\infty\}$ there holds

$$\begin{pmatrix} I \\ T(i\omega) \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} I \\ T(i\omega) \end{pmatrix} \prec 0.$$

(c) *there exists* $K = K^\top \in \mathbb{R}^{n\times n}$ *such that*

$$
\begin{aligned}
F(K) &= \begin{pmatrix} A^\top K + KA & KB \\ B^\top K & 0 \end{pmatrix} + \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} = \\
&= \begin{pmatrix} I & 0 \\ A & B \end{pmatrix}^\top \begin{pmatrix} 0 & K \\ K & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} + \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} = \\
&= \left(\begin{array}{cc} I & 0 \\ A & B \\ \hline 0 & I \\ C & D \end{array}\right)^\top \left(\begin{array}{cc|cc} 0 & K & 0 & 0 \\ K & 0 & 0 & 0 \\ \hline 0 & 0 & Q & S \\ 0 & 0 & S^\top & R \end{array}\right) \left(\begin{array}{cc} I & 0 \\ A & B \\ \hline 0 & I \\ C & D \end{array}\right) \\
&\prec 0.
\end{aligned}
$$

This result characterizes *quadratic performance* of the system (3.3.1) in the sense that it provides necessary and sufficient conditions for the quadratic performance function $J := \int_0^\infty s(w, z)dt$ to be strictly negative for all square integrable trajectories of the system. Proposition (3.8) provides an equivalent condition in terms of a frequency domain inequality and an equivalent linear matrix inequality for this. This very general result proves useful in a number of important special cases, which we describe below.

3.3.2 H_∞ nominal performance

A popular performance measure of a stable linear time-invariant system is the H_∞ norm of its transfer function. It is defined as follows. Consider the system (3.3.1) together with its transfer function T. Assume the system to be asymptotically stable. In that case, $T(s)$ is bounded for all $s \in \mathbb{C}$ with positive real part. By this, we mean that the largest singular value $\sigma_{\max}(T(s))$ is finite for all $s \in \mathbb{C}$ with $\operatorname{Re} s > 0$. This is an example of an H_∞ function. To be slightly more formal on this class of functions, let $\mathbb{C}^+$ denote the set of complex numbers with positive real part. The *Hardy space* H_∞ consists of all complex valued functions $T : \mathbb{C}^+ \to \mathbb{C}^{p\times m}$ which are analytic and for which

$$\|T\|_\infty := \sup_{s\in\mathbb{C}^+} \sigma_{\max}(T(s)) < \infty.$$

The left-hand side of this expression satisfies the axioms of a norm and defines the H_∞ norm of T. Although H_∞ functions are defined on the right-half complex plane, it can be shown that each such function has a unique extension to the imaginary axis (which is usually also denoted by T) and that the H_∞ norm is given by

$$\|T\|_\infty = \sup_{\omega\in\mathbb{R}} \sigma_{\max}(T(i\omega)).$$

In words, the H_∞ norm of a transfer function is the supremum of the maximum singular value of the frequency response of the system.

Remark 3.9 Various graphical representations of frequency responses are illustrative to investigate system properties like bandwidth, gains, etc. Probably the most important one is a plot of the singular values $\sigma_j(T(i\omega))$ ($j = 1, \dots, \min(m, p)$) viewed as function of the frequency $\omega \in \mathbb{R}$. For single-input single-output systems there is only one singular value and $\sigma(T(i\omega)) = |T(i\omega)|$. A *Bode diagram* of the system is a plot of the mapping $\omega \mapsto |T(i\omega)|$ and provides useful information to what extent the system amplifies purely harmonic input signals with frequencies $\omega \in \mathbb{R}$. In order to interpret these diagrams one usually takes logarithmic scales on the ω axis and plots $20^{10}\log(T(j\omega))$ to get units in decibels dB. The H_∞ norm of a transfer function is then nothing else than the highest peak value which occurs in the Bode plot. In other words it is the largest gain if the system is fed with harmonic input signals.

The H_∞ norm of a stable linear system admits an interpretation in terms of dissipativity of the system with respect to a specific quadratic supply function. This is expressed in the following result.

Proposition 3.10 *Let the system* (3.3.1) *be asymptotically stable and* $\gamma > 0$. *Then the following statements are equivalent.*

(a) $\|T\|_\infty < \gamma$.

(b) *for all* w *there holds that*

$$\sup_{0<\|w\|_2<\infty} \frac{\|z\|_2}{\|w\|_2} < \gamma$$

where z *is the output of* (3.3.1) *subject to input* w *and initial condition* $x(0) = 0$.

(c) *The system* (3.3.1) *is strictly dissipative with respect to the supply function* $s(w, z) = \gamma^2\|w\|^2 - \|z\|^2$.

(d) *there exists a solution* $K = K^\top$ *to the LMI*

$$\begin{pmatrix} A^\top K + KA + C^\top C & KB + C^\top D \\ B^\top K + D^\top C & D^\top D - \gamma^2 I \end{pmatrix} \prec 0. \tag{3.3.3}$$

Proof. Apply Proposition 3.8 with

$$\begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} = \begin{pmatrix} -\gamma^2 I & 0 \\ 0 & I \end{pmatrix}.$$

■

For a stable system, the H_∞ norm of the transfer function therefore coincides with the $\mathscr{L}_2$-induced norm of the input-output operator associated with the system. Using the Kalman-Yakubovich-Popov lemma, this yields an LMI feasibility test to verify whether or not the H_∞ norm of the transfer function T is bounded by γ.

3.3.3 H_2 nominal performance

The Hardy space H_2 consists of the class of complex valued functions which are analytic in $\mathbb{C}^+$ and for which

$$\|T\|_{H_2} := \sqrt{\frac{1}{2\pi} \sup_{\sigma>0} \text{trace} \int_{-\infty}^{\infty} T(\sigma + i\omega)[T(\sigma + i\omega)]^* \, d\omega}$$

is finite. This defines the H_2 norm of T. This 'cold-blooded' definition may seem little appealing at first sight, but in fact, it has nice and important system theoretic interpretations. As in H_∞, it can be shown that each function in H_2 has a unique extension to the imaginary axis, which we also denote by T, and that in fact the H_2 norm satisfies

$$\|T\|_{H_2}^2 = \frac{1}{2\pi} \text{trace} \int_{-\infty}^{\infty} T(i\omega)T(i\omega)^* \, d\omega \tag{3.3.4}$$

We will first give an interpretation of the H_2 norm of a system in terms of its *impulsive behavior*. Consider the system (3.3.1) and suppose that we are interested only in the impulse responses of this system. This means, that we take impulsive inputs[1] of the form

$$w(t) = \delta(t)e_j$$

where e_j the jth basis vector in the standard basis of the input space $\mathbb{R}^m$, ($j = 1, \dots, m$). The output z^j which corresponds to the input w and initial condition $x(0) = 0$ is uniquely defined and given by

$$z^j(t) = \begin{cases} C\exp(At)Be_j & \text{for } t > 0 \\ De_j\delta(t) & \text{for } t = 0 \\ 0 & \text{for } t < 0 \end{cases}.$$

Since the system is assumed to be stable, the outputs z^j are square integrable for all $j = 1, \dots, m$, provided that $D = 0$. In that case

$$\begin{aligned} \sum_{j=1}^{m} \|z^j\|_2^2 &= \text{trace} \int_0^{\infty} B^\top \exp(A^\top t)C^\top C\exp(At)B \, dt \\ &= \text{trace} \int_0^{\infty} C\exp(At)BB^\top \exp(A^\top t)C^\top \, dt. \end{aligned}$$

Long ago, Parseval taught us that the latter expression is equal to

$$\frac{1}{2\pi} \text{trace} \int_{-\infty}^{\infty} T(i\omega)T(i\omega)^* \, d\omega$$

which is $\|T\|_{H_2}^2$. Infer that the squared H_2 norm of T coincides with the total 'output energy' in the impulse responses of the system. What is more, this observation provides a straightforward

[1] Formally, the impulse δ is not a function and for this reason it is neither a signal. It requires a complete introduction to distribution theory to make these statements more precise, but we will not do this at this place.

algorithm to determine the H_2 norm of a stable rational transfer function. Indeed, associate with the system (3.3.1) the symmetric non-negative matrices

$$W := \int_0^\infty \exp(At)BB^\top \exp(A^\top t)\, dt$$
$$M := \int_0^\infty \exp(A^\top t)C^\top C \exp(At)\, dt.$$

Then W is usually referred to as the *controllability gramian* and M the *observability gramian* of the system (3.3.1). The gramians satisfy the matrix equations

$$AW + WA^\top + BB^\top = 0, \qquad A^\top M + MA + C^\top C = 0$$

and are, in fact, the unique solutions to these equations whenever A has its eigenvalues in $\mathbb{C}^-$ (as is assumed here). Consequently,

$$\|T\|_{H_2}^2 = \operatorname{trace}(CWC^\top) = \operatorname{trace}(B^\top MB).$$

A second interpretation of the H_2 norm makes use of stochastics. Consider the system (3.3.1) and assume that the components of the input w are independent zero-mean, white noise processes. If we take $x(0) = 0$ as initial condition, the state variance matrix

$$W(t) := \mathcal{E}(x(t)x^\top(t))$$

is the solution of the matrix differential equation

$$\dot{W} = AW + WA^\top + BB^\top, \qquad W(0) = 0.$$

Consequently, with $D = 0$, the output variance

$$\begin{aligned}\mathcal{E}(z^\top z(t)) &= \mathcal{E}(x^\top C^\top Cx(t)) = \mathcal{E}\operatorname{trace}(Cx(t)x^\top(t)C^\top) = \\ &= \operatorname{trace} C\mathcal{E}(x(t)x^\top(t))C^\top = \operatorname{trace}(CW(t)C^\top).\end{aligned}$$

Since A is asymptotically stable, the limit $W := \lim_{t\to\infty} W(t)$ exists and is equal to the controllability gramian of the system (3.3.1). Consequently, the asymptotic output variance

$$\lim_{t\to\infty} \mathcal{E}(z(t)z^\top(t)) = \operatorname{trace}(CWC^\top)$$

which is the square of the H_2 norm of the system. The H_2 norm therefore has an interpretation in terms of the asymptotic output variance of the system when it is excited by white noise input signals.

The following theorem characterizes the H_2 norm in terms of linear matrix inequalities.

Proposition 3.11 *Suppose that the system* (3.3.1) *is asymptotically stable and let* $T(s) = C(Is - A)^{-1}B + D$ *denote its transfer function. Then*

(a) $\|T\|_2 < \infty$ *if and only if* $D = 0$.

(b) If $D = 0$ then the following statements are equivalent

(i) $\|T\|_2 < \gamma$

(ii) there exists $X \succ 0$ such that

$$AX + XA^\top + BB^\top \prec 0, \qquad \mathrm{trace}(CXC^\top) < \gamma^2.$$

(iii) there exists $Y \succ 0$ such that

$$A^\top Y + YA + C^\top C \prec 0, \qquad \mathrm{trace}(B^\top YB) < \gamma^2.$$

(iv) there exists $K = K^\top \succ 0$ and Z such that

$$\begin{pmatrix} A^\top K + KA & KB \\ B^\top K & -\gamma I \end{pmatrix} \prec 0; \quad \begin{pmatrix} K & C^\top \\ C & Z \end{pmatrix} \succ 0; \quad \mathrm{trace}(Z) < \gamma \tag{3.3.5}$$

(v) there exists $K = K^\top \succ 0$ and Z such that

$$\begin{pmatrix} AK + KA^\top & KC^\top \\ CK & -\gamma I \end{pmatrix} \prec 0; \quad \begin{pmatrix} K & B \\ B^\top & Z \end{pmatrix} \succ 0; \quad \mathrm{trace}(Z) < \gamma. \tag{3.3.6}$$

Proof. The first claim is immediate from the definition of the H_2 norm. To prove the second part, note that $\|T\|_2 < \gamma$ is equivalent to requiring that the controllability gramian W satisfies $\mathrm{trace}(CWC^\top) < \gamma^2$. Since the controllability gramian is the unique positive definite solution of the Lyapunov equation $AW + WA^\top + BB^\top = 0$ this is equivalent to saying that there exists $X \succ 0$ such that

$$AX + XA^\top + BB^\top \prec 0; \qquad \mathrm{trace}(CXC^\top) < \gamma^2.$$

In turn, with a change of variables $K := X^{-1}$, this is equivalent to the existence of $K > 0$ and Z such that

$$A^\top K + KA + KBB^\top K \prec 0; \quad CK^{-1}C^\top \prec Z; \quad \mathrm{trace}(Z) < \gamma^2.$$

Now, using Schur complements for the first two inequalities yields that $\|T\|_2 < \gamma$ is equivalent to the existence of $K \succ 0$ and Z such that

$$\begin{pmatrix} A^\top K + KA & KB \\ B^\top K & -I \end{pmatrix} \prec 0; \quad \begin{pmatrix} K & C^\top \\ C & Z \end{pmatrix} \succ 0; \quad \mathrm{trace}(Z) < \gamma^2$$

which is (3.3.5) as desired. The equivalence with (3.3.6) and the matrix inequalities in Y are obtained by a direct dualization and the observation that $\|T\|_2 = \|T^\top\|_2$. ■

Interpretation 3.12 The smallest possible upperbound of the H_2-norm of the transfer function can be calculated by minimizing the criterion $\mathrm{trace}(Z)$ over the variables $K \succ 0$ and Z that satisfy the LMI's defined by the first two inequalities in (3.3.5) or (3.3.6).

3.3.4 Generalized H_2 nominal performance

Consider again the system (3.3.1) and suppose that $x(0) = 0$ and A has its eigenvalues in $\mathbb{C}^-$. Recall that $\|T\|_{H_2} < \infty$ if and only if $D = 0$. The system then defines a bounded operator from $\mathscr{L}_2$ inputs to $\mathscr{L}_\infty$ outputs. That is, for any input w for which $\|w\|_2^2 := \int_0^\infty \|w(t)\|^2 dt < \infty$ the corresponding output z belongs to $\mathscr{L}_\infty$, the space of signals $z : \mathbb{R}_+ \to \mathbb{R}^p$ of finite *amplitude*[2]

$$\|z\|_\infty := \sup_{t \geq 0} \sqrt{\langle z(t), z(t) \rangle}.$$

The $\mathscr{L}_2$-$\mathscr{L}_\infty$ induced norm (or 'energy to peak' norm) of the system is defined as

$$\|T\|_{2,\infty} := \sup_{0 < \|w\|_2 < \infty} \frac{\|z\|_\infty}{\|w\|_2}$$

and satisfies ([26])

$$\|T\|_{2,\infty}^2 = \frac{1}{2\pi} \lambda_{\max} \left(\int_{-\infty}^{\infty} T(i\omega) T(i\omega)^* \, d\omega \right) \tag{3.3.7}$$

where $\lambda_{\max}(\cdot)$ denotes maximum eigenvalue. Note that when z is scalar valued, the latter expression reduces to the H_2 norm, i.e, for systems with scalar valued output variables

$$\|T\|_{2,\infty} = \|T\|_{H_2},$$

which is the reason why we refer to (3.3.7) as a *generalized H_2 norm*. The following result characterizes an upperbound on this quantity.

Proposition 3.13 *Suppose that the system* (3.3.1) *is asymptotically stable and that* $D = 0$*. Then* $\|T\|_{2,\infty} < \gamma$ *if and only if there exists a solution* $K = K^\top > 0$ *to the LMI's*

$$\begin{pmatrix} A^\top K + KA & KB \\ B^\top K & -\gamma I \end{pmatrix} \prec 0; \qquad \begin{pmatrix} K & C^\top \\ C & \gamma I \end{pmatrix} \succ 0 \tag{3.3.8}$$

Proof. Firstly, infer from Theorem 2.9 that the existence of $K > 0$ with

$$\begin{pmatrix} A^\top K + KA & KB \\ B^\top K & -I \end{pmatrix} \prec 0$$

is equivalent to the dissipativity of the system (3.3.1) with respect to the supply function $s(w, z) = w^\top w$. Equivalently, for all $w \in \mathscr{L}_2$ and $t \geq 0$ there holds

$$x(t)^\top K x(t) \leq \int_0^t w(\tau)^\top w(\tau) \, d\tau.$$

[2] An alternative and more common definition for the $\mathscr{L}_\infty$ norm of a signal $z : \mathbb{R} \to \mathbb{R}^p$ is $\|z\|_\infty := \max_{j=1,\ldots,p} \sup_{t \geq 0} |z_j(t)|$. For scalar valued signals this coincides with the given definition, but for non-scalar signals this is a different signal norm. When equipped with this alternative amplitude norm of output signals, the characterization (3.3.7) still holds with $\lambda_{\max}(\cdot)$ redefined as the *maximal entry on the diagonal* of its argument. See [26] for details.

Secondly, using Schur complements, the LMI

$$\begin{pmatrix} K & C^\top \\ C & \gamma^2 I \end{pmatrix} \succ 0$$

is equivalent to the existence of an $\epsilon > 0$ such that $C^\top C \prec (\gamma^2 - \epsilon^2)K$. Together, this yields that for all $t \geq 0$

$$\begin{aligned} \langle z(t), z(t)\rangle &= x(t)^\top C^\top C x(t) \leq (\gamma^2 - \epsilon^2) x(t)^\top K x(t) \\ &\leq (\gamma^2 - \epsilon^2) \int_0^t w(\tau)^\top w(\tau)\, d\tau. \\ &\leq \int_0^\infty w(\tau)^\top w(\tau)\, d\tau. \end{aligned}$$

Take the supremum over $t \geq 0$ yields the existence of $\epsilon > 0$ such that for all $w \in \mathcal{L}_2$

$$\|z\|_\infty^2 \;\leq\; (\gamma^2 - \epsilon^2)\|w\|_2^2.$$

Dividing the latter expression by $\|w\|_2^2$ and taking the supremum over all $w \in \mathcal{L}_2$ then yields the result. ■

3.3.5 L_1 or peak-to-peak nominal performance

Consider the system (3.3.1) and assume again that the system is stable. For fixed initial condition $x(0) = 0$ this system defines a mapping from bounded amplitude inputs $w \in \mathcal{L}_\infty$ to bounded amplitude outputs $z \in \mathcal{L}_\infty$ and a relevant performance criterion is the 'peak-to-peak' or $\mathcal{L}_\infty$-induced norm of this mapping

$$\|T\|_{\infty,\infty} := \sup_{0<\|w\|_\infty<\infty} \frac{\|z\|_\infty}{\|w\|_\infty}.$$

The following result gives a sufficient condition for an upperbound γ of the peak-to-peak gain of the system.

Proposition 3.14 *If there exists $K > 0$, $\lambda > 0$ and $\mu > 0$ such that*

$$\begin{pmatrix} A^\top K + KA + \lambda K & KB \\ B^\top K & -\mu I \end{pmatrix} \prec 0; \qquad \begin{pmatrix} \lambda K & 0 & C^\top \\ 0 & (\gamma - \mu)I & D^\top \\ C & D & \gamma I \end{pmatrix} \succ 0 \tag{3.3.9}$$

then the peak-to-peak (or $\mathcal{L}_\infty$ induced) norm of the system is smaller than γ, i.e., $\|T\|_{\infty,\infty} < \gamma$.

Proof. The first inequality in (3.3.9) implies that

$$\frac{d}{dt} x(t)^\top K x(t) + \lambda x(t) K x(t) - \mu w(t)^\top w(t) \;<\; 0.$$

for all w and x for which $\dot{x} = Ax + Bw$. Now assume that $x(0) = 0$ and $w \in \mathcal{L}_\infty$ with $\|w\|_\infty \leq 1$. Then, since $K > 0$, we obtain (pointwise in $t \geq 0$) that

$$x(t)^\top K x(t) \leq \frac{\mu}{\lambda}.$$

Taking a Schur complement of the second inequality in (3.3.9) yields that

$$\begin{pmatrix} \lambda K & 0 \\ 0 & (\gamma - \mu) I \end{pmatrix} - \frac{1}{\gamma - \epsilon} \begin{pmatrix} C & D \end{pmatrix}^\top \begin{pmatrix} C & D \end{pmatrix} \succ 0$$

so that, pointwise in $t \geq 0$ and for all $\|w\|_\infty \leq 1$ we can write

$$\begin{aligned} \langle z(t), z(t) \rangle &\leq (\gamma - \epsilon)[\lambda x(t)^\top K x(t) + (\gamma - \mu) w(t)^\top w(t)] \\ &\leq \gamma(\gamma - \epsilon) \end{aligned}$$

Consequently, the peak-to-peak gain of the system is smaller than γ. ∎

Remark 3.15 We emphasize that Proposition 3.14 gives only a sufficient condition for an upperbound γ of the peak-to-peak gain of the system. The minimal $\gamma \geq 0$ for which the there exist $K > 0$, $\lambda > 0$ and $\mu \geq 0$ such that (3.3.9) is satisfied is usually only an upperbound of the real peak-to-peak gain of the system.

3.4 The generalized plant concept

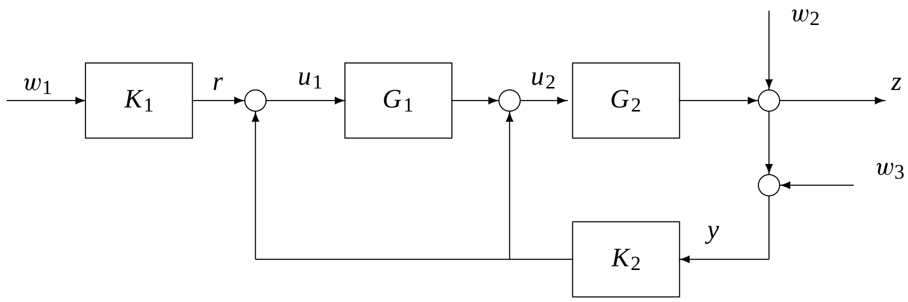


Figure 3.1: A system interconnection

Consider a configuration of interconnected systems such as the one given in Figure 3.1. The subsystems G_j, called *plant components*, are assumed to be given, whereas the subsystems K_j are to-be-designed and generally referred to as *controller components*. In this section, we would like to illustrate how one embeds a specific system interconnection (such as the one in Figure 3.1) into a so-called *generalized plant*. The first step is to identify the following (possibly multivariable) signals:

- **Generalized disturbances**. The collection w of all exogenous signals that affect the interconnection and cannot be altered or influenced by any of the controller components.
- **Controlled outputs**. The collection z of all signals such that the channel $w \mapsto z$ allows to characterize whether the controlled system has user-desired or user specified properties.
- **Control inputs**. The collection u of all signals that are actuated (or generated) by all controller components.
- **Measured outputs**. The collection y of all signals that enter the controller components.

After having specified these four signals, one can disconnect the controller components to arrive at the *open-loop interconnection*. Algebraically, the open-loop interconnection is described by

$$\begin{pmatrix} z \\ y \end{pmatrix} = P \begin{pmatrix} w \\ u \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} w \\ u \end{pmatrix} \tag{3.4.1}$$

where P defines the specific manner in which the plant components interact. A transfer function or state-space description of (3.4.1) can be computed manually or with the help of basic routines in standard modeling toolboxes. The *controlled interconnection* is obtained by connecting the controller as

$$u = Ky.$$

With $y = P_{21}w + P_{22}Ky$ or $(I - P_{22}K)y = P_{21}w$, let us assume that K is such that $I - P_{22}K$ has a proper inverse. Then $y = (I - P_{22}K)^{-1}P_{21}w$ and hence

$$z = [P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}]w = (P \star K)w$$

such that the transfer function $w \mapsto z$ is proper. The general closed-loop interconnection is depicted in Figure 3.2. Clearly, $I - P_{22}K$ has a proper inverse if and only if its direct feed-through matrix

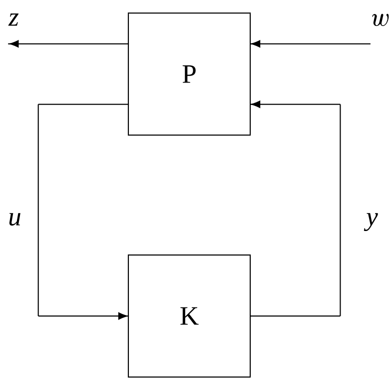


Figure 3.2: General closed-loop interconnection.

$I - P_{22}(\infty)K(\infty)$ is non-singular, a test which is easily verifiable. Equivalently,

$$\begin{pmatrix} I & -K \\ -P_{22} & I \end{pmatrix} \quad \text{has a proper inverse.} \tag{3.4.2}$$

In general, the controller K is required to stabilize the controlled interconnection. For most applications, it is not sufficient that K just renders $P \star K$ stable. Instead, we will say that K *stabilizes* P (or P *is stabilized by* K) if the interconnection depicted in Figure **??** and defined through the relations

$$\begin{pmatrix} z \\ y \end{pmatrix} = P \begin{pmatrix} w \\ u \end{pmatrix}, \qquad u = Kv + v_1, \quad v = y + v_2$$

or, equivalently, through the relations

$$\begin{pmatrix} z \\ \hline v_1 \\ v_2 \end{pmatrix} = \left(\begin{array}{c|cc} P_{11} & P_{12} & 0 \\ \hline 0 & I & -K \\ -P_{21} & -P_{22} & I \end{array}\right) \begin{pmatrix} w \\ \hline u \\ v \end{pmatrix} \tag{3.4.3}$$

defines a *proper and stable transfer matrix* mapping $\mathrm{col}(w, v_1, v_2)$ to $\mathrm{col}(z, u, v)$. Note that the mapping $\mathrm{col}(u, v) \mapsto \mathrm{col}(v_1, v_2)$ is represented by (3.4.2). Consequently, (3.4.2) holds whenever K stabilizes P. In that case, the mapping $\mathrm{col}(v_1, v_2) \mapsto \mathrm{col}(u, v)$ can be explicitly computed as

$$\begin{pmatrix} I & -K \\ -P_{22} & I \end{pmatrix}^{-1} = \begin{pmatrix} (I - KP_{22})^{-1} & K(I - P_{22}K)^{-1} \\ (I - P_{22}K)^{-1}P_{22} & (I - P_{22}K)^{-1} \end{pmatrix}.$$

Consequently, this transfer matrix is explicitly given as

$$\left(\begin{array}{c|c} P_{11} + \begin{pmatrix} P_{12} & 0 \end{pmatrix} \begin{pmatrix} I & -K \\ -P_{22} & I \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ P_{21} \end{pmatrix} & \begin{pmatrix} P_{12} & 0 \end{pmatrix} \begin{pmatrix} I & -K \\ -P_{22} & I \end{pmatrix}^{-1} \\ \hline \begin{pmatrix} I & -K \\ -P_{22} & I \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ P_{21} \end{pmatrix} & \begin{pmatrix} I & -K \\ -P_{22} & I \end{pmatrix}^{-1} \end{array}\right) =$$

$$= \left(\begin{array}{c|cc} P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21} & P_{12}(I - KP_{22})^{-1} & P_{12}K(I - P_{22}K)^{-1} \\ \hline K(I - P_{22}K)^{-1}P_{21} & (I - KP_{22})^{-1} & K(I - P_{22}K)^{-1} \\ (I - P_{22}K)^{-1}P_{21} & (I - P_{22}K)^{-1}P_{22} & (I - P_{22}K)^{-1} \end{array}\right). \tag{3.4.4}$$

Hence, K stabilizes P if it satisfies (3.4.2) and if all nine block transfer matrices (3.4.4) are stable.

Let us now assume that P and K admit the state-space descriptions

$$\begin{aligned} \dot{x} &= Ax + B_1 w + B_2 u \\ z &= C_1 x + D_{11} w + D_{12} u \\ y &= C_2 x + D_{21} w + D_{22} u \end{aligned} \tag{3.4.5}$$

and

$$\begin{aligned} \dot{x}_K &= A_K x_K + B_K u_K \\ y_K &= C_K x_K + D_K u_K. \end{aligned} \tag{3.4.6}$$

Clearly, (3.4.3) admits the state-space realization

$$\begin{pmatrix} \dot{x} \\ \dot{x}_K \\ \hline z \\ \hline v_1 \\ v_2 \end{pmatrix} = \left(\begin{array}{cc|c|cc} A & 0 & B_1 & B_2 & 0 \\ 0 & A_K & 0 & 0 & B_K \\ \hline C_1 & 0 & D_{11} & D_{12} & 0 \\ \hline 0 & -C_K & 0 & I & -D_K \\ -C_2 & 0 & -D_{21} & -D_{22} & I \end{array}\right) \begin{pmatrix} x \\ x_K \\ \hline w \\ \hline u \\ v \end{pmatrix}. \tag{3.4.7}$$

Obviously, (3.4.2) just translates into the property that

$$\begin{pmatrix} I & -D_K \\ -D_{21} & I \end{pmatrix} \text{ is non-singular.} \tag{3.4.8}$$

We can write (3.4.7) in more compact format as

$$\begin{pmatrix} \dot{x} \\ \dot{x}_K \\ \hline z \\ \hline v_1 \\ v_2 \end{pmatrix} = \left(\begin{array}{c|c|c} \boldsymbol{A} & \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \hline \boldsymbol{C}_1 & \boldsymbol{D}_{11} & \boldsymbol{D}_{12} \\ \hline \boldsymbol{C}_2 & \boldsymbol{D}_{21} & \boldsymbol{D}_{22} \end{array}\right) \begin{pmatrix} x \\ x_K \\ \hline w \\ \hline u \\ v \end{pmatrix}$$

This hypothesis allows to derive the following state-space realization of (**??**):

$$\begin{pmatrix} \dot{x} \\ \dot{x}_K \\ \hline z \\ \hline u \\ v \end{pmatrix} = \left(\begin{array}{c|c|c} \boldsymbol{A} - \boldsymbol{B}_2\boldsymbol{D}_{22}^{-1}\boldsymbol{C}_2 & \boldsymbol{B}_1 - \boldsymbol{B}_2\boldsymbol{D}_{22}^{-1}\boldsymbol{D}_{21} & \boldsymbol{B}_2\boldsymbol{D}_{22}^{-1} \\ \hline \boldsymbol{C}_1 - \boldsymbol{D}_{12}\boldsymbol{D}_{22}^{-1}\boldsymbol{C}_2 & \boldsymbol{D}_1 - \boldsymbol{D}_{12}\boldsymbol{D}_{22}^{-1}\boldsymbol{D}_{21} & \boldsymbol{D}_{12}\boldsymbol{D}_{22}^{-1} \\ \hline \boldsymbol{D}_{22}^{-1}\boldsymbol{C}_2 & \boldsymbol{D}_{22}^{-1}\boldsymbol{D}_{21} & \boldsymbol{D}_{22}^{-1} \end{array}\right) \begin{pmatrix} x \\ x_K \\ \hline w \\ \hline v_1 \\ v_2 \end{pmatrix}. \tag{3.4.9}$$

Let us now assume that the realizations of P and K are stabilizable and detectable. It is a simple exercise to verify that the realizations (3.4.7) and (3.4.9) are then stabilizable and detectable. This allows to conclude that the transfer matrix (**??**) with stabilizable and detectable realization (3.4.9) is stable if and only if $\boldsymbol{A} - \boldsymbol{B}_2\boldsymbol{D}_{22}^{-1}\boldsymbol{C}_2$ is Hurwitz. More explicitly

$$\begin{pmatrix} A & 0 \\ 0 & A_K \end{pmatrix} + \begin{pmatrix} B_2 & 0 \\ 0 & B_K \end{pmatrix} \begin{pmatrix} I & -D_K \\ -D_{22} & I \end{pmatrix}^{-1} \begin{pmatrix} 0 & C_K \\ C_2 & 0 \end{pmatrix} \text{ is Hurwitz.} \tag{3.4.10}$$

In summary K stabilizies P iff (for stabilizable and detectable realizations) (3.4.8) and (3.4.10) hold true.

Finally we call P a **generalized plant** if it indeed admits a stabilizing controller. This property can be easily tested both in terms of the transfer matrix P and its state-space realization.

Suppose that P admits a stabilizing controller. In terms of (stabilizable/detectable) representations this means (3.4.10) which in turn implies, just with e.g. the Hautus test, that

$$(A, B_2) \text{ is stabilizable and } (A, C_2) \text{ is detectable.} \tag{3.4.11}$$

Conversely, (3.4.11) allow to determine F and L such that $A+B_2F$ and $A+LC_2$ are Hurwitz. Then the standard observer based controller for the system $\dot{x} = Ax + B_2u$, $y = C_2x + D_{22}u$ which is defined as

$$\dot{x}_K = (A + B_2F + LC_2 + LD_{22}F)x_K - Ly, \quad u = Fx_K$$

does indeed stabilize P. The realization $(A_K, B_K, C_K, D_K) = (A+B_2F+LC_2+LD_{22}F, -L, F, 0)$ is obviously stabilizable and detectable. Since $D_K = 0$ we infer (3.4.8). Finally a simple computation shows

$$\begin{pmatrix} A & 0 \\ 0 & A_K \end{pmatrix} + \begin{pmatrix} B_2 & 0 \\ 0 & B_K \end{pmatrix} \begin{pmatrix} I & 0 \\ -D_{22} & I \end{pmatrix}^{-1} \begin{pmatrix} 0 & C_K \\ C_2 & 0 \end{pmatrix} = \begin{pmatrix} A & B_2F \\ -LC_2 & A + B_2F + LC_2 \end{pmatrix}$$

and the similarity transformation (error dynamics)

$$\begin{pmatrix} I & 0 \\ I & -I \end{pmatrix} \begin{pmatrix} A & B_2F \\ -LC_2 & A + B_2F + LC_2 \end{pmatrix} \begin{pmatrix} I & 0 \\ I & -I \end{pmatrix}^{-1} = \begin{pmatrix} A + B_2F & B_2F \\ 0 & A + LC_2 \end{pmatrix}$$

reveals that (3.4.10) is true as well. Hence (3.4.11) is necessary and sufficient for P being a generalized plant.

Let us finally derive a test directly in terms of transfer matrices. If K stabilizes P we conclude that

$$\begin{pmatrix} I & -K \\ -P_{22} & I \end{pmatrix}^{-1} \text{ is proper and stable.} \tag{3.4.12}$$

Note that this comes down to K stabilizing P_{22} in the sense of our definition for empty w_p and z_p. We conclude that the set of controllers which stabilize P is a subset of the controller which stabilize P_{22}. Generalized plants are characterized by the fact that these two sets of controllers do actually coincide. This can be most easily verified in terms of state-space realizations: With stabilizable and detectable realizations (3.4.5) and (3.4.6) suppose that K renders (3.4.12) satisfied. In exactly the same fashion as above we observe that (3.4.12) admits the realization

$$\begin{pmatrix} \dot{x} \\ \dot{x}_K \\ \hline u \\ v \end{pmatrix} = \left(\begin{array}{c|c} \boldsymbol{A} - \boldsymbol{B}_2 \boldsymbol{D}_{22}^{-1} \boldsymbol{C}_2 & \boldsymbol{B}_2 \boldsymbol{D}_{22}^{-1} \\ \hline \boldsymbol{D}_{22}^{-1} \boldsymbol{C}_2 & \boldsymbol{D}_{22}^{-1} \end{array} \right) \begin{pmatrix} x \\ x_K \\ \hline v_1 \\ v_2 \end{pmatrix}$$

which is stabilizable and detectable. This implies that $\boldsymbol{A} - \boldsymbol{B}_2 \boldsymbol{D}_{22}^{-1} \boldsymbol{C}_2$ is Hurwitz, which in turn means that K also stabilizes P.

This leads us to the following two easily verifiable input-output and state-space tests for P being a generalized plant.

- Determine a K which stabilizes P_{22}. Then P is a generalized plant if and only if K also stabilizes P.
- If P admits a stabilizable and detectable representation (3.4.5), then it is a generalized plant if and only if (A, B_2) is stabilizable and (A, C_2) is detectable.

3.5 Further reading

Lyapunov theory: [5, 16, 27, 47]
More details on the generalized H_2 norm, see [26]. A first variation of Theorem 3.7 appeared in [3].

3.6 Exercises

Exercise 1
Consider the non-linear differential equation

$$\begin{cases} \dot{x}_1 &= x_1(1 - x_1) \\ \dot{x}_2 &= 1 \end{cases} \tag{3.6.1}$$

(a) Show that there exists one and only one solution of (3.6.1) which passes through $x_0 \in \mathbb{R}^2$ at time $t = 0$.

(b) Show that (3.6.1) admits only one equilibrium point.

(c) Show that the set $S := \{x \in \mathbb{R}^2 \mid |x_1| = 1\}$ is a positive and negative invariant set of (3.6.1).

Exercise 2
Show that

(a) the quadratic function $V(x) := x^\top K x$ is positive definite if and only if K is a positive definite matrix.

(b) the function $V : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ defined as $V(x, t) := e^{-t} x^2$ is decrescent but not positive definite.

Exercise 3
A pendulum of mass m is connected to a servo motor which is driven by a voltage u. The angle which the pendulum makes with respect to the upright vertical axis through the center of rotation is denoted by θ (that is, $\theta = 0$ means that the pendulum is in upright position). The system is described by the equations

$$J \frac{d^2\theta}{dt^2} = mlg \sin(\theta) + u \tag{3.6.2}$$

$$y = \theta \tag{3.6.3}$$

where l denotes the distance from the axis of the servo motor to the center of mass of the pendulum, J is the inertia and g is the gravitation constant. The system is specified by the constants $J = 0.03$, $m = 1$, $l = 0.15$ and $g = 10$.

(a) Determine the equilibrium points of this system.

(b) Are the equilibrium points Lyapunov stable? If so, determine a Lyapunov function.

(c) Linearize the system around the equilibrium points and provide a state space representation of the linearized systems.

(d) Verify whether the linearized systems are stable, unstable or asymptotically stable.

(e) A *proportional feedback* controller is a controller of the form $u = ky$ where $k \in \mathbb{R}$. Does there exists a proportional feedback controller such that the unstable equilibrium point of the system becomes asymptotically stable?

Exercise 4
Let a stability region $\mathbb{C}_{\text{stab}}$ be defined as those complex numbers $s \in \mathbb{C}$ which satisfy

$$\begin{cases} \operatorname{Re}(s) < -\alpha & \text{and} \\ |s - c| < r & \text{and} \\ |\operatorname{Im}(s)| < |\operatorname{Re}(s)|. \end{cases}$$

where $\alpha > 0$, $c > 0$ and $r > 0$. Specify a real symmetric matrix $P \in \mathbb{S}^{2m \times 2m}$ such that $\mathbb{C}_{\text{stab}}$ coincides with the LMI region L_P as specified in Definition (3.6).

Exercise 5
Let $0 \leq \alpha \leq \pi$ and consider the Lyapunov equation $A^\top X + XA + I = 0$ where

$$A = \begin{pmatrix} \sin(\alpha) & \cos(\alpha) \\ -\cos(\alpha) & \sin(\alpha) \end{pmatrix}$$

Show that the solution X of the Lyapunov equation diverges in the sense that $\det(X) \longrightarrow \infty$ whenever $\alpha \longrightarrow 0$.

Exercise 6
Consider the the suspension system in Exercise 4 of Chapter 2. Recall that the variable q_0 represents the road profile.

(a) Consider the case where $F = 0$ and $q_0 = 0$ (thus no active force between chassis and axle and a 'flat' road characteristic). Verify whether this system is asymptotically stable.

(b) Determine a Lyapunov function $V : \mathcal{X} \to \mathbb{R}$ of this system (with $F = 0$ and $q_0 = 0$) and show that its derivative is negative along solutions of the autonomous behavior of the system (i.e. $F = 0$ and $q_0 = 0$).

(c) Design your favorite road profile q_0 in MATLAB and simulate the response of the system to this road profile (the force F is kept 0). Plot the variables q_1 and q_2. What are your conclusions?

(d) Consider, with $F = 0$, the transfer function T which maps the road profile q_0 to the output $\operatorname{col}(q_1, q_2)$ of the system. Determine the norms $\|T\|_{H_\infty}$ and $\|T\|_{H_2}$.

Exercise7
Consider the system $\dot{x} = Ax + Bw$, $z = Cx + Dw$ with

$$A = \begin{pmatrix} -1 & 0 & 0 & 1 \\ 0 & -1 & 4 & -3 \\ 1 & -3 & -1 & -3 \\ 0 & 4 & 2 & -1 \end{pmatrix} B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ -1 & 0 \\ 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

(a) Write a program to compute the H_∞ norm of this system.

(b) Check the eigenvalues of the Lyapunov matrix. What can you conclude from this?

Exercise 8
Consider a batch chemical reactor with a constant volume V of liquids. Inside the reactor the series reaction

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C$$

takes place. Here k_1 and k_2 represent the kinetic rate constants (1/sec.) for the conversions $A \to B$ and $B \to C$, respectively. The conversions are assumed to be irreversible which leads to the model equations

$$\begin{aligned} \dot{C}_A &= -k_1 C_A \\ \dot{C}_B &= k_1 C_A - k_2 C_B \\ \dot{C}_C &= k_2 C_B \end{aligned}$$

where C_A, C_B and C_C denote the concentrations of the components A, B and C, respectively, and k_1 and k_2 are positive constants. Reactant B is the desired product and we will be interested in the evolution of its concentration.

(a) Show that the system which describes the evolution of C_B is asymptotically stable.

(b) Determine a Lyapunov function for this system.

(c) Suppose that at time $t = 0$ the reactor is injected with an initial concentration $C_A(0) = 10$ (mol/liter) of reactant A and that $C_B(0) = C_C(0) = 0$. Plot the time evolution of the concentration C_B of reactant B if $(k_1, k_2) = (0.2, 0.4)$ and if $(k_1, k_2) = (0.3, 0.3)$.

Exercise 9
Consider the nonlinear scalar differential equation

$$\dot{x} = \sqrt{x}.$$

with initial condition $x(0) = 0$.

(a) Show that this differential equation does not satisfy the Lipschitz condition to guarantee uniqueness of solutions $x : \mathbb{R}_+ \to \mathbb{R}$.

(b) Show that the differential equation has at least two solutions $x(t)$, $t \geq 0$, with $x(0) = 0$.

Exercise 10
Consider the discrete time system

$$x(t+1) = f(x(t)). \tag{3.6.4}$$

(a) Show that x^* is an equilibrium point of (3.6.4) if and only if $f(x^*) = x^*$.

(b) Let $x(t+1) = Ax(t)$ be the linearization of (3.6.4) around the equilibrium point x^*. Derive an LMI feasibility test which is necessary and sufficient for A to have its eigenvalues in $\{z \in \mathbb{C} \mid |z| < 1\}$. (*Hint: apply Theorem* **??**).

Exercise 11
In chemical process industry, distillation columns play a key role to split an input stream of chemical species into two or more output streams of desired chemical species. Distillation is usually the most economical method for separating liquids, and consists of a process of multi-stage equilibrium separations. Figure 3.3 illustrates a typical distillation column.

Separation of input components, the *feed*, is achieved by controlling the transfer of components between the various *stages* (also called *trays* or *plates*), within the column, so as to produce output products at the bottom and at the top of the column. In a typical distillation system, two recycle streams are returned to the column. A *condenser* is added at the top of the column and a fraction of the overhead vapor V is condensed to form a liquid recycle L. The liquid recycle provides the liquid stream needed in the tower. The remaining fraction of V, is the *distillate*- or *top product*. A vaporizer or *reboiler* is added to the bottom of the column and a portion of the bottom liquid, L_{b}, is vaporized and recycled to the tower as a vapor stream V_{b}. This provides the vapor stream needed in the tower, while the remaining portion of L_{b} is the *bottom product*.

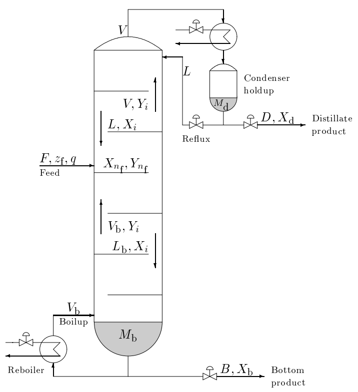


Figure 3.3: Binary distillation column

The column consists of n *stages*, numbered from top to bottom. The *feed* enters the column at stage n_{f}, with $1 < n_{\mathrm{f}} < n$. The *feed flow*, F [kmol/hr], is a saturated liquid with composition z_F [mole fraction]. L [kmole/hr] denotes the *reflux flow rate* of the condenser, V_b [kmole/hr] is the *boilup flow rate* of the reboiler. The variable

$$w = \begin{pmatrix} L \\ V_\mathrm{b} \\ F \end{pmatrix}$$

is taken as input of the plant. The top product consists of a *distillate stream* D [kmol/hr], with composition X_d [mole fraction]. Likewise, the bottom product consists of a *bottom stream* B, with composition X_B [mole fraction]. The output of the system is taken to be

$$z = \begin{pmatrix} X_\mathrm{d} \\ X_\mathrm{b} \end{pmatrix}$$

and therefore consists of the distillate composition and bottom composition, respectively.

A model for this type of reactors is obtained as follows. The stages above the feed stage (index $i < n_\mathrm{f}$) define the *enriching section* and those below the feed stage (index $i > n_\mathrm{f}$) the *stripping section* of the column. The liquid flow rate in the stripping section is defined as $L_\mathrm{b} = L + qF$ where $0 \leq q \leq 1$ is a constant. The vapor flow rate in the enriching section is given by $V = V_\mathrm{b} + (1-q)F$. The distillate and bottom product flow rates are $D = V - L$ and $B = L_\mathrm{b} - V_\mathrm{b}$, respectively. Denote by X_i and Y_i [mole fraction] the liquid and vapor compositions of stage i, respectively. For constant liquid holdup conditions, the *material balances* of the column are given as follows.

$$\begin{aligned}
M_\mathrm{d}\frac{dX_1}{dt} &= VY_2 - (L + D)X_1 && \text{Condenser stage} \\
M\frac{dX_i}{dt} &= L(X_{i-1} - X_i) + V(Y_{i+1} - Y_i) && 1 < i < n_\mathrm{f} \\
M\frac{dX_\mathrm{f}}{dt} &= LX_{\mathrm{f}-1} - L_\mathrm{b}X_\mathrm{f} + V_\mathrm{b}Y_{\mathrm{f}+1} - VY_\mathrm{f} + Fz_\mathrm{f} && \text{Feed stage} \\
M\frac{dX_i}{dt} &= L_\mathrm{b}(X_{i-1} - X_i) + V_\mathrm{b}(Y_{i+1} - Y_i) && n_\mathrm{f} < i < n \\
M_\mathrm{b}\frac{dX_n}{dt} &= L_\mathrm{b}X_{n-1} - V_\mathrm{b}Y_n - BX_n. && \text{Reboiler stage}
\end{aligned}$$

Here, M, M_d and M_b [kmol] denote the nominal stage hold-up of material at the trays, the condenser and the bottom, respectively. The vapor-liquid equilibrium describes the relation between the vapor and liquid compositions Y_i and X_i on each stage i of the column and is given by the non-linear expression:

$$Y_i = \frac{aX_i}{1 + (a-1)X_i}, \qquad i = 1, \ldots, n.$$

where a is the so called *relative volatility* (dependent on the product). With x denoting the vector of components X_i, these equations yield a nonlinear model of the form

$$\dot{x} = f(x, w), \qquad z = g(x, w)$$

of state dimension n (equal to the number of stages).

We consider a medium-sized propane-butane distillation column whose physical parameters are given in Table 3.1. The last three entries in this table define w^*.

n	Number of stages	20
n_f	Feed stage	6
$M\mathrm{d}$	Condenser holdup	200 [kmol]
$M\mathrm{b}$	Reboiler holdup	400 [kmol]
M	Stage holdup	50 [kmol]
z_f	Feed composition	0.5 [mole fraction]
q	Feed liquid fraction	1
a	Relative volatility	2.46
L^*	Reflux flow	1090 [kmol/hr]
V_b^*	Boilup vapor flow	1575 [kmol/hr]
F^*	Feed flow	1000 [kmol/hr]

Table 3.1: Operating point data

(a) Calculate an equilibrium state x^* of the model if the input is set to w^*. The equilibrium point (w^*, x^*, z^*) represents a steady-state or *nominal operating point* of the column.

(b) Construct (or compute) a linear model of the column when linearized around the equilibrium point (w^*, x^*, z^*).

(c) Is the linear model stable?

(d) Is the linear model asymptotically stable?

(e) Make a Bode diagram of the 2×3 transfer function of this system.

(f) Determine the responses of the system if the set-points of, respectively, L, V_b and F undergo a $+10\%$ step-change with respect to their nominal values L^*, V_b^* and F^*. Can you draw any conclusion on the (relative) sensitivity of the system with respect to its inputs w? Are there any major differences in the settling times?

Chapter 4

Controller synthesis

In this chapter we provide a powerful result that allows to step in a straightforward manner from the performance analysis conditions formulated in terms of matrix inequalities to the corresponding matrix inequalities for controller synthesis. This is achieved by a nonlinear and essentially bijective transformation of the controller parameters.

4.1 The setup

Suppose a linear time-invariant system is described as

$$\begin{pmatrix} \dot{x} \\ \hline z_1 \\ \vdots \\ z_q \\ y \end{pmatrix} = \left(\begin{array}{c|cccc} A & B_1 & \cdots & B_q & B \\ \hline C_1 & D_1 & \cdots & D_{1q} & E_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_q & D_{q1} & \cdots & D_q & E_q \\ C & F_1 & \cdots & F_q & 0 \end{array}\right) \begin{pmatrix} x \\ \hline w_1 \\ \vdots \\ w_q \\ u \end{pmatrix}. \tag{4.1.1}$$

We denote by u the control input, by y the measured output available for control, and by $w_j \to z_j$ the channels on which we want to impose certain robustness and/or performance objectives. Since we want to extend the design technique to mixed problems with various performance specifications on various channels, we already start at this point with a multi-channel system description. Sometimes we collect the signals as

$$z = \begin{pmatrix} z_1 \\ \vdots \\ z_q \end{pmatrix}, \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_q \end{pmatrix}.$$

Remark 4.1 Note that we do not exclude the situation that some of the signals w_j or z_j are identical. Therefore, we only need to consider an equal number of input- and output-signals. Moreover, it might seem restrictive to only consider the diagonal channels and neglect the channels $w_j \to z_k$ for $j \neq k$. This is not the case. As a typical example, suppose we intend to impose for $z = Tw$ specifications on $L_j T R_j$ where L_j, R_j are arbitrary matrices that pick out certain linear combinations of the signals z, w (or of the rows/columns of the transfer matrix if T is described by an LTI system). If we set $w = R_j w_j$, $z_j = L_j z$, we are hence interested in specifications on the diagonal channels of

$$\begin{pmatrix} z_1 \\ z_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} L_1 \\ L_2 \\ \vdots \end{pmatrix} T \begin{pmatrix} R_1 & R_2 & \dots \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \end{pmatrix}.$$

If T is LTI, the selection matrices L_j and R_j can be easily incorporated into the realization to arrive at the description (4.1.1).

A *controller* is any finite dimensional linear time invariant system described as

$$\begin{pmatrix} \dot{x}_c \\ u \end{pmatrix} = \begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \begin{pmatrix} x_c \\ y \end{pmatrix} \tag{4.1.2}$$

that has y as its input and u as its output. Controllers are hence simply parameterized by the matrices A_c, B_c, C_c, D_c.

The controlled or closed-loop system then admits the description

$$\begin{pmatrix} \dot{\xi} \\ z \end{pmatrix} = \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{pmatrix} \begin{pmatrix} \xi \\ w \end{pmatrix} \text{ or } \begin{pmatrix} \dot{\xi} \\ z_1 \\ \vdots \\ z_q \end{pmatrix} = \left(\begin{array}{c|ccc} \mathcal{A} & \mathcal{B}_1 & \cdots & \mathcal{B}_q \\ \hline \mathcal{C}_1 & \mathcal{D}_1 & \cdots & \mathcal{D}_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{C}_q & \mathcal{D}_{q1} & \cdots & \mathcal{D}_q \end{array}\right) \begin{pmatrix} \xi \\ w_1 \\ \vdots \\ w_q \end{pmatrix}. \tag{4.1.3}$$

The corresponding input-output mappings (or transfer matrices) are denoted as

$$w = \mathcal{T} z \text{ or } \begin{pmatrix} z_1 \\ \vdots \\ z_q \end{pmatrix} = \begin{pmatrix} \mathcal{T}_1 & & * \\ & \ddots & \\ * & & \mathcal{T}_q \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_q \end{pmatrix}.$$

respectively.

One can easily calculate a realization of $\mathcal{T}_j$ as

$$\begin{pmatrix} \dot{\xi} \\ z_j \end{pmatrix} = \begin{pmatrix} \mathcal{A} & \mathcal{B}_j \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix} \begin{pmatrix} \xi \\ w_j \end{pmatrix} \tag{4.1.4}$$

where

$$\left(\begin{array}{c|c} \mathcal{A} & \mathcal{B}_j \\ \hline \mathcal{C}_j & \mathcal{D}_j \end{array}\right) = \left(\begin{array}{cc|c} A + BD_cC & BC_c & B_j + BD_cF_j \\ B_cC & A_c & B_cF_j \\ \hline C_j + E_jD_cC & E_jC_c & D_j + E_jD_cF_j \end{array}\right).$$

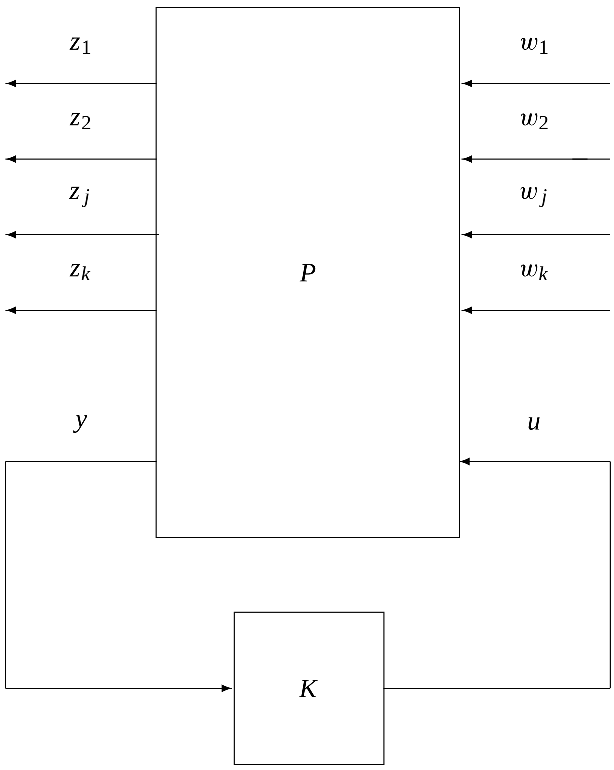


Figure 4.1: Multi-channel closed-loop system

It simplifies some calculations if we use the equivalent alternative formula

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}_j \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix} = \left(\begin{array}{cc|c} A & 0 & B_j \\ 0 & 0 & 0 \\ \hline C_j & 0 & D_j \end{array}\right) + \left(\begin{array}{cc} 0 & B \\ I & 0 \\ \hline 0 & E_j \end{array}\right) \begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \left(\begin{array}{cc|c} 0 & I & 0 \\ C & 0 & F_j \end{array}\right). \tag{4.1.5}$$

From this notation it is immediate that the left-hand side is an affine function of the controller parameters.

4.2 From analysis to synthesis – a general procedure

As a paradigm example let us consider the design of a controller that achieves stability and quadratic performance in the channel $w_j \to z_j$. For that purpose we suppose that we have given a performance index

$$P_j = \begin{pmatrix} Q_j & S_j \\ S_j^T & R_j \end{pmatrix} \text{ with } R_j \geq 0.$$

In Chapter 2 we have revealed that the following conditions are equivalent: The controller (4.1.2) renders (4.1.4) internally stable and leads to

$$\int_0^\infty \begin{pmatrix} w_j(t) \\ z_j(t) \end{pmatrix}^T P_j \begin{pmatrix} w_j(t) \\ z_j(t) \end{pmatrix} dt \leq -\epsilon \int_0^\infty w_j(t)^T w_j(t)\, dt$$

for some $\epsilon > 0$ if and only if

$$\sigma(\mathcal{A}) \subset \mathbb{C}^- \text{ and } \begin{pmatrix} I \\ \mathcal{T}_j(i\omega) \end{pmatrix}^* P_j \begin{pmatrix} I \\ \mathcal{T}_j(i\omega) \end{pmatrix} < 0 \text{ for all } \omega \in \mathbb{R} \cup \{\infty\}$$

if and only if there exists a symmetric $\mathcal{X}$ satisfying

$$\mathcal{X} > 0, \begin{pmatrix} \mathcal{A}^T\mathcal{X} + \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \mathcal{B}_j^T\mathcal{X} & 0 \end{pmatrix} + \begin{pmatrix} 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix}^T P_j \begin{pmatrix} 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix} < 0. \tag{4.2.1}$$

The corresponding *quadratic performance synthesis problem* amounts to finding controller parameters $\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}$ and an $\mathcal{X} > 0$ such that (4.2.1) holds.

Obviously, $\mathcal{A}$ depends on the controller parameters. Since $\mathcal{X}$ is also a variable, we observe that $\mathcal{X}\mathcal{A}$ depends non-linearly on the variables to be found.

It has been observed only quite recently [18, 36] that there exist a nonlinear transformation

$$\left(\mathcal{X}, \begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \right) \to v = \left(X, \quad Y, \begin{pmatrix} K & L \\ M & N \end{pmatrix} \right) \tag{4.2.2}$$

and a $\mathcal{Y}$ such that, with the functions

$$\left.\begin{aligned} \boldsymbol{X}(v) &:= \begin{pmatrix} Y & I \\ I & X \end{pmatrix} \\ \begin{pmatrix} \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{pmatrix} &:= \left(\begin{array}{cc|c} AY+BM & A+BNC & B_j+BNF_j \\ K & AX+LC & XB_j+LF_j \\ \hline C_jY+E_jM & C_j+E_jNC & D_j+E_jNF_j \end{array}\right) \end{aligned}\right\} \tag{4.2.3}$$

one has

$$\left.\begin{aligned} \mathcal{Y}^T\mathcal{X}\mathcal{Y} &= \boldsymbol{X}(v) \\ \begin{pmatrix} \mathcal{Y}^T\mathcal{X}\mathcal{A}\mathcal{Y} & \mathcal{Y}^T\mathcal{X}\mathcal{B}_j \\ \mathcal{C}_j\mathcal{Y} & \mathcal{D}_j \end{pmatrix} &= \begin{pmatrix} \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{pmatrix} \end{aligned}\right\} \tag{4.2.4}$$

Hence, under congruence transformations with the matrices

$$\mathcal{Y} \text{ and } \begin{pmatrix} \mathcal{Y} & 0 \\ 0 & I \end{pmatrix}, \tag{4.2.5}$$

the blocks transform as

$$\mathcal{X} \to \boldsymbol{X}(v), \quad \begin{pmatrix} \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix} \to \begin{pmatrix} \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{pmatrix}.$$

Therefore, the original blocks that depend non-linearly on the decision variables $\mathcal{X}$ and $\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}$ are transformed into blocks that are *affine* functions of the new variables v.

If $\mathcal{Y}$ is nonsingular, we can perform a congruence transformation on the two inequalities in (4.2.1) with the nonsingular matrices (4.2.5) to obtain

$$\mathcal{Y}^T\mathcal{X}\mathcal{Y} > 0, \quad \begin{pmatrix} \mathcal{Y}^T[\mathcal{A}^T\mathcal{X}+\mathcal{X}\mathcal{A}]\mathcal{Y} & \mathcal{Y}^T\mathcal{X}\mathcal{B}_j \\ \mathcal{B}_j^T\mathcal{X}\mathcal{Y} & 0 \end{pmatrix} + \begin{pmatrix} 0 & I \\ \mathcal{C}_j\mathcal{Y} & \mathcal{D}_j \end{pmatrix}^T P_j \begin{pmatrix} 0 & I \\ \mathcal{C}_j\mathcal{Y} & \mathcal{D}_j \end{pmatrix} < 0 \tag{4.2.6}$$

what is nothing but

$$\boldsymbol{X}(v) > 0, \quad \begin{pmatrix} \boldsymbol{A}(v)^T + \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{B}_j(v)^T & 0 \end{pmatrix} + \begin{pmatrix} 0 & I \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{pmatrix}^T P_j \begin{pmatrix} 0 & I \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{pmatrix} < 0. \tag{4.2.7}$$

For $R_j = 0$ (as it happens in the positive real performance index), we infer $P_j = \begin{pmatrix} Q_j & S_j \\ S_j^T & 0 \end{pmatrix}$ what implies that the inequalities (4.2.7) are affine in v. For a general performance index with $R_j \geq 0$, the second inequality in (4.2.7) is non-linear but convex in v. It is straightforward to transform it to a genuine LMI with a Schur complement argument. Since it is more convenient to stay with the inequalities in the form (4.2.7), we rather formulate a general auxiliary result that displays how to perform the linearization whenever it is required for computational purposes.

Lemma 4.2 (Linearization Lemma) *Suppose that A and S are constant matrices, that $B(v)$, $Q(v) = Q(v)^T$ depend affinely on a parameter v, and that $R(v)$ can be decomposed as $TU(v)^{-1}T^T$ with $U(v)$ being affine. Then the non-linear matrix inequalities*

$$U(v) > 0, \quad \begin{pmatrix} A \\ B(v) \end{pmatrix}^T \begin{pmatrix} Q(v) & S \\ S' & R(v) \end{pmatrix} \begin{pmatrix} A \\ B(v) \end{pmatrix} < 0$$

are equivalent to the linear matrix inequality

$$\begin{pmatrix} A^T Q(v)A + A^T SB(v) + B(v)^T S^T A & B(v)^T T \\ T^T B(v) & -U(v) \end{pmatrix} < 0.$$

In order to apply this lemma we rewrite the second inequality of (4.2.7) as

$$\left(\begin{array}{cc} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & Q_j & S_j \\ 0 & 0 & S_j^T & R_j \end{array}\right) \left(\begin{array}{cc} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{array}\right) < 0 \tag{4.2.8}$$

what is, after a simple permutation, nothing but

$$\left(\begin{array}{cc} I & 0 \\ 0 & I \\ \hline \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & 0 & I & 0 \\ 0 & Q_j & 0 & S_j \\ \hline I & 0 & 0 & 0 \\ 0 & S_j^T & 0 & R_j \end{array}\right) \left(\begin{array}{cc} I & 0 \\ 0 & I \\ \hline \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{array}\right) < 0. \tag{4.2.9}$$

This inequality can be linearized according to Lemma 4.2 with an arbitrary factorization

$$R_j = T_j T_j^T \text{ leading to } \begin{pmatrix} 0 & 0 \\ 0 & R_j \end{pmatrix} = \begin{pmatrix} 0 \\ T_j \end{pmatrix} \begin{pmatrix} 0 & T_j^T \end{pmatrix}.$$

So far we have discussed how to derive the synthesis inequalities (4.2.7). Let us now suppose that we have verified that these inequalities do have a solution, and that we have computed some solution v. If we can find a preimage $\left(\mathcal{X}, \begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \right)$ of v under the transformation (4.2.2) and a nonsingular $\mathcal{Y}$ for which (4.2.4) holds, then we can simply reverse all the steps performed above to reveal that (4.2.7) is equivalent to (4.2.1). Therefore, the controller defined by $\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}$ renders (4.2.1) satisfied and, hence, leads to the desired quadratic performance specification for the controlled system.

Before we comment on the resulting design procedure, let us first provide a proof of the following result that summarizes the discussion.

Theorem 4.3 *There exists a controller $\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}$ and an $\mathcal{X}$ satisfying* (4.2.1) *iff there exists an v that solves the inequalities* (4.2.7). *If v satisfies* (4.2.7)*, then $I - XY$ is nonsingular and there exist* *nonsingular U, V with $I - XY = UV^T$. The unique $\mathcal{X}$ and* $\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}$ *with*

$$\begin{pmatrix} Y & V \\ I & 0 \end{pmatrix} \mathcal{X} = \begin{pmatrix} I & 0 \\ X & U \end{pmatrix} \text{ and}$$
$$\begin{pmatrix} K & L \\ M & N \end{pmatrix} = \begin{pmatrix} U & XB \\ 0 & I \end{pmatrix} \begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \begin{pmatrix} V^T & 0 \\ CY & I \end{pmatrix} + \begin{pmatrix} XAY & 0 \\ 0 & 0 \end{pmatrix} \tag{4.2.10}$$

satisfy the LMI's (4.2.1).

Note that U and V are square and nonsingular so that (4.2.10) leads to the formulas

$$\mathcal{X} = \begin{pmatrix} Y & V \\ I & 0 \end{pmatrix}^{-1} \begin{pmatrix} I & 0 \\ X & U \end{pmatrix} \text{ and}$$
$$\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} = \begin{pmatrix} U & XB \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} K - XAY & L \\ M & N \end{pmatrix} \begin{pmatrix} V^T & 0 \\ CY & I \end{pmatrix}^{-1}.$$

Due to the zero blocks in the inverses, the formulas can be rendered even more explicit. Of course, numerically it is better to directly solve the equations (4.2.10) by a stable technique.

Proof. Suppose a controller and some $\mathcal{X}$ satisfy (4.2.1). Let us partition

$$\mathcal{X} = \begin{pmatrix} X & U \\ U^T & * \end{pmatrix} \text{ and } \mathcal{X}^{-1} = \begin{pmatrix} Y & V \\ V^T & * \end{pmatrix}$$

according to $\mathcal{A}$. Define

$$\mathcal{Y} = \begin{pmatrix} Y & I \\ V^T & 0 \end{pmatrix} \text{ and } \mathcal{Z} = \begin{pmatrix} I & 0 \\ X & U \end{pmatrix} \text{ to get } \mathcal{Y}^T \mathcal{X} = \mathcal{Z}. \tag{4.2.11}$$

Without loss of generality we can assume that the dimension of A_c is larger than that of A. Hence, U has more columns than rows, and we can perturb this block (since we work with strict inequalities) such that it has full row rank. Then $\mathcal{Z}$ has full row rank and, hence, $\mathcal{Y}$ has full column rank.

Due to $XY + UV^T = I$, we infer

$$\mathcal{Y}^T \mathcal{X} \mathcal{Y} = \begin{pmatrix} Y & I \\ I & X \end{pmatrix} = \boldsymbol{X}(v)$$

what leads to the first relation in (4.2.4). Let us now consider

$$\begin{pmatrix} \mathcal{Y} & 0 \\ 0 & I \end{pmatrix}^T \begin{pmatrix} \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix} \begin{pmatrix} \mathcal{Y} & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} \mathcal{Y}^T \mathcal{X}\mathcal{A}\mathcal{Y} & \mathcal{Y}^T \mathcal{X}\mathcal{B}_j \\ \mathcal{C}_j \mathcal{Y} & \mathcal{D}_j \end{pmatrix}.$$

Using (4.1.5), a very brief calculation (do it!) reveals that

$$\begin{pmatrix} \mathcal{Y}^T\mathcal{X}\mathcal{A}\mathcal{Y} & \mathcal{Y}^T\mathcal{X}\mathcal{B}_j \\ \mathcal{C}_j\mathcal{Y} & \mathcal{D}_j \end{pmatrix} = \begin{pmatrix} \mathcal{Z}\mathcal{A}\mathcal{Y} & \mathcal{Z}\mathcal{B}_j \\ \mathcal{C}_j\mathcal{Y} & \mathcal{D}_j \end{pmatrix} = \left(\begin{array}{cc|c} AY & A & B_j \\ 0 & XA & XB_j \\ \hline C_jY & C_j & D_j \end{array}\right) +$$

$$+\left(\begin{array}{cc} 0 & B \\ I & 0 \\ \hline 0 & E_j \end{array}\right)\left[\begin{pmatrix} U & XB \\ 0 & I \end{pmatrix}\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}\begin{pmatrix} V^T & 0 \\ CY & I \end{pmatrix} + \begin{pmatrix} XAY & 0 \\ 0 & 0 \end{pmatrix}\right]\left(\begin{array}{cc|c} I & 0 & 0 \\ 0 & C & F_j \end{array}\right).$$

If we introduce the new parameters $\begin{pmatrix} K & L \\ M & N \end{pmatrix}$ as in (4.2.10), we infer

$$\begin{pmatrix} \mathcal{Y}^T\mathcal{X}\mathcal{A}\mathcal{Y} & \mathcal{Y}^T\mathcal{X}\mathcal{B}_j \\ \mathcal{C}_j\mathcal{Y} & \mathcal{D}_j \end{pmatrix} =$$

$$= \left(\begin{array}{cc|c} AY & A & B_j \\ 0 & XA & XB_j \\ \hline C_jY & C_j & D_j \end{array}\right) + \left(\begin{array}{cc} 0 & B \\ I & 0 \\ \hline 0 & E_j \end{array}\right)\begin{pmatrix} K & L \\ M & N \end{pmatrix}\left(\begin{array}{cc|c} I & 0 & 0 \\ 0 & C & F_j \end{array}\right) =$$

$$= \left(\begin{array}{cc|c} AY + BM & A + BNC & B_j + BNF_j \\ K & AX + LC & XB_j + LF_j \\ \hline C_jY + E_jM & C_j + E_jNC & D_j + E_jNF_j \end{array}\right) = \begin{pmatrix} \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{pmatrix}.$$

Hence the relations (4.2.4) are valid. Since $\mathcal{Y}$ has full column rank, (4.2.1) implies (4.2.6), and by (4.2.4), (4.2.6) is identical to (4.2.7). This proves necessity.

To reverse the arguments we assume that v is a solution of (4.2.7). Due to $\boldsymbol{X}(v) > 0$, we infer that $I - XY$ is nonsingular. Hence we can factorize $I - XY = UV^T$ with square and nonsingular U, V. Then $\mathcal{Y}$ and $\mathcal{Z}$ defined in (4.2.11) are, as well, square and nonsingular. Hence we can choose $\mathcal{X}$, $\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}$ such that (4.2.10) hold true; this implies that, again, the relations (4.2.4) are valid. Therefore, (4.2.7) and (4.2.6) are identical. Since $\mathcal{Y}$ is nonsingular, a congruence transformation with $\mathcal{Y}^{-1}$ and $\mathrm{diag}(\mathcal{Y}^{-1}, I)$ leads from (4.2.6) back to (4.2.1) and the proof is finished. ■

We have obtained a general procedure for deriving from analysis inequalities the corresponding synthesis inequalities and for construction corresponding controllers as follows:

- Rewrite the analysis inequalities in the blocks $\mathcal{X}$, $\mathcal{X}\mathcal{A}$, $\mathcal{X}\mathcal{B}_j$, $\mathcal{C}_j$, $\mathcal{D}_j$ in order to be able to find a (formal) congruence transformation involving $\mathcal{Y}$ which leads to inequalities in the blocks $\mathcal{Y}^T\mathcal{X}\mathcal{Y}$, $\mathcal{Y}^T\mathcal{X}\mathcal{A}\mathcal{Y}$, $\mathcal{Y}^T\mathcal{X}\mathcal{B}_j$, $\mathcal{C}_j\mathcal{Y}$, $\mathcal{D}_j$.

- Perform the substitution (4.2.4) to arrive at matrix inequalities in the variables v.

- After having solved the synthesis inequalities for v, one factorizes $I - XY$ into non-singular blocks UV^T and solves the equations (4.2.10) to obtain the controller parameters A_c, B_c, C_c, D_c and a Lyapunov matrix $\mathcal{X}$ which render the analysis inequalities satisfied.

The power of this procedure lies in its simplicity and its generality. Virtually all controller design methods that are based on matrix inequality analysis results can be converted with ease into the corresponding synthesis result. In the subsequent section we will include an extensive discussion of how to apply this technique to the various analysis results that have been obtained in the present notes.

Remark 4.4 (controller order) In Theorem 4.3 we have not restricted the order of the controller. In proving necessity of the solvability of the synthesis inequalities, the size of A_c was arbitrary. The specific construction of a controller in proving sufficiency leads to an A_c that has *the same size* as A. Hence Theorem 4.3 also include the side result that controllers of order larger than that of the plant offer no advantage over controllers that have the same order as the plant. The story is very different in reduced order control: Then the intention is to include a constraint $\dim(A_c) \leq k$ for some k that is smaller than the dimension of A. It is not very difficult to derive the corresponding synthesis inequalities; however, they include *rank constraints* that are hard if not impossible to treat by current optimization techniques. We will only briefly comment on a concrete result later.

Remark 4.5 (strictly proper controllers) Note that the direct feed-through of the controller D_c is actually not transformed; we simply have $D_c = N$. If we intend to design a strictly proper controller (i.e. $D_c = 0$), we can just set $N = 0$ to arrive at the corresponding synthesis inequalities. The construction of the other controller parameters remains the same. Clearly, the same holds if one wishes to impose an arbitrary more refined structural constraint on the direct feed-through term as long as it can be expressed in terms of LMI's.

Remark 4.6 (numerical aspects) After having verified the solvability of the synthesis inequalities, we recommend to take some precautions to improve the conditioning of the calculations to reconstruct the controller out of the decision variable v. In particular, one should avoid that the parameters v get too large, and that $I - XY$ is close to singular what might render the controller computation ill-conditioned. We have observed good results with the following two-step procedure:

- Add to the feasibility inequalities the bounds

  $$\|X\| < \alpha, \ \|Y\| < \alpha, \ \left\| \begin{pmatrix} K & L \\ M & N \end{pmatrix} \right\| < \alpha$$

  as extra constraints and minimize α. Note that these bounds are equivalently rewritten in LMI form as

  $$X < \alpha I, \ Y < \alpha I, \ \left(\begin{array}{cc|cc} \alpha I & 0 & K & L \\ 0 & \alpha I & M & N \\ \hline K^T & M^T & \alpha I & 0 \\ L^T & N^T & 0 & \alpha I \end{array} \right) > 0.$$

  Hence they can be easily included in the feasibility test, and one can directly minimize α to compute the smallest bound α_*.

- In a second step, one adds to the feasibility inequalities and to the bounding inequalities for some enlarged but fixed $\alpha > \alpha_*$ the extra constraint

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

Of course, the resulting LMI system is feasible for $\beta = 1$. One can hence maximize β to obtain a supremal $\beta_* > 1$. The value β_* gives an indication of the conditioning of the controller reconstruction procedure. In fact, the extra inequality is equivalent to $X - \beta^2 Y^{-1} > 0$. Hence, maximizing β amounts to 'pushing X away from Y^{-1}'. Therefore, this step is expected to push the smallest singular value of $I - XY$ away from zero. The larger the smaller singular value of $I - XY$, the larger one can choose the smallest singular values of both U and V in the factorization $I - XY = UV^T$. This improves the conditioning of U and V, and renders the calculation of the controller parameters more reliable.

4.3 Other performance specifications

4.3.1 H_∞ Design

The optimal value of the H_∞ problem is defined as

$$\gamma_j^* = \inf_{A_c, B_c, C_c, D_c \text{ such that } \sigma(\mathcal{A}) \subset \mathbb{C}^-} \|\mathcal{T}_j\|_\infty.$$

Clearly, the number γ_j is larger than γ_j^* iff there exists a controller which renders

$$\sigma(\mathcal{A}) \subset \mathbb{C}^- \text{ and } \|\mathcal{T}_j\|_\infty < \gamma_j$$

satisfied. These two properties are equivalent to stability and quadratic performance for the index

$$P_j = \begin{pmatrix} Q_j & S_j \\ S_j^T & R_j \end{pmatrix} = \begin{pmatrix} -\gamma_j I & 0 \\ 0 & (\gamma_j I)^{-1} \end{pmatrix}.$$

The corresponding synthesis inequalities (4.2.7) are rewritten with Lemma 4.2 to

$$\boldsymbol{X}(v) > 0, \quad \begin{pmatrix} \boldsymbol{A}(v)^T + \boldsymbol{A}(v) & \boldsymbol{B}_j(v) & \boldsymbol{C}_j(v)^T \\ \boldsymbol{B}_j(v)^T & -\gamma_j I & \boldsymbol{D}_j(v)^T \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) & -\gamma_j I \end{pmatrix} < 0.$$

Note that the the optimal H_∞ value γ_j^* is then just given by the minimal γ_j for which these inequalities are feasible; one can directly compute γ_j^* by a standard LMI algorithm.

For the controller reconstruction, one should improve the conditioning (as described in the previous section) by an additional LMI optimization. We recommend not to perform this step with the optimal value γ_j^* itself but with a slightly increased value $\gamma_j > \gamma_j^*$. This is motivated by the observation that, at optimality, the matrix $\boldsymbol{X}(v)$ is often (but not always!) close to singular; then $I - XY$ is close to singular and it is expected to be difficult to render it better conditioned if γ_j is too close to the optimal value γ_j^*.

4.3.2 Positive real design

In this problem the goal is to test whether there exists a controller which renders the following two conditions satisfied:

$$\sigma(\mathcal{A}) \subset \mathbb{C}^-, \quad \mathcal{T}_j(i\omega)^* + \mathcal{T}_j(i\omega) > 0 \ \text{ for all } \ \omega \in \mathbb{R} \cup \{\infty\}.$$

This is equivalent to stability and quadratic performance for

$$P_j = \begin{pmatrix} Q_j & S_j \\ S_j^T & R_j \end{pmatrix} = \begin{pmatrix} 0 & -I \\ -I & 0 \end{pmatrix},$$

and the corresponding synthesis inequalities read as

$$\boldsymbol{X}(v) > 0, \quad \begin{pmatrix} \boldsymbol{A}(v)^T + \boldsymbol{A}(v) & \boldsymbol{B}_j(v) - \boldsymbol{C}_j(v)^T \\ \boldsymbol{B}_j(v)^T - \boldsymbol{C}_j(v) & -\boldsymbol{D}_j(v) - \boldsymbol{D}_j(v)^T \end{pmatrix} < 0.$$

4.3.3 H_2-problems

Let us define the linear functional

$$f_j(Z) := \text{trace}(Z).$$

Then we recall that $\mathcal{A}$ is stable and $\|\mathcal{T}_j\|_2 < \gamma_j$ iff there exists a symmetric $\mathcal{X}$ with

$$\mathcal{D}_j = 0, \quad \mathcal{X} > 0, \quad \begin{pmatrix} \mathcal{A}^T\mathcal{X} + \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \mathcal{B}_j^T\mathcal{X} & -\gamma_j I \end{pmatrix} < 0, \quad f_j(\mathcal{C}_j\mathcal{X}^{-1}\mathcal{C}_j^T) < \gamma_j. \tag{4.3.1}$$

The latter inequality is rendered affine in $\mathcal{X}$ and $\mathcal{C}_j$ by introducing the auxiliary variable (or slack variable) Z_j. Indeed, the analysis test is equivalent to

$$\mathcal{D}_j = 0, \quad \begin{pmatrix} \mathcal{A}^T\mathcal{X} + \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \mathcal{B}_j^T\mathcal{X} & -\gamma_j I \end{pmatrix} < 0, \quad \begin{pmatrix} \mathcal{X} & \mathcal{C}_j^T \\ \mathcal{C}_j & Z_j \end{pmatrix} > 0, \quad f_j(Z_j) < \gamma_j. \tag{4.3.2}$$

This version of the inequalities is suited to simply read-off the corresponding synthesis inequalities.

Corollary 4.7 *There exists a controller that renders* (4.3.2) *for some $\mathcal{X}$, Z_j satisfied iff there exist v and Z_j with*

$$\boldsymbol{D}_j(v) = 0, \quad \begin{pmatrix} \boldsymbol{A}(v)^T + \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{B}_j(v)^T & -\gamma_j I \end{pmatrix} < 0, \quad \begin{pmatrix} \boldsymbol{X}(v) & \boldsymbol{C}_j(v)^T \\ \boldsymbol{C}_j(v) & Z_j \end{pmatrix} > 0, \quad f_j(Z_j) < \gamma_j. \tag{4.3.3}$$

The proof of this statement and the controller construction are literally the same as for quadratic performance.

For the generalized H_2-norm $\|\mathcal{T}_j\|_{2g}$, we recall that $\mathcal{A}$ is stable and $\|\mathcal{T}_j\|_{2g} < \gamma_j$ iff

$$\mathcal{D}_j = 0, \quad \mathcal{X} > 0, \quad \begin{pmatrix} \mathcal{A}^T\mathcal{X} + \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \mathcal{B}_j^T\mathcal{X} & -\gamma_j I \end{pmatrix} < 0, \quad \mathcal{C}_j\mathcal{X}^{-1}\mathcal{C}_j^T < \gamma_j I.$$

These conditions are nothing but

$$\mathcal{D}_j = 0, \quad \begin{pmatrix} \mathcal{A}^T\mathcal{X} + \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \mathcal{B}_j^T\mathcal{X} & -\gamma_j I \end{pmatrix} < 0, \quad \begin{pmatrix} \mathcal{X} & \mathcal{C}_j^T \\ \mathcal{C}_j & \gamma_j I \end{pmatrix} > 0$$

and it is straightforward to derive the synthesis LMI's.

Note that the corresponding inequalities are equivalent to (4.3.3) for the function

$$f_j(Z) = Z.$$

In contrast to the genuine H_2-problem, there is no need for the extra variable Z_j to render the inequalities affine.

Remarks.

- If f assigns to Z its diagonal $\mathrm{diag}(z_1, \ldots, z_m)$ (where m is the dimension of Z), one characterizes a bound on the gain of $L_2 \ni w_j \to z_j \in L_\infty$ if equipping L_∞ with the norm $\|x\|_\infty := \operatorname{ess\,sup}_{t\geq 0} \max_k |x_k(t)|$ [26,29]. Note that the three concrete H_2-like analysis results for $f_j(Z) = \mathrm{trace}(Z)$, $f_j(Z) = Z$, $f_j(Z) = \mathrm{diag}(z_1, \ldots, z_m)$ are exact characterizations, and that the corresponding synthesis results do not involve any conservatism.
- In fact, Corollary 4.7 holds for any affine function f that maps symmetric matrices into symmetric matrices (of possibly different dimension) and that has the property $Z \geq 0 \Rightarrow f(Z) \geq 0$. Hence, Corollary 4.7 admits many other specializations.
- Similarly as in the H_∞ problem, we can directly minimize the bound γ_j to find the optimal H_2-value or the optimal generalized H_2-value that can be achieved by stabilizing controllers.
- We observe that it causes no trouble in our general procedure to derive the synthesis inequalities if the underlying analysis inequalities involve certain auxiliary parameters (such as Z_j) as extra decision variables.
- It is instructive to equivalently rewrite (4.3.2) as $\mathcal{X} > 0, \;\; Z_j > 0, \;\; f_j(Z_j) < \gamma_j$ and

$$\left(\begin{array}{cc} I & 0 \\ \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \hline 0 & I \end{array}\right)^T \left(\begin{array}{cc|c} 0 & I & 0 \\ I & 0 & 0 \\ \hline 0 & 0 & -\gamma_j I \end{array}\right) \left(\begin{array}{cc} I & 0 \\ \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \hline 0 & I \end{array}\right) < 0,$$

$$\left(\begin{array}{cc} I & 0 \\ \hline 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{array}\right)^T \left(\begin{array}{c|cc} -\mathcal{X} & 0 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & Z_j^{-1} \end{array}\right) \left(\begin{array}{cc} I & 0 \\ \hline 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{array}\right) \leq 0.$$

Note that the last inequality is non-strict and includes the algebraic constraint $\mathcal{D}_j = 0$. It can be equivalently replaced by

$$\begin{pmatrix} I \\ \mathcal{C}_j \end{pmatrix}^T \begin{pmatrix} -\mathcal{X} & 0 \\ 0 & Z_j^{-1} \end{pmatrix} \begin{pmatrix} I \\ \mathcal{C}_j \end{pmatrix} < 0, \quad \mathcal{D}_j = 0.$$

The synthesis relations then read as $\boldsymbol{X}(v) > 0, \quad Z_j > 0, \quad f_j(Z_j) < \gamma_j$ and

$$\left(\begin{array}{cc} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \end{array}\right)^T \left(\begin{array}{cc|c} 0 & I & 0 \\ I & 0 & 0 \\ \hline 0 & 0 & -\gamma_j I \end{array}\right) \left(\begin{array}{cc} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \end{array}\right) < 0, \tag{4.3.4}$$

$$\begin{pmatrix} I \\ \boldsymbol{C}_j(v) \end{pmatrix}^T \begin{pmatrix} -\boldsymbol{X}(v) & 0 \\ 0 & Z_j^{-1} \end{pmatrix} \begin{pmatrix} I \\ \boldsymbol{C}_j(v) \end{pmatrix} < 0, \quad \boldsymbol{D}_j(v) = 0. \tag{4.3.5}$$

The first inequality is affine in v, whereas the second one can be rendered affine in v and Z_j with Lemma 4.2.

4.3.4 Upper bound on peak-to-peak norm

The controller (4.1.2) renders $\mathcal{A}$ stable and the bound

$$\|w_j\|_\infty \le \gamma_j \|z_j\|_\infty \ \text{ for all } \ z_j \in L_\infty$$

satisfied **if** there exist a symmetric $\mathcal{X}$ and real parameters λ, μ with

$$\lambda > 0, \quad \begin{pmatrix} \mathcal{A}^T\mathcal{X} + \mathcal{X}\mathcal{A} + \lambda\mathcal{X} & \mathcal{X}\mathcal{B}_j \\ \mathcal{B}_j^T\mathcal{X} & 0 \end{pmatrix} + \begin{pmatrix} 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix}^T \begin{pmatrix} -\mu I & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix} < 0$$

$$\begin{pmatrix} 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix}^T \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\gamma_j} I \end{pmatrix} \begin{pmatrix} 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix} < \begin{pmatrix} \lambda\mathcal{X} & 0 \\ 0 & (\gamma_j - \mu)I \end{pmatrix}.$$

(Note that $\mathcal{X} > 0$ is built in. Where?) The inequalities are obviously equivalent to

$$\lambda > 0, \quad \begin{pmatrix} \mathcal{A}^T\mathcal{X} + \mathcal{X}\mathcal{A} + \lambda\mathcal{X} & \mathcal{X}\mathcal{B}_j \\ \mathcal{B}_j^T\mathcal{X} & -\mu I \end{pmatrix} < 0, \quad \begin{pmatrix} \lambda\mathcal{X} & 0 & \mathcal{C}_j^T \\ 0 & (\gamma_j - \mu)I & \mathcal{D}_j^T \\ \mathcal{C}_j & \mathcal{D}_j & \gamma_j I \end{pmatrix} > 0,$$

and the corresponding synthesis inequalities thus read as

$$\lambda > 0, \quad \begin{pmatrix} \boldsymbol{A}(v)^T + \boldsymbol{A}(v) + \lambda\boldsymbol{X}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{B}_j(v)^T & -\mu I \end{pmatrix} < 0, \quad \begin{pmatrix} \lambda\boldsymbol{X}(v) & 0 & \boldsymbol{C}_j(v)^T \\ 0 & (\gamma_j - \mu)I & \boldsymbol{D}_j(v)^T \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) & \gamma_j I \end{pmatrix} > 0.$$

If these inequalities are feasible, one can construct a stabilizing controller which bounds the peak-to-peak norm of $z_j = \mathcal{T}_j z_j$ by γ_j. We would like to stress that the converse of this statement is not true since the analysis result involves conservatism.

Note that the synthesis inequalities are formulated in terms of the variables v, λ, and μ; hence they are *non-linear* since $\lambda \boldsymbol{X}(v)$ depends quadratically on λ and v. This problem can be overcome as follows: For fixed $\lambda > 0$, test whether the resulting *linear* matrix inequalities are feasible; if yes, one can stop since the bound γ_j on the peak-to-peak norm has been assured; if the LMI's are infeasible, one has to pick another $\lambda > 0$ and repeat the test.

In practice, it might be advantageous to find the best possible upper bound on the peak-to-peak norm that can be assured with the present analysis result. This would lead to the problem of minimizing γ_j under the synthesis inequality constraints as follows: Perform a line-search over $\lambda > 0$ to minimize $\gamma_j^*(\lambda)$, the minimal value of γ_j if $\lambda > 0$ is held fixed; note that the calculation of $\gamma_j^*(\lambda)$ indeed amounts to solving a genuine LMI problem. The line-search leads to the best achievable upper bound

$$\gamma_j^u = \inf_{\lambda > 0} \gamma_j^*(\lambda).$$

To estimate the conservatism, let us recall that $\|\mathcal{T}_j\|_\infty$ is a lower bound on the peak-to-peak norm of $\mathcal{T}_j$. If we calculate the minimal achievable H_∞-norm, say γ_j^l, of $\mathcal{T}_j$, we know that the actual optimal peak-to-peak gain must be contained in the interval

$$[\gamma_j^l, \gamma_j^u].$$

If the length of this interval is small, we have a good estimate of the actual optimal peak-to-peak gain that is achievable by control, and if the interval is large, this estimate is poor.

4.4 Multi-objective and mixed controller design

In a realistic design problem one is usually not just confronted with a single-objective problem but one has to render various objectives satisfied. As a typical example, one might wish to keep the H_∞ norm of $z_1 = \mathcal{T}_1 w_1$ below a bound γ_1 to ensure robust stability against uncertainties entering as $w_1 = \Delta z_1$ where the stable mapping Δ has L_2-gain smaller than $1/\gamma_1$, and render, at the same time, the H_2-norm of $z_2 = \mathcal{T}_2 w_2$ as small as possible to ensure good performance measured in the H_2-norm (such as guaranteeing small asymptotic variance of z_j against white noise inputs w_j or small energy of the output z_j against pulses as inputs w_j.)

Such a problem would lead to minimizing γ_2 over all controllers which render

$$\sigma(\mathcal{A}) \subset \mathbb{C}^-, \quad \|\mathcal{T}_1\|_\infty < \gamma_1, \quad \|\mathcal{T}_2\|_2 < \gamma_2 \tag{4.4.1}$$

satisfied. This is a **multi-objective** H_2/H_∞ control problem with two performance specifications.

Note that it is often interesting to investigate the trade-off between the H_∞-norm and the H_2-norm constraint. For that purpose one plots the curve of optimal values if varying γ_1 in some interval $[\gamma_1^l, \gamma_1^u]$ where the lower bound γ_1^l could be taken close to the smallest achievable H_∞-norm of $\mathcal{T}_1$. Note that the optimal value will be non-increasing if increasing γ_1. The actual curve will provide insight in how far one can improve performance by giving up robustness. In practice, it might be numerically advantageous to give up the hard constraints and proceed, alternatively, as follows: For fixed real weights α_1 and α_2, minimize

$$\alpha_1\gamma_1 + \alpha_2\gamma_2$$

over all controllers that satisfy (4.4.1). The larger α_j, the more weight is put on penalizing large values of γ_j, the more the optimization procedure is expected to reduce the corresponding bound γ_j.

Multi-objective control problems as formulated here are hard to solve. Let us briefly sketch one line of approach. The Youla parameterization [17] reveals that the set of all $\mathcal{T}_j$ that can be obtained by internally stabilizing controllers can be parameterized as

$$T_1^j + T_2^j Q T_3^j \ \text{ with } Q \text{ varying freely in } RH_\infty^{p\times q}.$$

Here T_1^j, T_2^j, T_3^j are real-rational proper and stable transfer matrices which can be easily computed in terms of the system description (4.1.1) and an arbitrary stabilizing controller. Recall also that $RH_\infty^{p\times q}$ denotes the algebra of real-rational proper and stable transfer matrices of dimension $p \times q$. With this re-parameterization, the multi-objective control problem then amounts to finding a $Q \in RH_\infty^{p\times q}$ that minimizes γ_2 under the constraints

$$\|T_1^1 + T_2^1 Q T_3^1\|_\infty < \gamma_1, \quad \|T_1^2 + T_2^2 Q T_3^2\|_2 < \gamma_2. \tag{4.4.2}$$

After this re-formulation, we are hence faced with a convex optimization problem in the parameter Q which varies in the infinite-dimensional space RH_∞. A pretty standard Ritz-Galerkin approximation scheme leads to finite-dimensional problems. In fact, consider for a fixed real parameter $a > 0$ the sequence of finite-dimensional subspaces

$$\mathcal{S}_\nu := \left\{ Q_0 + Q_1\frac{s-a}{s+a} + Q_2\frac{(s-a)^2}{(s+a)^2} + \cdots + Q_\nu\frac{(s-a)^\nu}{(s+a)^\nu} : \ Q_0,\ldots,Q_\nu \in \mathbb{R}^{p\times q} \right\}$$

of the space $RH_\infty^{p\times q}$. Let us now denote the infimum of all γ_2 satisfying the constraint (4.4.2) for $Q \in RH_\infty^{p\times q}$ by γ_2^*, and that for $Q \in \mathcal{S}_\nu$ by $\gamma_2(\nu)$. Since $\mathcal{S}_\nu \subset RH_\infty^{p\times q}$, we clearly have

$$\gamma_2^* \le \gamma_2(\nu+1) \le \gamma_2(\nu) \ \text{ for all } \ \nu = 0,1,2\ldots.$$

Hence solving the optimization problems for increasing ν leads to a non-increasing sequence of values $\gamma(\nu)$ that are all upper bounds on the actual optimum γ_2^*. If we now note that any element of Q can be approximated in the H_∞-norm with arbitrary accuracy by an element in $\mathcal{S}_\nu$ if ν is chosen sufficiently large, it is not surprising that $\gamma_2(\nu)$ actually converges to γ_2^* for $\nu \to \infty$. To be more precise, we need to assume that the strict constraint $\|T_1^1 + T_2^1 Q T_3^1\|_\infty < \gamma_1$ is feasible for $Q \in \mathcal{S}_\nu$ and some ν, and that T_1^1 and T_2^2 or T_2^3 are strictly proper such that $\|T_1^2 + T_2^2 Q T_3^2\|_2$ is finite for all $Q \in RH_\infty^{p\times q}$. Then it is not difficult to show that

$$\lim_{\nu\to\infty} \gamma_2(\nu) = \gamma_2^*.$$

Finally, we observe that computing $\gamma_2(\nu)$ is in fact an LMI problem. For more information on this and related problems the reader is referred to [7,30,38].

We observe that the approach that is sketched above suffers from two severe disadvantages: First, if improving the approximation accuracy by letting ν grow, the size of the LMI's and the number of variables that are involved grow drastically what renders the corresponding computations slow. Second, increasing ν amounts to a potential increase of the McMillan degree of $Q \in \mathcal{S}_\nu$ what leads to controllers whose McMillan degree cannot be bounded a priori.

In view of these difficulties, it has been proposed to replace the multi-objective control problem by the mixed control problem. To prepare its definition, recall that the conditions (4.4.1) are guaranteed by the existence of symmetric matrices $\mathcal{X}_1$, $\mathcal{X}_2$, Z_2 satisfying

$$\mathcal{X}_1 > 0, \quad \begin{pmatrix} \mathcal{A}^T\mathcal{X}_1 + \mathcal{X}_1\mathcal{A} & \mathcal{X}_1\mathcal{B}_1 & \mathcal{C}_1^T \\ \mathcal{X}_1\mathcal{B}_1 & -\gamma_1 I & \mathcal{D}_1^T \\ \mathcal{C}_1 & \mathcal{D}_1 & -\gamma_1 I \end{pmatrix} < 0$$

$$\mathcal{D}_2 = 0, \quad \begin{pmatrix} \mathcal{A}^T\mathcal{X}_2 + \mathcal{X}_2\mathcal{A} & \mathcal{X}_2\mathcal{B}_2 \\ \mathcal{B}_2^T\mathcal{X}_2 & -\gamma_2 I \end{pmatrix} < 0, \quad \begin{pmatrix} \mathcal{X}_2 & \mathcal{C}_2^T \\ \mathcal{C}_2 & Z_2 \end{pmatrix} > 0, \quad \text{trace}(Z_2) < \gamma_2.$$

If trying to apply the general procedure to derive the synthesis inequalities, there is some trouble since the controller parameter transformation depends on the closed-loop Lyapunov matrix; here two such matrices $\mathcal{X}_1$, $\mathcal{X}_2$ do appear such that the technique breaks down. This observation itself motivates a remedy: Just force the two Lyapunov matrices to be equal. This certainly introduces conservatism that is, in general, hard to quantify. On the positive side, if one can find a common matrix

$$\mathcal{X} = \mathcal{X}_1 = \mathcal{X}_2$$

that satisfies the analysis relations, we can still guarantee (4.4.1) to hold. However, the converse is not true, since (4.4.1) does not imply the existence of common Lyapunov matrix to satisfy the above inequalities.

This discussion leads to the definition of the **mixed** H_2/H_∞ control problem: Minimize γ_2 subject to the existence of $\mathcal{X}$, Z_2 satisfying

$$\begin{pmatrix} \mathcal{A}^T\mathcal{X} + \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_1 & \mathcal{C}_1^T \\ \mathcal{B}_1^T\mathcal{X} & -\gamma_1 I & \mathcal{D}_1^T \\ \mathcal{C}_1 & \mathcal{D}_1 & -\gamma_1 I \end{pmatrix} < 0$$

$$\mathcal{D}_2 = 0, \quad \begin{pmatrix} \mathcal{A}^T\mathcal{X} + \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_2 \\ \mathcal{B}_2^T\mathcal{X} & -\gamma_2 I \end{pmatrix} < 0, \quad \begin{pmatrix} \mathcal{X} & \mathcal{C}_2^T \\ \mathcal{C}_2 & Z_2 \end{pmatrix} > 0, \quad \text{trace}(Z_2) < \gamma_2.$$

This problem is amenable to our general procedure. One proves as before that the corresponding synthesis LMI's are

$$\begin{pmatrix} \boldsymbol{A}(v)^T + \boldsymbol{A}(v) & \boldsymbol{B}_1(v) & \boldsymbol{C}_1(v)^T \\ \boldsymbol{B}_1(v)^T & -\gamma_1 I & \boldsymbol{D}_1(v)^T \\ \boldsymbol{C}_1(v) & \boldsymbol{D}_1(v) & -\gamma_1 I \end{pmatrix} < 0$$

$$\boldsymbol{D}_2(v) = 0, \quad \begin{pmatrix} \boldsymbol{A}(v)^T + \boldsymbol{A}(v) & \boldsymbol{B}_2(v) \\ \boldsymbol{B}_2(v)^T & -\gamma_2 I \end{pmatrix} < 0, \quad \begin{pmatrix} \boldsymbol{X}(v) & \boldsymbol{C}_2(v)^T \\ \boldsymbol{C}_2(v) & Z_2 \end{pmatrix} > 0, \quad \text{trace}(Z_2) < \gamma_2,$$

and the controller construction remains unchanged.

Let us conclude this section with some important remarks.

- After having solved the synthesis inequalities corresponding to the mixed problem for v and Z_2, one can construct a controller which satisfies (4.4.1) and which has a McMillan degree (size of A_c) that is not larger than (equal to) the size of A.
- For the controller resulting from mixed synthesis one can perform an analysis with different Lyapunov matrices $\mathcal{X}_1$ and $\mathcal{X}_2$ without any conservatism. In general, the actual H_∞-norm of $\mathcal{T}_1$ will be strictly smaller than γ_1, and the H_2-norm will be strictly smaller than the optimal value obtained from solving the mixed problem. Judging a mixed controller should, hence, rather be based on an additional non-conservative and direct analysis.
- Performing synthesis by searching for a common Lyapunov matrix introduces conservatism. Little is known about how to estimate this conservatism a priori. However, the optimal value of the mixed problem is always an upper bound of the optimal value of the actual multi-objective problem.
- Starting from a mixed controller, it has been suggested in [33, 34] how to compute sequences of upper and lower bounds, on the basis of solving LMI problems, that approach the actual optimal value. This allows to provide an a posteriori estimate of the conservatism that is introduced by setting $\mathcal{X}_1$ equal to $\mathcal{X}_2$.
- If starting from different versions of the analysis inequalities (e.g. through scaling the Lyapunov matrix), the artificial constraint $\mathcal{X}_1 = \mathcal{X}_2$ might lead to a different mixed control problem. Therefore, it is recommended to choose those analysis tests that are expected to lead to Lyapunov matrices which are close to each other. However, there is no general rule how to guarantee this property.
- In view of the previous remark, let us sketch one possibility to reduce the conservatism in mixed design. If we multiply the analysis inequalities for stability of $\mathcal{A}$ and for $\|\mathcal{T}_1\|_\infty < \gamma_1$ by an arbitrary real parameter $\alpha > 0$, we obtain

  $$\alpha\mathcal{X}_1 > 0, \quad \begin{pmatrix} \mathcal{A}^T(\alpha\mathcal{X}_1) + (\alpha\mathcal{X}_1)\mathcal{A} & (\alpha\mathcal{X}_1)\mathcal{B}_1 & \alpha\mathcal{C}_1^T \\ \mathcal{B}_1^T(\alpha\mathcal{X}_1) & -\alpha\gamma_1 I & \alpha\mathcal{D}_1^T \\ \alpha\mathcal{C}_1 & \alpha\mathcal{D}_1 & -\alpha\gamma_1 I \end{pmatrix} < 0.$$

  If we multiply the last row and the last column of the second inequality with $\frac{1}{\alpha}$ (what is a congruence transformation) and if we introduce $\mathcal{Y}_1 := \alpha\mathcal{X}_1$, we arrive at the following equivalent version of the analysis inequality for the H_∞-norm constraint:

  $$\mathcal{Y}_1 > 0, \quad \begin{pmatrix} \mathcal{A}^T\mathcal{Y}_1 + \mathcal{Y}_1\mathcal{A} & \mathcal{Y}_1\mathcal{B}_1 & \mathcal{C}_1^T \\ \mathcal{B}_1^T\mathcal{Y}_1 & -\gamma_1\alpha I & \mathcal{D}_1^T \\ \mathcal{C}_1 & \mathcal{D}_1 & -\gamma_1/\alpha I \end{pmatrix} < 0.$$

  Performing mixed synthesis with this analysis inequality leads to optimal values of the mixed H_2/H_∞ problem that depend on α. Each of these values form an upper bound on the actual optimal value of the multi-objective problem such that the best bound is found by performing a line-search over $\alpha > 0$.

- Contrary to previous approaches to the mixed problem, the one presented here does not require identical input- or output-signals of the H_∞ or H_2 channel. In view of their interpretation (uncertainty for H_∞ and performance for H_2), such a restriction is, in general, very unnatural. However, due to this flexibility, it is even more crucial to suitably scale the Lyapunov matrices.

- We can incorporate with ease various other performance or robustness specifications (formulated in terms of linear matrix inequalities) on other channels. Under the constraint of using for all desired specifications the same Lyapunov matrix, the design of a mixed controller is straightforward. Hence, one could conceivably consider a mixture of H_∞, H_2, generalized H_2, and peak-to-peak upper bound requirements on more than one channel. In its flexibility and generality, this approach is unique; however, one should never forget the conservatism that is involved.

- Using the same Lyapunov function might appear less restrictive if viewing the resulting procedure as a *Lyapunov shaping technique*. Indeed, one can start with the most important specification to be imposed on the controller. This amounts to solving a single-objective problem without conservatism. Then one keeps the already achieved property as a constraint and systematically imposes other specifications on other channels of the system to exploit possible additional freedom that is left in designing the controller. Hence, the Lyapunov function is shaped to realize additional specifications.

- Finally, constraints that are not necessarily related to input- output-specifications can be incorporated as well. As a nice example we mention the possibility to place the eigenvalues of $\mathscr{A}$ into an arbitrary LMI region $\{z : \; Q + Pz + P^T\bar{z} < 0\}$. For that purpose one just has to include

$$\begin{pmatrix} p_{11}\boldsymbol{X}(v) + q_{11}\boldsymbol{A}(v) + q_{11}\boldsymbol{A}(v)^T & \dots & p_{1k}\boldsymbol{X}(v) + q_{1k}\boldsymbol{A}(v) + q_{k1}\boldsymbol{A}(v)^T \\ \vdots & \ddots & \vdots \\ p_{k1}\boldsymbol{X}(v) + q_{k1}\boldsymbol{A}(v) + q_{1k}\boldsymbol{A}(v)^T & \dots & p_{kk}\boldsymbol{X}(v) + q_{kk}\boldsymbol{A}(v) + q_{kk}\boldsymbol{A}(v)^T \end{pmatrix} < 0$$

  in the set of synthesis LMI (see Chapter 2).

4.5 Elimination of parameters

The general procedure described in Section 4.2 leads to synthesis inequalities in the variables K, L, M, N and X, Y as well as some auxiliary variables. For specific problems it is often possible to eliminate some of these variables in order to reduce the computation time. For example, since K has the same size as A, eliminating K for a system with McMillan degree 20 would save 400 variables. In view of the fact that, in our experience, present-day solvers are practical for solving problems up to about one thousand variables, parameter elimination might be of paramount importance to be able to solve realistic design problems.

In general, one cannot eliminate any variable that appears in at least two synthesis inequalities. Hence, in mixed design problems, parameter elimination is typically only possible under specific circumstances. In single-objective design problems one has to distinguish various information structures. In output-feedback design problems, it is in general not possible to eliminate X, Y but it might be possible to eliminate some of the variables K, L, M, N if they only appear in one inequality. For example, in quadratic performance problems one can eliminate all the variables K, L, M, N. In state-feedback design, one can typically eliminate in addition X, and for estimation problems one can eliminate Y.

To understand which variables can be eliminated and how this is performed, we turn to a discussion of two topics that will be of relevance, namely the dualization of matrix inequalities and explicit solvability tests for specifically structured LMI's [9, 31].

4.5.1 Dualization

The synthesis inequalities for quadratic performance can be written in the form (4.2.9). The second inequality has the structure

$$\begin{pmatrix} I \\ M \end{pmatrix}^T \begin{pmatrix} Q & S \\ S' & R \end{pmatrix} \begin{pmatrix} I \\ M \end{pmatrix} < 0 \text{ and } R \geq 0. \tag{4.5.1}$$

Let us re-formulate these conditions in geometric terms. For that purpose we abbreviate

$$P = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \in \mathbb{R}^{(k+l)\times(k+l)}$$

and observe that (4.5.1) is nothing but

$$P < 0 \text{ on } \operatorname{im}\begin{pmatrix} I \\ M \end{pmatrix} \quad \text{and} \quad P \geq 0 \text{ on } \operatorname{im}\begin{pmatrix} 0 \\ I \end{pmatrix}.$$

Since the direct sum of im $\begin{pmatrix} I \\ M \end{pmatrix}$ and im $\begin{pmatrix} 0 \\ I \end{pmatrix}$ spans the whole $\mathbb{R}^{(k+l)\times(k+l)}$, we can apply the following dualization lemma if P is non-singular.

Lemma 4.8 (Dualization Lemma) *Let P be a non-singular symmetric matrix in $\mathbb{R}^{n\times n}$, and let $\mathcal{U}$, $\mathcal{V}$ be two complementary subspaces whose sum equals $\mathbb{R}^n$. Then*

$$x^T P x < 0 \text{ for all } x \in \mathcal{U}\setminus\{0\} \quad \text{and} \quad x^T P x \geq 0 \text{ for all } x \in \mathcal{V} \tag{4.5.2}$$

is equivalent to

$$x^T P^{-1} x > 0 \text{ for all } x \in \mathcal{U}^\perp \setminus\{0\} \quad \text{and} \quad x^T P^{-1} x \leq 0 \text{ for all } x \in \mathcal{V}^\perp. \tag{4.5.3}$$

Proof. Since $\mathcal{U} \oplus \mathcal{V} = \mathbb{R}^n$ is equivalent to $\mathcal{U}^\perp \oplus \mathcal{V}^\perp = \mathbb{R}^n$, it suffices to prove that (4.5.2) implies (4.5.3); the converse implication follows by symmetry. Let us assume that (4.5.2) is true.

Moreover, let us assume that $\mathcal{U}$ and $\mathcal{V}$ have dimension k and l respectively. We infer from (4.5.2) that P has at least k negative eigenvalues and at least l non-negative eigenvalues. Since $k + l = n$ and since P is non-singular, we infer that P has exactly k negative and l positive eigenvalues. We first prove that P^{-1} is positive definite on $\mathcal{U}^\perp$. We assume, to the contrary, that there exists a vector $y \in \mathcal{U}^\perp \setminus \{0\}$ with $y^T P^{-1} y \geq 0$. Define the non-zero vector $z = P^{-1} y$. Then z is not contained in $\mathcal{U}$ since, otherwise, we would conclude from (4.5.2) on the one hand $z^T P z < 0$, and on the other hand $z \perp y = Pz$ what implies $z^T P z = 0$. Therefore, the space $\mathcal{U}_e := \mathrm{span}(z) + \mathcal{U}$ has dimension $k + 1$. Moreover, P is positive semi-definite on this space: for any $x \in \mathcal{U}$ we have

$$(z + x)^T P (z + x) = y^T P^{-1} y + y^T x + x^T y + x^T P x = y^T P^{-1} y + x^T P x \geq 0.$$

This implies that P has at least $k + 1$ non-negative eigenvalues, a contradiction to the already established fact that P has exactly k positive eigenvalues and that 0 is not an eigenvalue of P.

Let us now prove that P^{-1} is negative semi-definite on $\mathcal{V}^\perp$. For that purpose we just observe that $P + \epsilon I$ satisfies

$$x^T (P + \epsilon I) x < 0 \ \text{ for all } \ x \in \mathcal{U} \setminus \{0\} \quad \text{and} \quad x^T (P + \epsilon I) x > 0 \ \text{ for all } \ x \in \mathcal{V} \setminus \{0\}$$

for all small $\epsilon > 0$. Due to what has been already proved, this implies

$$x^T (P + \epsilon I)^{-1} x > 0 \text{ for all } x \in \mathcal{U}^\perp \setminus \{0\} \ \text{ and } \ x^T (P + \epsilon I)^{-1} x < 0 \text{ for all } x \in \mathcal{V}^\perp \setminus \{0\}$$

for all small ϵ. Since P is non-singular, $(P + \epsilon I)^{-1}$ converges to P^{-1} for $\epsilon \to 0$. After taking the limit, we end up with

$$x^T P^{-1} x \geq 0 \ \text{ for all } \ x \in \mathcal{U}^\perp \setminus \{0\} \quad \text{and} \quad x^T P^{-1} x \leq 0 \ \text{ for all } \ x \in \mathcal{V}^\perp \setminus \{0\}.$$

Since we already know that the first inequality must be strict, the proof is finished. ∎

Let us hence introduce

$$P^{-1} = \begin{pmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{pmatrix} \in \mathbb{R}^{(k+l)\times(k+l)}$$

and observe that

$$\mathrm{im} \begin{pmatrix} I \\ M \end{pmatrix}^\perp = \ker \begin{pmatrix} I & M^T \end{pmatrix} = \mathrm{im} \begin{pmatrix} -M^T \\ I \end{pmatrix} \ \text{ as well as } \ \mathrm{im} \begin{pmatrix} 0 \\ I \end{pmatrix}^\perp = \mathrm{im} \begin{pmatrix} I \\ 0 \end{pmatrix}.$$

Hence Lemma 4.8 implies that (4.5.1) is equivalent to

$$\begin{pmatrix} -M^T \\ I \end{pmatrix}^T \begin{pmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{pmatrix} \begin{pmatrix} -M^T \\ I \end{pmatrix} > 0 \ \text{ and } \ \tilde{Q} \leq 0. \tag{4.5.4}$$

As an immediate consequence, we arrive at the following dual version of the quadratic performance synthesis inequalities.

Corollary 4.9 *Let* $P_j := \begin{pmatrix} Q_j & S_j \\ S_j^T & R_j \end{pmatrix}$ *be non-singular, and abbreviate* $P_j^{-1} := \begin{pmatrix} \tilde{Q}_j & \tilde{S}_j \\ \tilde{S}_j^T & \tilde{R}_j \end{pmatrix}$. *Then*

$$\begin{pmatrix} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{pmatrix}^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & Q_j & S_j \\ 0 & 0 & S_j^T & R_j \end{array}\right) \begin{pmatrix} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{pmatrix} < 0, \quad R_j \geq 0$$

is equivalent to

$$\begin{pmatrix} -\boldsymbol{A}(v)^T & -\boldsymbol{C}(v)^T \\ I & 0 \\ \hline -\boldsymbol{B}(v)^T & -\boldsymbol{D}(v)^T \\ 0 & I \end{pmatrix}^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & \tilde{Q}_j & \tilde{S}_j \\ 0 & 0 & \tilde{S}_j^T & \tilde{R}_j \end{array}\right) \begin{pmatrix} -\boldsymbol{A}(v)^T & -\boldsymbol{C}(v)^T \\ I & 0 \\ \hline -\boldsymbol{B}(v)^T & -\boldsymbol{D}(v)^T \\ 0 & I \end{pmatrix} > 0, \quad \tilde{Q}_j \leq 0.$$

Remark. Any non-singular performance index $P_j = \begin{pmatrix} Q_j & S_j \\ S_j^T & R_j \end{pmatrix}$ can be inverted to $P_j^{-1} = \begin{pmatrix} \tilde{Q}_j & \tilde{S}_j \\ \tilde{S}_j^T & \tilde{R}_j \end{pmatrix}$. Recall that we required P_j to satisfy $R_j \geq 0$ since, otherwise, the synthesis inequalities may not be convex. The above discussion reveals that any non-singular performance index has to satisfy as well $\tilde{Q}_j \leq 0$ since, otherwise, we are sure that the synthesis inequalities are not feasible. We stress this point since, in general, $R_j \geq 0$ does not imply $\tilde{Q}_j \leq 0$. (Take e.g. $P_j > 0$ such that $P_j^{-1} > 0$.)

Similarly, we can dualize the H_2-type synthesis inequalities as formulated in (4.3.4)-(4.3.5).

Corollary 4.10 *For* $\gamma_j > 0$,

$$\begin{pmatrix} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \end{pmatrix}^T \left(\begin{array}{cc|c} 0 & I & 0 \\ I & 0 & 0 \\ \hline 0 & 0 & -\gamma_j I \end{array}\right) \begin{pmatrix} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \end{pmatrix} < 0$$

if and only if

$$\begin{pmatrix} -\boldsymbol{A}(v)^T \\ -\boldsymbol{B}_j(v)^T \\ \hline I \end{pmatrix}^T \left(\begin{array}{cc|c} 0 & I & 0 \\ I & 0 & 0 \\ \hline 0 & 0 & -\frac{1}{\gamma_j} I \end{array}\right) \begin{pmatrix} -\boldsymbol{A}(v)^T \\ -\boldsymbol{B}_j(v)^T \\ \hline I \end{pmatrix} > 0.$$

For $\boldsymbol{X}(v) > 0$ *and* $Z_j > 0$,

$$\begin{pmatrix} I \\ \boldsymbol{C}_j(v) \end{pmatrix}^T \begin{pmatrix} -\boldsymbol{X}(v) & 0 \\ 0 & Z_j^{-1} \end{pmatrix} \begin{pmatrix} I \\ \boldsymbol{C}_j(v) \end{pmatrix} < 0$$

if and only if

$$\begin{pmatrix} -\boldsymbol{C}_j(v)^T \\ I \end{pmatrix}^T \begin{pmatrix} -\boldsymbol{X}(v)^{-1} & 0 \\ 0 & Z_j \end{pmatrix} \begin{pmatrix} -\boldsymbol{C}_j(v)^T \\ I \end{pmatrix} > 0.$$

Again, Lemma 4.2 allows to render the first and the second dual inequalities affine in γ_j and $\boldsymbol{X}(v)$ respectively.

4.5.2 Special linear matrix inequalities

Let us now turn to specific linear matrix inequalities for which one can easily derive explicit solvability tests.

We start by a trivial example that is cited for later reference.

Lemma 4.11 *The inequality*

$$\begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22}+X & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} < 0$$

in the symmetric unknown X has a solution if and only if

$$\begin{pmatrix} P_{11} & P_{13} \\ P_{31} & P_{33} \end{pmatrix} < 0.$$

Proof. The direction 'only if' is obvious by cancelling the second row/column. To prove the converse implication, we just need to observe that any X with

$$X < -P_{22} + \begin{pmatrix} P_{11} & P_{13} \end{pmatrix} \begin{pmatrix} P_{11} & P_{13} \\ P_{31} & P_{33} \end{pmatrix}^{-1} \begin{pmatrix} P_{12} \\ P_{32} \end{pmatrix} < 0$$

(such as $X = -\alpha I$ for sufficiently large $\alpha > 0$) is a solution (Schur). ■

Remark. This result extends to finding a common solution to a whole system of LMI's, due to the following simple fact: For finitely matrices $Q_1, \ldots, Q_m$, there exists an X with $X < Q_j$, $j = 1, \ldots, m$.

The first of three more advanced results in this vain is just a simple consequence of a Schur complement argument and it can be viewed as a powerful variant of what is often called the technique of 'completing the squares'.

Lemma 4.12 (Projection Lemma) *Let P be a symmetric matrix partitioned into three rows/columns and consider the LMI*

$$\begin{pmatrix} P_{11} & P_{12}+X^T & P_{13} \\ P_{21}+X & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} < 0 \tag{4.5.5}$$

in the unstructured matrix X. There exists a solution X of this LMI iff

$$\begin{pmatrix} P_{11} & P_{13} \\ P_{31} & P_{33} \end{pmatrix} < 0 \text{ and } \begin{pmatrix} P_{22} & P_{23} \\ P_{32} & P_{33} \end{pmatrix} < 0. \tag{4.5.6}$$

If (4.5.6) *hold, one particular solution is given by*

$$X = P_{32}^T P_{33}^{-1} P_{31} - P_{21}. \tag{4.5.7}$$

Proof. If (4.5.5) has a solution then (4.5.6) just follow from (4.5.5) by canceling the first or second block row/column.

Now suppose that (4.5.6) holds what implies $P_{33} < 0$. We observe that (4.5.5) is equivalent to (Schur complement)

$$\begin{pmatrix} P_{11} & P_{12} + X^T \\ P_{21} + X & P_{22} \end{pmatrix} - \begin{pmatrix} P_{13} \\ P_{23} \end{pmatrix} P_{33}^{-1} \begin{pmatrix} P_{31} & P_{32} \end{pmatrix} < 0.$$

Due to (4.5.6), the diagonal blocks are negative definite. X defined in (4.5.7) just renders the off-diagonal block zero such that it is a solution of the latter matrix inequality. ■

An even more powerful generalization is the so-called projection lemma.

Lemma 4.13 (Projection Lemma) *For arbitrary A, B and a symmetric P, the LMI*

$$A^T X B + B^T X^T A + P < 0 \tag{4.5.8}$$

in the unstructured X has a solution if and only if

$$Ax = 0 \text{ or } Bx = 0 \text{ imply } x^T P x < 0 \text{ or } x = 0. \tag{4.5.9}$$

If $A_\perp$ and $B_\perp$ denote arbitrary matrices whose columns form a basis of $\ker(A)$ *and* $\ker(B)$ *respectively,* (4.5.9) *is equivalent to*

$$A_\perp^T P A_\perp < 0 \text{ and } B_\perp^T P B_\perp < 0. \tag{4.5.10}$$

We give a full proof of the Projection Lemma since it provides a scheme for constructing a solution X if it exists. It also reveals that, in suitable coordinates, Lemma 4.13 reduces to Lemma 4.12 if the kernels of A and B together span the whole space.

Proof. The proof of ‘only if’ is trivial. Indeed, let us assume that there exists some X with $A^T X B + B^T X^T A + P < 0$. Then $Ax = 0$ or $Bx = 0$ with $x \neq 0$ imply the desired inequality $0 > x^T(A^T X B + B^T X^T A + P)x = x^T P x$.

For proving ‘if’, let $S = (S_1\ S_2\ S_3\ S_4)$ be a nonsingular matrix such that the columns of S_3 span $\ker(A) \cap \ker(B)$, those of $(S_1\ S_3)$ span $\ker(A)$, and those of $(S_2\ S_3)$ span $\ker(B)$. Instead of (4.5.8), we consider the equivalent inequality $S^T(4.5.8)S < 0$ which reads as

$$(AS)^T X (BS) + (BS)^T X^T (AS) + S^T P S < 0. \tag{4.5.11}$$

Now note that AS and BS have the structure $(0\ A_2\ 0\ A_4)$ and $(B_1\ 0\ 0\ B_4)$ where $(A_2\ A_4)$ and $(B_1\ B_4)$ have full column rank respectively. The rank properties imply that the equation

$$(AS)^T X(BS) = \begin{pmatrix} 0 \\ A_2^T \\ 0 \\ A_4^T \end{pmatrix} X \begin{pmatrix} B_1 & 0 & 0 & B_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ Z_{21} & 0 & 0 & Z_{24} \\ 0 & 0 & 0 & 0 \\ Z_{41} & 0 & 0 & Z_{44} \end{pmatrix}$$

has a solution X for arbitrary Z_{21}, Z_{24}, Z_{41}, Z_{44}. With $Q := S^T PS$ partitioned accordingly, (4.5.11) hence reads as

$$\left(\begin{array}{ccc|c} Q_{11} & Q_{12} + Z_{21}^T & Q_{13} & Q_{14} + Z_{41}^T \\ Q_{21} + Z_{21} & Q_{22} & Q_{23} & Q_{24} + Z_{24} \\ Q_{31} & Q_{32} & Q_{33} & Q_{34} \\ \hline Q_{41} + Z_{41} & Q_{42} + Z_{24}^T & Q_{43} & Q_{44} + Z_{44} + Z_{44}^T \end{array}\right) < 0 \tag{4.5.12}$$

with *free* blocks Z_{21}, Z_{24}, Z_{41}, Z_{44}. Since

$$\ker(AS) = \mathrm{im} \begin{pmatrix} I & 0 \\ 0 & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \ker(BS) = \mathrm{im} \begin{pmatrix} 0 & 0 \\ I & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix},$$

the hypothesis (4.5.9) just amounts to the conditions

$$\begin{pmatrix} Q_{11} & Q_{13} \\ Q_{31} & Q_{33} \end{pmatrix} < 0 \ \text{ and } \ \begin{pmatrix} Q_{22} & Q_{23} \\ Q_{32} & Q_{33} \end{pmatrix} < 0.$$

By Lemma 4.12, we can hence find a matrix Z_{21} which renders the marked 3×3 block in (4.5.12) negative definite. The blocks Z_{41} and Z_{24} can be taken arbitrary. After having fixed Z_{21}, Z_{41}, Z_{24}, we can choose Z_{44} according to Lemma 4.11 such that the whole matrix on the left-hand side of (4.5.12) is negative definite. ■

Remark. We can, of course, replace $<$ everywhere by $>$. It is important to recall that the unknown X is unstructured. If one requires X to have a certain structure (such as being symmetric), the tests, if existing at all, are much more complicated. There is, however, a generally valid extension of the Projection Lemma to block-triangular unknowns X [28]. Note that the results do not hold true as formulated if just replacing the strict inequalities by non-strict inequalities (as it is sometimes erroneously claimed in the literature)! Again, it is possible to provide a full generalization of the Projection Lemma to non-strict inequalities.

Let

$$P = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \text{ with } R \geq 0 \quad \text{have the inverse} \quad P^{-1} = \begin{pmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{pmatrix} \text{ with } \tilde{Q} \leq 0 \tag{4.5.13}$$

and let us finally consider the quadratic inequality

$$\begin{pmatrix} I \\ A^T XB + C \end{pmatrix}^T P \begin{pmatrix} I \\ A^T XB + C \end{pmatrix} < 0 \tag{4.5.14}$$

in the unstructured unknown X. According to Lemma 4.8, we can dualize this inequality to

$$\begin{pmatrix} -B^T X^T A - C' \\ I \end{pmatrix}^T P^{-1} \begin{pmatrix} -B^T X^T A - C' \\ I \end{pmatrix} > 0. \tag{4.5.15}$$

It is pretty straightforward to derive necessary conditions for the solvability of (4.5.14). Indeed, let us assume that (4.5.14) holds for some X. If $A_\perp$ and $B_\perp$ denote basis matrices of $\ker(A)$ and $\ker(B)$ respectively, we infer

$$\begin{pmatrix} I \\ A^T X B + C \end{pmatrix} B_\perp = \begin{pmatrix} I \\ C \end{pmatrix} B_\perp \text{ and } \begin{pmatrix} -B^T X^T A - C' \\ I \end{pmatrix} A_\perp = \begin{pmatrix} -C' \\ I \end{pmatrix} A_\perp.$$

Since $B_\perp^T(4.5.14)B_\perp < 0$ and $A_\perp^T(4.5.15)A_\perp > 0$, we arrive at the two easily verifiable inequalities

$$B_\perp^T \begin{pmatrix} I \\ C \end{pmatrix}^T P \begin{pmatrix} I \\ C \end{pmatrix} B_\perp < 0 \text{ and } A_\perp^T \begin{pmatrix} -C^T \\ I \end{pmatrix}^T P^{-1} \begin{pmatrix} -C^T \\ I \end{pmatrix} A_\perp > 0 \tag{4.5.16}$$

which are necessary for a solution of (4.5.14) to exist. One can constructively prove that they are sufficient [35].

Lemma 4.14 (Elimination Lemma) *Under the hypotheses* (4.5.13) *on* P, *the inequality* (4.5.14) *has a solution if and only if* (4.5.16) *hold true.*

Proof. It remains to prove that (4.5.16) implies the existence of a solution of (4.5.14).

Let us first reveal that one can assume without loss of generality that $R > 0$ and $\tilde{Q} < 0$. For that purpose we need to have information about the inertia of P. Due to $R \geq 0$, P and P^{-1} have size(R) positive eigenvalues (since none of the eigenvalues can vanish). Similarly, $\tilde{Q} \leq 0$ implies that P^{-1} and P have size($\tilde{Q}$) $=$ size(Q) negative eigenvalues. Let us now consider (4.5.14) with the perturbed data

$$P_\epsilon := \begin{pmatrix} Q & S \\ S^T & R + \epsilon I \end{pmatrix} \text{ where } \epsilon > 0$$

is fixed sufficiently small such that (4.5.16) persist to hold for P_ϵ, and such that P_ϵ and P have the same number of positive and negative eigenvalues. Trivially, the right-lower block of P_ϵ is positive definite. The Schur complement $Q - S(R + \epsilon I)^{-1} S^T$ of this right-lower block must be negative definite since P_ϵ has size(Q) negative and size(R) positive eigenvalues. Hence the left-upper block of P_ϵ^{-1} which equals $[Q - S(R + \epsilon I)^{-1} S^T]^{-1}$ is negative definite as well. If the result is proved with $R > 0$ and $\tilde{Q} < 0$, we can conclude that (4.5.14) has a solution X for the perturbed data P_ϵ. Due to $P_0 \leq P_\epsilon$, the very same X also satisfies the original inequality for P_0.

Let us hence assume from now on $R > 0$ and $\tilde{Q} < 0$. We observe that the left-hand side of (4.5.14) equals

$$\begin{pmatrix} I \\ C \end{pmatrix}^T P \begin{pmatrix} I \\ C \end{pmatrix} + (A^T X B)^T (S^T + RC) + (S^T + RC)^T (A^T X B) + (A^T X B)^T R (A^T X B).$$

Hence (4.5.14) is equivalent to (Schur)

$$\begin{pmatrix} \begin{pmatrix} I \\ C \end{pmatrix}^T P \begin{pmatrix} I \\ C \end{pmatrix} + (A^T XB)^T(S^T + RC) + (S^T + RC)^T(A^T XB) & (A^T XB)^T \\ (A^T XB) & -R^{-1} \end{pmatrix} < 0$$

or

$$\begin{pmatrix} \begin{pmatrix} I \\ C \end{pmatrix}^T P \begin{pmatrix} I \\ C \end{pmatrix} & 0 \\ 0 & -R^{-1} \end{pmatrix} +$$

$$+ \begin{pmatrix} A(S^T + RC)^T \\ A \end{pmatrix}^T X \begin{pmatrix} B & 0 \end{pmatrix} + \begin{pmatrix} B^T \\ 0 \end{pmatrix} X^T \begin{pmatrix} A(S^T + RC) & A \end{pmatrix} < 0. \quad (4.5.17)$$

The inequality (4.5.17) has the structure as required in the Projection Lemma. We need to show that

$$\begin{pmatrix} B & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0, \quad \begin{pmatrix} x \\ y \end{pmatrix} \neq 0 \quad (4.5.18)$$

or

$$\begin{pmatrix} A(S^T + RC) & A \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0, \quad \begin{pmatrix} x \\ y \end{pmatrix} \neq 0 \quad (4.5.19)$$

imply

$$\begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} \begin{pmatrix} I \\ C \end{pmatrix}^T P \begin{pmatrix} I \\ C \end{pmatrix} & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^T \begin{pmatrix} I \\ C \end{pmatrix}^T P \begin{pmatrix} I \\ C \end{pmatrix} x - y^T y < 0. \quad (4.5.20)$$

In a first step we show that (4.5.17) and hence (4.5.14) have a solution if $A = I$. Let us assume (4.5.18). Then (4.5.20) is trivial if $x = 0$. For $x \neq 0$ we infer $Bx = 0$ and the first inequality in (4.5.16) implies

$$x^T \begin{pmatrix} I \\ C \end{pmatrix}^T P \begin{pmatrix} I \\ C \end{pmatrix} x < 0$$

what shows that (4.5.20) is true. Let us now assume (4.5.19) with $A = I$. We infer $x \neq 0$ and $y = -(S^T + RC)x$. The left-hand side of (4.5.20) is nothing but

$$x^T \begin{pmatrix} I \\ C \end{pmatrix}^T P \begin{pmatrix} I \\ C \end{pmatrix} x - x^T(S^T + RC)^T R^{-1}(S^T + RC)x =$$

$$= x^T \begin{pmatrix} I \\ C \end{pmatrix}^T P \begin{pmatrix} I \\ C \end{pmatrix} x - x^T \begin{pmatrix} I \\ C \end{pmatrix}^T \begin{pmatrix} S \\ R \end{pmatrix} R^{-1} \begin{pmatrix} S^T & R \end{pmatrix} \begin{pmatrix} I \\ C \end{pmatrix} x =$$

$$= x^T \begin{pmatrix} I \\ C \end{pmatrix}^T \left[P - \begin{pmatrix} SR^{-1}S^T & S \\ S^T & R \end{pmatrix} \right] \begin{pmatrix} I \\ C \end{pmatrix} x = x^T(Q - SR^{-1}S^T)x$$

what is indeed negative since $\tilde{Q}^{-1} = Q - SR^{-1}S^T < 0$ and $x \neq 0$. We conclude that, for $A = I$, (4.5.17) and hence

$$\begin{pmatrix} I \\ XB + C \end{pmatrix}^T P \begin{pmatrix} I \\ XB + C \end{pmatrix} < 0$$

have a solution.

By symmetry –since one can apply the arguments provided above to the dual inequality (4.5.15)– we can infer that

$$\begin{pmatrix} I \\ A^T X + C \end{pmatrix}^T P \begin{pmatrix} I \\ A^T X + C \end{pmatrix} < 0$$

has a solution X. This implies that (4.5.17) has a solution for $B = I$. Therefore, with the Projection Lemma, (4.5.19) implies (4.5.20) for a general A.

In summary, we have proved for general A and B that (4.5.18) or (4.5.19) imply (4.5.20). We can infer the solvability of (4.5.17) or that of (4.5.14). ■

4.5.3 The quadratic performance problem

For the performance index

$$P_j = \begin{pmatrix} Q_j & S_j \\ S_j' & R_j \end{pmatrix}, \quad R_j \geq 0 \quad \text{with inverse} \quad P_j^{-1} = \begin{pmatrix} \tilde{Q}_j & \tilde{S}_j \\ \tilde{S}_j' & \tilde{R}_j \end{pmatrix}, \quad \tilde{Q}_j \leq 0, \tag{4.5.21}$$

we have derived the following synthesis inequalities:

$$\boldsymbol{X}(v) > 0, \quad \left(\begin{array}{cc} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & Q_j & S_j \\ 0 & 0 & S_j^T & R_j \end{array}\right) \left(\begin{array}{cc} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{array}\right) < 0. \tag{4.5.22}$$

Due to the specific structure

$$\left(\begin{array}{c|c} \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{array}\right) = \left(\begin{array}{cc|c} AY & A & B_j \\ 0 & XA & XB_j \\ \hline C_jY & C_j & D_j \end{array}\right) + \left(\begin{array}{cc} 0 & B \\ I & 0 \\ \hline 0 & E_j \end{array}\right) \begin{pmatrix} K & L \\ M & N \end{pmatrix} \left(\begin{array}{cc|c} I & 0 & 0 \\ 0 & C & F_j \end{array}\right), \tag{4.5.23}$$

it is straightforward to apply Lemma 4.14 to eliminate all the variables $\begin{pmatrix} K & L \\ M & N \end{pmatrix}$. For that purpose it suffices to compute basis matrices

$$\Phi_j = \begin{pmatrix} \Phi_j^1 \\ \Phi_j^2 \end{pmatrix} \text{ of } \ker\begin{pmatrix} B^T & E_j^T \end{pmatrix} \quad \text{and} \quad \Psi_j = \begin{pmatrix} \Psi_j^1 \\ \Psi_j^2 \end{pmatrix} \text{ of } \ker\begin{pmatrix} C & F_j \end{pmatrix}.$$

Corollary 4.15 *For a performance index with* (4.5.21)*, there exists a solution v of* (4.5.22) *if and only if there exist symmetric X and Y that satisfy*

$$\begin{pmatrix} Y & I \\ I & X \end{pmatrix} > 0, \tag{4.5.24}$$

$$\Psi^T \left(\begin{array}{cc} I & 0 \\ A & B_j \\ \hline 0 & I \\ C_j & D_j \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & X & 0 & 0 \\ X & 0 & 0 & 0 \\ \hline 0 & 0 & Q_j & S_j \\ 0 & 0 & S_j^T & R_j \end{array}\right) \left(\begin{array}{cc} I & 0 \\ A & B_j \\ \hline 0 & I \\ C_j & D_j \end{array}\right) \Psi < 0, \tag{4.5.25}$$

$$\Phi^T \left(\begin{array}{cc} -A^T & -C_j^T \\ I & 0 \\ \hline -B_j^T & -D_j^T \\ 0 & I \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & Y & 0 & 0 \\ Y & 0 & 0 & 0 \\ \hline 0 & 0 & \tilde{Q}_j & \tilde{S}_j \\ 0 & 0 & \tilde{S}_j^T & \tilde{R}_j \end{array}\right) \left(\begin{array}{cc} -A^T & -C_j^T \\ I & 0 \\ \hline -B_j^T & -D_j^T \\ 0 & I \end{array}\right) \Phi > 0. \tag{4.5.26}$$

Remark. Note that the columns of $\begin{pmatrix} B \\ E_j \end{pmatrix}$ indicate in how far the right-hand side of (4.1.1) can be modified by control, and the rows of $\begin{pmatrix} C & F_j \end{pmatrix}$ determine those functionals that provide information about the system state and the disturbance that is available for control. Roughly speaking, the columns of Φ_j or of Ψ_j indicate what *cannot* be influenced by control or which information *cannot* be extracted from the measured output. Let us hence compare (4.5.24)-(4.5.26) with the synthesis inequalities that would be obtained for

$$\left(\begin{array}{c} \dot{x} \\ \hline z_1 \\ \vdots \\ z_q \end{array}\right) = \left(\begin{array}{c|ccc} A & B_1 & \cdots & B_q \\ \hline C_1 & D_1 & \cdots & D_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ C_q & D_{q1} & \cdots & D_q \end{array}\right) \left(\begin{array}{c} x \\ \hline w_1 \\ \vdots \\ w_q \end{array}\right) \tag{4.5.27}$$

without control input and measurement output. For this system we could choose $\Phi = I$ and $\Psi = I$ to arrive at the synthesis inequalities

$$\begin{pmatrix} Y & I \\ I & X \end{pmatrix} > 0, \tag{4.5.28}$$

$$\left(\begin{array}{cc} I & 0 \\ A & B_j \\ \hline 0 & I \\ C_j & D_j \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & X & 0 & 0 \\ X & 0 & 0 & 0 \\ \hline 0 & 0 & Q_j & S_j \\ 0 & 0 & S_j^T & R_j \end{array}\right) \left(\begin{array}{cc} I & 0 \\ A & B_j \\ \hline 0 & I \\ C_j & D_j \end{array}\right) < 0, \tag{4.5.29}$$

$$\left(\begin{array}{cc} -A^T & -C_j^T \\ I & 0 \\ \hline -B_j^T & -D_j^T \\ 0 & I \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & Y & 0 & 0 \\ Y & 0 & 0 & 0 \\ \hline 0 & 0 & \tilde{Q}_j & \tilde{S}_j \\ 0 & 0 & \tilde{S}_j^T & \tilde{R}_j \end{array}\right) \left(\begin{array}{cc} -A^T & -C_j^T \\ I & 0 \\ \hline -B_j^T & -D_j^T \\ 0 & I \end{array}\right) > 0. \tag{4.5.30}$$

Since there is no control and no measured output, these could be viewed as analysis inequalities for (4.5.27). Hence we have very nicely displayed in how far controls or measurements do influence the synthesis inequalities through Φ_j and Ψ_j. Finally, we note that (4.5.28)-(4.5.30) are equivalent to $X > 0$, (4.5.29) or to $Y > 0$, (4.5.30). Moreover, if dualizing $X > 0$, (4.5.29), we arrive at $Y > 0$, (4.5.30) for $Y := X^{-1}$.

Proof of Corollary 4.15. The first inequality (4.5.24) is just $\boldsymbol{X}(v) > 0$. The inequalities (4.5.25)-(4.5.26) are obtained by simply applying Lemma 4.14 to the second inequality of (4.5.22), viewed as a quadratic matrix inequality in the unknowns $\begin{pmatrix} K & L \\ M & N \end{pmatrix}$. For that purpose we first observe that

$$\ker\left(\begin{array}{cc|c} 0 & I & 0 \\ B^T & 0 & E_j^T \end{array}\right), \quad \ker\left(\begin{array}{cc|c} I & 0 & 0 \\ 0 & C & F_j \end{array}\right) \text{ have the basis matrices } \left(\begin{array}{c} \Phi_j^1 \\ 0 \\ \hline \Phi_j^2 \end{array}\right), \quad \left(\begin{array}{c} 0 \\ \Psi_j^1 \\ \hline \Psi_j^2 \end{array}\right)$$

respectively. Due to

$$\left(\begin{array}{cc} I & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \hline 0 & I \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{array}\right)\left(\begin{array}{c} 0 \\ \Psi_j^1 \\ \hline \Psi_j^2 \end{array}\right) = \left(\begin{array}{cc|c} I & 0 & 0 \\ 0 & I & 0 \\ AY & A & B_j \\ 0 & XA & XB_j \\ \hline 0 & 0 & I \\ C_jY & C_j & D_j \end{array}\right)\left(\begin{array}{c} 0 \\ \Psi_j^1 \\ \hline \Psi_j^2 \end{array}\right) = \left(\begin{array}{cc} 0 & 0 \\ I & 0 \\ A & B_j \\ XA & XB_j \\ \hline 0 & I \\ C_j & D_j \end{array}\right)\Psi,$$

the solvability condition that corresponds to the first inequality in (4.5.16) reads as

$$\Psi^T \left(\begin{array}{cc} 0 & 0 \\ I & 0 \\ A & B_j \\ XA & XB_j \\ \hline 0 & I \\ C_j & D_j \end{array}\right)^T \left(\begin{array}{cccc|cc} 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ I & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & Q_j & S_j \\ 0 & 0 & 0 & 0 & S_j^T & R_j \end{array}\right) \left(\begin{array}{cc} 0 & 0 \\ I & 0 \\ A & B_j \\ XA & XB_j \\ \hline 0 & I \\ C_j & D_j \end{array}\right)\Psi < 0$$

what simplifies to

$$\Psi^T \left(\begin{array}{cc} I & 0 \\ XA & XB_j \\ \hline 0 & I \\ C_j & D_j \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & Q_j & S_j \\ 0 & 0 & S_j^T & R_j \end{array}\right) \left(\begin{array}{cc} I & 0 \\ XA & XB_j \\ \hline 0 & I \\ C_j & D_j \end{array}\right)\Psi < 0.$$

This is clearly nothing but (4.5.25). The very same simple steps lead to (4.5.26). Indeed, we have

$$\left(\begin{array}{cc} \boldsymbol{A}(v)^T & \boldsymbol{C}_j(v)^T \\ I & 0 \\ \hline \boldsymbol{B}_j(v)^T & \boldsymbol{D}_j(v)^T \\ 0 & I \end{array}\right)\left(\begin{array}{c} \Phi_j^1 \\ 0 \\ \hline \Phi_j^2 \end{array}\right) =$$

$$= \left(\begin{array}{cc|c} -YA^T & 0 & -YC_j^T \\ -A^T & -A^TX & -C_j^T \\ I & 0 & 0 \\ 0 & I & 0 \\ \hline -B_j^T & -XB_j^T & -D_j^T \\ 0 & 0 & I \end{array}\right)\left(\begin{array}{c} \Phi_j^1 \\ 0 \\ \hline \Phi_j^2 \end{array}\right) = \left(\begin{array}{cc} -YA^T & -YC_1^T \\ -A^T & -C_1^T \\ I & 0 \\ 0 & 0 \\ \hline -B_1^T & -D_1^T \\ 0 & I \end{array}\right)\Phi$$

such that the solvability condition that corresponds to the second inequality in (4.5.16) is

$$\Phi^T \left(\begin{array}{cc} -YA^T & -YC_1^T \\ -A^T & -C_1^T \\ I & 0 \\ 0 & 0 \\ \hline -B_1^T & -D_1^T \\ 0 & I \end{array}\right)^T \left(\begin{array}{cccc|cc} 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ I & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \tilde{Q}_j & \tilde{S}_j \\ 0 & 0 & 0 & 0 & \tilde{S}_j^T & \tilde{R}_j \end{array}\right) \left(\begin{array}{cc} -YA^T & -YC_1^T \\ -A^T & -C_1^T \\ I & 0 \\ 0 & 0 \\ \hline -B_1^T & -D_1^T \\ 0 & I \end{array}\right)\Phi < 0$$

what simplifies to

$$\Phi^T \left(\begin{array}{cc} -YA^T & -YC_1^T \\ I & 0 \\ \hline -B_1^T & -D_1^T \\ 0 & I \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & \tilde{Q}_j & \tilde{S}_j \\ 0 & 0 & \tilde{S}_j^T & \tilde{R}_j \end{array}\right) \left(\begin{array}{cc} -YA^T & -YC_1^T \\ I & 0 \\ \hline -B_1^T & -D_1^T \\ 0 & I \end{array}\right)\Phi < 0$$

and we arrive at (4.5.26). ■

Starting from the synthesis inequalities (4.5.22) in the variables X, Y, $\begin{pmatrix} K & L \\ M & N \end{pmatrix}$, we have derived the equivalent inequalities (4.5.24)-(4.5.26) in the variables X, Y only. Testing feasibility of these latter inequalities can hence be accomplished much faster. This is particularly advantageous if optimizing an additional parameter, such as minimizing the sup-optimality level γ in the H_∞ problem.

To conclude this section, let us comment on how to compute the controller after having found solutions X, Y of (4.5.24)-(4.5.26). One possibility is to explicitly solve the quadratic inequality (4.5.22) in $\begin{pmatrix} K & L \\ M & N \end{pmatrix}$ along the lines of the proof of Lemma 4.14, and reconstruct the controller parameters as earlier. One could as well proceed directly: Starting from X and Y, we can compute non-singular U and V with $UV^T = I - XY$, and determine $\mathcal{X} > 0$ by solving the first equation in (4.2.10). Due to (4.1.5), we can apply Lemma 4.14 directly to the analysis inequality

$$\left(\begin{array}{cc} I & 0 \\ \mathcal{A} & \mathcal{B}_j \\ \hline 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & \mathcal{X} & 0 & 0 \\ \mathcal{X} & 0 & 0 & 0 \\ \hline 0 & 0 & Q_j & S_j \\ 0 & 0 & S_j^T & R_j \end{array}\right) \left(\begin{array}{cc} I & 0 \\ \mathcal{A} & \mathcal{B}_j \\ \hline 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{array}\right) < 0$$

if viewing $\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}$ as variables. It is not difficult (and you should provide the details!) to verify the solvability conditions for this quadratic inequality, and to construct an explicit solution along the lines of the proof of Lemma 4.14. Alternatively, one can transform the quadratic inequality to a linear matrix inequality with Lemma 4.2, and apply the Projection Lemma to reconstruct the controller parameters. For the latter step the LMI-Lab offers a standard routine. We conclude that there are many basically equivalent alternative ways to compute a controller once one has determined X and Y.

4.5.4 H_2-problems

If recalling (4.2.3), we observe that both inequalities in the H_2-synthesis conditions (4.3.3) involve the variables M and N, but only the first one

$$\begin{pmatrix} \boldsymbol{A}(v)^T + \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{B}_j(v)^T & -\gamma_j I \end{pmatrix} < 0 \tag{4.5.31}$$

is affected by K and L. This might suggest that the latter two variables can be eliminated in the synthesis conditions. Since (4.5.31) is affine in $\begin{pmatrix} K & L \end{pmatrix}$, we can indeed apply the Projection Lemma to eliminate these variables. It is not difficult to arrive at the following alternative synthesis conditions for H_2-type criteria.

Corollary 4.16 *There exists a controller that renders* (4.3.2) *for some $\mathcal{X}$, Z_j satisfied iff there exist X, Y, M, N, Z_j with $f_j(Z_j) < \gamma_j$, $D_j + E_j N F_j = 0$ and*

$$\begin{pmatrix} Y & I & (C_jY + E_jM)^T \\ I & X & (C_j + E_jNC)^T \\ C_jY + E_jM & C_j + E_jNC & Z_j \end{pmatrix} > 0,$$

$$\Psi^T \begin{pmatrix} A^TX + XA & XB_j \\ B_j^TX & -\gamma_j I \end{pmatrix} \Psi < 0, \quad \begin{pmatrix} (AY + BM) + (AY + BM)^T & B_j + BNF_j \\ (B_j + BNF_j)^T & -\gamma_j I \end{pmatrix} < 0. \tag{4.5.32}$$

Proof. We only need to show that the elimination of K and L in (4.5.31) leads to the two inequalities (4.5.32). Let us recall

$$\begin{pmatrix} \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \end{pmatrix} = \left(\begin{array}{cc|c} AY & A & B_j \\ 0 & XA & XB_j \end{array}\right) + \begin{pmatrix} 0 & B \\ I & 0 \end{pmatrix}\begin{pmatrix} K & L \\ M & N \end{pmatrix}\left(\begin{array}{cc|c} I & 0 & 0 \\ 0 & C & F_j \end{array}\right) =$$
$$= \left(\begin{array}{cc|c} AY + BM & A + BNC & B_j + BNF_j \\ 0 & XA & XB_j \end{array}\right) + \begin{pmatrix} 0 \\ I \end{pmatrix}\begin{pmatrix} K & L \end{pmatrix}\left(\begin{array}{cc|c} I & 0 & 0 \\ 0 & C & F_j \end{array}\right).$$

Therefore, (4.5.31) is equivalent to

$$\left(\begin{array}{cc|c} AY + YA^T & A & B_j \\ A^T & A^TX + XA & XB_j \\ \hline B_j^T & B_j^TX & -\gamma_j I \end{array}\right) + \operatorname{sym}\left(\left(\begin{array}{c} B \\ 0 \\ \hline 0 \end{array}\right)\begin{pmatrix} M & N \end{pmatrix}\left(\begin{array}{cc|c} I & 0 & 0 \\ 0 & C & F_j \end{array}\right)\right) +$$
$$+ \operatorname{sym}\left(\left(\begin{array}{c} 0 \\ I \\ \hline 0 \end{array}\right)\begin{pmatrix} K & L \end{pmatrix}\left(\begin{array}{cc|c} I & 0 & 0 \\ 0 & C & F_j \end{array}\right)\right) < 0$$

where $\operatorname{sym}(M) := M + M^T$ is just an abbreviation to shorten the formulas. Now note that

$$\ker\left(\begin{array}{cc|c} 0 & I & 0 \end{array}\right), \quad \ker\left(\begin{array}{cc|c} I & 0 & 0 \\ 0 & C & F_j \end{array}\right) \quad \text{have the basis matrices} \quad \left(\begin{array}{cc} I & 0 \\ 0 & 0 \\ \hline 0 & I \end{array}\right), \quad \left(\begin{array}{c} 0 \\ \Psi_j^1 \\ \hline \Psi_j^2 \end{array}\right)$$

respectively. Therefore, the Projection Lemma leads to the two inequalities

$$\left(\begin{array}{c} 0 \\ \Psi_j^1 \\ \hline \Psi_j^2 \end{array}\right)^T \left(\begin{array}{cc|c} AY + YA^T & A & B_j \\ A^T & A^TX + XA & XB_j \\ \hline B_j^T & B_j^TX & -\gamma_j I \end{array}\right)\left(\begin{array}{c} 0 \\ \Psi_j^1 \\ \hline \Psi_j^2 \end{array}\right) < 0$$

and

$$\begin{pmatrix} AY + YA^T & B_j \\ B_j^T & -\gamma_j I \end{pmatrix} + \operatorname{sym}\left(\begin{pmatrix} B \\ 0 \end{pmatrix}\begin{pmatrix} M & N \end{pmatrix}\begin{pmatrix} I & 0 \\ 0 & F_j \end{pmatrix}\right) < 0$$

that are easily rewritten to (4.5.32). ■

If it happens that E_j vanishes, we can also eliminate all variables $\begin{pmatrix} K & L \\ M & N \end{pmatrix}$ from the synthesis inequalities. The corresponding results are obtained in a straightforward fashion and their proof is left as an exercise.

Corollary 4.17 *Suppose that* $E_j = 0$. *Then there exists a controller that renders* (4.3.2) *for some* $\mathcal{X}$, Z_j *satisfied iff* $D_j = 0$ *and there exist* X, Y Z_j *with* $f_j(Z_j) < \gamma_j$ *and*

$$\begin{pmatrix} Y & I & (C_jY)^T \\ I & X & C_j^T \\ C_jY & C_j & Z_j \end{pmatrix} > 0,$$

$$\Psi^T\begin{pmatrix} A^TX + XA & XB_j \\ B_j^TX & -\gamma_j I \end{pmatrix}\Psi < 0, \quad \begin{pmatrix} \widehat{\Phi} & 0 \\ 0 & I \end{pmatrix}^T\begin{pmatrix} AY + YA^T & B_j \\ B_j^T & -\gamma_j I \end{pmatrix}\begin{pmatrix} \widehat{\Phi} & 0 \\ 0 & I \end{pmatrix} < 0$$

where $\widehat{\Phi}$ *is a basis matrix of* $\ker(B)$.

Remarks.

- Once the synthesis inequalities have been solved, the computation of $\begin{pmatrix} K & L \end{pmatrix}$ or of $\begin{pmatrix} K & L \\ M & N \end{pmatrix}$ can be performed along the lines of the proof of the Projection Lemma.
- It was our main concern to perform the variable elimination with as little computations as possible. They should be read as examples how one can proceed in specific circumstances, and they can be easily extended to various other performance specifications. As an exercise, the reader should eliminate variables in the peak-to-peak upper bound synthesis LMI's.

4.6 State-feedback problems

The state-feedback problem is characterized by

$$y = x \;\text{ or }\; \begin{pmatrix} C & F_1 & \cdots & F_q \end{pmatrix} = \begin{pmatrix} I & 0 & \cdots & 0 \end{pmatrix}.$$

Then the formulas (4.2.3) read as

$$\begin{pmatrix} \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{C}_j(v) & \boldsymbol{D}_j(v) \end{pmatrix} = \left(\begin{array}{cc|c} AY + BM & A + BN & B_j \\ K & AX + L & XB_j \\ \hline C_jY + E_jM & C_j + E_jN & D_j \end{array}\right).$$

Note that the variable L only appears in the $(2,2)$-block, and that we can assign an arbitrary matrix in this position by suitably choosing L. Therefore, by varying L, the $(2,2)$ block of

$$\begin{pmatrix} \boldsymbol{A}(v)^T + \boldsymbol{A}(v) & \boldsymbol{B}_j(v) \\ \boldsymbol{B}_j(v) & 0 \end{pmatrix} =$$
$$= \left(\begin{array}{cc|c} (AY + BM) + (AY + BM)^T & (A + BN) + K^T & B_j \\ K + (A + BN)^T & (AX + L) + (AX + L)^T & XB_j \\ \hline B_j^T & B_j^T X & 0 \end{array}\right)$$

varies in the set of all symmetric matrices. This allows to apply Lemma 4.11 in order to eliminate L in synthesis inequalities what leads to a drastic simplification.

Let us illustrate all this for the quadratic performance problem. The corresponding synthesis inequalities (4.2.7) read as

$$\begin{pmatrix} Y & I \\ I & X \end{pmatrix} > 0, \quad \left(\begin{array}{cc|c} (AY + BM) + (AY + BM)^T & (A + BN) + K^T & B_j \\ K + (A + BN)^T & (AX + L) + (AX + L)^T & XB_j \\ \hline B_j^T & B_j^T X & 0 \end{array}\right) +$$
$$+ \left(\begin{array}{cc|c} 0 & 0 & I \\ C_jY + E_jM & C_j + E_jN & D_j \end{array}\right)^T P_j \left(\begin{array}{cc|c} 0 & 0 & I \\ C_jY + E_jM & C_j + E_jN & D_j \end{array}\right) < 0.$$

These imply, just by cancelling the second block row/column,

$$Y > 0,\ \begin{pmatrix} (AY+BM)+(AY+BM)^T & B_j \\ B_j^T & 0 \end{pmatrix} + \\ + \begin{pmatrix} 0 & I \\ C_jY+E_jM & D_j \end{pmatrix}^T P_j \begin{pmatrix} 0 & I \\ C_jY+E_jM & D_j \end{pmatrix} < 0$$

or

$$Y > 0,\ \left(\begin{array}{cc} I & 0 \\ AY+BM & B_j \\ \hline 0 & I \\ C_jY+E_jM & D_j \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & Q_j & S_j \\ 0 & 0 & S_j^T & R_j \end{array}\right) \left(\begin{array}{cc} I & 0 \\ AY+BM & B_j \\ \hline 0 & I \\ C_jY+E_jM & D_j \end{array}\right) < 0. \quad (4.6.1)$$

This is a drastic simplification since only the variables Y and M do appear in the resulting inequalities. It is no problem to reverse the arguments in order to show that the reduced inequalities are equivalent to the full synthesis inequalities.

However, proceeding in a different fashion leads to another fundamental insight: With solutions Y and M of (4.6.1), one can in fact design a *static* controller which solves the quadratic performance problem. Indeed, we just choose

$$D_c := MY^{-1}$$

to infer that the static controller $y = D_c u$ leads to a controlled system with the describing matrices

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}_j \\ \mathcal{C}_j & \mathcal{D}_j \end{pmatrix} = \begin{pmatrix} A+BD_c & B_j \\ C_j+E_jD_c & D_j \end{pmatrix} = \begin{pmatrix} (AY+BM)Y^{-1} & B_j \\ (C_jY+E_jM)Y^{-1} & D_j \end{pmatrix}.$$

We infer that (4.6.1) is identical to

$$Y > 0,\ \left(\begin{array}{cc} I & 0 \\ \mathcal{A}Y & \mathcal{B}_j \\ \hline 0 & I \\ \mathcal{C}_jY & \mathcal{D}_j \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & Q_j & S_j \\ 0 & 0 & S_j^T & R_j \end{array}\right) \left(\begin{array}{cc} I & 0 \\ \mathcal{A}Y & \mathcal{B}_j \\ \hline 0 & I \\ \mathcal{C}_jY & \mathcal{D}_j \end{array}\right) < 0.$$

If we perform congruence transformations with Y^{-1} and $\begin{pmatrix} Y^{-1} & 0 \\ 0 & I \end{pmatrix}$, we arrive with $\mathcal{X} := Y^{-1}$ at

$$\mathcal{X} > 0,\ \left(\begin{array}{cc} I & 0 \\ \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \hline 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & Q_j & S_j \\ 0 & 0 & S_j^T & R_j \end{array}\right) \left(\begin{array}{cc} I & 0 \\ \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_j \\ \hline 0 & I \\ \mathcal{C}_j & \mathcal{D}_j \end{array}\right) < 0.$$

Hence the static gain $\mathcal{D}$ indeed defines a controller which solves the quadratic performance problem.

Corollary 4.18 *Under the state-feedback information structure, there exists a dynamic controller* $\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}$ *and some* $\mathcal{X}$ *which satisfy* (4.2.1) *iff there exist solutions* Y *and* M *of the inequalities* (4.6.1). *If* Y *and* M *solve* (4.6.1), *the* static *state-feedback controller gain*

$$D_c = MY^{-1}$$

and the Lyapunov matrix $\mathcal{X} := Y^{-1}$ *render* (4.2.1) *satisfied.*

In literally the same fashion as for output-feedback control, we arrive at the following general procedure to proceed from analysis inequalities to synthesis inequalities, and to construct a static state-feedback controller:

- Rewrite the analysis inequalities in the blocks $\mathcal{X}$, $\mathcal{X}\mathcal{A}$, $\mathcal{X}\mathcal{B}_j$, $\mathcal{C}_j$, $\mathcal{D}_j$ in order to be able to find a (formal) congruence transformation involving $\mathcal{Y}$ which leads to inequalities in the blocks $\mathcal{Y}^T\mathcal{X}\mathcal{Y}$, $\mathcal{Y}^T\mathcal{X}\mathcal{A}\mathcal{Y}$, $\mathcal{Y}^T\mathcal{X}\mathcal{B}_j$, $\mathcal{C}_j\mathcal{Y}$, $\mathcal{D}_j$.
- Perform the substitutions
$$\mathcal{Y}^T\mathcal{X}\mathcal{Y} \to Y \ \text{ and } \ \begin{pmatrix} \mathcal{Y}^T\mathcal{X}\mathcal{A}\mathcal{Y} & \mathcal{Y}^T\mathcal{X}\mathcal{B}_j \\ \mathcal{C}_j\mathcal{Y} & \mathcal{D}_j \end{pmatrix} \to \begin{pmatrix} AY + BM & B_j \\ C_jY + E_jM & D_j \end{pmatrix}$$
to arrive at matrix inequalities in the variables Y and M.
- After having solved the synthesis inequalities for Y and M, the static controller gain and the Lyapunov matrix
$$\mathcal{D} = MY^{-1} \ \text{ and } \ \mathcal{X} = Y^{-1}$$
render the analysis inequalities satisfied.

As an illustration, starting form the analysis inequalities (4.3.2) for H_2-type problems, the corresponding state-feedback synthesis conditions read as

$$\begin{pmatrix} (AY + BM)^T + (AY + BM) & B_j \\ B_j^T & -\gamma_j I \end{pmatrix} < 0,$$
$$\begin{pmatrix} Y & (C_jY + E_jM)^T \\ C_jY + E_jM & Z_j \end{pmatrix} > 0, \quad f_j(Z_j) < \gamma_j, \quad D_j = 0.$$

All our previous remarks pertaining to the (more complicated) procedure for the output-feedback information structure apply without modification.

In general we can conclude that dynamics in the controller do not offer any advantage over static controllers for state-feedback problems. This is also true for mixed control problems. This statements requires extra attention since our derivation was based on eliminating the variable L which might occur in several matrix inequalities. At this point the remark after Lemma 4.11 comes into play:

This particular elimination result also applies to systems of matrix inequalities such that, indeed, the occurrence of L is various inequalities will not harm the arguments.

As earlier, in the single-objective quadratic performance problem by state-feedback, it is possible to eliminate the variable M in (4.6.1). Alternatively, one could as well exploit the particular structure of the system description to simplify the conditions in Theorem 4.15. Both approaches lead to the following result.

Corollary 4.19 *For the state-feedback quadratic performance problem with index satisfying* (4.5.21)*, there exists dynamic controller and some* $\mathcal{X}$ *with* (4.2.1) *iff there exists a symmetric* Y *which solves*

$$Y > 0,\quad \Phi^T \left(\begin{array}{cc} -A^T & -C_j^T \\ I & 0 \\ \hline -B_j^T & -D_j^T \\ 0 & I \end{array}\right)^T \left(\begin{array}{cc|cc} 0 & Y & 0 & 0 \\ Y & 0 & 0 & 0 \\ \hline 0 & 0 & \tilde{Q}_j & \tilde{S}_j \\ 0 & 0 & \tilde{S}_j^T & \tilde{R}_j \end{array}\right) \left(\begin{array}{cc} -A^T & -C_j^T \\ I & 0 \\ \hline -B_j^T & -D_j^T \\ 0 & I \end{array}\right) \Phi > 0. \qquad (4.6.2)$$

Remarks. All these results should be viewed as illustrations how to proceed for specific system descriptions. Indeed, another popular choice is the so-called *full information* structure in which both the state and the disturbance are measurable:

$$y = \begin{pmatrix} x \\ w \end{pmatrix}.$$

Similarly, one could consider the corresponding dual versions that are typically related to *estimation problems*, such as e.g.

$$\begin{pmatrix} B \\ E_1 \\ \vdots \\ E_q \end{pmatrix} = \begin{pmatrix} I \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

We have collected all auxiliary results that allow to handle these specific problems without any complications.

4.7 Discrete-Time Systems

Everything what has been said so far can be easily extended to discrete time-design problems. This is particularly surprising since, in the literature, discrete-time problem solutions often seem much more involved and harder to master than their continuous-time counterparts.

Our general procedure to step from analysis to synthesis as well as the technique to recover the controller need no change at all; in particular, the concrete formulas for the block substitutions do not change. The elimination of transformed controller parameters proceeds in the same fashion on the basis of the Projection Lemma or the Elimination Lemma and the specialized version thereof.

Only as a example we consider the problem discussed in [12]: the mixed H_2/H_∞ problem with different disturbance inputs and controlled outputs in discrete-time.

It is well-known [12] that $\mathcal{A}$ has all its eigenvalues in the unit disk, that the discrete time H_2-norm of

$$\mathcal{C}_1(zI - \mathcal{A})^{-1}\mathcal{B}_1 + \mathcal{D}_1$$

is smaller than γ_1, and that the discrete time H_∞-norm of

$$\mathcal{C}_2(zI - \mathcal{A})^{-1}\mathcal{B}_2 + \mathcal{D}_2$$

is smaller than γ_2 iff there exist symmetric matrices $\mathcal{X}_1$, $\mathcal{X}_2$, and Z with $\mathrm{trace}(Z) < \gamma_1$ and

$$\begin{pmatrix} \mathcal{X}_1 & \mathcal{X}_1\mathcal{A} & \mathcal{X}_1\mathcal{B}_1 \\ \mathcal{A}^T\mathcal{X}_1 & \mathcal{X}_1 & 0 \\ \mathcal{B}_1^T\mathcal{X}_1 & 0 & \gamma_1 I \end{pmatrix} > 0, \begin{pmatrix} \mathcal{X}_1 & 0 & \mathcal{C}_1^T \\ 0 & I & \mathcal{D}_1^T \\ \mathcal{C}_1 & \mathcal{D}_1 & Z \end{pmatrix} > 0, \begin{pmatrix} \mathcal{X}_2 & 0 & \mathcal{A}^T\mathcal{X}_2 & \mathcal{C}_2^T \\ 0 & \gamma_2 I & \mathcal{B}_2^T\mathcal{X}_2 & \mathcal{D}_2^T \\ \mathcal{X}_2\mathcal{A} & \mathcal{X}_2\mathcal{B}_2 & \mathcal{X}_2 & 0 \\ \mathcal{C}_2 & \mathcal{D}_2 & 0 & \gamma_2 I \end{pmatrix} > 0. \tag{4.7.1}$$

Note that we have transformed these analysis LMI's such that they are affine in the blocks that will be transformed for synthesis.

The mixed problem consists of searching for a controller that renders these inequalities satisfied with a common Lyapunov function $\mathcal{X} = \mathcal{X}_1 = \mathcal{X}_2$. The solution is immediate: Perform congruence transformations of (4.7.1) with

$$\mathrm{diag}(\mathcal{Y}, \mathcal{Y}, I), \quad \mathrm{diag}(\mathcal{Y}, I, I), \quad \mathrm{diag}(\mathcal{Y}, I, \mathcal{Y}, I)$$

and read off the synthesis LMI's using (4.2.3). After solving the synthesis LMI's, we stress again that the controller construction proceeds along *the same steps* as in Theorem 4.3. The inclusion of pole constraints for arbitrary LMI regions (related, of course, to discrete time stability) and other criteria poses no extra problems.

4.8 Exercises

Exercise 1 Derive an LMI solution of the H_∞-problem for the system

$$\begin{pmatrix} \dot{x} \\ z_1 \\ y \end{pmatrix} = \begin{pmatrix} A & B_1 & B \\ C_1 & D_1 & E \\ C & F & 0 \end{pmatrix} \begin{pmatrix} x \\ w_1 \\ u \end{pmatrix}$$

with

$$C = \begin{pmatrix} I \\ 0 \end{pmatrix}, \quad F = \begin{pmatrix} 0 \\ I \end{pmatrix} \text{ such that } y = \begin{pmatrix} x \\ w_1 \end{pmatrix}.$$

(This is the so-called full information problem.)

Exercise 2 (Nominal and robust estimation) Consider the system

$$\begin{pmatrix} \dot{x} \\ z \\ y \end{pmatrix} = \begin{pmatrix} A & B_1 \\ C_1 & D_1 \\ C & F \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix}$$

and inter-connect it with the estimator

$$\begin{pmatrix} \dot{x}_c \\ \hat{z} \end{pmatrix} = \begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \begin{pmatrix} x_c \\ y \end{pmatrix} \tag{4.8.1}$$

where both A and A_c are Hurwitz. The goal in optimal estimation is to design an estimator which keeps $z - \widehat{z}$ as small as possible for all disturbances w in a certain class. Out of the multitude of possibilities, we choose the L_2-gain of $w \to z - \widehat{z}$ (for zero initial condition of both the system and the estimator) as the measure of the estimation quality.

This leads to the following problem formulation: Given $\gamma > 0$, test whether there exists an estimator which renders

$$\sup_{w \in L_2,\ w \neq 0} \frac{\|z - \widehat{z}\|_2}{\|w\|_2} < \gamma \tag{4.8.2}$$

satisfied. If yes, reveal how to design an estimator that leads to this property.

(a) Show that the estimation problem is a specialization of the general output-feedback H_∞-design problem.

(b) Due to the specific structure of the open-loop system, show that there exists a linearizing transformation of the estimator parameters which does not involve any matrices that describe the open-loop system.

Hint: To find the transformation, proceed as in the proof of Theorem 4.3 with the factorization

$$\mathcal{Y}^T \mathcal{X} = \mathcal{Z} \text{ where } \mathcal{Y}^T = \begin{pmatrix} I & Y^{-1}V \\ I & 0 \end{pmatrix}, \quad \mathcal{Z} = \begin{pmatrix} Y^{-1} & 0 \\ X & U \end{pmatrix},$$

and consider as before the blocks $\mathcal{Y}^T \mathcal{X} \mathcal{A} \mathcal{Y}$, $\mathcal{Y}^T \mathcal{X} \mathcal{B}$, $\mathcal{C} \mathcal{Y}$.

(c) Now assume that the system is affected by time-varying uncertain parameters as

$$\begin{pmatrix} \dot{x} \\ z \\ y \end{pmatrix} = \begin{pmatrix} A(\Delta(t)) & B_1(\Delta(t)) \\ C_1(\Delta(t)) & D_1(\Delta(t)) \\ C(\Delta(t)) & F(\Delta(t)) \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix}$$

where

$$\begin{pmatrix} A(\Delta) & B_1(\Delta) \\ C_1(\Delta) & D_1(\Delta) \\ C(\Delta) & F(\Delta) \end{pmatrix} \text{ is affine in } \Delta \text{ and } \Delta(t) \in \mathrm{conv}\{\Delta_1, \ldots, \Delta_N\}.$$

Derive LMI conditions for the existence of an estimator that guarantees (4.8.2) for all uncertainties, and show how to actually compute such an estimator if the LMI's are feasible.

Hint: Recall what we have discussed for the state-feedback problem in Section 8.1.3.

(d) Suppose that the uncertainty enters rationally, and that it has been pulled out to arrive at the LFT representation

$$\begin{pmatrix} \dot{x} \\ z_1 \\ z \\ y \end{pmatrix} = \begin{pmatrix} A & B_1 & B_2 \\ C_1 & D_1 & D_{12} \\ C_2 & D_{21} & D_2 \\ C & F_1 & F_2 \end{pmatrix} \begin{pmatrix} x \\ w_1 \\ w \end{pmatrix}, \quad w_1(t) = \Delta(t) z_1(t), \quad \Delta(t) \in \mathrm{conv}\{\Delta_1, ..., \Delta_N\}$$

of the uncertain system. Derive synthesis inequalities with full-block scalings that guarantee the existence of an estimator that guarantees (4.8.2) for all uncertainties and reveal how to actually compute such an estimator if the LMI's are feasible. What happens if $D_1 = 0$ such that the uncertainty enters affinely?

Hint: The results should be formulated analogously to what we have done in Section 8.1.2. There are two possibilities to proceed: You can either just use the transformation (4.2.10) to obtain synthesis inequalities that can be rendered convex by an additional congruence transformation, or you can employ the alternative parameter transformation as derived in part 2 of this exercise to directly obtain a convex test.

Exercise 3

This is a simulation exercise that involves the synthesis of an active controller for the suspension system in Exercise 3 of Chapter 2. We consider the rear wheel of a tractor-trailer combination as is depicted in Figure 4.2. Here m_1 represents tire, wheel and rear axle mass, m_2 denotes a fraction

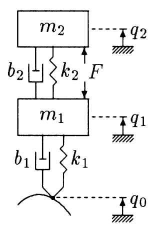


Figure 4.2: Active suspension system

of the semitrailer mass. The deflection variables q_i are properly scaled so that $q_2 - q_1 = 0$ and $q_1 - q_0 = 0$ in steady state. The system is modeled by the state space equations

$$\dot{x} = Ax + B \begin{pmatrix} q_0 \\ F \end{pmatrix}$$
$$z = Cx + D \begin{pmatrix} q_0 \\ F \end{pmatrix}$$

where

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{k_1+k_2}{m_1} & \frac{k_2}{m_1} & -\frac{b_1+b_2}{m_1} & \frac{b_2}{m_1} \\ \frac{k_2}{m_2} & -\frac{k_2}{m_2} & \frac{b_2}{m_2} & -\frac{b_2}{m_2} \end{pmatrix}; \qquad B = \begin{pmatrix} \frac{b_1}{m_1} & 0 \\ 0 & 0 \\ \frac{k_1}{m_1} - \frac{b_1}{m_1}\frac{b_1+b_2}{m_1} & -\frac{1}{m_1} \\ \frac{b_1 b_2}{m_1 m_2} & \frac{1}{m_2} \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{k_2}{m_2} & -\frac{k_2}{m_2} & \frac{b_2}{m_2} & -\frac{b_2}{m_2} \\ -1 & 1 & 0 & 0 \end{pmatrix}; \qquad D = \begin{pmatrix} -1 & 0 \\ 0 & 1 \\ \frac{b_1 b_2}{m_1 m_2} & \frac{1}{m_2} \\ 0 & 0 \end{pmatrix}.$$

Here, $x = \mathrm{col}(q_1, q_2, \dot{q}_1 - b_1 q_0/m_1, \dot{q}_2)$ and $z = \mathrm{col}(q_1 - q_0, F, \ddot{q}_2, q_2 - q_1)$ define the state and the to-be-controlled output, respectively. The control input is the force F, the exogenous input is the road profile q_0.

Let the physical parameters be specified as in Table 2.1 in Chapter 2 and let $b_1 = 1.7 \times 10^3$. The aim of the exercise is to design an active suspension control system that generates the force F as a (causal) function of the measured variable $y = \mathrm{col}(\ddot{q}_2, q_2 - q_1)$.

The main objective of the controller design is to achieve low levels of acceleration throughout the vehicle ($\ddot{q}_2$), bounded suspension deflection ($q_2 - q_1$ and $q_1 - q_0$) and bounded dynamic tire force (F).

(a) Let the road profile be represented by $q_0 = W_{q_0}\tilde{q}_0$ where $\tilde{q}_0 \in \mathcal{L}_2$ is equalized in frequency and W_{q_0} is the transfer function

$$W_{q_0}(s) = \frac{0.01}{0.4s + 1}$$

reflecting the quality of the road when the vehicle drives at constant speed. Define the to-be-controlled output $\tilde{z} = W_z z$ where W_z is a weighting matrix with transfer function

$$W_z(s) = \begin{pmatrix} 200 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & \frac{0.0318s+0.4}{3.16\times10^{-4}s^2+0.0314s+1} & 0 \\ 0 & 0 & 0 & 100 \end{pmatrix}.$$

The weight on the chassis acceleration reflects the human sensitivity to vertical accelerations. Use the routines `ltisys`, `smult` and `sdiag` (from the LMI toolbox) to implement the generalized plant

$$P : \begin{pmatrix} \tilde{q}_0 \\ F \end{pmatrix} \mapsto \begin{pmatrix} \tilde{z} \\ y \end{pmatrix}$$

and synthesize with the routine `hinflmi` a controller which minimizes the H_∞ norm of the closed-loop transfer function $\mathcal{T} : \tilde{q}_0 \mapsto \tilde{z}$.

(b) Construct with this controller the closed-loop system which maps q_0 to z (not $\tilde{q}_0$ to $\tilde{z}$!) and validate the controlled system by plotting the four frequency responses of the closed-loop system and the four responses to a step with amplitude 0.2 (meter). (See the routines `slft`, `ssub` and `splot`). What are your conclusions about the behavior of this active suspension system?

(c) Partition the output z of the system into

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}; \quad z_1 = \begin{pmatrix} q_1 - q_0 \\ F \end{pmatrix}; \quad z_2 = \begin{pmatrix} \ddot{q}_2 \\ q_2 - q_1 \end{pmatrix}.$$

and let the weights on the signal components be as in the first part of this exercise. Let $\mathcal{T}_i$, $i = 1, 2$ be the transfer function mapping $\tilde{q}_0 \mapsto \tilde{z}_i$. We wish to obtain insight in the achievable trade-offs between upper bounds of $\|\mathcal{T}_1\|_\infty$ and $\|\mathcal{T}_2\|_2$. To do this,

(i) Calculate the minimal achievable H_∞ norm of $\mathcal{T}_1$.

(ii) Calculate the minimal achievable H_2 norm of $\mathcal{T}_2$.

(iii) Calculate the minimal achievable H_2 norm of $\mathcal{T}_2$ subject to the bound $\|\mathcal{T}_1\|_\infty < \gamma_1$ where γ_1 takes some[1] values in the interval [0.15, 0.30].

Make a plot of the Pareto optimal performances, i.e, plot the minimal achievable H_2 norm of $\mathcal{T}_2$ as function of γ_1. (See the routine `hinfmix` for details).

[1] Slightly depending on your patience and the length of your coffee breaks, I suggest about 5.

Chapter 5

Robust stability and robust performance

First principle models of physical systems are often represented by state space descriptions in which the various components of the state variable represent a well defined physical quantity. Variations, perturbations or uncertainties in specific physical parameters lead to uncertainty in the model. Often, this uncertainty is reflected by variations in well distinguished parameters or coefficients in the model, while in addition the nature and/or range of the uncertain parameters may be known, or partially known. Since very small parameter variations may have a major impact on the dynamics of a system it is of evident importance to analyse parametric uncertainties of dynamical systems. This will be the subject of this chapter. We reconsider the notions of nominal stability and nominal performance introduced in Chapter 3 in the light of different types of parametric uncertainty that effect the behavior of the system. We aim to derive robust stability tests and robust performance tests for systems with time-varying rate-bounded parametric uncertainties.

5.1 Parametric uncertainties

Suppose that $\delta = (\delta_1, \ldots, \delta_p)$ is the vector which expresses the ensemble of all uncertain quantities in a given dynamical system. There are at least two distinct cases which are of independent interest:

(a) **time-invariant parametric uncertainties:** the vector δ is a fixed but unknown element of a uncertainty set $\Delta \subseteq \mathbb{R}^p$.

(b) **time-varying parametric uncertainties:** the vector δ is an unknown time varying function $\delta : \mathbb{R} \to \mathbb{R}^k$ whose values $\delta(t)$ belong to an uncertainty set $\Delta \subseteq \mathbb{R}^p$, and possibly satisfy additional constraints on rates of variation, continuity, spectral content, etc.

The first case typically appears in models in which the physical parameters are fixed but only approximately known up to some level of accuracy. The second case typically captures models in which the uncertainty in parameters, coefficients, or other physical quantities is time-dependent. One may object that in many practical situations both time-varying and time-invariant uncertainties occur so that the distinction between the two cases may seem somewhat artificial. This is true, but since time-invariant uncertainties can equivalently be viewed as time-varying uncertainties with a zero rate constraint, combined time-varying and time-invariant uncertainties are certainly not excluded.

A rather general class of *uncertain* continuous time, dynamical systems is described by the state space equations

$$\dot{x} = f(x, w, \delta) \tag{5.1.1a}$$
$$z = g(x, w, \delta) \tag{5.1.1b}$$

where δ may or may not be time-varying, x, w and z are the state, the input and the output which take values in the state space X, the input space W and the output space Z, respectively. This constitutes a generalization of the model described in (2.2.1) defined in Chapter 2. If the uncertainties δ are fixed but unknown elements of an uncertainty set $\Delta \subseteq \mathbb{R}^p$ then one way to think of equations of this sort is to view them as a set of time-invariant systems, parametrized by $\delta \in \Delta$. However, if δ is time-varying, then (5.1.1a) is to be interpreted as $\dot{x}(t) = f(x(t), w(t), \delta(t))$ and (5.1.1) is better viewed as a time-varying dynamical system. If the components of $\delta(t)$ coincide, for example, with state components then (5.1.1) defines a non-linear system, even when the mappings f and g are linear. If $\delta(t)$ is scalar valued and assumes values in a finite set $\Delta = \{1, \ldots, K\}$ then (5.1.1) defines a hybrid system of K modes whose kth mode is defined by the dynamics

$$\dot{x} = f_k(x, w) := f(x, w, k)$$
$$z = g_k(x, w) := g(x, w, k)$$

and where the time-varying behavior of $\delta(t)$ defines the switching events between the various modes. In any case, the system (5.1.1) is of considerable theoretical and practical interest as it covers quite some relevant classes of dynamical systems.

5.1.1 Affine parameter dependent systems

If f and g in (5.1.1) are linear in x and w then the uncertain model assumes a representation

$$\begin{pmatrix} \dot{x} \\ z \end{pmatrix} = \begin{pmatrix} A(\delta) & B(\delta) \\ C(\delta) & D(\delta) \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix} \tag{5.1.2}$$

in which δ may or may not be time-varying.

Of particular interest will be those systems in which the system matrices *affinely depend on* δ. This means that

$$\begin{aligned} A(\delta) &= A_0 + \delta_1 A_1 + \ldots + \delta_p A_p \\ B(\delta) &= B_0 + \delta_1 B_1 + \ldots + \delta_p B_p \\ C(\delta) &= C_0 + \delta_1 C_1 + \ldots + \delta_p C_p \\ D(\delta) &= D_0 + \delta_1 D_1 + \ldots + \delta_p D_p \end{aligned}$$

or, written in a more compact form,

$$S(\delta) = S_0 + \delta_1 S_1 + \ldots + \delta_p S_p$$

where

$$S(\delta) = \begin{pmatrix} A(\delta) & B(\delta) \\ C(\delta) & D(\delta) \end{pmatrix}$$

is the system matrix associated with (5.1.2). We call these models *affine parameter dependent models*.

5.1.2 Polytopic parameter dependent systems

In Definition 1.3 of Chapter 1 we introduced the notion of a convex combination of a finite set of points. This notion gets relevance in the context of dynamical systems if ‘points’ become systems. Consider a time-varying dynamical system

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)w(t) \\ z(t) &= C(t)x(t) + D(t)w(t) \end{aligned}$$

with input w, output z and state x. Suppose that its *system matrix*

$$S(t) := \begin{pmatrix} A(t) & B(t) \\ C(t) & D(t) \end{pmatrix}$$

is a time varying object which for any time instant $t \in \mathbb{R}$ can be written as a *convex combination* of the p system matrices $S_1, \ldots, S_p$. This means that there exist functions $\alpha_j : \mathbb{R} \to [0, 1]$ such that for any time instant $t \in \mathbb{R}$ we have that

$$S(t) = \sum_{j=1}^{p} \alpha_j(t) S_j$$

where $\sum_{j=1}^{p} \alpha_j(t) = 1$ and

$$S_j = \begin{pmatrix} A_j & B_j \\ C_j & D_j \end{pmatrix}, \quad j = 1, \ldots, p$$

are constant system matrices of equal dimension. In particular, this implies that the system matrices $S(t)$, $t \in \mathbb{R}$ belong to the convex hull of $S_1, \ldots, S_p$, i.e.,

$$S(t) \in \operatorname{conv}(S_1, \ldots, S_p), \qquad t \in \mathbb{R}.$$

Such models are called *polytopic linear differential inclusions* and arise in a wide variety of modeling problems.

5.2 Robust stability

5.2.1 Time-invariant parametric uncertainty

An important issue in the design of control systems involves the question as to what extend the stability and performance of the controlled system is robust against perturbations and uncertainties in the parameters of the system. In this section we consider the linear time-invariant system defined by

$$\dot{x} = A(\delta)x \tag{5.2.1}$$

where the state matrix $A(\cdot)$ is a *continuous* function of a real valued time-invariant parameter vector $\delta = \operatorname{col}(\delta_1, \ldots, \delta_p)$ which we assume to be contained in an *uncertainty set* $\Delta \subseteq \mathbb{R}^p$. Let $X = \mathbb{R}^n$ be the state space of this system. We will analyze the *robust stability* of the equilibrium point $x^* = 0$ of this system. Precisely, we address the question when the equilibrium point $x^* = 0$ of (5.2.1) is asymptotically stable in the sense of Definition 3.1 for all $\delta \in \Delta$.

Example 5.1 As an example, let

$$A(\delta) = \begin{pmatrix} -1 & 2\delta_1 & 2 \\ \delta_2 & -2 & 1 \\ 3 & -1 & \frac{\delta_3 - 10}{\delta_1 + 1} \end{pmatrix}$$

where δ_1, δ_2 and δ_3 are bounded as $\delta_1 \in [-0.5, 1]$, $\delta_2 \in [-2, 1]$ and $\delta_3 \in [-0.5, 2]$. Then the uncertainty set Δ is polytopic and defined by

$$\Delta = \{\operatorname{col}(\delta_1, \delta_2, \delta_3) \mid \delta_1 \in [-0.5, 1],\ \delta_2 \in [-2, 1],\ \delta_3 \in [-0.5, 2]\}$$

and $\Delta = \operatorname{conv}(\Delta_0)$ with

$$\Delta_0 = \{\operatorname{col}(\delta_1, \delta_2, \delta_3) \mid \delta_1 \in \{-0.5, 1\},\ \delta_2 \in \{-2, 1\},\ \delta_3 \in \{-0.5, 2\}\}$$

the set of vertices (or generators) of Δ.

In the case of time-invariant parametric uncertainties, the system $\dot{x} = A(\delta)x$ is asymptotically stable if and only if $A(\delta)$ is Hurwitz for all $\delta \in \Delta$. That is, if and only if the eigenvalues of $A(\delta)$ lie in the open left-half complex plane for all admissible perturbations $\delta \in \Delta$. Hence, using Proposition 3.5, the verification of robust stability amounts to checking whether

$$\rho(A(\delta)) := \max_{\lambda \in \lambda(A(\delta))} \operatorname{Re}(\lambda) < 0 \quad \text{for all } \delta \in \Delta.$$

If Δ is a continuum in $\mathbb{R}^p$, this means verifying an inequality at an infinite number of points. Even if Δ is a polytope, it will not suffice to check the above inequality on the vertices of the uncertainty set only. Since $\rho(A(\delta)$ is not a convex or concave function of δ (except in a few trivial cases) it will be difficult to find its global maximum over Δ.

Quadratic stability with time-invariant uncertainties

We will apply Theorem 3.4 to infer the asymptotic stability of the equilibrium point $x^* = 0$ of (5.2.1).

Definition 5.2 (Quadratic stability) The system (5.2.1) is said to be *quadratically stable for perturbations* $\delta \in \Delta$ if there exists a matrix $K = K^\top \succ 0$ such that

$$A(\delta)^\top K + KA(\delta) \prec 0 \quad \text{for all } \delta \in \Delta. \tag{5.2.2}$$

The importance of this definition becomes apparent when considering quadratic Lyapunov functions $V(x) = x^\top Kx$. Indeed, if $K \succ 0$ satisfies (5.2.2) then there exists an $\varepsilon > 0$ such that

$$A(\delta)^\top K + KA(\delta) + \varepsilon K \preceq 0 \quad \text{for all } \delta \in \Delta.$$

so that the time-derivative V' of the composite function $V(x(t)) = x(t)^\top Kx(t)$ along solutions of (5.2.1), defined in (3.1.5), satisfies

$$V'(x) + \varepsilon V(x) = \partial_x V(x) A(\delta)x + \varepsilon V(x) = x^\top \left[A(\delta)^\top K + KA(\delta) + \varepsilon K \right] x \leq 0$$

for all $x \in X$ and all $\delta \in \Delta$. But then the composite function $V^*(t) = V(x(t))$ with $x(\cdot)$ a solution of (5.2.1) satisfies

$$\dot{V}^*(t) + \varepsilon V^*(t) = V'(x(t)) + \varepsilon V(x(t)) = x(t)^\top \left[A(\delta)^\top K + KA(\delta) + \varepsilon K \right] x(t) \leq 0$$

for all $t \in \mathbb{R}$ and all $\delta \in \Delta$. Integrating this expression over an interval $[t_0, t_1]$ shows that V^* has exponential decay according to $V^*(t_1) \leq V^*(t_0) e^{-\varepsilon(t_1 - t_0)}$ for all $t_1 \geq t_0$ and all $\delta \in \Delta$. Now use that

$$\lambda_{\min}(K) \|x\|^2 \leq x^\top Kx \leq \lambda_{\max}(K) \|x\|^2$$

to see that

$$\|x(t_1)\|^2 \leq \frac{1}{\lambda_{\min}(K)} V(x(t_1)) \leq \frac{1}{\lambda_{\min}(K)} V(x(t_0)) e^{-\varepsilon(t_1 - t_0)} \leq \frac{\lambda_{\max}(K)}{\lambda_{\min}(K)} \|x(t_0)\|^2 e^{-\varepsilon(t_1 - t_0)}.$$

That is,

$$\|x(t)\| \;\le\; \|x(t_0)\|\sqrt{\frac{\lambda_{\max}(K)}{\lambda_{\min}(K)}}e^{-\frac{\varepsilon}{2}(t-t_0)} \quad \text{for all } \delta \in \Delta$$

which shows that the origin is globally exponentially stable (and hence globally asymptotically stable) for all $\delta \in \Delta$. Moreover, the exponential decay rate, $\varepsilon/2$ does not depend on δ.

It is worthwhile to understand (and perhaps appreciate) the arguments in this reasoning as they are at the basis of more general results to come.

Verifying quadratic stability with time-invariant uncertainties

The verification of quadratic stability of the system places an infinite number of constraints on the symmetric matrix K in (5.2.2). It is the purpose of this section to make additional assumptions on the way the uncertainty enters the system, so as to convert (5.2.2) into a numerically tractable condition.

Suppose that $A(\delta)$ in (5.2.1) is an *affine* function of the parameter vector δ. That is, suppose that there exist real matrices $A_0, \ldots A_p$, all of dimension $n \times n$, such that

$$A(\delta) = A_0 + \delta_1 A_1 + \cdots + \delta_p A_p \tag{5.2.3}$$

with $\delta \in \Delta$. Then (5.2.1) is referred to as an *affine parameter dependent model*. In addition, let us suppose that the uncertainty set Δ is convex and coincides with the convex hull of a set $\Delta_0 \subset \mathbb{R}^p$. With this structure on A and Δ we state the following result.

Proposition 5.3 *If $A(\cdot)$ is an affine function and $\Delta = \mathrm{conv}(\Delta_0)$ with $\Delta_0 \subset \mathbb{R}^p$, then the system* (5.2.1) *is quadratically stable if and only if there exists $K = K^\top \succ 0$ such that*

$$A(\delta)^\top K + KA(\delta) \prec 0 \quad \textit{for all } \delta \in \Delta_0.$$

Proof. The proof of this result is an application of Proposition 1.14 in Chapter 1. Indeed, fix $x \in \mathbb{R}^n$ and consider the mapping $f_x : \Delta \to \mathbb{R}$ defined by

$$f_x(\delta) := x^\top[A(\delta)^\top K + KA(\delta)]x.$$

The domain Δ of this mapping is a convex set and by assumption $\Delta = \mathrm{conv}(\Delta_0)$. Since $A(\cdot)$ is affine, it follows that $f_x(\cdot)$ is a *convex* function. In particular, Proposition 1.14 (Chapter 1) yields that $f_x(\delta) < 0$ for all $\delta \in \Delta$ if and only if $f_x(\delta) < 0$ for all $\delta \in \Delta_0$. Since x is arbitrary, it follows that the inequality $A(\delta)^\top K + KA(\delta) \prec 0$ holds for all $\delta \in \Delta$ if and only if it holds for all $\delta \in \Delta_0$, which yields the result. ■

Obviously, the importance of this result lies in the fact that quadratic stability can be concluded from a *finite test* of matrix inequalities whenever Δ_0 consists of a finite number of elements. That is, when the uncertainty set is the convex hull of a finite number of points in $\mathbb{R}^p$. In that case, the condition stated in Proposition 5.3 is a feasibility test of a (finite) system of LMI's.

Example 5.4 Continuing Example 5.1, the matrix $A(\delta)$ is not affine in δ, but by setting $\delta_4 = \frac{\delta_3 - 10}{\delta_1 + 1} + 12$ we obtain that

$$A(\delta) = \begin{pmatrix} -1 & 2\delta_1 & 2 \\ \delta_2 & -2 & 1 \\ 3 & -1 & \delta_4 - 12 \end{pmatrix}, \qquad \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_4 \end{pmatrix} \in \Delta = [-0.5, 1] \times [-2, 1] \times [-9, 8].$$

Since $\Delta = \text{conv}(\Delta_0)$ with $\Delta_0 = \{\delta^1, \ldots, \delta^N\}$ consisting of the $N = 2^3 = 8$ corner points of the uncertainty set, the verification of the quadratic stability of (5.2.1) is a feasibility test of the (finite) system of LMI's

$$K \succ 0, \qquad A(\delta^j)^\top K + KA(\delta^j) \prec 0, \quad j = 1, \ldots, 8.$$

The test will not pass. By Proposition 5.3, the system is not quadratically stable for the given uncertainty set.

Example 5.5 Consider the uncertain control system $\dot{x} = A(\delta)x + B(\delta)u$ where we wish to construct a feedback law $u = Fx$ such that the controlled system $\dot{x} = (A(delta) + B(\delta)F)x$ is quadratically stable for all δ in some uncertainty set $\Delta = \text{conv}(\Delta_0)$ with Δ_0 a finite set. By Proposition 5.3, this is equivalent of finding F and $K \succ 0$ such that

$$(A(\delta) + B(\delta)F)^\top K + K(A(\delta) + B(\delta)F) \prec 0, \quad \text{for all} \quad \delta \in \Delta_0$$

This is not a system of LMI's. However, with $X = K^{-1}$ and $L = FK^{-1}$ and assuming that $A(\cdot)$ and $B(\cdot)$ are affine, we can transform the latter to an LMI feasibility test to find $X \succ 0$ such that

$$A(\delta)X + XA(\delta) + B(\delta)L + (B(\delta)L)^\top \prec 0, \quad \text{for all} \quad \delta \in \Delta_0.$$

Whenever the test passes, the feedback law is given by $F = LX^{-1}$.

Parameter-dependent Lyapunov functions with time-invariant uncertainties

The main disadvantage in searching for one quadratic Lyapunov function for a class of uncertain models is the conservatism of the test. Indeed, the last example shows that (5.2.1) may not be quadratically stable, but no conclusions can been drawn from this observation concerning the instability of the uncertain system. To reduce conservatism of the quadratic stability test we will consider quadratic Lyapunov functions for the system (5.2.1) which are *parameter dependent*, i.e., Lyapunov functions $V : X \times \Delta \to \mathbb{R}$ of the form

$$V(x, \delta) := x^\top K(\delta)x$$

where $K(\delta)$ is a matrix valued function that is allowed to dependent on the uncertain parameter δ. A sufficient condition for robust asymptotic stability can be stated as follows.

Proposition 5.6 *Let the uncertainty set Δ be compact and suppose that $K(\delta)$ is continuously differentiable on Δ and satisfies*

$$K(\delta) \succ 0 \quad \textit{for all } \delta \in \Delta \tag{5.2.4a}$$

$$A(\delta)^\top K(\delta) + K(\delta)A(\delta) \prec 0 \quad \textit{for all } \delta \in \Delta. \tag{5.2.4b}$$

Then the system (5.2.1) *is globally asymptotically stable for all $\delta \in \Delta$.*

Proof. Let $K(\delta)$ satisfy (5.2.4) and consider $V(x,\delta) = x^\top K(\delta)x$ as candidate Lyapunov function. There exists $\varepsilon > 0$ such that $K(\delta)$ satisfies $A(\delta)^\top K(\delta) + K(\delta)A(\delta) + \varepsilon K(\delta) \preceq 0$. Take the time derivative of the composite function $V^*(t) := V(x(t),\delta)$ along solutions of (5.2.1) to infer that $\dot{V}^*(t)+\varepsilon V^*(t) \leq 0$ for all $t \in \mathbb{R}$ and all $\delta \in \Delta$. This means that for all $\delta \in \Delta$, $V^*(\cdot)$ is exponentially decaying along solutions of (5.2.1) according to $V^*(t) \leq V^*(0)e^{-\varepsilon t}$. Define $a := \inf_{\delta\in\Delta} \lambda_{\min}(K(\delta))$ and $b := \sup_{\delta\in\Delta} \lambda_{\max}(K(\delta))$. If Δ is compact, the positive definiteness of K implies that both a and b are positive and we have that $a\|x\|^2 \leq V(x,\delta) \leq b\|x\|^2$ for all $\delta \in \Delta$ and all $x \in X$. Together with the exponential decay of V^* this yields that $\|x(t)\|^2 \leq \frac{b}{a}\|x(0)\|^2 e^{-\varepsilon t}$ for all $\delta \in \Delta$ which proves the exponential and asymptotic stability of (5.2.1). ∎

The search for matrix valued *functions* that satisfy the conditions 5.2.4 is much more involved and virtually intractable from a computational point of view. We therefore wish to turn Proposition 5.6 into a numerically efficient scheme to compute parameter varying Lyapunov functions. For this, first consider Lyapunov functions that are *affine* in the parameter δ, i.e.,

$$K(\delta) = K_0 + \delta_1 K_1 + \ldots + \delta_p K_p$$

where $K_0, \ldots, K_p$ are real symmetric matrices of dimension $n \times n$ and $\delta = \operatorname{col}(\delta_1, \ldots, \delta_p)$ is the time-invariant uncertainty vector. Clearly, with $K_1 = \ldots = K_p = 0$ we are back to the case of parameter *independent* quadratic Lyapunov functions as discussed in the previous subsection. The system (5.2.1) is called *affine quadratically stable* if there exists matrices $K_0, \ldots, K_p$ such that $K(\delta)$ satisfies the conditions (5.2.4) of Proposition 5.6.

Let $A(\cdot)$ be affine and represented by (5.2.3). Suppose that Δ is convex with $\Delta = \operatorname{conv}(\Delta_0)$ where Δ_0 is a finite set of vertices of Δ. Then the expression

$$L(\delta) := A(\delta)^\top K(\delta) + K(\delta)A(\delta)$$

in (5.2.4b) is, in general, not affine in δ. As a consequence, for fixed $x \in \mathbb{R}^n$, the function $f_x : \Delta \to \mathbb{R}$ defined by

$$f_x(\delta) := x^\top L(\delta)x \tag{5.2.5}$$

will not be convex so that the implication

$$\{f_x(\delta) < 0 \text{ for all } \delta \in \Delta_0\} \quad \Longrightarrow \quad \{f_x(\delta) < 0 \text{ for all } \delta \in \Delta\} \tag{5.2.6}$$

used in the previous section (proof of Proposition 5.3), will not hold. Expanding $L(\delta)$ yields

$$\begin{aligned} L(\delta) &= [A_0 + \sum_{j=1}^{p} \delta_j A_j]^\top [K_0 + \sum_{j=1}^{p} \delta_j K_j] + [K_0 + \sum_{j=1}^{p} \delta_j K_j][A_0 + \sum_{j=1}^{p} \delta_j A_j] \\ &= \sum_{i=0}^{p} \sum_{j=0}^{p} \delta_i \delta_j [A_i^\top K_j + K_j A_i] \end{aligned}$$

where, for the latter compact notation, we introduced $\delta_0 = 1$. Consequently, (5.2.5) takes the form

$$f_x(\delta) = c_0 + \sum_{j=1}^{p} \delta_j c_j + \sum_{j=1}^{p} \sum_{i=1}^{j-1} \delta_i \delta_j c_{ij} + \sum_{j=1}^{p} \delta_j^2 d_j$$

where c_0, c_j, c_{ij} and d_j are constants. Consider the following uncertainty sets

$$\Delta = \{\delta \in \mathbb{R}^p \mid \delta_k \in [\underline{\delta}_k, \bar{\delta}_k]\}, \qquad \Delta_0 = \{\delta \in \mathbb{R}^p \mid \delta_k \in \{\underline{\delta}_k, \bar{\delta}_k\}\} \tag{5.2.7}$$

where $\underline{\delta}_k \leq \bar{\delta}_k$. Then $\Delta = \operatorname{conv}(\Delta_0)$ and it is easily seen that a sufficient condition for the implication (5.2.6) to hold for the uncertainty sets (5.2.7) is that $f_x(\delta_1, \ldots, \delta_j, \ldots, \delta_p)$ is *partially convex*, that is convex in each of its arguments δ_j, $j = 1, \ldots, p$. Since f_x is a twice differentiable function, this is the case when

$$d_j = \frac{1}{2} \frac{\partial^2 f_x}{\partial \delta_j^2} = x^\top [A_j^\top K_j + K_j A_j] x \geq 0$$

for $j = 1, \ldots, p$. Since x is arbitrary, we therefore obtain that

$$A_j^\top K_j + K_j A_j \succeq 0, \qquad j = 1, \ldots, p$$

is a sufficient condition for (5.2.6) to hold on the uncertainty sets (5.2.7). This leads to the following main result.

Theorem 5.7 *If $A(\cdot)$ is an affine function described by* (5.2.3) *and $\Delta = \operatorname{conv}(\Delta_0)$ assumes the form* (5.2.7)*, then the system* (5.2.1) *is affine quadratically stable if there exist real matrices $K_0, \ldots, K_p$ such that $K(\delta) = K_0 + \sum_{j=1}^{p} \delta_j K_j$ satisfies*

$$A(\delta)^\top K(\delta) + K(\delta) A(\delta) \prec 0 \quad \textit{for all } \delta \in \Delta_0 \tag{5.2.8a}$$

$$K(\delta) \succ 0 \quad \textit{for all } \delta \in \Delta_0 \tag{5.2.8b}$$

$$A_j^\top K_j + K_j A_j \succeq 0 \quad \textit{for } j = 1, \ldots, p. \tag{5.2.8c}$$

In that case, the parameter varying function satisfies the conditions (5.2.4) *and $V(x, \delta) := x^\top K(\delta) x$ is a quadratic parameter-dependent Lyapunov function of the system.*

Proof. It suffices to prove that (5.2.8) implies (5.2.4) for all $\delta \in \Delta$. Let x be a non-zero fixed but arbitrary element of $\mathbb{R}^n$. Since $K(\delta)$ is affine in δ, the mapping

$$\delta \mapsto x^\top K(\delta) x$$

with $\delta \in \Delta$ is convex. Consequently, $x^\top K(\delta)x$ is larger than zero for all $\delta \in \Delta$ if it is larger than zero for all $\delta \in \Delta_0$. As x is arbitrary, this yields that (5.2.4a) is implied by (5.2.8b). It now suffices to prove that (5.2.8a) and (5.2.8c) imply (5.2.4b) for all $\delta \in \Delta$. This however, we showed in the arguments preceding the theorem. ■

Theorem 5.7 reduces the problem to verify affine quadratic stability of the system (5.2.1) with box-type uncertainties to a feasibility problem of a (finite) set of linear matrix inequalities.

5.2.2 Time-varying parametric uncertainty

Robust stability against time-varying perturbations is generally a more demanding requirement than robust stability against time-invariant parameter uncertainties. In this section we consider the question of robust stability for the system

$$\dot{x}(t) = A(\delta(t))x(t) \tag{5.2.9}$$

where the values of the time-varying parameter vector $\delta(t)$ belong to the uncertainty set $\Delta \subset \mathbb{R}^p$ for all time $t \in \mathbb{R}$. In this section we assess the *robust stability* of the fixed point $x^* = 0$. It is important to remark that, unlike the case with time-invariant uncertainties, robust stability of the origin of the time-varying system (5.2.9) is *not equivalent* to the condition that the (time-varying) eigenvalues $\lambda(A(\delta(t)))$ belong to the stability region $\mathbb{C}^-$ for all admissible perturbations $\delta(t) \in \Delta$.

Proposition 5.8 *The uncertain system* (5.2.9) *with time-varying uncertainties* $\delta(\cdot) \in \Delta$ *is asymptotically stable if there exists a matrix* $K = K^\top \succ 0$ *such that* (5.2.2) *holds.*

The inequality (5.2.2) is a *time-invariant* condition that needs to be verified for all points in the uncertainty set Δ so as to conclude asymptotic stability of the *time-varying* uncertainties that occur in (5.2.9). Since Proposition 5.8 is obtained as a special case of Theorem 5.9 below, we defer its proof.

An interesting observation related to Proposition 5.8 is that the existence of a real symmetric matrix $K = K^\top \succ 0$ satisfying (5.2.2) not only yields quadratic stability of the system (5.2.1) with $\delta \in \Delta$ but also the asymptotic stability of (5.2.9) with $\delta(t) \in \Delta$. Hence, the existence of such K implies that *arbitrary fast variations* in the time-varying parameter vector $\delta(\cdot)$ may occur so as to guarantee asymptotic stability of (5.2.9). If additional *a priori* information on the time-varying parameters is known, the result of Proposition 5.8 may become too conservative and we therefore may like to resort to different techniques to incorporate information about the parameter trajectories $\delta(\cdot)$.

One way to do this is to assume that the trajectories $\delta(\cdot)$ are continuously differentiable while

$$\delta(t) \in \Delta, \qquad \dot{\delta}(t) \in \Lambda \quad \text{for all time } t \in \mathbb{R}. \tag{5.2.10}$$

Here, Δ and Λ are subsets of $\mathbb{R}^p$ and we will assume that these sets are compact. We will therefore assume that not only the *values* but also the *rates* of the parameter trajectories are constrained.

Quadratic stability with time-varying uncertainties

A central result for achieving robust stability of the system (5.2.9) against all uncertainties (5.2.10) is given in the following theorem.

Theorem 5.9 *Suppose that the function $K : \Delta \to \mathbb{S}^n$ is continuously differentiable on a compact set Δ and satisfies*

$$K(\delta) \succ 0 \quad \text{for all } \delta \in \Delta \tag{5.2.11a}$$

$$\partial_\delta K(\delta)\lambda + A(\delta)^\top K(\delta) + K(\delta)A(\delta) \prec 0 \quad \text{for all } \delta \in \Delta \text{ and } \lambda \in \Lambda. \tag{5.2.11b}$$

Then the origin of the system (5.2.9) *is exponentially stable against all time-varying uncertainties $\delta : \mathbb{R} \to \mathbb{R}^p$ that satisfy* (5.2.10). *Moreover, in that case $V(x, \delta) := x^\top K(\delta)x$ is a quadratic parameter depending Lyapunov function for the system* (5.2.9).

Proof. The proof follows very much the same lines as the proof of Proposition 5.6, but now includes time-dependence of the parameter functions. Suppose that $K(\delta)$ satisfies the hypothesis. Consider $V(x, \delta) = x^\top K(\delta)x$ as candidate Lyapunov function. Let $a := \inf_{\delta\in\Delta} \lambda_{\min} K(\delta)$ and $b := \sup_{\delta\in\Delta} \lambda_{\max} K(\delta)$. If Δ is compact, the positive definiteness of $K(\delta)$ for all $\delta \in \Delta$ implies that both a and b are positive. In addition, we can find $\varepsilon > 0$ such that $K(\delta)$ satisfies

$$aI \preceq K(\delta) \preceq bI, \quad \partial_\delta K(\delta)\lambda + A(\delta)^\top K(\delta) + K(\delta)A(\delta) + \varepsilon K(\delta) \preceq 0$$

for all $\delta \in \Delta$ and $\lambda \in \Lambda$. Take the time derivative of the composite function $V^*(t) := V(x(t), \delta(t))$ along solutions of (5.2.9) to infer that

$$\begin{aligned}\dot{V}^*(t) + \varepsilon V^*(t) &= x(t)^\top A(\delta(t))^\top K(\delta(t))x(t) + x(t)^\top K(\delta(t))A(\delta(t))x(t)+ \\ &+ \varepsilon x(t)^\top K(\delta(t))x(t) + x(t)^\top \left\{ \sum_{k=1}^{p} \partial_k K(\delta(t))\dot{\delta}_k(t) \right\} x(t) \ \le \ 0\end{aligned}$$

for all $t \in \mathbb{R}$, all $\delta(t) \in \Delta$ and all $\dot{\delta}(t) \in \Lambda$. This means that for this class of uncertainties $V^*(\cdot)$ is exponentially decaying along solutions of (5.2.1) according to $V^*(t) \le V^*(0)e^{-\varepsilon t}$. Moreover, since $a\|x\|^2 \le V(x, \delta) \le b\|x\|^2$ for all $\delta \in \Delta$ and all $x \in X$ we infer that $\|x(t)\|^2 \le \frac{b}{a}\|x(0)\|^2 e^{-\varepsilon t}$ for all $t \ge 0$ and all uncertainties $\delta(t)$ satisfying (5.2.10). Hence, (5.2.9) is exponentially stable against uncertainties (5.2.10). ∎

Theorem 5.9 involves a search for matrix functions satisfying the inequalities (5.2.11) to guarantee robust asymptotic stability. Note that the result is a sufficient algebraic test only that provides a quadratic parameter dependent Lyapunov function, when the test passes. The result is not easy to apply or verify by a computer program as it involves (in general) an infinite number of conditions on the inequalities (5.2.11). We will therefore focus on a number of special cases that convert Theorem 5.9 in a feasible algorithm.

For this, first consider the case where the parameters are time-invariant. This is equivalent to saying that $\Lambda = \{0\}$. The conditions (5.2.11) then coincide with (5.2.4) and we therefore obtain Proposition 5.6 as a special case. In particular, the sufficient condition (5.2.11) for robust stability in Theorem 5.9 is also necessary in this case.

If we assume arbitrary fast time-variations in $\delta(t)$ then we consider rate constraints of the form $\Lambda = [-r, r]^p$ with $r \to \infty$. For (5.2.11b) to hold for any λ with $|\lambda_k| > r$ and $r \to \infty$ it is immediate that $\partial_\delta K(\delta)$ needs to vanish for all $\delta \in \Delta$. Consequently, in this case K can not depend on δ and Theorem 5.9 reduces to Proposition 5.8. In particular, this argument proves Proposition 5.8 as a special case.

Verifying quadratic stability with time-varying uncertainties

In this section we will assume that $A(\cdot)$ in (5.2.9) is an *affine function* of $\delta(t)$. The uncertainty sets Δ and Λ are assumed to be convex sets defined by the 'boxes'

$$\Delta = \{\delta \in \mathbb{R}^p \mid \delta_k \in [\underline{\delta}_k, \bar{\delta}_k]\}, \qquad \Lambda = \{\lambda \in \mathbb{R}^p \mid \lambda_k \in [\underline{\lambda}_k, \bar{\lambda}_k]\} \tag{5.2.12}$$

Stated otherwise, the uncertainty regions are the convex hulls of the sets

$$\Delta_0 = \{\delta \in \mathbb{R}^p \mid \delta_k \in \{\underline{\delta}_k, \bar{\delta}_k\}\ \}, \qquad \Lambda_0 = \{\lambda \in \mathbb{R}^p \mid \lambda_k \in \{\underline{\lambda}_k, \bar{\lambda}_k\}\ \}.$$

In addition, the search of a parameter dependent $K(\delta)$ will be restricted to the class of *affine functions* $K(\delta)$ represented by

$$K(\delta) = K_0 + \delta_1 K_1 + \cdots + \delta_p K_p$$

where $K_j \in \mathbb{S}^n$, $j = 1, \ldots, p$ is symmetric. For this class of parameter dependent functions we have that $\partial_{\delta_k} K(\delta) = K_k$ so that (5.2.11b) reads

$$\sum_{k=1}^{p} \partial_{\delta_k} K(\delta)\lambda_k \,+\, A(\delta)^\top K(\delta) \,+\, K(\delta)A(\delta) \;=\; \sum_{k=1}^{p} K_k\lambda_k + \sum_{\nu=0}^{p}\sum_{\mu=0}^{p} \delta_\nu\delta_\mu (A_\nu^\top K_\mu \,+\, K_\mu A_\nu).$$

Here, we set $\delta_0 = 1$ to simplify notation. The latter expression is

- affine in $K_1, \ldots, K_p$
- affine in $\lambda_1, \ldots, \lambda_p$
- quadratic in $\delta_1, \ldots, \delta_p$ due to the mixture of constant, linear and quadratic terms.

Again, similar to (5.2.5), introduce for arbitrary $x \in X$ the function $f_x : \Delta \times \Lambda \to \mathbb{R}$ defined by

$$f_x(\delta, \lambda) := x^\top \left[\sum_{k=1}^{p} K_k\lambda_k + \sum_{\nu=0}^{p}\sum_{\mu=0}^{p} \delta_\nu\delta_\mu (A_\nu^\top K_\mu + K_\mu A_\nu) \right] x.$$

A sufficient condition for the implication

$$\{ f_x(\delta,\lambda) < 0 \text{ for all } (\delta,\lambda) \in \Delta_0 \times \Lambda_0\} \Longrightarrow \{ f_x(\delta,\lambda) < 0 \text{ for all } (\delta,\lambda) \in \Delta \times \Lambda\}$$

is that f_x is *partially convex* in each of its arguments δ_j, $j = 1, \ldots, p$. As in subsection 5.2.1, this condition translates to the requirement that

$$A_j^\top K_j + K_j A_j \succeq 0, \qquad j = 1, \ldots, p,$$

which brings us to the following main result.

Theorem 5.10 *Suppose that $A(\cdot)$ is affine as described by* (5.2.3)*, and assume that $\delta(t)$ satisfies the constraints* (5.2.10) *with Δ and Λ compact sets specified in* (5.2.12)*. Then the origin of the system* (5.2.9) *is robustly asymptotically stable against all time-varying uncertainties that satisfy* (5.2.10) *if there exist real matrices $K_0, \ldots, K_p$ such that $K(\delta) = K_0 + \sum_{j=1}^p \delta_j K_j$ satisfies*

$$\sum_{k=1}^{p} K_k \lambda_k + \sum_{k=0}^{p}\sum_{\ell=0}^{p} \delta_k \delta_\ell \left(A_k^\top K_\ell + K_\ell A_k \right) \prec 0 \quad \text{for all } \delta \in \Delta_0 \text{ and } \lambda \in \Lambda_0 \tag{5.2.13a}$$

$$K(\delta) \succ 0 \quad \text{for all } \delta \in \Delta_0 \tag{5.2.13b}$$

$$A_j^\top K_j + K_j A_j \succeq 0 \quad \text{for } j = 1, \ldots, p. \tag{5.2.13c}$$

Moreover, in that case, $V(x,\delta) := x^\top K(\delta) x$ defines a quadratic parameter-dependent Lyapunov function for the system.

Theorem 5.10 provides an LMI feasibility test to verify robust asymptotic stability against uncertainty described by (5.2.10).

It is interesting to compare the numerical complexity of the conditions of Theorem 5.7 with the conditions mentioned in Theorem 5.10. If the uncertainty vector δ is p-dimensional then the vertex set Δ_0 has dimension 2^p so that the verification of conditions (5.2.11) requires a feasibility test of

$$2^p + 2^p + p$$

linear matrix inequalities. In this case, also the vertex set Λ_0 has dimension 2^k which implies that the condition of Theorem 5.10 require a feasibility test of

$$2^{2p} + 2^p + p = 4^p + 2^p + p$$

linear matrix inequalities.

Generalizations

Assuming an affine structure for the state evolution map $A(\cdot)$ and for the matrix function $K(\cdot)$, we were able (Theorem 5.10) to restrict the search of a parameter dependent Lyapunov function for the system (5.2.9) to a finite dimensional subspace. The central idea that led to Theorem 5.10 allows many generalizations to non-affine structures. For example, if $b_1(\delta)$, $b_2(\delta)$, $\ldots$, $b_p(\delta)$ denote a set of scalar, continuously differentiable basis functions in the uncertain parameter δ, we may assume that $A(\delta)$ and $K(\delta)$ allow for expansions

$$A(\delta) = A_1 b_1(\delta) + \cdots + A_p b_p(\delta)$$
$$K(\delta) = K_1 b_1(\delta) + \cdots + K_p b_p(\delta).$$

The condition (5.2.11b) in Theorem 5.9 then involves the partial derivative

$$\partial_k K(\delta) = \sum_{j=1}^{p} K_j \partial_k b_j(\delta).$$

The robust stability conditions (5.2.11) in Theorem 5.9 then translate to

$$\sum_{k=1}^{p} K_k b_k(\delta) \succ 0 \quad \text{for all } \delta \in \Delta \tag{5.2.14a}$$

$$\sum_{k=1}^{p} \left(\sum_{j=1}^{p} K_j \partial_k b_j(\delta) \lambda_k + [A(\delta)^\top K_j + K_j A(\delta)] b_j(\delta) \right) \prec 0 \quad \text{for all } \delta \in \Delta \text{ and } \lambda \in \Lambda. \tag{5.2.14b}$$

which is finite dimensional but not yet an LMI feasibility test. Possible basis functions are

(a) **natural basis**

$$b_1(\delta) = \delta_1, \ldots, b_p(\delta) = \delta_p$$

where $\delta_j = \langle e_j, \delta \rangle$ is the jth component of δ in the natural basis $\{e_j\}_{j=1}^p$ of $\mathbb{R}^p$.

(b) **polynomial basis**

$$b_{k_1,\ldots,k_p}(\delta) = \delta_1^{k_1} \cdots \delta_p^{k_p}, \quad k_\nu = 0, 1, 2, \ldots, \quad \text{and } \nu = 1, \ldots, p.$$

As a conclusion, in this section we obtained from the nominal stability characterizations in Chapter 3 the corresponding robust stability tests against both time-invariant and time-varying and rate-bounded parametric uncertainties. We continue this chapter to also generalize the performance characterizations in Chapter 3 to robust performance.

5.3 Robust dissipativity

Among the various refinements and generalizations of the notion of a dissipative dynamical system, we mentioned in Section 2.2 the idea of a robustly dissipative system. To make this more precise, let $s : W \times Z \to \mathbb{R}$ be a *supply function* associated with the uncertain system (5.1.1) where the uncertain parameter $\delta(\cdot)$ satisfies the constraints (5.2.10).

Definition 5.11 (Robust dissipativity) The system (5.1.1) with supply function s is said to be *robustly dissipative* against time-varying uncertainties (5.2.10) if there exists a function $V : X \times \Delta \to \mathbb{R}$ such that the dissipation inequality

$$V(x(t_0), \delta(t_0)) + \int_{t_0}^{t_1} \mathrm{s}(w(t), z(t))dt \;\geq\; V(x(t_1), \delta(t_1)) \tag{5.3.1}$$

holds for all $t_0 \leq t_1$ and all signals (w, x, z, δ) that satisfy (5.1.1) and (5.2.10).

Any function V that satisfies (5.3.1) is called a *(parameter dependent) storage function* and (5.3.1) is referred to as the *robust dissipation inequality*. If the composite function $V(x(t), \delta(t))$ is differentiable as a function of time t, then it is easily seen that the system (5.3.1) is robustly dissiptive if and only if

$$\partial_x V(x, \delta) f(x, w, \delta) + \partial_\delta V(x, \delta)v \leq \mathrm{s}(w, g(x, w, \delta)) \tag{5.3.2}$$

holds for all points $(x, w, \delta, v) \in \mathbb{R}^n \times \mathbb{R}^m \times \Delta \times \Lambda$. The latter *robust differential dissipation inequality* (5.3.2) makes robust dissipativity a *de facto* local property of the functions f, g, the supply function s and the uncertainty sets Δ and Λ.

As in Chapter 2, we will specialize this to linear systems with quadratic supply functions and derive explicit tests for the verification of robust dissipativity. Suppose that f and g in (5.1.1) are linear in x and w. This results in the model (5.1.2) in which the uncertainty δ will be time-varying. In addition, suppose that the supply function is a general quadratic form in (w, z) given by (2.3.2) in Chapter 2. A sufficient condition for robust dissipativity is then given as follows.

Theorem 5.12 *Suppose that the function $K : \Delta \to \mathbb{S}^n$ is continuously differentiable on a compact set Δ and satisfies*

$$F(K, \delta, \lambda) := \begin{pmatrix} \partial_\delta K(\delta)\lambda + A(\delta)^\top K(\delta) + K(\delta)A(\delta) & K(\delta)B(\delta) \\ B(\delta)^\top K(\delta) & 0 \end{pmatrix} - \begin{pmatrix} 0 & I \\ C(\delta) & D(\delta) \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} 0 & I \\ C(\delta) & D(\delta) \end{pmatrix} \preceq 0. \tag{5.3.3}$$

for all $\delta \in \Delta$ and all $\lambda \in \Lambda$. Then the uncertain system

$$\begin{cases} \dot{x}(t) = A(\delta(t))x(t) + B(\delta(t))w(t) \\ z(t) = C(\delta(t))x(t) + D(\delta(t))w(t) \end{cases} \tag{5.3.4}$$

where $\delta : \mathbb{R} \to \mathbb{R}^p$ is in the class of continuously differentiable functions satisfying the value and rate constraints (5.2.10) *is robustly dissipative with respect to the quadratic supply function*

$$\mathrm{s}(w, z) = \begin{pmatrix} w \\ z \end{pmatrix}^\top \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix} \begin{pmatrix} w \\ z \end{pmatrix}.$$

Moreover, in that case, $V(x, \delta) := x^\top K(\delta)x$ is a parameter dependent storage function.

Proof. With $V(x, \delta) := x^\top K(\delta)x$, the robust differential dissipation inequality (5.3.2) reads $\mathrm{col}(x, w)F(K, \delta, v)\,\mathrm{col}(x, w) < 0$ for all x, w, $\delta \in \Delta$ and $\lambda \in \Lambda$. In particular, if (5.3.3) holds for all $(\delta, \lambda) \in \Delta \times \Lambda$, then (5.1.2) is robustly dissipative. ■

Theorem 5.12 provides a sufficient condition for robust dissipativity. We remark that the condition (5.3.3) is also necessary if the class of storage function $V(x,\delta)$ that satisfy the dissipation inequality (5.3.1) is restricted to functions that are *quadratic* in x. A result similar to Theorem 5.12 can be derived for the notion of a robust *strictly* dissipative system. Furthermore, if the uncertainty parameter δ is time-invariant, then $\Lambda = \{0\}$ and $F(K,\delta,\lambda)$ simplifies to an expression that does not depend on λ. Only in this case, one can apply the Kalman-Yakubovich-Popov Lemma 2.12 to infer that (5.3.3) is equivalent to to the frequency domain characterization: ...

(to be continued)

5.4 Robust performance

5.4.1 Robust quadratic performance

Consider the uncertain input-output system as described by (5.3.4) and suppose that $\delta : \mathbb{R} \to \mathbb{R}^p$ is in the class of continuously differentiable functions satisfying the value and rate constraints (5.2.10) with Δ and Λ both compact subsets of $\mathbb{R}^p$.

Proposition 5.13 *Consider the uncertain system* (5.3.4) *where δ satisfies the value and rate-constraints* (5.2.10) *with Δ and Λ compact. Let*

$$P = \begin{pmatrix} Q & S \\ S^\top & R \end{pmatrix}$$

be a real symmetric matrix with $R \succeq 0$. Suppose there exists a continuously differentiable Hermitian-valued function $K(\delta)$ such that $K(\delta) \succ 0$ and

$$\begin{pmatrix} \partial_\delta K(\delta)\lambda + A(\delta)^\top K(\delta) + K(\delta)A(\delta) & K(\delta)B(\delta) \\ B(\delta)^\top K(\delta) & 0 \end{pmatrix} + \begin{pmatrix} 0 & I \\ C(\delta) & D(\delta) \end{pmatrix}^\top P \begin{pmatrix} 0 & I \\ C(\delta) & D(\delta) \end{pmatrix} \prec 0$$

for all $\delta \in \Delta$ and $\lambda \in \Lambda$. Then

(a) *the uncertain system* (5.3.4) *is exponentially stable and*

(b) *there exists $\varepsilon > 0$ such that for $x(0) = 0$ and for all $w \in \mathcal{L}_2$ and all uncertain parameter functions $\delta(\cdot)$ satisfying* (5.2.10) *we have*

$$\int_0^\infty \begin{pmatrix} w(t) \\ z(t) \end{pmatrix}^\top P \begin{pmatrix} w(t) \\ z(t) \end{pmatrix} dt \leq -\varepsilon^2 \int_0^\infty w^\top(t)w(t)dt \tag{5.4.1}$$

5.5 Further reading

5.6 Exercises

Exercise 1
Consider the system $\dot{x} = A(\delta)x$ with $A(\delta)$ and Δ defined in Example 5.4.

(a) Show that $A(\delta)$ is quadratically stable on the set 0.4Δ.

(b) Derive a quadratic stability test for the eigenvalues of $A(\delta)$ being located in a disc of radius r around 0.

(c) Can you find the smallest radius r by convex optimization?

Exercise 2
Time-invariant perturbations and arbitrary fast perturbations can be viewed as two extreme cases of time-varying uncertainty sets of the type (5.2.10). These two extreme manifestations of time-varying perturbations reduce Theorem 5.9 to two special cases.

(a) Show that the result of Theorem 5.7 is obtained as a special case of Theorem 5.10 if $\Lambda = \{0\}$.

(b) Show that if $\Lambda = [-r, r]^p$ with $r \to \infty$ then the matrices $K_0, \ldots, K_k$ satisfying the conditions of Theorem 5.10 necessarily satisfy $K_1 = \ldots = K_k = 0$.

The latter property implies that with arbitrary fast perturbations the only affine parameter-dependent Lyapunov matrices $K(\delta) = K_0 + \sum_{j=1}^{k} \delta_j K_j$ are the constant (parameter-independent) ones. It is in this sense that Theorem 5.10 reduces to Proposition 5.3 for arbitrarily fast perturbations.

Exercise 3
Reconsider the suspension system of Exercise 3 in Chapter 4. Suppose that the road profile $q_0 = 0$ and the active suspension force $F = 0$. Let $\bar{k} = 50000$ and $\bar{b} = 5000$. The suspension damping is a time-varying uncertain quantity with

$$b_2(t) \in [50 \times 10^3 - \bar{b}, 50 \times 10^3 + \bar{b}], \quad t \geq 0 \tag{5.6.1}$$

and the suspension stiffness is a time-varying uncertainty parameter with

$$k_2(t) \in [500 \times 10^3 - \bar{k}, 500 \times 10^3 + \bar{k}], \quad t \geq 0. \tag{5.6.2}$$

Let

$$\delta = \begin{pmatrix} b_2 \\ k_2 \end{pmatrix}$$

be the vector containing the uncertain physical parameters.

(a) Let $x = \mathrm{col}(q_1, q_2, \dot{q}_1, \dot{q}_2)$ denote the state of this system and write this system in the form (5.2.1). Verify whether $A(\delta)$ is affine in the uncertainty parameter δ.

(b) Use Proposition 5.3 to verify whether this system is quadratically stable. If so, give a quadratic Lyapunov function for this system.

(c) Calculate vertex matrices $A_1, \ldots, A_k$ such that

$$A(\delta) \in \mathrm{conv}(A_1, \ldots, A_k)$$

for all δ satisfying the specifications.

(d) Suppose that b_2 and k_2 are time-varying and that their rates of variation satisfy

$$|\dot{b}_2| \leq \beta \tag{5.6.3a}$$

$$|\dot{k}_2| \leq \kappa \tag{5.6.3b}$$

where $\beta = 1$ and $\kappa = 3.7$. Use Theorem 5.10 to verify whether there exists a parameter dependent Lyapunov function that proves affine quadratic stability of the uncertain system. If so, give such a Lyapunov function.

Exercise 4
In Exercise 8 of Chapter 3 we considered the batch chemical reactor where the series reaction

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C$$

takes place. k_1 and k_2 are the kinetic rate constants of the conversions from product A to B and from product B to product C, respectively. We will be interested in the concentration C_B of product B.

(a) Show that C_B satisfies the differential equation $\ddot{C}_B + (k_1 + k_2)\dot{C}_B + k_1 k_2 C_B = 0$ and represent this system in state space form with state $x = \mathrm{col}(C_A, C_B)$.

(b) Show that the state space system is of the form (5.2.1) where A is an affine function of the kinetic rate constants.

(c) Verify whether this system is quadratically stable in view of jointly uncertain kinetic constants k_1 and k_2 in the range $[.1, 1]$. If so, calculate a Lyapunov function for the uncertain system.

(d) At time $t = 0$ the reactor is injected with an initial concentration $C_{A0} = 10$ (mol/liter) of reactant A while the concentrations $C_B(0) = C_C(0) = 0$. Plot the time evolution of the concentration C_B of reactant B if

$$k_1(t) = 1 - 0.9\exp(-t); \qquad k_2(t) = 0.1 + 0.9\exp(-t)$$

Exercise 5
Let $f : \Delta \to \mathbb{R}$ be partially convex and suppose that $\Delta = \{\delta \mid \delta_k \in [\underline{\delta}_k, \bar{\delta}_k], k = 1, \ldots, p\}$ with

$\underline{\delta}_k \leq \bar{\delta}_k$. Let $\Delta_0 = \{\delta \mid \delta_k \in \{\underline{\delta}_k, \bar{\delta}_k\}, k = 1, \ldots, p\}$ be the corresponding set of corner points. Show that for all $\gamma \in \mathbb{R}$ we have that

$$f(\delta) \leq \gamma \text{ for all } \delta \in \Delta$$

if and only if

$$f(\delta) \leq \gamma \text{ for all } \delta \in \Delta_0.$$

Chapter 6

Robust linear algebra problems

Control systems need to operate well in the face of many types of uncertainties. That is, stability and performance of controlled systems need to be as robust as possible against perturbations, model uncertainties and un-modelled dynamics. These uncertainties may be known, partly known or completely unknown. It will be one of the fundamental insights in this chapter that guarantees for robust stability and robust performance against linear time invariant uncertainties of complex systems can be reduced to robust linear algebra problems. This has the main advantage that complex and diverse robustness questions in control can be transformed to a specific algebraic problem and subsequently be solved. It is of quite some independent interest to understand the emergence of robustness in algebraic problems and to convey the essential techniques how to handle them. We therefore turn in this chapter to a practically most relevant problem in linear algebra, the problem of robust approximation by least-squares.

6.1 Robust least-squares

The classical least squares approximation problem is perfectly understood when it comes to questions about existence, uniqueness, and representations of its solutions. Today, highly efficient algorithms exist for solving this problem, and it has found widespread applications in diverse fields of applications. The purpose of this section is to introduce the robust counterpart of least squares approximation.

Given matrices $M \in \mathbb{R}^{q\times r}$ and $N \in \mathbb{R}^{p\times r}$ and with the Frobenius norm $\|\cdot\|_F$, the least squares approximation problem amounts to solving

$$\min_{X\in\mathbb{R}^{p\times q}} \|XM - N\|_F. \tag{6.1.1}$$

It will be explained below why we start with the somewhat unusual representation $XM - N$ of the residual. First note that

$$\|XM - N\|_F < \gamma \quad \text{if and only if} \quad \sum_{j=1}^{r} \frac{1}{\gamma} e_j^\top (XM - N)^\top (XM - N) e_j < \gamma.$$

Therefore, if we define

$$G(X) := \begin{pmatrix} XMe_1 - Ne_1 \\ \vdots \\ XMe_r - Ne_r \end{pmatrix} \quad \text{and} \quad P_\gamma := \begin{pmatrix} -\gamma I & 0 \\ 0 & \frac{1}{\gamma} I \end{pmatrix},$$

then G is an affine function and it follows that

$$\|XM - N\|_F < \gamma \qquad \text{if and only if} \qquad \begin{pmatrix} I \\ G(X) \end{pmatrix}^\top P_\gamma \begin{pmatrix} I \\ G(X) \end{pmatrix} < 0.$$

Therefore, we can compute X which minimizes γ by solving an LMI-problem. In practice, it often happens that the matrices M and N depend (continuously) on an uncertain parameter vector δ that is known to be contained in some (compact) parameter set $\boldsymbol{\delta} \subset \mathbb{R}^m$. To express this dependence, we write $M(\delta)$ and $N(\delta)$ for M and N and for a given matrix $X \in \mathbb{R}^{p\times q}$, the *worst case squared residual* is the number

$$\max_{\delta\in\boldsymbol{\delta}} \|XM(\delta) - N(\delta)\|_F^2. \tag{6.1.2}$$

The *robust least squares problem* then amounts to finding X which renders this worst-case residual as small as possible. That is, it requires solving the min-max problem

$$\inf_{X\in\mathbb{R}^{p\times q}} \max_{\delta\in\boldsymbol{\delta}} \|XM(\delta) - N(\delta)\|_F^2.$$

With

$$G(\delta, X) := \begin{pmatrix} XM(\delta)e_1 - N(\delta)e_1 \\ \vdots \\ XM(\delta)e_r - N(\delta)e_r \end{pmatrix}$$

optimal or almost optimal solutions (see Chapter 1) to this problem are obtained by minimizing γ such that the inequality

$$\begin{pmatrix} I \\ G(\delta, X) \end{pmatrix}^\top P_\gamma \begin{pmatrix} I \\ G(\delta, X) \end{pmatrix} < 0 \qquad \text{for all} \quad \delta \in \boldsymbol{\delta}$$

is feasible in X. We thus arrive at a semi-infinite LMI problem. The main goal of this chapter is to show how to re-formulate this semi-infinite LMI-problem into a finite-dimensional problem if $M(\delta)$ and $N(\delta)$ depend *rationally* on the parameter δ.

6.2 Linear fractional representations of rational functions

Suppose that $G(\delta)$ is a real matrix-valued function which depends on a parameter vector $\delta = \mathrm{col}(\delta_1, \ldots, \delta_m) \in \mathbb{R}^m$. Throughout this section it will be essential to identify, as usual, the *matrix* $G(\delta)$ with the *linear function* which maps a vector ξ into the vector $\eta = G(\delta)\xi$.

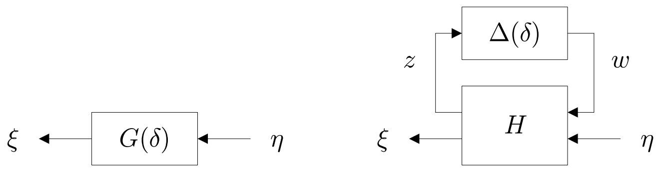


Figure 6.1: Linear Fractional Representation

Definition 6.1 A *linear fractional representation* (LFR) of $G(\delta)$ is a pair $(H, \Delta(\delta))$ where

$$H = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

is a constant partitioned matrix and Δ a *linear* function of δ such that for all δ for which $I - A\Delta(\delta)$ is invertible and all (ξ, η) there holds

$$\eta = G(\delta)\xi \tag{6.2.1}$$

if and only if there exist vectors w and z such that

$$\begin{pmatrix} z \\ \eta \end{pmatrix} = \underbrace{\begin{pmatrix} A & B \\ C & D \end{pmatrix}}_{H} \begin{pmatrix} w \\ \xi \end{pmatrix}, \qquad w = \Delta(\delta)z. \tag{6.2.2}$$

This relation is pictorially expressed in Figure 6.1. We will sometimes call $\Delta(\delta)$ the *parameter-block* of the LFR, and the LFR is said to be *well-posed* at δ if $I - A\Delta(\delta)$ is invertible. If (6.2.2) is a well posed LFR of $G(\delta)$ at δ then

$$\begin{aligned}
\exists\, w, z : (6.2.2) \quad &\iff \quad \exists\, z :\ (I - A\Delta(\delta))z = B\xi \quad \eta = C\Delta(\delta)z + D\xi \\
&\iff \quad \exists\, z :\ z = (I - A\Delta(\delta))^{-1}B\xi \quad \eta = C\Delta(\delta)z + D\xi \\
&\iff \quad \eta = [D + C\Delta(\delta)(I - A\Delta(\delta))^{-1}B]\xi \\
&\iff: \quad \eta = [\Delta(\delta) \star H]\xi.
\end{aligned}$$

Hence, in that case, the equations (6.2.2) define a linear mapping in that $G(\delta) = \Delta(\delta) \star H$. Here, the 'star product' $\Delta(\delta) \star H$ is defined by the latter expression and will implicitly mean that $I - A\Delta(\delta)$ is invertible. Since $\Delta(\delta)$ is linear in δ, this formula reveals that any $G(\delta)$ which admits an LFR $(H, \Delta(\delta))$ must be a *rational function* of δ. Moreover, since $G(0) = \Delta(0) \star H = D$ it is clear that zero is not a pole of G. The main point of this section is to prove that the converse is true as well. That is, we will show that *any matrix-valued multivariable rational function without pole in zero admits an LFR.*

Before proving this essential insight we first derive some elementary operations that allow manipulations with LFR's. This wil prove useful especially in later chapters where all operations on LFR's are fully analogous to manipulating (feedback) interconnections of linear systems and their

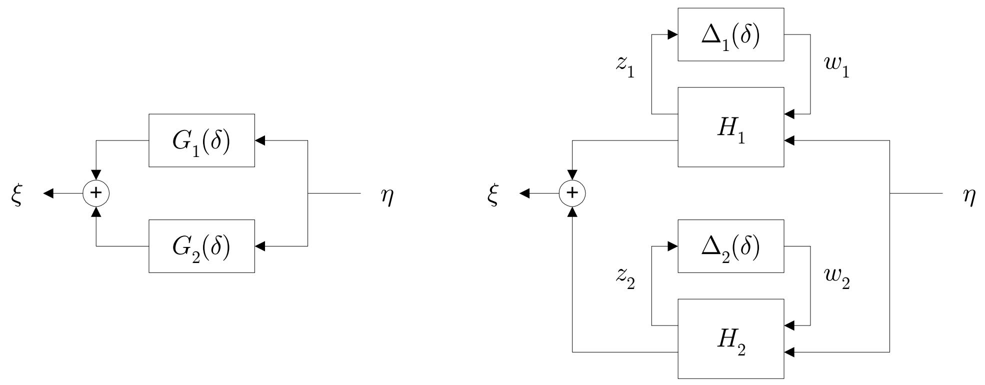


Figure 6.2: Sum of LFR's

state-space realizations. Let us summarize some of the most important operations for a single LFR $G(\delta) = \Delta(\delta) \star H$ or for two LFR's $G_1(\delta) = \Delta_1(\delta) \star H_1$ and $G_2(\delta) = \Delta_2(\delta) \star H_2$ assuming compatible matrix dimensions. We refer to [**?**, Chapter 10] for a nice collection of a variety of other operations and configurations that can all be easily derived.

Summation
As depicted in Figure 6.2, $\eta = [G_1(\delta) + G_2(\delta)]\xi$ admits the LFR

$$\begin{pmatrix} z_1 \\ z_2 \\ \hline \eta \end{pmatrix} = \left(\begin{array}{cc|c} A_1 & 0 & B_1 \\ 0 & A_2 & B_2 \\ \hline C_1 & C_2 & D \end{array}\right) \begin{pmatrix} w_1 \\ w_2 \\ \hline \xi \end{pmatrix}, \qquad \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} \Delta_1(\delta) & 0 \\ 0 & \Delta_2(\delta) \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

Multiplication
As shown in Figure 6.3, $\eta = [G_1(\delta) G_2(\delta)]\xi$ admits the LFR

$$\begin{pmatrix} z_1 \\ z_2 \\ \hline \eta \end{pmatrix} = \left(\begin{array}{cc|c} A_1 & B_1 C_2 & B_1 D_2 \\ 0 & A_2 & B_2 \\ \hline C_1 & D_1 C_2 & D_1 D_2 \end{array}\right) \begin{pmatrix} w_1 \\ w_2 \\ \hline \xi \end{pmatrix}, \qquad \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} \Delta_1(\delta) & 0 \\ 0 & \Delta_2(\delta) \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

Augmentation
For arbitrary matrices L and R, the augmented system $\eta = [L G(\delta) R]\xi$ admits the LFR

$$\begin{pmatrix} z \\ \eta \end{pmatrix} = \begin{pmatrix} A & BR \\ LC & LDR \end{pmatrix} \begin{pmatrix} w \\ \xi \end{pmatrix}, \qquad w = \Delta(\delta) z.$$

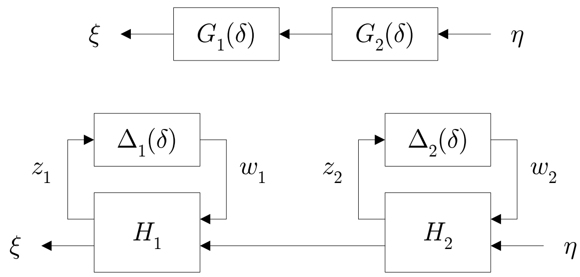


Figure 6.3: Product of LFR's

This allows to construct LFR's of row, column and diagonal augmentations by summation:

$$\begin{aligned}
\begin{pmatrix} G_1(\delta) & G_2(\delta) \end{pmatrix} &= G_1(\delta)\begin{pmatrix} I & 0 \end{pmatrix} + G_2(\delta)\begin{pmatrix} 0 & I \end{pmatrix}\\
\begin{pmatrix} G_1(\delta) \\ G_2(\delta) \end{pmatrix} &= \begin{pmatrix} I \\ 0 \end{pmatrix} G_1(\delta) + \begin{pmatrix} 0 \\ I \end{pmatrix} G_2(\delta)\\
\begin{pmatrix} G_1(\delta) & 0 \\ 0 & G_2(\delta) \end{pmatrix} &= \begin{pmatrix} I \\ 0 \end{pmatrix} G_1(\delta)\begin{pmatrix} I & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ I \end{pmatrix} G_2(\delta)\begin{pmatrix} 0 & I \end{pmatrix}.
\end{aligned}$$

Inversion
If D is invertible then $\xi = G(\delta)^{-1}\eta$ admits the LFR

$$\begin{pmatrix} z \\ \xi \end{pmatrix} = \begin{pmatrix} A - BD^{-1}C & -BD^{-1} \\ D^{-1}C & D^{-1} \end{pmatrix}\begin{pmatrix} w \\ \eta \end{pmatrix}, \qquad w = \Delta(\delta)z.$$

LFR of LFR is LFR
As illustrated in Figure 6.4, if $I - D_1A_2$ is invertible then $\eta = [G_1(\delta) \star H_2]\xi$ admits the LFR

$$\begin{pmatrix} z_1 \\ \xi_2 \end{pmatrix} = \underbrace{\begin{pmatrix} A_1 + B_1A_2(I - D_1A_2)^{-1}C_1 & B_1(I - A_2D_1)^{-1}B_2 \\ C_2(I - D_1A_2)^{-1}C_1 & D_2 + C_2D_1(I - A_2D_1)^{-1}B_2 \end{pmatrix}}_{H_{12}}\begin{pmatrix} w_1 \\ \eta_2 \end{pmatrix}$$

$$w_1 = \Delta_1(\delta)z_1.$$

In other words, $(\Delta_1(\delta) \star H_1) \star H_2 = \Delta_1(\delta) \star H_{12}$ with H_{12} as defined here. This expression is relevant for the shifting and scaling of parameters. Indeed, with $A_1 = 0$ the LFR

$$G_1(\delta) = \Delta_1(\delta) \star H_1 = D_1 + C_1\Delta_1(\delta)B_1$$

is a joint scaling and shifting of the uncertainty block $\Delta_1(\delta)$. We thus find that $G_1(\delta) \star H_2 = \Delta_1(\delta) \star H_{12}$. If the parameter block $\Delta_1(\delta)$ is not shifted, then in addition $D_1 = 0$ which leads to the even more special case:

$$[C_1\Delta_1(\delta)B_1] \star H_2 = \Delta_1(\delta) \star \tilde{H}_2 \;\text{ with }\; \tilde{H}_2 = \begin{pmatrix} B_1A_2C_1 & B_1B_2 \\ C_2C_1 & D_2 \end{pmatrix}. \tag{6.2.3}$$

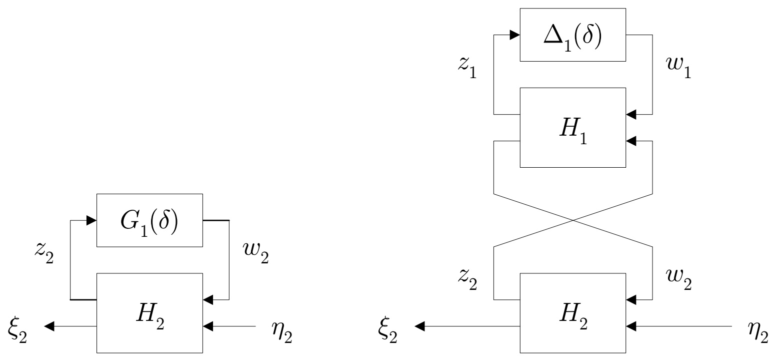


Figure 6.4: LFR of LFR is LFR

If $\Delta_1(\delta)$ is of smaller dimension than $C_1\Delta_1(\delta)B_1$, this formula reveals that one can compress $[C_1\Delta_1(\delta)B_1] \star H_2$ to the more compact LFR $\Delta_1(\delta) \star \tilde{H}_2$.

On the other hand, (6.2.3) easily allows to convince ourselves that any LFR with a general linear parameter-block $\Delta(\delta)$ can be transformed into one with the concrete parameter dependence

$$\Delta_d(\delta) = \begin{pmatrix} \delta_1 I_{d_1} & & 0 \\ & \ddots & \\ 0 & & \delta_m I_{d_m} \end{pmatrix} \tag{6.2.4}$$

where $I_{d_1}, \ldots, I_{d_m}$ are identity matrices of sizes $d_1, \ldots, d_m$ respectively. In this case we call the positive integer vector $d = (d_1, \ldots, d_m)$ the *order* of the LFR. Indeed if $\Delta(\delta)$ depends linearly on δ, there exist suitable coefficient matrices Δ_j such that

$$w = \Delta(\delta)z = [\delta_1\Delta_1 + \cdots + \delta_m\Delta_m]z. \tag{6.2.5}$$

Again we view the matrices $\delta_j I$ as non-dynamic system components and simply name the input signals and output signals of each of these subsystems. Then $w = \Delta(\delta)z$ is equivalent to $w = w_1 + \cdots + w_m$, $w_j = \delta_j z_j = [\delta_j I_j]z_j$, $z_j = \Delta_j z$ which is the desired alternative LFR. We observe that the size of I_j, which determines how often δ_j has to be repeated, corresponds to the number of components of the signals w_j and z_j. Since the coefficient matrices Δ_j have often small rank in practice, this procedure can be adapted to reduce the order of the LFR. One just needs to perform the factorizations

$$\Delta_1 = L_1 R_1, \quad \ldots \quad , \Delta_m = L_m R_m$$

such that the number of columns and rows of L_j and R_j equal the rank of Δ_j and are, therefore, as small as possible. These full-rank factorizations can be easily computed by the Gauss-elimination algorithm or by singular value decompositions. Then (6.2.5) reads as

$$w = \left[L_1(\delta_1 I_{d_1})R_1 + \cdots + L_m(\delta_m I_{d_m})R_m\right] z$$

which leads to the following smaller order LFR of $w = \Delta(\delta)z$:

$$\begin{pmatrix} z_1 \\ \vdots \\ z_m \\ \hline w \end{pmatrix} = \left(\begin{array}{ccc|c} 0 & \cdots & 0 & R_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & R_m \\ \hline L_1 & \cdots & L_m & 0 \end{array}\right) \begin{pmatrix} w_1 \\ \vdots \\ w_m \\ \hline z \end{pmatrix}, \quad \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix} = \begin{pmatrix} \delta_1 I_{d_1} & & 0 \\ & \ddots & \\ 0 & & \delta_m I_{d_m} \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_m \end{pmatrix}.$$

Since an LFR of an LFR is an LFR, it is straightforward to combine this LFR of $w = \Delta(\delta)z$ with that of $\eta = G(\delta)\xi$ to obtain the following new LFR of $\eta = G(\delta)\xi$ with a parameter-block that admits the desired diagonal structure:

$$\begin{pmatrix} z_1 \\ \vdots \\ z_m \\ \hline \eta \end{pmatrix} = \left(\begin{array}{ccc|c} R_1AL_1 & \cdots & R_1AL_m & R_1B \\ \vdots & \ddots & \vdots & \vdots \\ R_mAL_1 & \cdots & R_mAL_m & R_mB \\ \hline CL_1 & \cdots & CL_m & D \end{array}\right) \begin{pmatrix} w_1 \\ \vdots \\ w_m \\ \hline \xi \end{pmatrix}, \quad w_j = [\delta_j I_{d_j}]z_j.$$

Let us now turn to the construction of LFR for rational functions. We begin with $G(\delta)$ depending affinely on δ. Then $G(\delta) = G_0 + \Delta(\delta)$ with $\Delta(\delta)$ linear, and an LFR of $\eta = G(\delta)\xi$ is simply given by

$$\begin{pmatrix} z \\ \eta \end{pmatrix} = \begin{pmatrix} 0 & I \\ I & G_0 \end{pmatrix} \begin{pmatrix} w \\ \xi \end{pmatrix}, \quad w = \Delta(\delta)z.$$

As just described, one can easily go one step further to obtain an LFR with a block-diagonal parameter-block as described for general LFR's.

We proceed with a one-variable polynomial dependence

$$\eta = \left[G_0 + \delta G_1 + \cdots + \delta^p G_p\right]\xi. \tag{6.2.6}$$

For achieving efficient LFR's we rewrite as

$$\eta = G_0\xi + \delta\left[G_1 + \delta\left[G_2 + \cdots + \delta\left[G_{p-1} + \delta G_p\right]\cdots\right]\right]\xi. \tag{6.2.7}$$

Then we can separate the uncertainties as

$$\eta = G_0\xi + w_1,\ w_1 = \delta z_1,\ z_1 = G_1\xi + w_2,\ w_2 = \delta z_2,\ z_2 = G_2\xi + w_3,\ \ldots$$
$$\ldots,\ w_{p-1} = \delta z_{p-1},\ z_{p-1} = G_{p-1}\xi + w_p,\ w_p = \delta z_p,\ z_p = G_p\xi.$$

With $w = \mathrm{col}(w_1, \ldots, w_p)$, $z = \mathrm{col}(z_1, \ldots, z_p)$, this LFR of (6.2.6) reads as

$$\begin{pmatrix} z_1 \\ \vdots \\ z_{p-1} \\ z_p \\ \hline \eta \end{pmatrix} = \left(\begin{array}{cccc|c} 0 & I & \cdots & 0 & G_1 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & I & G_{p-1} \\ 0 & \cdots & 0 & 0 & G_p \\ \hline I & \cdots & 0 & 0 & G_0 \end{array}\right) \begin{pmatrix} w_1 \\ \vdots \\ w_{p-1} \\ w_p \\ \hline \xi \end{pmatrix}, \quad w = \delta z.$$

Now assume that $G(\delta)$ is a one-parameter matrix-valued rational function without pole in zero. It is straightforward to construct a square polynomial matrix D such that $DG = N$ is itself polynomial. For example, one can determine the common multiple of the element's denominators in each row and collect these scalar polynomials on the diagonal of D. Since none of the elements' denominators of G vanish in zero this procedure guarantees that $D(0)$ is nonsingular. Therefore, $\det(D(\delta))$ does not vanish identically which implies $G(\delta) = D(\delta)^{-1}N(\delta)$ for all δ for which $D(\delta)$ is non-singular. In summary, one can construct polynomial matrices

$$D(\delta) = D_0 + \delta D_1 + \cdots + \delta^p D_p \ \text{ and } \ N(\delta) = N_0 + \delta N_1 + \cdots + \delta^p N_p$$

with either $D_p \neq 0$ or $N_p \neq 0$ and with the following properties:

- D is square and $D(0) = D_0$ is non-singular.
- If δ is chosen such that $D(\delta)$ is invertible then G has no pole in δ, and $\eta = G(\delta)\xi$ is equivalent to $D(\delta)\eta = N(\delta)\xi$.

Clearly $D(\delta)\eta = N(\delta)\xi$ reads as

$$\left[D_0 + \delta D_1 + \cdots + \delta^p D_p\right]\eta = \left[N_0 + \delta N_1 + \cdots + \delta^p N_p\right]\xi$$

which is equivalent to

$$\eta = D_0^{-1}N_0\xi + \delta(D_0^{-1}N_1\xi - D_0^{-1}D_1\eta) + \cdots + \delta^p(D_0^{-1}N_p\xi - D_0^{-1}D_p\eta)$$

and hence to

$$\eta = D_0^{-1}N_0\xi + w_1,\ w_1 = \delta z_1,\ z_1 = \sum_{j=1}^{p}\delta^{j-1}[N_j\xi - D_jD_0^{-1}N_0\xi - D_0^{-1}D_jw_1].$$

Exactly as described for polynomial dependence this leads to the LFR

$$\left(\begin{array}{c} z_1 \\ \vdots \\ z_{p-1} \\ z_p \\ \hline \eta \end{array}\right) = \left(\begin{array}{cccc|c} -D_1D_0^{-1} & I & \cdots & 0 & N_1 - D_1D_0^{-1}N_0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ -D_{p-1}D_0^{-1} & \cdots & 0 & I & N_{p-1} - D_{p-1}D_0^{-1}N_0 \\ -D_pD_0^{-1} & \cdots & 0 & 0 & N_p - D_pD_0^{-1}N_0 \\ \hline I & \cdots & 0 & 0 & N_0 \end{array}\right)\left(\begin{array}{c} w_1 \\ \vdots \\ w_{p-1} \\ w_p \\ \hline \xi \end{array}\right),$$

$$w_1 = \delta z_1,\ \ldots,\ w_m = \delta z_m.$$

If this LFR is well-posed at δ, then $\xi = 0$ implies $\eta = 0$, which in turn shows that $D(\delta)$ is invertible. Therefore we can conclude that (**??**) is indeed an LFR of $\eta = G(\delta)\xi$ as desired.

Remark 6.2 We have essentially recapitulated how to construct a realization for the rational function $G(1/s)$ in the variable s which is proper since G has no pole in zero. There exists a large body of literature how to compute efficient minimal realizations, either by starting with polynomial fractional representations that are coprime, or by reducing (**??**) on the basis of state-space techniques. If constructing a minimal realization of $G(1/s)$ with any of these (numerically stable) techniques it is thus possible to determine a minimally sized matrix A with

$$G(\delta) = C(\frac{1}{\delta}I - A)^{-1}B + D = D + C\delta(I - A\delta)^{-1}B = [\delta I_{\dim(A)}] \star \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

This leads to the minimal sized LFR

$$\begin{pmatrix} z \\ \eta \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} w \\ \xi \end{pmatrix}, \quad w = \delta z$$

of (6.2.1) with rational one-parameter dependence.

Remark 6.3 Let us choose an arbitrary mapping $\pi : \{1, \ldots, p\} \to \{1, \ldots, p\}$. If replacing $w_j = [\delta I] z_j$ in (**??**) with $w_j = [\delta_{\pi(j)} I] z_j$, we can follow the above derivation backwards to observe that one obtains an LFR of

$$[D_0 + \delta_{\pi(1)} D_1 + (\delta_{\pi(1)}\delta_{\pi(2)}) D_2 + \cdots + (\delta_{\pi(1)} \cdots \delta_{\pi(p)}) D_p]\eta =$$
$$= [N_0 + \delta_{\pi(1)} N_1 + (\delta_{\pi(1)}\delta_{\pi(2)}) N_2 + \cdots + (\delta_{\pi(1)} \cdots \delta_{\pi(p)}) N_p]\xi.$$

with a specific multi-linear dependence of $D(\delta)$ and $N(\delta)$ on δ.

Let us finally consider a general multivariable matrix-valued rational function $G(\delta)$ without pole in zero. This just means that each of its elements can be represented as

$$\frac{\displaystyle\sum_{j_1=0}^{p_1} \cdots \sum_{j_m=0}^{p_m} \alpha_{j_1,\ldots,j_m}\, \delta_1^{j_1} \cdots \delta_m^{j_m}}{\displaystyle\sum_{j_1=0}^{l_1} \cdots \sum_{j_m=0}^{l_m} \beta_{j_1,\ldots,j_m}\, \delta_1^{j_1} \cdots \delta_m^{j_m}} \quad \text{with} \quad \beta_{0,\ldots,0} \neq 0.$$

In literally the same fashion as described for one-variable polynomials on can easily construct multivariable polynomial matrices D and N with $DG = N$ and nonsingular $D(0)$. Again, for all δ for which $D(\delta)$ is non-singular one concludes that $\eta = G(\delta)\xi$ is equivalent to $D(\delta)\eta = N(\delta)\xi$ or to the kernel representation

$$0 = H(\delta) \begin{pmatrix} \xi \\ \eta \end{pmatrix} \quad \text{with} \quad H(\delta) = \begin{pmatrix} N(\delta) & -D(\delta) \end{pmatrix}. \tag{6.2.8}$$

One can decompose as

$$H(\delta) = H_0 + \delta_1 H_1^1(\delta) + \cdots + \delta_m H_m^1(\delta)$$

with easily determined polynomial matrices H_j^1 whose degrees are strictly smaller than those of H. Then (6.2.8) is equivalent to

$$0 = H_0 \begin{pmatrix} \xi \\ \eta \end{pmatrix} + \sum_{j=1}^{m} w_j^1, \quad w_j^1 = \delta_j z_j^1, \quad z_j^1 = H_j^1(\delta) \begin{pmatrix} \xi \\ \eta \end{pmatrix}. \tag{6.2.9}$$

Let us stress that the channel j should be considered absent (empty) if H_j^1 is the zero polynomial matrix. With $w^1 = \mathrm{col}(w_1^1, \ldots, w_m^1)$, $z^1 = \mathrm{col}(z_1^1, \ldots, z_m^1)$, $E_0 := (I \ \cdots \ I)$, $H^1 = \mathrm{col}(H_1^1, \ldots, H_m^1)$, (6.2.9) is more compactly written as

$$0 = H_0 \begin{pmatrix} \xi \\ \eta \end{pmatrix} + E_0 w^1, \ \ z^1 = H^1(\delta) \begin{pmatrix} \xi \\ \eta \end{pmatrix}, \ \ w^1 = \mathrm{diag}(\delta_1 I, \ldots, \delta_m I) z^1. \tag{6.2.10}$$

We are now in the position to iterate. For this purpose one decomposes

$$H^1(\delta) = H_1 + \delta_1 H_1^2(\delta) + \cdots + \delta_m H_m^2(\delta)$$

and pulls parameters out of the middle relation in (6.2.10) as

$$z_1 = H_1 \begin{pmatrix} \xi \\ \eta \end{pmatrix} + w_1^2 + \cdots + w_m^2, \ \ w_j^2 = \delta_j z_j^2, \ \ z_j^2 = H_j^2(\delta) \begin{pmatrix} \xi \\ \eta \end{pmatrix}.$$

Again using compact notations, (6.2.10) is equivalent to

$$0 = H_0 \begin{pmatrix} \xi \\ \eta \end{pmatrix} + E_0 w^1, \ \ z^1 = H_1 \begin{pmatrix} \xi \\ \eta \end{pmatrix} + E_1 w^2, \ \ z^2 = H^2(\delta) \begin{pmatrix} \xi \\ \eta \end{pmatrix},$$
$$w^1 = \mathrm{diag}(\delta_1 I, \ldots, \delta_m I) z^1, \ \ w^2 = \mathrm{diag}(\delta_1 I, \ldots, \delta_m I) z^2.$$

If continuing in this fashion, the degree of H^j is strictly decreased in each step, and the iteration stops after k steps if $H^k(\delta) = H_k$ is just a constant matrix. One arrives at

$$0 = H_0 \begin{pmatrix} \xi \\ \eta \end{pmatrix} + E_0 w^1, \ z^j = H_j \begin{pmatrix} \xi \\ \eta \end{pmatrix} + E_j w^{j+1}, \ w^j = \mathrm{diag}(\delta_1 I, \ldots, \delta_m I) z^j, \ j = 1, \ldots, k$$

with E_k and w^{k+1} being empty. In a last step we turn this implicit relation into a genuine input output relation. For this purpose we partition

$$H_j = \begin{pmatrix} N_j & -D_j \end{pmatrix}$$

conformably with $H = (N \ -D)$. Let us recall that $D_0 = D(0)$ is non-singular and hence

$$0 = H_0 \begin{pmatrix} \xi \\ \eta \end{pmatrix} + E_0 w^1 \ \text{ is equivalent to } \ \eta = D_0^{-1} N_0 \xi + D_0^{-1} E_0 w^1.$$

After performing this substitution we indeed directly arrive at the desired explicit LFR of the input-output relation $\eta = G(\delta)\xi$:

$$\eta = D_0^{-1} N_0 \xi + D_0^{-1} E_0 w^1, \ \ z^j = [N_j - D_j D_0^{-1} N_0]\xi - [D_j D_0^{-1} E_0] w^1 + E_j w^{j+1},$$
$$w^j = \mathrm{diag}(\delta_1 I, \ldots, \delta_m I) z^j, \ \ j = 1, \ldots, k.$$

Of course this LFR can be again modified to one with the parameter-block (6.2.4), which just amounts to reordering in the present situation. We have constructively proved the desired representation theorem for rational $G(\delta)$, where the presented technique is suited to lead to reasonably sized LFR's in practice.

Theorem 6.4 *Suppose that all entries of the matrix $G(\delta)$ are rational functions of $\delta = (\delta_1, \ldots, \delta_m)$ whose denominators do not vanish at $\delta = 0$. Then there exist matrices A, B, C, D and non-negative integers $d_1, \ldots, d_m$ such that, with $\Delta(\delta) = \mathrm{diag}(\delta_1 I_{d_1}, \ldots, \delta_m I_{d_m})$,*

$$\eta = G(\delta)\xi \quad \textit{if and only if there exists} \quad w, z : \begin{pmatrix} z \\ \eta \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} w \\ \xi \end{pmatrix}, \quad w = \Delta(\delta) z$$

for all δ with $\det(I - A\Delta(\delta)) \neq 0$ and for all ξ.

Example 6.5 Consider

$$\eta = \underbrace{\frac{1}{(\delta_2 - 1)(\delta_2 + \delta_1^2 + \delta_1 - 1)} \begin{pmatrix} 2\delta_2^2 - \delta_2 + 2\delta_1^2\delta_2 - \delta_1^2 + 2\delta_1 & \delta_1\delta_2 + \delta_1^2 + \delta_2 + \delta_1 \\ (-2\delta_1\delta_2 + 3\delta_1 + 2\delta_2 - 1)\delta_1 & (\delta_1\delta_2 + 1)\delta_1 \end{pmatrix}}_{G(\delta)} \xi$$

or

$$\underbrace{\begin{pmatrix} (\delta_2 - 1)(\delta_2 + \delta_1^2 + \delta_1 - 1) & 0 \\ 0 & (\delta_2 - 1)(\delta_2 + \delta_1^2 + \delta_1 - 1) \end{pmatrix}}_{D(\delta)} \eta =$$
$$= \underbrace{\begin{pmatrix} 2\delta_2^2 - \delta_2 + 2\delta_1^2\delta_2 - \delta_1^2 + 2\delta_1 & \delta_1\delta_2 + \delta_1^2 + \delta_2 + \delta_1 \\ (-2\delta_1\delta_2 + 3\delta_1 + 2\delta_2 - 1)\delta_1 & (\delta_1\delta_2 + 1)\delta_1 \end{pmatrix}}_{N(\delta)} \xi.$$

Sorting for common powers leads to

$$\eta = \delta_1 \left[\begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \xi + \eta \right] + \delta_1^2 \left[\begin{pmatrix} -1 & 1 \\ 3 & 0 \end{pmatrix} \xi + \eta \right] + \delta_1^2\delta_2 \left[\begin{pmatrix} 2 & 0 \\ -2 & 1 \end{pmatrix} \xi - \eta \right] +$$
$$+ \delta_2 \left[\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \xi + 2\eta \right] + \delta_1\delta_2 \left[\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} \xi - \eta \right] + \delta_2^2 \left[\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \xi - \eta \right].$$

In a first step we pull out the parameters as

$$\eta = w_1 + w_4, \quad w_1 = \delta_1 z_1, \quad w_4 = \delta_2 z_4$$
$$z_1 = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \xi + \eta + \delta_1 \left[\begin{pmatrix} -1 & 1 \\ 3 & 0 \end{pmatrix} \xi + \eta \right] + \delta_1\delta_2 \left[\begin{pmatrix} 2 & 0 \\ -2 & 1 \end{pmatrix} \xi - \eta \right]$$
$$z_4 = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \xi + 2\eta + \delta_1 \left[\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} \xi - \eta \right] + \delta_2 \left[\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \xi - \eta \right].$$

We iterate with the second step as

$$\eta = w_1 + w_4, \quad w_1 = \delta_1 z_1, \quad w_2 = \delta_1 z_2, \quad w_3 = \delta_1 z_3, \quad w_4 = \delta_2 z_4, \quad w_5 = \delta_2 z_5$$
$$z_1 = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \xi + \eta + w_2, \quad z_2 = \begin{pmatrix} -1 & 1 \\ 3 & 0 \end{pmatrix} \xi + \eta + \delta_2 \left[\begin{pmatrix} 2 & 0 \\ -2 & 1 \end{pmatrix} \xi - \eta \right]$$
$$z_4 = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \xi + 2\eta + w_3 + w_5, \quad z_3 = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} \xi - \eta, \quad z_5 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \xi - \eta.$$

It just requires one last step to complete the separation as

$$\eta = w_1 + w_4,\ \ w_1 = \delta_1 z_1,\ \ w_2 = \delta_2 z_2,\ \ w_3 = \delta_1 z_3,\ \ w_4 = \delta_1 z_4,\ \ w_5 = \delta_2 z_5,\ \ w_6 = \delta_2 z_6$$
$$z_1 = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \xi + \eta + w_2,\ \ z_2 = \begin{pmatrix} -1 & 1 \\ 3 & 0 \end{pmatrix} \xi + \eta + w_6,\ \ z_6 = \begin{pmatrix} 2 & 0 \\ -2 & 1 \end{pmatrix} \xi - \eta$$
$$z_4 = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \xi + 2\eta + w_3 + w_5,\ \ z_3 = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} \xi - \eta,\ \ z_5 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \xi - \eta.$$

In matrix format this can be expressed as

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \hline z_4 \\ z_5 \\ z_6 \\ \hline \eta \end{pmatrix} = \left(\begin{array}{cc|cc|cc|cc|cc|cc||cc}
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 1 \\ \hline
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & -1 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 3 & 0 \\ \hline
-1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 2 & 0 \\ \hline
2 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 & -1 & 1 \\
0 & 2 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline
-1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
-1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -2 & 1 \\ \hline\hline
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right) \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \hline w_4 \\ w_5 \\ w_6 \\ \hline \xi \end{pmatrix}$$
$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \hline w_4 \\ w_5 \\ w_6 \end{pmatrix} = \left(\begin{array}{ccc|ccc}
\delta_1 I_2 & 0 & 0 & 0 & 0 & 0 \\
0 & \delta_1 I_2 & 0 & 0 & 0 & 0 \\
0 & 0 & \delta_1 I_2 & 0 & 0 & 0 \\ \hline
0 & 0 & 0 & \delta_2 I_2 & 0 & 0 \\
0 & 0 & 0 & 0 & \delta_2 I_2 & 0 \\
0 & 0 & 0 & 0 & 0 & \delta_2 I_2
\end{array}\right) \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \hline z_4 \\ z_5 \\ z_6 \end{pmatrix}.$$

The parameter-block turns out to be of size 12. We stress that, even for the rather simple case at hand, a less careful construction can easily lead to much larger-sized LFR's. For example with SISO realization techniques it is not difficult to find LFR's of the elements of D, N of sizes 4 and 3 respectively. If using the operations augmentation, inversion and multiplication one obtains an LFR with parameter-block of size $(4+4)+(4+3+3+3) = 21$. We will briefly comment on various aspects around the practical construction of LFR's in the next section.

6.3 On the practical construction of LFR's

LFR-Toolbox. The practical construction of LFR's is supported by a very helpful, professionally composed and freely available Matlab LFR-Toolbox developed by Francois Magni [**?**]. The toolbox is complemented with an excellent manual which contains numerous theoretical and practical hints, a whole variety of examples and many pointers to the recent literature. The purpose of this section is to provide some additional remarks that might not be as easily accessible.

Reduction. It is of crucial importance to construct LFR's with reasonably sized parameter-block. After having found one LFR $\Delta(\delta) \star H$, it is hence desirable to systematically construct an LFR $\Delta_r(\delta) \star H_r$ whose parameter-block $\Delta_r(\delta)$ is as small as possible and such that $\Delta(\delta) \star H = \Delta_r(\delta) \star H_r$ for all δ for which both are well-posed. Unfortunately this minimal representation problem does not admit a computationally tractable solution. Two reduction schemes, channel-by-channel [**?**] multi-channel [**?**] reduction, have been suggested in the literature and are implemented in the LFR-Toolbox [?]. To explain the channel-by-channel scheme consider the LFR

$$\left(\begin{array}{c} z_1 \\ \vdots \\ z_m \\ \hline \eta \end{array}\right) = \left(\begin{array}{ccc|c} A_{11} & \cdots & A_{1m} & B_1 \\ \vdots & \ddots & \vdots & \vdots \\ A_{m1} & \cdots & A_{mm} & B_m \\ \hline C_1 & \cdots & C_m & D \end{array}\right) \left(\begin{array}{c} w_1 \\ \vdots \\ w_m \\ \hline \xi \end{array}\right), \quad w_j = \delta_j z_j, \;\; j = 1, \ldots, m. \tag{6.3.1}$$

For each j we associate with this LFR (via the substitution $w_j \to x_j$, $z_j \to \dot{x}_j$) the LTI system

$$\left(\begin{array}{c} z_1 \\ \vdots \\ \dot{x}_j \\ \vdots \\ z_m \\ \hline \eta \end{array}\right) = \left(\begin{array}{ccccc|c} A_{11} & \cdots & A_{1j} & \cdots & A_{1m} & B_1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ A_{j1} & \cdots & A_{jj} & \cdots & A_{jm} & B_j \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ A_{m1} & \cdots & A_{mj} & \cdots & A_{mm} & B_j \\ \hline C_1 & \cdots & C_j & \cdots & C_m & D \end{array}\right) \left(\begin{array}{c} w_1 \\ \vdots \\ x_j \\ \vdots \\ w_m \\ \hline \xi \end{array}\right).$$

If non-minimal, this realization can be reduced to be controllable and observable to arrive at

$$\left(\begin{array}{c} z_1 \\ \vdots \\ \dot{\hat{x}}_j \\ \vdots \\ z_m \\ \hline \eta \end{array}\right) = \left(\begin{array}{ccccc|c} A_{11} & \cdots & \hat{A}_{1j} & \cdots & A_{1m} & B_1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ \hat{A}_{j1} & \cdots & \hat{A}_{jj} & \cdots & \hat{A}_{jm} & \hat{B}_j \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ A_{m1} & \cdots & \hat{A}_{mj} & \cdots & A_{mm} & B_j \\ \hline C_1 & \cdots & \hat{C}_j & \cdots & C_m & D \end{array}\right) \left(\begin{array}{c} w_1 \\ \vdots \\ \hat{x}_j \\ \vdots \\ w_m \\ \hline \xi \end{array}\right)$$

with $\hat{x}_j$ of smaller size than x_j. If sequentially performing this reduction step for all $j = 1, \ldots, m$ we arrive at the LFR

$$\left(\begin{array}{c} \hat{z}_1 \\ \vdots \\ \hat{z}_j \\ \vdots \\ \hat{z}_m \\ \hline \eta \end{array}\right) = \left(\begin{array}{ccccc|c} \hat{A}_{11} & \cdots & \hat{A}_{1j} & \cdots & \hat{A}_{1m} & \hat{B}_1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ \hat{A}_{j1} & \cdots & \hat{A}_{jj} & \cdots & \hat{A}_{jm} & \hat{B}_j \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ \hat{A}_{m1} & \cdots & \hat{A}_{mj} & \cdots & \hat{A}_{mm} & \hat{B}_j \\ \hline \hat{C}_1 & \cdots & \hat{C}_j & \cdots & \hat{C}_m & D \end{array}\right) \left(\begin{array}{c} w_1 \\ \vdots \\ \hat{w}_j \\ \vdots \\ w_m \\ \hline \xi \end{array}\right), \quad \hat{w}_j = \delta_j \hat{z}_j \tag{6.3.2}$$

with $\hat{z}_j$ and $\hat{w}_j$ of smaller size than z_j and w_j. A little reflection convinces us that (6.3.1) is equivalent to (6.3.2) for all δ for which both LFR's are well-posed. Note that the algebraic variety of δ for which the original LFR and the new LFR are well-posed are generally different. For a multi-channel extension we refer to [?], and for an illustration of how these schemes work in practice we consider an example.

Example 6.6 Let us continue Example 6.5 by constructing an LFR of the rational matrix with the LFR-Toolbox. The block structure is $(d_1, d_2) = (32, 23)$ and after SISO and MIMO reduction it amounts to $(d_1, d_2) = (23, 10)$ and $(d_1, d_2) = (13, 10)$ respectively. Starting with the Horner decomposition of G even leads to the larger reduced size $(d_1, d_2) = (14, 12)$. Even after reduction, all these LFR's are much larger than the one found in Exercise 6.5. Constructing an LFR of the polynomial matrix $(D - N)$ (after Horner decomposition and reduction) leads to $(d_1, d_2) = (7, 6)$ which allows to find an LFR of $D^{-1}N$ with the same parameter-block (Exercise 1). Reducing the manually constructed LFR from Exercise 6.5 leads to the smallest sizes $(d_1, d_2) = (6, 5)$. We stress that this is still far from optimal!

Approximation. If exact reduction of a given LFR $G(\delta) = \Delta(\delta) \star H$ is hard or impossible, one might construct an approximation $\Delta_a(\delta) \star H_a$ with smaller-sized $\Delta_a(\delta)$. For this purpose we stress that LFR's are typically used for parameters δ in some a priori given set $\boldsymbol{\delta}$. If both the original and the approximate LFR's are well-posed for all $\delta \in \boldsymbol{\delta}$, we can use

$$\sup_{\delta \in \boldsymbol{\delta}} \|\Delta(\delta) \star H - \Delta_a(\delta) \star H_a\|$$

(with any matrix norm $\|.\|$) as a measure for the distance of the two LFR's, and hence as an indicator for the approximation quality. For LMI-based approximate reduction algorithms that are similar to the balanced truncation reduction of transfer matrices we refer to []. We stress that the evaluation of the distance just requires knowledge of the values $G(\delta) = \Delta(\delta) \star H$ at $\delta \in \boldsymbol{\Delta}$. Given $G(\delta)$, this leads to the viable idea to construct LFR's by non-linear optimization []. Among the various possibilities we would like to sketch one scheme that will be successfully applied to our example. Choose some parameter-block $\Delta_a(\delta)$ (of small size) which depends linearly on δ together with finitely many distinct points $\delta_1, \ldots, \delta_N \in \boldsymbol{\delta}$ (which should be well-distributed in $\boldsymbol{\delta}$). Then the goal is to find A_a, B_a, C_a, D_a such that

$$\max_{j=1,\ldots,N} \left\| G(\delta_j) - \Delta_a(\delta_j) \star \begin{pmatrix} A_a & B_a \\ C_a & D_a \end{pmatrix} \right\| \tag{6.3.3}$$

is as small as possible. Clearly

$$V_j = G(\delta_j) - \Delta_a(\delta_j) \star \begin{pmatrix} A_a & B_a \\ C_a & D_a \end{pmatrix} \iff$$
$$\iff \exists\, U_j : \ V_j = G(\delta_j) - D_a - C_a U_j, \ \ U_j = (I - \Delta_a(\delta_j) A_a)^{-1} \Delta_a(\delta_j) B_a \iff$$
$$\iff \exists\, U_j : \ [\Delta_a(\delta_j) A_a - I] U_j + \Delta_a(\delta_j) B_a = 0, \ \ C_a U_j + D_a - G(\delta_j) = V_j.$$

Therefore we have to solve the nonlinear optimization problem

$$\begin{array}{ll} \text{minimize} & \max_{j=1,\ldots,N} \|C_a U_j + D_a - G(\delta_j)\| \\ \text{subject to} & [\Delta_a(\delta_j) A_a - I] U_j + \Delta_a(\delta_j) B_a = 0, \quad j = 1, \ldots, N. \end{array} \tag{6.3.4}$$

in the variables A_a, B_a, C_a, D_a, U_j, $j = 1, \ldots, N$. It might be relevant to explicitly include the constraint that $\Delta_a(\delta_j)A_a - I$ has to be invertible. Moreover we stress that it is immediate to formulate variations which might considerably simplify the corresponding computations. For example if striving for LFR's which are exact in $\delta_1, \ldots, \delta_N$, one could solve the non-linear least-squares problem

$$\inf_{\substack{A_a, B_a, C_a, D_a \\ U_j, \; j=1,\ldots,N}} \sum_{j=1}^{N} \left\| \begin{pmatrix} \Delta_a(\delta_j)A_a - I \\ C_a \end{pmatrix} U_j + \begin{pmatrix} \Delta_a(\delta_j)B_a \\ D_a - G(\delta_j) \end{pmatrix} \right\|_F . \tag{6.3.5}$$

Example 6.7 We continue Example 6.5 by constructing an LFR with nonlinear least-squares optimization. It is simple to verify that the denominator polynomial $(\delta_2 - 1)(\delta_2 + \delta_1^2 + \delta_1 - 1)$ is nonzero on the box $\boldsymbol{\delta} = \{(\delta_1, \delta_2) : \ -0.3 \le \delta_1 \le 0.3, \ -0.3 \le \delta_1 \le 0.3\}$. With $\Delta(\delta) = \operatorname{diag}(\delta_1, \delta_1, \delta_1, \delta_2, \delta_2, \delta_2)$ and the 49 points $\{-0.3 + 0.1\,j \ : \ j = 0, \ldots, 6\}^2$ we solve (6.3.5) to achieve a cost smaller than $1.3\,10^{-9}$. The deviation (6.3.3) turns out to be not larger than $8\,10^{-5}$ on the denser grid $\{-0.3 + 0.01\,j \ : \ j = 0, \ldots, 60\}^2$ with 3721 points. We have thus constructed an approximate LFR of considerably reduced order $(d_1, d_2) = (3, 3)$. It can be actually verified that

$$G(\delta) = \begin{pmatrix} \delta_1 & 0 & 0 & 0 \\ 0 & \delta_1 & 0 & 0 \\ 0 & 0 & \delta_2 & 0 \\ 0 & 0 & 0 & \delta_2 \end{pmatrix} \star \left(\begin{array}{cccc|cc} 1 & 1 & 1 & 2 & 2 & 1 \\ 1 & 0 & 1 & 1 & -1 & 1 \\ 1 & 0 & 1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ \hline 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{array} \right)$$

which could not be reconstructed with any of the suggested techniques! The Examples 6.5-6.7 demonstrate that it might be difficult to reconstruct this small-sized LFR from the corresponding rational function with reasonable computational effort. As the main lesson to be learnt for practice, one should keep track of how large-sized LFR's are actually composed by smaller ones through interconnection in order not to loose essential structural information.

Approximate LFR's of non-rational functions. We observe that the suggested approximate reduction technique is only relying on the availability of $G(\delta_j)$ for $j = 1, \ldots, N$. This opens the path to determine LFR approximations for non-rational matrix valued mappings or even for mappings that are just defined through look-up tables. As we have seen, the direct construction of LFR's then requires the solution of a generally non-convex optimization problem.

Alternatively one could first interpolate the data matrices with multivariable polynomial or rational functions, and then construct an LFR of the interpolant. With multivariable polynomial matrices $D(\delta)$ and $N(\delta)$ of fixed degree and to-be-determined coefficients, this requires to solve the equation $D(\delta_j)^{-1}N(\delta_j) = G(\delta_j)$ which can be reduced to $D(\delta_j)G(\delta_j) - N(\delta_j) = 0$ (for all $j = 1, \ldots, N$) and hence turns out to be a linear system of equations in the coefficient matrices.

As a further alternative one can rely on polynomial or rational approximation. This amounts to finding polynomial matrices D, N which minimize $\max_{j=1,\ldots,N} \|D(\delta_j)^{-1}N(\delta_j) - G(\delta_j)\|$. For a fixed denominator polynomial matrix D and for a fixed degree of N this is clearly a convex optimization problem. Indeed, as an immediate consequence of the proof of Theorem 6.4, any such fixed denominator function can be parameterized as $\Delta(\delta) \star \begin{pmatrix} A_a & B_a \\ C_a & D_a \end{pmatrix}$ with *fixed* A_a, C_a and *free* B_a, D_a. Then it is obvious that (6.3.4) is a convex optimization problem which can be reduced to an LMI-problem if using any matrix norm whose sublevel sets admit an LMI-representation. For general rational approximation one has to rely on solving the non-convex problem (6.3.4) with *free* A_a, B_a, C_a, D_a, or one could turn to the multitude of existing alternatives such as Padé approximation for which we refer to the approximation literature.

Non-diagonal parameter-blocks. We have proved how to construct LFR's of rational matrices with diagonal parameter-blocks in a routine fashion. It is often possible to further reduce the size of LFR's with the help of full non-diagonal parameter-blocks which are only required to depend linearly on the parameters.

Example 6.8 At this point we disclose that the rational matrix in Examples 6.5-6.7 was actually constructed as

$$G(\delta) = \begin{pmatrix} \delta_1 & \delta_2 & \delta_1 \\ 0 & \delta_1 & \delta_2 \\ 0 & 0 & \delta_2 \end{pmatrix} \star \left(\begin{array}{ccc|cc} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \end{array}\right)$$

which is still smaller than all the LFR's constructed so far.

Example 6.9 Let us illustrate how to actually find such reduced-sized LFR's by means of the example

$$G(\delta) = \frac{3\delta_1^2 - 2\delta_2}{1 - 4\delta_1 + 2\delta_1\delta_2 - \delta_2}.$$

A key role is played by finding suitable (matrix) factorizations. Indeed $\eta = G(\delta)\xi$ iff

$$\eta = (3\delta_1^2 - 2\delta_2)\xi + (4\delta_1 - 2\delta_1\delta_2 + \delta_2)\eta = \begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix} \begin{pmatrix} 3\delta_1 & 4 - 2\delta_2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} \iff$$

$$\iff \eta = w_1,\ w_1 = \begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix} z_1,\ z_1 = \begin{pmatrix} 4\eta \\ -2\xi + \eta \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix} \begin{pmatrix} 3\xi \\ -2\eta \end{pmatrix} \iff$$

$$\iff \eta = w_1,\ w_1 = \begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix} z_1,\ z_1 = \begin{pmatrix} 4\eta + w_2 \\ -2\xi + \eta \end{pmatrix},\ w_2 = \begin{pmatrix} \delta_1 & \delta_2 \end{pmatrix} z_2,\ z_2 = \begin{pmatrix} 3\xi \\ -2\eta \end{pmatrix}$$

which leads to the LFR

$$\left(\begin{array}{c} z_1 \\ \hline z_2 \\ \hline \eta \end{array}\right) = \left(\begin{array}{cc|c} 4 & 1 & 0 \\ 1 & 0 & -2 \\ \hline 0 & 0 & 3 \\ -2 & 0 & 0 \\ \hline 1 & 0 & 0 \end{array}\right) \left(\begin{array}{c} w_1 \\ w_2 \\ \hline \xi \end{array}\right), \quad \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \left(\begin{array}{cc|cc} \delta_1 & \delta_2 & 0 & 0 \\ 0 & 0 & \delta_1 & \delta_2 \end{array}\right) \left(\begin{array}{c} z_1 \\ \hline z_2 \end{array}\right). \quad (6.3.6)$$

We observe that this LFR has the structure

$$\begin{pmatrix} z \\ \hat{z} \\ \hline \eta \end{pmatrix} = \left(\begin{array}{c|c} A & B \\ LA & LB \\ \hline C & D \end{array}\right) \begin{pmatrix} w \\ \hline \xi \end{pmatrix}, \quad w = \begin{pmatrix} \Delta(\delta) & \hat{\Delta}(\delta) \end{pmatrix} \begin{pmatrix} z \\ \hat{z} \end{pmatrix}$$

with $L = \begin{pmatrix} 0 & -2 & -4/3 \end{pmatrix}$ and is, due to $\hat{z} = Lz$, hence equivalent to

$$\begin{pmatrix} z \\ \eta \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} w \\ \xi \end{pmatrix}, \quad w = [\Delta(\delta) + \hat{\Delta}(\delta)L]z.$$

Therefore (6.3.6) can be compressed to

$$\begin{pmatrix} z_1 \\ \hline z_2 \\ \hline \eta \end{pmatrix} = \left(\begin{array}{cc|c} 4 & 1 & 0 \\ 1 & 0 & -2 \\ \hline 0 & 0 & 3 \\ \hline 1 & 0 & 0 \end{array}\right) \begin{pmatrix} w_1 \\ w_2 \\ \hline \xi \end{pmatrix} \quad \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \left(\begin{array}{cc|c} \delta_1 & \delta_2 & 0 \\ 0 & -2\delta_2 & \delta_1 - 4/3\delta_2 \end{array}\right) \begin{pmatrix} z_1 \\ \hline z_2 \end{pmatrix}$$

with a parameter-block of size 2×3. Standard LFR's can be obtained with orders $(d_1, d_2) = (3, 1)$ or $(d_1, d_2) = (2, 2)$ which involves parameter-blocks of larger size 4×4.

Nonlinear parameter-blocks and parameter transformations. System descriptions of interconnected non-linear components very often involve highly structured matrix-valued mappings such as

$$G(\delta) = \begin{pmatrix} 0 & \delta_1\delta_4 & \delta_2\delta_4 + \cos(\delta_3) \\ -\delta_1\delta_4 & -\delta_2\delta_4 - \cos(\delta_3) & -\delta_1\delta_4^2 + \delta_4\cos(\delta_3) \end{pmatrix}.$$

One can construct a fractional representation with the nonlinear parameter-block $\Delta(\delta) = \mathrm{diag}(\delta_1 I_2, \delta_2 I_2, \cos(\delta_3) I_2, \delta_4 I_4)$ of size 10×10. If performing the parameter transformation

$$\hat{\delta}_1 := \delta_1\delta_4, \quad \hat{\delta}_2 := \delta_2\delta_4 + \cos(\delta_3), \quad \hat{\delta}_3 = \cos(\delta_3), \quad \hat{\delta}_4 = \delta_4$$

one has to pull the transformed parameters out of

$$\hat{G}(\hat{\delta}) = \begin{pmatrix} 0 & \hat{\delta}_1 & \hat{\delta}_2 \\ -\hat{\delta}_1 & -\hat{\delta}_2 & (\hat{\delta}_3 - \hat{\delta}_1)\hat{\delta}_4 \end{pmatrix}$$

with only one bilinear element. Now one can construct an LFR with parameter-block $\hat{\Delta}(\hat{\delta}) = \mathrm{diag}(\hat{\delta}_1 I_2, \hat{\delta}_2 I_2, \hat{\delta}_3, \hat{\delta}_4)$ of smaller size 6×6. In concrete applications, if one has to analyze $G(\delta)$ for $\delta \in \boldsymbol{\delta}$, it remains to determine the image $\hat{\boldsymbol{\delta}}$ of $\boldsymbol{\delta}$ under the non-linear parameter transformation and to perform the desired analysis for $G(\hat{\delta})$ with $\hat{\delta} \in \hat{\boldsymbol{\delta}}$.

6.4 Robust linear algebra problems

Recall that we translated the robust counterpart of least-squares approximation into the min-max problem

$$\inf_{X \in \mathbb{R}^{p \times q}} \max_{\delta \in \boldsymbol{\delta}} \|XM(\delta) - N(\delta)\|_F^2.$$

Let us now assume that $M(\delta)$ and $N(\delta)$ are rational functions of δ without pole in zero, and determine an LFR of the stacked columns of $M(\delta)$ and $N(\delta)$ as

$$\begin{pmatrix} M(\delta)e_1 \\ N(\delta)e_1 \\ \vdots \\ M(\delta)e_l \\ N(\delta)e_l \end{pmatrix} = \Delta(\delta) \star \left(\begin{array}{c|c} A & B \\ \hline C_1^M & D_1^M \\ C_1^N & D_1^N \\ \vdots & \vdots \\ C_l^M & D_l^M \\ C_l^N & D_l^N \end{array}\right).$$

Then it is clear that

$$G(\delta, X) := \begin{pmatrix} XM(\delta)e_1 - N(\delta)e_1 \\ \vdots \\ XM(\delta)e_r - N(\delta)e_r \end{pmatrix}$$

admits the LFR

$$\Delta(\delta) \star \left(\begin{array}{c|c} A & B \\ \hline XC_1^M - C_1^N & XD_1^M - D_1^N \\ \vdots & \vdots \\ XC_l^M - C_l^N & XD_l^M - D_l^N \end{array}\right) = \Delta(\delta) \star \begin{pmatrix} A & B \\ C(X) & D(X) \end{pmatrix} = \Delta(\delta) \star H(X).$$

Robust least-squares approximation then requires to find $X \in \mathbb{R}^{p\times q}$ which minimizes γ under the semi-infinite constraint

$$\begin{pmatrix} I \\ \Delta(\delta) \star H(X) \end{pmatrix}^T P_\gamma \begin{pmatrix} I \\ \Delta(\delta) \star H(X) \end{pmatrix} < 0 \text{ for all } \delta \in \boldsymbol{\delta}.$$

——————— Formulate better

Let us distinguish analysis from synthesis ... This structure is the starting-point for deriving LMI relaxations on the basis of the so-called full block S-procedure in the next section.

——————— Formulate better

Example 6.10 As a concrete example we consider robust interpolation as suggested in [**?**]. Given data points (a_i, b_i), $i = 1, \ldots, k$, standard polynomial interpolation requires to find the coefficients of the polynomial $p(t) = z_0 + z_1 t + \cdots z_n t^n$ such that it satisfies $p(a_i) = b_i$ for all $i = 1, \ldots, k$. If such an interpolant does not exist one can instead search for a least squares approximant, a polynomial for which $\sum_{i=1}^k |p(a_i) - b_i|^2$ is as small as possible. Hence we require to solve

$$\min_{z\in\mathbb{R}^{n+1}} \left\| z^T \begin{pmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_k \\ \vdots & & \vdots \\ a_1^n & \cdots & a_k^n \end{pmatrix} - \begin{pmatrix} b_1 & \cdots & b_k \end{pmatrix} \right\|^2.$$

Let us now assume that (a_i, b_i) are only known to satisfy $|a_i - \bar{a}_i| \leq r_i$ and $|b_i - \bar{b}_i| \leq s_i$ for $i = 1, \ldots, k$ (since they are acquired through error-prone experiments). Then one should search the coefficient vector z such that it minimizes the worst case least squares error

$$\max_{|a_i - \bar{a}_i| \leq r_i,\ |b_i - \bar{b}_i| \leq s_i,\ i=1,\ldots,k} \left\| z^T \begin{pmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_k \\ \vdots & & \vdots \\ a_1^n & \cdots & a_k^n \end{pmatrix} - \begin{pmatrix} b_1 & \cdots & b_k \end{pmatrix} \right\|^2 .$$

Note that it is rather simple to determine the LFR

$$\begin{pmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_k \\ \vdots & & \vdots \\ a_1^n & \cdots & a_k^n \\ b_1 & \cdots & b_k \end{pmatrix} = \mathrm{diag}(a_1 I_n, \ldots, a_k I_n, b_1, \ldots, b_k) \star \underbrace{\left(\begin{array}{cccc|ccc} N & & & & e_1 & & 0 \\ & \ddots & & & & \ddots & \\ & & N & & 0 & & e_1 \\ & & & 0_k & e_1 & \cdots & e_k \\ \hline M & \cdots & M & 0 & e_1 & \cdots & e_1 \\ 0 & \cdots & 0 & e & 0 & \cdots & 0 \end{array}\right)}_{\begin{pmatrix} A & B \\ C & D \end{pmatrix}}$$

with the standard unit vector e_j, the all-ones row vector e and with

$$N = \begin{pmatrix} 0 & 0 \\ I_{n-1} & 0 \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad M = \begin{pmatrix} 0 \\ I_n \end{pmatrix} \in \mathbb{R}^{(n+1) \times n},$$

just by combining (through augmentation) the LFR's for one-variable polynomials. Then it is straightforward to derive an LFR's in the the variables

$$(\delta_1, \ldots, \delta_k, \delta_{k+1}, \delta_{2k}) := (\frac{1}{r_1}(a_1 - \bar{a}_1), \ldots, \frac{1}{r_k}(a_k - \bar{a}_k), \frac{1}{s_1}(b_1 - \bar{b}_1), \ldots, \frac{1}{s_k}(b_k - \bar{b}_k))$$

by affine substitution. Indeed with

$$\Delta_0 = \mathrm{diag}(\bar{a}_1 I_n, \ldots, \bar{a}_k I_n, \bar{b}_1, \ldots, \bar{b}_k), \quad W = \mathrm{diag}(r_1 I_n, \ldots, r_k I_n, s_1, \ldots, s_k)$$

we infer

$$\mathrm{diag}(a_1 I_n, \ldots, a_k I_n, b_1, \ldots, b_k) = \Delta_0 + W\,\mathrm{diag}(\delta_1 I_n, \ldots, \delta_k I_n, \delta_{k+1}, \ldots, \delta_{2k})$$

which leads to the LFR

$$\begin{pmatrix} 1 & \cdots & 1 \\ \bar{a}_1 + \delta_1 & \cdots & \bar{a}_k + \delta_k \\ \vdots & & \vdots \\ (\bar{a}_1 + \delta_1)^n & \cdots & (\bar{a}_k + \delta_k)^n \\ \bar{b}_1 + \delta_{k+1} & \cdots & \bar{b}_k + \delta_{2k} \end{pmatrix} = \mathrm{diag}(\delta_1 I_n, \ldots, \delta_k I_n, \delta_{k+1}, \ldots, \delta_{2k}) \star H$$

with
$$H = \begin{pmatrix} A(I-\Delta_0 A)^{-1}W & (I-A\Delta_0)^{-1}B \\ C(I-\Delta_0 A)^{-1}W & D + C\Delta_0(I-\Delta_0 A)^{-1}B \end{pmatrix},$$
just by using the formula for computing the LFR of an LFR. Hence robust polynomial interpolation nicely fits into the general robust least squares framework as introduced in this section.

6.5 Full block S-procedure

If not specified all matrices and vectors in this section are tacitly assumed to have their elements in $\mathbb{K}$ where $\mathbb{K}$ either equals $\mathbb{R}$ or $\mathbb{C}$.

Many robust linear algebra problems, in particular those appearing in robust control, can be converted to testing whether

$$\Delta \star H \text{ is well-posed and } \begin{pmatrix} I \\ \Delta \star H \end{pmatrix}^* P_p \begin{pmatrix} I \\ \Delta \star H \end{pmatrix} < 0 \ \text{ for all } \ \Delta \in \mathbf{\Delta}_c. \tag{6.5.1}$$

Here $\mathbf{\Delta}_c$ is some set of $\mathbb{K}$-matrices and the LFR coefficient matrix H as well as the Hermitian P_p to express the desired bound on $\Delta \star H$ are partitioned as

$$H = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \ \text{ and } \ P_p = \begin{pmatrix} Q_p & S_p \\ S_p^* & R_p \end{pmatrix} \ \text{ with } \ R_p \geq 0.$$

Recall that $\Delta \star H$ is well-posed, by definition, if $I - A\Delta$ is invertible. Here is the fundamental result which allows to design relaxations of (6.5.1) that are computationally tractable with LMI's.

Theorem 6.11 (Concrete full block S-procedure) *If there exist matrices $Q = Q^*$, S, $R = R^*$ with*

$$\begin{pmatrix} \Delta \\ I \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^* & R \end{pmatrix} \begin{pmatrix} \Delta \\ I \end{pmatrix} \geq 0 \ \text{ for all } \ \Delta \in \mathbf{\Delta}_c \tag{6.5.2}$$

and

$$\begin{pmatrix} I & 0 \\ A & B \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^* & R \end{pmatrix} \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} + \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^* \begin{pmatrix} Q_p & S_p \\ S_p^* & R_p \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} < 0 \tag{6.5.3}$$

then (6.5.1) *is true. The converse holds in case that $\mathbf{\Delta}_c$ is compact.*

As will be explained below it is elementary to show that (6.5.2) and (6.5.3) are sufficient for (6.5.1) whereas the proof of necessity is more difficult. Before we embed this concrete version of the S-procedure into a more versatile abstract formulation we intend to address the practical benefit of this reformulation. With the help of introducing the auxiliary variables Q, R, S, which are often called

multipliers or **scalings**, it has been possible to equivalently rephrase (6.5.1) involving a multivariable rational function into the LMI conditions (6.5.2) and (6.5.3). Let us introduce, for the purpose of clarity, the obviously convex set of all multipliers that satisfy the infinite family of LMI's (6.5.2):

$$\boldsymbol{P_{\mathrm{all}}} := \left\{ P = \begin{pmatrix} Q & S \\ S^* & R \end{pmatrix} : \begin{pmatrix} \Delta \\ I \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^* & R \end{pmatrix} \begin{pmatrix} \Delta \\ I \end{pmatrix} \geq 0 \text{ for all } \Delta \in \boldsymbol{\Delta}_c \right\}.$$

Testing (6.5.1) then simply amounts to finding an element in $\boldsymbol{P_{\mathrm{all}}}$ which fulfills the LMI (6.5.3). Unfortunately, the actual implementation of this test requires a description of $\boldsymbol{P_{\mathrm{all}}}$ through finitely many LMI's, or at least a close approximation thereof. The need for such an approximation is the fundamental reason for conservatism in typical robust stability and robust performance tests in control!

Let us assume that $\boldsymbol{P_i} \subset \boldsymbol{P_{\mathrm{all}}}$ is an inner approximation with an LMI description. Then the computation of $P \in \boldsymbol{P_i}$ with (6.5.3) is computationally tractable. If the existence of such a multiplier can be verified, it is clear that (6.5.1) has been verified. On the other hand let $\boldsymbol{P_{\mathrm{all}}} \subset \boldsymbol{P_o}$ be an outer approximation with an LMI description. If one can computationally confirm that there does not exist any $P \in \boldsymbol{P_o}$ with (6.5.3), it is guaranteed that (6.5.1) is not true. We stress that the non-existence of $P \in \boldsymbol{P_i}$ or the existence of $P \in \boldsymbol{P_o}$ satisfying (6.5.1) does not allow to draw any conclusion without any additional knowledge about the quality of the inner or outer approximations respectively. In the sequel we discuss a selection of possible choices for inner approximations that have been suggested in the literature.

- **Spectral-Norm Multipliers.** The simplest and crudest inner approximation is obtained by neglecting any structural properties of the matrices in $\boldsymbol{\Delta}_c$ and just exploiting a known bound on their spectral norm. Indeed suppose that $r > 0$ is any (preferably smallest) number with

  $$\|\Delta\| \leq r \text{ for all } \Delta \in \boldsymbol{\Delta}_c.$$

  Since

  $$\|\Delta\| \leq r \text{ iff } \frac{1}{r}\Delta^*\Delta \leq rI \text{ iff } \begin{pmatrix} \Delta \\ I \end{pmatrix}^* \begin{pmatrix} -1/r & 0 \\ 0 & r \end{pmatrix} \begin{pmatrix} \Delta \\ I \end{pmatrix} \geq 0$$

  we can choose the set

  $$\boldsymbol{P_n} := \left\{ \tau \begin{pmatrix} -1/r & 0 \\ 0 & r \end{pmatrix} : \tau \geq 0 \right\}$$

  as an inner approximation, and checking (6.5.3) amounts to solving a one-parameter LMI problem.

- **Full-Block Multipliers.** For a more refined inner approximation let us assume that the set $\boldsymbol{\Delta}_c$ is described as

  $$\boldsymbol{\Delta}_c = \mathrm{conv}(\boldsymbol{\Delta}_g) = \mathrm{conv}\{\Delta_1, ..., \Delta_N\} \text{ with } \boldsymbol{\Delta}_g = \{\Delta_1, ..., \Delta_N\}$$

  as finitely many generators. Then define

  $$\boldsymbol{P_f} := \left\{ P = \begin{pmatrix} Q & S \\ S^* & R \end{pmatrix} : Q \leq 0, \begin{pmatrix} \Delta \\ I \end{pmatrix}^* P \begin{pmatrix} \Delta \\ I \end{pmatrix} \geq 0 \text{ for all } \Delta \in \boldsymbol{\Delta}_g \right\}.$$

Since any $P \in \boldsymbol{P_f}$ has a left upper block which is negative semi-definite, the mapping

$$\Delta \to \begin{pmatrix} \Delta \\ I \end{pmatrix}^* P \begin{pmatrix} \Delta \\ I \end{pmatrix} \text{ is concave,}$$

and hence positivity of its values at the generators $\Delta \in \boldsymbol{\Delta}_g$ implies positivity for all $\Delta \in \boldsymbol{\Delta}_c$. We conclude that $\boldsymbol{P_f} \subset \boldsymbol{P}_{\mathbf{all}}$. On the other hand, $\boldsymbol{P_f}$ is described by finitely many LMI's; hence searching for $P \in \boldsymbol{P_f}$ satisfying (6.5.3) is a standard LMI problem that can be easily implemented.

The construction of LFR's revealed that $\boldsymbol{\Delta}_c$ is often parameterized with a linear parameter-block $\Delta(\delta)$ in $\delta \in \boldsymbol{\delta} \subset \mathbb{K}^m$ as

$$\boldsymbol{\Delta}_c = \{\Delta(\delta) : \ \delta \in \boldsymbol{\delta}\}.$$

Just by linearity of $\Delta(\delta)$ we infer for convex finitely generated parameter sets that

$$\boldsymbol{\Delta}_c = \mathrm{conv}\{\Delta(\delta^j) : \ j = 1, \ldots, N\} \text{ in case of } \boldsymbol{\delta} = \mathrm{conv}\{\delta^1, \ldots, \delta^N\}.$$

For parameter boxes (products of intervals) defined as

$$\boldsymbol{\delta} = \{\delta \in \mathbb{R}^m : \ \delta_j \in [a_j, b_j], \ j = 1, \ldots, m\}$$

we can use the representation

$$\boldsymbol{\delta} = \mathrm{conv}(\boldsymbol{\delta}_g) \text{ with } \boldsymbol{\delta}_g := \mathrm{conv}\{\delta \in \mathbb{R}^m : \ \delta_j \in \{a_j, b_j\}, \ j = 1, \ldots, m\}.$$

All these choices do not depend upon a specific structure of $\Delta(\delta)$ and hence allow for non-diagonal parameter-blocks.

- **Extended Full-Block Multipliers.** Can we be more specific if $\boldsymbol{\delta}$ is a box generated by $\boldsymbol{\delta}_g$ and the parameter-block $\Delta(\delta)$ is diagonal, as in (6.2.4)? With the column partition $(E_1 \cdots E_m) = I$ of the identity matrix conformable with that of $\Delta(\delta)$, we infer that

$$\Delta(\delta) = \sum_{k=1}^{m} E_k[\delta_k I]E_k^T.$$

Let us then define

$$\boldsymbol{P_{fe}} := \left\{ P : \ E_k^T Q E_k \le 0, \ k = 1, \ldots, m, \ \begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix}^* P \begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix} \ge 0, \ \delta \in \boldsymbol{\delta}_g \right\}.$$

Due to $E_k^T Q E_k \le 0$, $k = 1, \ldots, m$, we conclude for any $P \in \boldsymbol{P_{fe}}$ that

$$\delta \to \begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^* & R \end{pmatrix} \begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix} \text{ is concave in } \delta_k \text{ for } k = 1, \ldots, m.$$

It is easily seen that this implies again that $\boldsymbol{P_{fe}} \subset \boldsymbol{P}_{\mathbf{all}}$. We stress that $\boldsymbol{P_f} \subset \boldsymbol{P_{fe}}$ such that $\boldsymbol{P_{fe}}$ is a potentially better inner approximation of $\boldsymbol{P}_{\mathbf{all}}$ which leads to less conservative numerical results. This comes at the expense of a more complex description of $\boldsymbol{P_{fe}}$ since it involves a larger number of LMI's. This observation leads us to the description of the last inner approximation of $\boldsymbol{P}_{\mathbf{all}}$ in our non-exhaustive list.

- **Diagonal Multipliers.** If $m_j = (a_j + b_j)/2$ denotes the mid-point of the interval $[a_j, b_j]$ and $d_j = (b_j - a_j)/2$ the half of its diameter we infer for $\delta_j \in \mathbb{R}$ that

$$\delta_j \in [a_j, b_j] \ \text{ iff } \ (\delta_j - m_j)^2 \leq d_j^2 \ \text{ iff } \ \begin{pmatrix} \delta_j \\ 1 \end{pmatrix}^* \underbrace{\begin{pmatrix} -1 & m_j \\ m_j & d_j^2 - m_j^2 \end{pmatrix}}_{\begin{pmatrix} q_j & s_j \\ s_j^* & r_j \end{pmatrix}} \begin{pmatrix} \delta_j \\ 1 \end{pmatrix} \geq 0.$$

Since

$$\begin{pmatrix} E_k & 0 \\ 0 & E_k \end{pmatrix}^T \begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix} = \begin{pmatrix} \sum_{l=1}^m E_k^T E_l [\delta_l I] E_l^T \\ E_k^T \end{pmatrix} = \begin{pmatrix} \delta_k I \\ I \end{pmatrix} E_k^T$$

(due to $E_k^T E_l = 0$ for $k \neq l$ and $E_k^T E_k = I$) we infer for $D_j \geq 0$ that

$$\begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix}^T \sum_{j=1}^m \begin{pmatrix} E_j & 0 \\ 0 & E_j \end{pmatrix} \begin{pmatrix} q_j D_j & s_j D_j \\ s_j D_j & r_j D_j \end{pmatrix} \begin{pmatrix} E_j & 0 \\ 0 & E_j \end{pmatrix}^T \begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix} =$$

$$= \sum_{j=1}^m E_j \begin{pmatrix} \delta_j I \\ I \end{pmatrix}^T \begin{pmatrix} q_j D_j & s_j D_j \\ s_j^* D_j & r_j D_j \end{pmatrix} \begin{pmatrix} \delta_j I \\ I \end{pmatrix} E_j^T =$$

$$= \sum_{j=1}^m E_j D_j E_j^T \begin{pmatrix} \delta_j \\ 1 \end{pmatrix}^T \begin{pmatrix} q_j & s_j \\ s_j^* & r_j \end{pmatrix} \begin{pmatrix} \delta_j \\ 1 \end{pmatrix} \geq 0.$$

The fact that δ_j is real can be expressed as

$$\delta_j^* + \delta_j = 0 \ \text{ or } \ \begin{pmatrix} \delta_j \\ 1 \end{pmatrix}^* \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \delta_j \\ 1 \end{pmatrix} = 0.$$

Exactly the same computation as above reveals for any $G_j = G_j^*$ that

$$\begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix}^* \sum_{j=1}^m \begin{pmatrix} E_j & 0 \\ 0 & E_j \end{pmatrix} \begin{pmatrix} 0 & G_j \\ -G_j & 0 \end{pmatrix} \begin{pmatrix} E_j & 0 \\ 0 & E_j \end{pmatrix}^T \begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix} = 0.$$

This implies for the set of block-diagonal multipliers

$$\boldsymbol{P_d} := \left\{ \sum_{j=1}^m \begin{pmatrix} E_j & 0 \\ 0 & E_j \end{pmatrix} \begin{pmatrix} q_j D_j & s_j D_j + G_j \\ s_j^* D_j - G_j & r_j D_j \end{pmatrix} \begin{pmatrix} E_j & 0 \\ 0 & E_j \end{pmatrix}^T : \ D_j \geq 0, \ G_j = G_j^* \right\}$$

that $\boldsymbol{P_d} \subset \boldsymbol{P_{\mathrm{all}}}$. Note that the set $\boldsymbol{P_d}$ is explicitly parametrized by finitely many linear matrix inequalities.

- **Mixtures.** Let us assume that $\Delta(\delta)$ is diagonal as in (6.2.4). Our discussion makes it possible to construct sets of multipliers that are mixtures of those that have been listed, as illustrated for the real-parameter example

$$\boldsymbol{\Delta}_c = \{\mathrm{diag}(\delta_1 I_2, \delta_2, \delta_3, \delta_4 I_4) : \ \delta_1 \in [-1, 1], \ \delta_2 \in [2, 4], \ |\delta_3| \leq 1 - \delta_4, \ \delta_4 \geq -1\}.$$

One can then work with the multiplier set of all

$$P = \left(\begin{array}{cccc|cccc} -D_1 & 0 & 0 & 0 & G_1 & 0 & 0 & 0 \\ 0 & -D_2 & 0 & 0 & 0 & 3D_2+G_2 & 0 & 0 \\ 0 & 0 & Q_{11} & Q_{12} & 0 & 0 & S_{11} & S_{12} \\ 0 & 0 & Q_{12}^* & Q_{22} & 0 & 0 & S_{21} & S_{22} \\ \hline -G_1 & 0 & 0 & 0 & D_1 & 0 & 0 & 0 \\ 0 & -3D_2-G_2 & 0 & 0 & 0 & -2D_2 & 0 & 0 \\ 0 & 0 & S_{11}^* & S_{21}^* & 0 & 0 & R_{11} & R_{12} \\ 0 & 0 & S_{12}^* & S_{22}^* & 0 & 0 & R_{12}^* & R_{22} \end{array}\right)$$

with

$$D_1, \geq 0,\ D_2 \geq 0,\ G_1 = G_1^*,\ G_2 = G_2^* \text{ and } Q_{11} \leq 0,\ Q_{22} \leq 0$$

as well as

$$\left(\begin{array}{cc} \delta_3 & 0 \\ 0 & \delta_4 I_4 \\ \hline 1 & 0 \\ 0 & 1 \end{array}\right)^* \left(\begin{array}{cc|cc} Q_{11} & Q_{12} & S_{11} & S_{12} \\ Q_{12}^* & Q_{22} & S_{21} & S_{22} \\ \hline S_{11}^* & S_{21}^* & R_{11} & R_{12} \\ S_{12}^* & S_{22}^* & R_{12}^* & R_{22} \end{array}\right) \left(\begin{array}{cc} \delta_3 & 0 \\ 0 & \delta_4 I_4 \\ \hline 1 & 0 \\ 0 & 1 \end{array}\right) \geq 0,\ \ (\delta_3, \delta_4) = \left\{ \begin{array}{l} (-2,-1) \\ (2,-1). \\ (0,1) \end{array}\right.$$

Of course it is straightforward to construct other block structures which all might lead to different numerical results in actual computations.

With the various classes of multipliers we arrive at the chain of inclusions

$$\boldsymbol{P_n} \subset \boldsymbol{P_d} \subset \boldsymbol{P_f} \subset \boldsymbol{P_{fe}} \subset \boldsymbol{P_{\mathbf{all}}}$$

which is briefly characterized as allowing the reduction of conservatism at the expense of increase in complexity of their descriptions. As the main distinction we stress that the number of LMI's to describe $\boldsymbol{P_d}$ grows linearly in the number m of parameters, whereas that for parametrizing $\boldsymbol{P_f}$ grows exponentially in m. This shows the practical relevance of allowing for mixed block structures to be able to reduce conservatism while avoiding the explosion of computational complexity.

Remark 6.12 In practice, instead of just testing feasibility of (6.5.3) for some multiplier class $P \in \boldsymbol{P}$, it is rather suggested to choose some $\epsilon > 0$ and to infimize γ over $P \in \boldsymbol{P}$ with

$$\begin{pmatrix} I & 0 \\ A & B \end{pmatrix}^* \begin{pmatrix} Q & S \\ S^* & R \end{pmatrix} \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} + \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^* \begin{pmatrix} Q_p & S_p \\ S_p^* & R_p \end{pmatrix} \begin{pmatrix} 0 & I \\ C & D \end{pmatrix} \leq \gamma \begin{pmatrix} \epsilon I & 0 \\ 0 & I \end{pmatrix}.$$

In this fashion the largest eigenvalue of the left-hand side is really pushed to smallest possible values. If there exists some $P \in \boldsymbol{P}$ with (6.5.3) then, trivially, there exists some $\epsilon > 0$ such that the infimal value γ_* is negative. It requires only slight modifications of the arguments to follow in order to show that

$$\Delta \star H \text{ is well-posed and } \begin{pmatrix} I \\ \Delta \star H \end{pmatrix}^* P_p \begin{pmatrix} I \\ \Delta \star H \end{pmatrix} \leq \gamma_* I \ \text{ for all } \ \Delta \in \boldsymbol{\Delta}_c.$$

Hence the value of γ_* provides an indication of distance to robust performance failure.

To conclude this section let us formulate a somewhat more abstract version of the full block S-procedure in terms of quadratic forms on parameter dependent subspaces which is considerably more powerful than the concrete version as will be illustrated by means of examples. Very similar to our general recipe for constructing LFR's, the abstraction relies on the mapping interpretation of the involved matrices. Indeed, if we fix $\Delta \in \mathbf{\Delta}_c$ such that $I - A\Delta$ is invertible we observe that

$$\begin{pmatrix} I \\ \Delta \star H \end{pmatrix}^* P_p \begin{pmatrix} I \\ \Delta \star H \end{pmatrix} < 0$$

iff

$$\xi \neq 0,\ \eta = [\Delta \star H]\xi \ \Rightarrow\ \begin{pmatrix} \xi \\ \eta \end{pmatrix}^* P_p \begin{pmatrix} \xi \\ \eta \end{pmatrix} < 0$$

iff (using the LFR)

$$\xi \neq 0,\ \begin{pmatrix} z \\ \eta \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} w \\ \xi \end{pmatrix},\ w = \Delta z \ \Rightarrow\ \begin{pmatrix} \xi \\ \eta \end{pmatrix}^* P_p \begin{pmatrix} \xi \\ \eta \end{pmatrix} < 0$$

iff (just by rearrangements)

$$\begin{pmatrix} 0 & I \end{pmatrix}\begin{pmatrix} w \\ \xi \end{pmatrix} \neq 0,\ \begin{pmatrix} I & 0 \\ A & B \end{pmatrix}\begin{pmatrix} w \\ \xi \end{pmatrix} \in \mathrm{im}\begin{pmatrix} \Delta \\ I \end{pmatrix} \Rightarrow$$
$$\Rightarrow \left[\begin{pmatrix} 0 & I \\ C & D \end{pmatrix}\begin{pmatrix} w \\ \xi \end{pmatrix}\right]^* P_p \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}\begin{pmatrix} w \\ \xi \end{pmatrix} < 0.$$

On the other hand $I - A\Delta$ is invertible iff $(I - A\Delta)z = 0$ implies $z = 0$ iff

$$z = Aw,\ w = \Delta z \ \Rightarrow\ w = 0,\ z = 0$$

iff

$$\begin{pmatrix} I \\ A \end{pmatrix} w \in \mathrm{im}\begin{pmatrix} \Delta \\ I \end{pmatrix} \Rightarrow w = 0$$

iff

$$\begin{pmatrix} 0 & I \end{pmatrix}\begin{pmatrix} w \\ \xi \end{pmatrix} = 0,\ \begin{pmatrix} I & 0 \\ A & B \end{pmatrix}\begin{pmatrix} w \\ \xi \end{pmatrix} \in \mathrm{im}\begin{pmatrix} \Delta \\ I \end{pmatrix} \Rightarrow \begin{pmatrix} w \\ \xi \end{pmatrix} = 0.$$

With the abbreviations

$$W = \begin{pmatrix} 0 & I \end{pmatrix},\ V = \begin{pmatrix} I & 0 \\ A & B \end{pmatrix},\ S(\Delta) = \begin{pmatrix} \Delta \\ I \end{pmatrix},\ T = \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}^* P_p \begin{pmatrix} 0 & I \\ C & D \end{pmatrix}$$

it is hence quite obvious that Theorem 6.11 is a direct consequence of the following result.

Theorem 6.13 (Abstract full block S-procedure) *Let $V \in \mathbb{C}^{v\times n}$, $W \in \mathbb{C}^{w\times n}$ and $T = T^* \in \mathbb{C}^{n\times n}$ where $x^*Tx \geq 0$ if $Wx = 0$. Moreover suppose that $S(\Delta) \in \mathbb{C}^{v\times *}$ depends continuously on $\Delta \in \mathbf{\Delta}_c$ and has constant rank on $\mathbf{\Delta}_c$. Then, for all $\Delta \in \mathbf{\Delta}_c$,*

$$Vx \in \mathrm{im}(S(\Delta)) \ \Rightarrow\ \begin{cases} x^*Tx < 0 & \text{if} \quad Wx \neq 0 \\ x = 0 & \text{if} \quad Wx = 0 \end{cases}$$

if and only if there exists some Hermitian matrix $P \in \mathbb{C}^{v\times v}$ with

$$S(\Delta)^* P S(\Delta) \geq 0 \ \text{ for all } \ \Delta \in \mathbf{\Delta}_c \ \text{ and } \ V^*PV + T < 0.$$

Remark 6.14 We observe that the proof of the 'if' direction is trivial. Suppose $Vx \in \mathrm{im}(S(\Delta))$. Then there exists z with $Vx = S(\Delta)z$ which in turn implies $x^*V^*PVx = z^*S(\Delta)^*PS(\Delta)z \geq 0$. Let $Wx \neq 0$ such that $x \neq 0$ and hence $x^*V^*PVx + x^*Tx < 0$ which implies $x^*Tx < 0$. Now suppose $Wx = 0$ such that $x^*Tx \geq 0$ by hypothesis. If $x \neq 0$ we infer $x^*V^*PVx < 0$ which contradicts $x^*V^*PVx \geq 0$; hence we can conclude $x = 0$. A proof of the other direction is somewhat less simple and can be found in [**?**].

Remark 6.15 Let us explicitly formulate the specialization of Theorem 6.13 with empty matrices Q and W respectively:

$$Vx \in \mathrm{im}(S(\Delta)) \;\Rightarrow\; x = 0 \;\text{ holds for all }\; \Delta \in \boldsymbol{\Delta}_c$$

iff there exists some Hermitian $P \in \mathbb{C}^{v \times v}$ with

$$S(\Delta)^*PS(\Delta) \geq 0 \;\text{ for all }\; \Delta \in \boldsymbol{\Delta}_c \;\text{ and }\; V^*PV < 0.$$

The latter specialization is particularly useful for guaranteeing well-posedness of LFR's. We have argued above that $I - A\Delta$ is non-singular iff

$$\begin{pmatrix} I \\ A \end{pmatrix} w \in \mathrm{im}\begin{pmatrix} \Delta \\ I \end{pmatrix} \;\Rightarrow\; w = 0.$$

This holds for all $\Delta \in \boldsymbol{\Delta}_c$ iff

$$\begin{pmatrix} I \\ A \end{pmatrix}^* P \begin{pmatrix} I \\ A \end{pmatrix} < 0 \;\text{ for some }\; P \in \boldsymbol{P}_{\mathbf{all}}.$$

To conclude this section let us illustrate the flexibility of the abstract version of the full-block S-procedure by considering the problem of guaranteeing that

$$N - \Delta M \;\text{ has full column rank for all }\; \Delta \in \boldsymbol{\Delta}_c.$$

Literally the same arguments as for non-singularity reveal that his holds for all $\Delta \in \boldsymbol{\Delta}_c$ iff

$$\begin{pmatrix} N \\ M \end{pmatrix}^* P \begin{pmatrix} N \\ M \end{pmatrix} < 0 \;\text{ for some }\; P \in \boldsymbol{P}_{\mathbf{all}}.$$

In the framework suggested it is straightforward to devise numerically verifiable robust full-rank tests, in contrast to the standard structured singular value theory which requires the development of dedicated algorithms [**?**].

6.6 Numerical solution of robust linear algebra problems

6.6.1 Relaxations for computing upper bounds

Let us assume that the decision variable is $v \in \mathbb{R}^n$, and suppose that

$$H(v) = \begin{pmatrix} A & B \\ C(v) & D(v) \end{pmatrix} \text{ with } C(v),\ D(v) \text{ depending affinely on } v,$$

and

$$P_p(v) = \begin{pmatrix} Q_p(v) & S_p \\ S_p^* & T_p U_p(v)^{-1} T_p^* \end{pmatrix} \text{ with } Q_p(v),\ U_p(v) \text{ depending affinely on } v.$$

Moreover let $\Delta(\delta)$ be a parameter-block which is linear in δ, and $\boldsymbol{\delta}$ is some subset of $\mathbb{R}^m$.

With the linear cost functional defined by $c \in \mathbb{R}^n$ we consider the following paradigm problem.

Problem Infimize $c^* v$ over all v which satisfy $U_p(v) > 0$ and

$$\det(I - A\Delta(\delta)) \neq 0,\ \begin{pmatrix} I \\ \Delta(\delta) \star H(v) \end{pmatrix}^T P_p(v) \begin{pmatrix} I \\ \Delta(\delta) \star H(v) \end{pmatrix} < 0 \text{ for all } \delta \in \boldsymbol{\delta}. \quad (6.6.1)$$

Denote the optimal value by $\gamma_{\text{opt}} \in [-\infty, \infty]$.

In various circumstances we have made explicit how to construct a set $\boldsymbol{P}$ with an LMI description and such that

$$\begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix}^* P \begin{pmatrix} \Delta(\delta) \\ I \end{pmatrix} \geq 0 \text{ for all } \delta \in \boldsymbol{\delta}. \quad (6.6.2)$$

This allows to consider the following relaxation.

Problem Infimize $c^* v$ over all v and $P \in \boldsymbol{P}$ with

$$U_p(v) > 0,\ \begin{pmatrix} I & 0 \\ A & B \end{pmatrix}^* P \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} + \begin{pmatrix} 0 & I \\ C(v) & D(v) \end{pmatrix}^* P_p(v) \begin{pmatrix} 0 & I \\ C(v) & D(v) \end{pmatrix} < 0.$$

Denote the optimal value by $\gamma_{\text{rel}} \in [-\infty, \infty]$.

Since $\boldsymbol{P}$ has an LMI description, the linearization lemma **??** reveals that we can compute γ_{rel} by solving a genuine LMI problem. The (simple) sufficiency conclusion in the full-block S-procedure implies that v is feasible for Problem 6.6.1 if v and $P \in \boldsymbol{P}$ are feasible for Problem 6.6.1. We can hence infer

$$\gamma_{\text{opt}} \leq \gamma_{\text{rel}},$$

including the cases $\gamma_{\text{opt}} \in \{-\infty, \infty\}$ and $\gamma_{\text{rel}} \in \{-\infty, \infty\}$ with the usual interpretations.

Unfortunately it is impossible to make a priori statements on the relaxation gap $\gamma_{\text{rel}} - \gamma_{\text{opt}}$ in the generality as discussed here. However, let us stress again that this gap can be possibly reduced by employing a larger (superset) of $\boldsymbol{P}$, and that $\gamma_{\rm opt} \leq \gamma_{\rm rel}$ if $\boldsymbol{\delta}$ is compact and if replacing $\boldsymbol{P}$ with the set of all multipliers P that satisfy 6.6.2, just due to the full-block S-procedure. Summarizing, the choice of an increasing family of multipliers $\boldsymbol{P}_1 \subset \boldsymbol{P}_2 \subset \cdots$ leads to a family of LMI-relaxations with non-increasing optimal values.

We conclude this section by stressing that the developed techniques allow an immediate extension to multiple-objectives expressed by finitely many constraints as

$$\begin{pmatrix} I \\ \Delta(\delta) \star H(v) \end{pmatrix}^T P_{p,j}(v) \begin{pmatrix} I \\ \Delta(\delta) \star H(v) \end{pmatrix} < 0 \text{ for all } \delta \in \boldsymbol{\delta},\ j = 1, \dots, p.$$

These constraints are relaxed as follows: For all $j = 1, \dots, p$, there exists $P_j \in \boldsymbol{P}$ with

$$\begin{pmatrix} I & 0 \\ A & B \end{pmatrix}^* P_j \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} + \begin{pmatrix} 0 & I \\ C(v) & D(v) \end{pmatrix}^* P_{p,j}(v) \begin{pmatrix} 0 & I \\ C(v) & D(v) \end{pmatrix} < 0.$$

It is hence essential to exploit the extra freedom to relax any individual constraint with its individual multiplier in order to keep conservatism subdued. We remark as well that one could even allow the LFR and the class of multipliers used for each relaxation vary from constraint to constraint without introducing any extra complications.

6.6.2 When are relaxations exact?

We have seen in the previous section that multiplier relaxations typically cause a gap $\gamma_{\rm rel} - \gamma_{\rm opt}$. In this section we formulate a general principle about a numerical verifiable sufficient condition for the absences of this gap and hence for exactness of the relaxation.

One situation is simple: If the relaxation is strictly feasible and has optimal value $\gamma_{\rm rel} = -\infty$. This certainly implies $\gamma_{\rm opt} = -\infty$ and there is no relaxation gap. Let us hence assume from now on that $\gamma_{\rm rel} > -\infty$.

The key is to consider the dual of the suggested relaxations. For this purpose we first apply the linearization lemma **??** in order to reformulate the relaxation as infimizing c^* over v and $P \in \boldsymbol{P}$ with

$$\begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix}^* P \begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix} + \begin{pmatrix} 0 & C(v)^* S_p^* & C(v)^* T_p \\ S_p C(v) & S_p D(v) + D(v)^* S_p^* + Q_p(v) & D(v)^* T_p \\ T_p^* C(v) & T_p^* D(v) & -U_p(v) \end{pmatrix} < 0.$$

With the standard basis vectors $e_1, \dots, e_n$ of $\mathbb{R}^n$ and with $e_0 = 0$ define

$$c_j = c(e_j),\ \ W_j = \begin{pmatrix} 0 & C(e_j)^* S_p^* & C(e_j)^* T_p \\ S_p C(e_j) & S_p D(e_j) + D(e_j)^* S_p^* + Q_p(e_j) & D(e_j)^* T_p \\ T_p^* C(e_j) & T_p^* D(e_j) & -U_p(e_j) \end{pmatrix},\ \ j = 0, 1, \dots, n.$$

Then we have to infimize c^*x over x and $P \in \boldsymbol{P}$ subject to

$$\begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix}^* P \begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix} + W_0 + \sum_j x_j W_j \leq 0.$$

For $\boldsymbol{\delta} = \operatorname{conv}\{\delta^1, \ldots, \delta^N\}$ let us now consider the concrete class $\boldsymbol{P}$ of full block multipliers P which are just implicitly described by the (strictly feasible) constraints

$$\begin{pmatrix} I \\ 0 \end{pmatrix}^* P \begin{pmatrix} I \\ 0 \end{pmatrix} \leq 0, \quad \begin{pmatrix} \Delta(\delta^j) \\ I \end{pmatrix}^* P \begin{pmatrix} \Delta(\delta^j) \\ I \end{pmatrix} \geq 0 \text{ for all } j = 1, \ldots, N.$$

With Lagrange multipliers M, $\hat{M}$, M_j, $j = 1, \ldots, N$, the Lagrangian hence reads as

$$\sum_{j=1}^n c_j x_j + \langle \begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix}^* P \begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix}, M \rangle + \langle W_0, M \rangle + \sum_{j=1}^n x_j \langle W_j, M \rangle +$$
$$+ \langle \begin{pmatrix} I \\ 0 \end{pmatrix}^* P \begin{pmatrix} I \\ 0 \end{pmatrix}, \hat{M} \rangle - \sum_{j=1}^N \langle \begin{pmatrix} \Delta(\delta^j) \\ I \end{pmatrix}^* P \begin{pmatrix} \Delta(\delta^j) \\ I \end{pmatrix}, M_j \rangle$$

or, after 'sorting for primal variables',

$$\langle W_0, M \rangle + \sum_{j=1}^n (c_j + \langle W_j, M \rangle) x_j +$$
$$+ \langle P, \begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix} M \begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix}^* + \begin{pmatrix} I \\ 0 \end{pmatrix} \hat{M} \begin{pmatrix} I \\ 0 \end{pmatrix}^* + \sum_{j=1}^N \begin{pmatrix} \Delta(\delta^j) \\ I \end{pmatrix} M_j \begin{pmatrix} \Delta(\delta^j) \\ I \end{pmatrix}^* \rangle.$$

By standard duality we can draw the following conclusions:

(a) The primal is not strictly feasible (which just means $\gamma_{\text{rel}} = \infty$) iff there exist $M \geq 0$, $\hat{M} \geq 0$, $M_j \geq 0$ with $\langle W_0, M \rangle \geq 0$, $\langle W_j, M \rangle = 0$, $j = 1, \ldots, N$ and

$$\begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix} M \begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix}^* + \begin{pmatrix} I \\ 0 \end{pmatrix} \hat{M} \begin{pmatrix} I \\ 0 \end{pmatrix}^* + \sum_{j=1}^N \begin{pmatrix} \Delta(\delta^j) \\ I \end{pmatrix} M_j \begin{pmatrix} \Delta(\delta^j) \\ I \end{pmatrix}^* = 0. \tag{6.6.3}$$

(b) If the primal is strictly feasible and has finite optimal value there exists $M \geq 0$, $\hat{M} \geq 0$, $M_j \geq 0$ which maximize $\langle W_0, M \rangle$ under the constraints $\langle W_j, M \rangle + c_j = 0$, $j = 1, \ldots, N$ and (6.6.3). The optimal value of this Lagrange dual problem equals γ_{rel}.

Theorem 6.16 *If $\gamma_{rel} = \infty$ and if M in 1. has rank one then $\gamma_{opt} = \infty$. If $\gamma_{rel} < \infty$ and if M in 2. has rank one then $\gamma_{opt} = \gamma_{rel}$.*

It is convenient to summarize this result as follows: If there exists either an infeasibility certificate or a dual optimizer such that M has rank one then the relaxation is exact.

Proof. If M has rank one, it can be decomposed as $M = mm^*$. Let us partition m as $\mathrm{col}(w, \xi, \xi_e)$ according to the columns of $(A\ B\ 0)$ and define $z = (A\ B\ 0)m = Aw + B\xi$. The essential point is to conclude from (6.6.3) that

$$w = \Delta(\delta^0)z \ \text{ for some } \ \delta^0 \in \boldsymbol{\delta}.$$

Indeed just by using the definitions we obtain with (6.6.3) the relation

$$\begin{pmatrix} w \\ z \end{pmatrix}\begin{pmatrix} w^* & z^* \end{pmatrix} + \begin{pmatrix} I \\ 0 \end{pmatrix}\hat{M}\begin{pmatrix} I & 0 \end{pmatrix} - \sum_{j=1}^{N}\begin{pmatrix} \Delta(\delta^j) \\ I \end{pmatrix} M_j \begin{pmatrix} \Delta(\delta^j) \\ I \end{pmatrix}^* = 0.$$

From $zz^* = \sum_{j=1}^N M_j$ we infer that $z^*x = 0$ implies $M_jx = 0$ for all j such that there exist α_j with $M_j = z\alpha_j z^*$. If $z \neq 0$ have $\alpha_j \geq 0$ and $\sum_{j=1}^N \alpha_j = 1$. Now $wz^* = \sum_{j=1}^N \Delta(\delta^j)M_j = 0$ allows to conclude by right-multiplication with z and division by $z^*z \neq 0$ that

$$w = \sum_{j=1}^{N}\Delta(\delta^j)z\alpha_j = \Delta\left(\sum_{j=1}^{N}\alpha_j\delta^j\right)z = \Delta(\delta^0)z \ \text{ with } \ \delta^0 := \sum_{j=1}^{N}\alpha_j\delta^j \in \boldsymbol{\delta}.$$

If $z = 0$ we infer $M_j = 0$ and hence $ww^* + \hat{M} = 0$ which implies $w = 0$ (and $\hat{M} = 0$) such that $w = \Delta(\delta^0)z$ holds for an arbitrary $\delta^0 \in \boldsymbol{\delta}$.

Let us now assume $\gamma_{\mathrm{rel}} = \infty$. In can happen that $I - A\Delta(\delta^0)$ is singular. This implies that Problem 6.6.1 is not feasible and hence $\gamma_{\mathrm{opt}} = \infty$.

Let us continue under the hypothesis that $I - A\Delta(\delta^0)$ is non-singular which implies $w = \Delta(\delta^0)(I - A\Delta(\delta^0))^{-1}B\xi$ (due to $z = Aw + B\xi$ and $w = \Delta(\delta^0)z$). We infer for all $x \in \mathbb{R}^n$:

$$\begin{aligned}
&\xi^*\begin{pmatrix} * \\ * \end{pmatrix}^*\begin{pmatrix} Q(x) & S \\ S^* & TU(x)^{-1}T^* \end{pmatrix}\begin{pmatrix} I \\ D(x) + C(x)\Delta(\delta^0)(I - A\Delta(\delta^0))^{-1}B \end{pmatrix}\xi = \\
&= \begin{pmatrix} w \\ \xi \end{pmatrix}^*\begin{pmatrix} 0 & I \\ C(x) & D(x) \end{pmatrix}^*\begin{pmatrix} Q(x) & S \\ S^* & TU(x)^{-1}T^* \end{pmatrix}\begin{pmatrix} 0 & I \\ C(x) & D(x) \end{pmatrix}\begin{pmatrix} w \\ \xi \end{pmatrix} = \\
&= \max_\eta \begin{pmatrix} w \\ \xi \\ \eta \end{pmatrix}^*\begin{pmatrix} 0 & C(x)^*S_p^* & C(x)^*T_p \\ S_pC(x) & S_pD(x) + D(x)^*S_p^* + Q_p(x) & D(x)^*T_p \\ T_p^*C(x) & T_p^*D(x) & -U_p(x) \end{pmatrix}\begin{pmatrix} w \\ \xi \\ \eta \end{pmatrix} \geq \\
&\geq \begin{pmatrix} w \\ \xi \\ \xi_e \end{pmatrix}^*\begin{pmatrix} 0 & C(x)^*S_p^* & C(x)^*T_p \\ S_pC(x) & S_pD(x) + D(x)^*S_p^* + Q_p(x) & D(x)^*T_p \\ T_p^*C(x) & T_p^*D(x) & -U_p(x) \end{pmatrix}\begin{pmatrix} w \\ \xi \\ \xi_e \end{pmatrix} = \\
&= m^*W_0m + \sum_{j=1}^{n} x_j m^*W_jm.
\end{aligned}$$

By hypothesis we have

$$0 \leq \langle W_0, M\rangle + \sum_{j=1}^{n} x_j \langle W_j, M\rangle = m^* W_0 m + \sum_{j=1}^{n} x_j m^* W_j m \text{ for all } x \in \mathbb{R}^n.$$

From the above-derived chain of inequalities we can hence infer that no $x \in \mathbb{R}^n$ can ever be feasible for (6.6.1), which implies again that $\gamma_{\rm opt} = \infty$ and thus $\gamma_{\rm opt} = \gamma_{\rm rel}$.

Let us now assume $\gamma_{\rm rel} < \infty$. Since problem 6.6.1 is strictly feasible it is guaranteed that $I - A\Delta(\delta^0)$ is nonsingular, and the above-derived chain of inequalities is again true. Let us apply this chain of inequalities for any x that is feasible for Problem 6.6.1 to infer

$$0 \geq m^* W_0 m + \sum_{j=1}^{n} x_j m^* W_j m.$$

Just due to $\langle W_0, M\rangle = \gamma_{\rm rel}$ and $c_j = \langle W_j, M\rangle$ we can conclude

$$\sum_{j=1}^{n} c_j x_j = \gamma_{\rm rel} - \langle W_0, M\rangle - \sum_{j=1}^{n} x_j \langle W_j, M\rangle = \gamma_{\rm rel} - m^* W_0 m - \sum_{j=1}^{n} x_j m^* W_j m$$

and hence $c^* x \geq \gamma_{\rm rel}$. Since v was an abitrary feasible point of Problem 6.6.1 we infer $\gamma_* \geq \gamma_{\rm rel}$ and thus $\gamma_* = \gamma_{\rm rel}$. ■

It is interesting to note that the proof reveals how one can construct a worst-case parameter uncertainty from a rank one dual multiplier! In a similar fashion it is possible to apply the novel principle to a whole variety of other problems which is reveals its charming character.

In one specific case with one-dimensional uncertainty one can in fact *always* construct a dual multiplier of rank one.

Theorem 6.17 *Suppose that* $\boldsymbol{\delta} = \mathrm{conv}\{\delta^1, \delta^2\}$. *Then* $\gamma_{rel} = \gamma_{opt}$.

Proof. Let us first investigate the constraint (6.6.3) in the specific case $N = 2$. Left-multiplication with $\begin{pmatrix} I & -\Delta(\delta^1) \end{pmatrix}$ and right-multiplication with $\begin{pmatrix} I & -\Delta(\delta^2) \end{pmatrix}^*$ leads with

$$U := \begin{pmatrix} I & -\Delta(\delta^1) \end{pmatrix} \begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix} \text{ and } V := \begin{pmatrix} I & -\Delta(\delta^2) \end{pmatrix} \begin{pmatrix} I & 0 & 0 \\ A & B & 0 \end{pmatrix}$$

to the relation $UMV^* + \hat{M} = 0$. This implies $UMV^* + V^*MU = -\hat{M} \leq 0$ and $UMV^* - V^*MU = 0$ which can be expressed as

$$\begin{pmatrix} U^* \\ V^* \end{pmatrix}^* \begin{pmatrix} 0 & -M \\ -M & 0 \end{pmatrix} \begin{pmatrix} U^* \\ V^* \end{pmatrix} \geq 0, \quad \begin{pmatrix} U^* \\ V^* \end{pmatrix}^* \begin{pmatrix} 0 & M \\ -M & 0 \end{pmatrix} \begin{pmatrix} U^* \\ V^* \end{pmatrix} = 0. \qquad (6.6.4)$$

Now choose an arbitrary $x \in \mathbb{R}^n$. Since $M \neq 0$ (???), we can apply Lemma **??** to infer the existence of a vector $m \neq 0$ (possibly depending on x) with

$$m^* \left(W_0 + \sum_{j=1}^{n} x_j W_j \right) m \geq \langle W_0 + \sum_{j=1}^{n} x_j W_j, M \rangle. \tag{6.6.5}$$

and

$$\begin{pmatrix} m^*U^* \\ m^*V^* \end{pmatrix}^* \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} m^*U^* \\ m^*V^* \end{pmatrix} \geq 0, \quad \begin{pmatrix} m^*U^* \\ m^*V^* \end{pmatrix}^* \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} m^*U^* \\ m^*V^* \end{pmatrix} = 0.$$

If we partition m again as $\mathrm{col}(w, \xi, \xi_e)$ and if we define $z = Aw + B\xi$ we infer $Um = w - \Delta(\delta^1)z$ and $Vm = w - \Delta(\delta^2)z$. If $Vm = 0$ we conclude $w = \Delta(\delta^2)z$. If $Vm \neq 0$ there exists, by Lemma **??**, some $\alpha \leq 0$ with $w - \Delta(\delta^1)z = \alpha w - \alpha\Delta(\delta^2)z$ or

$$w = \left(\frac{1}{1-\alpha}\Delta(\delta^1) - \frac{\alpha}{1-\alpha}\Delta(\delta^2) \right) z = \Delta\left(\frac{1}{1-\alpha}\delta^1 - \frac{\alpha}{1-\alpha}\delta^2 \right) z.$$

Therefore in both cases $w = \Delta(\delta^0)z$ for some $\delta^0 \in \mathrm{conv}\{\delta^1, \delta^2\}$.

If $\gamma_{\mathrm{rel}} = \infty$, we can infer $\langle W_0 + \sum_{j=1}^{n} x_j W_j, M \rangle \geq 0$ and hence with (6.6.5) we obtain $m^* W_0 m + \sum_{j=1}^{n} x_j m^* W_j m \geq 0$. This allows to finish the proof as above.

If $\gamma_{\mathrm{rel}} < \infty$, and if x is chosen feasible for Problem **??**, we infer again $0 \geq m^* W_0 m + \sum_{j=1}^{n} x_j m^* W_j m$. Now we exploit (6.6.5) to conclude $0 \geq \langle W_0 + \sum_{j=1}^{n} x_j W_j, M \rangle$ or $c^* x \geq \gamma_{\mathrm{rel}}$. We conclude again that $\gamma_{\mathrm{opt}} \geq \gamma_{\mathrm{rel}}$, and thus equality. ■

Remarks. This is a novel result on the absence of a relaxation gap for robust semi-definite programs with implicitly described full-block multipliers. Only slight modifications of our arguments lead to the same insights for block-diagonal multipliers, which reveals that we have provided generalizations of the result from [**?**] and [**?**]. It is important to stress that full-block multipliers are in general expected to lead to less conservative results. We can conclude, however, that the one parameter case allows the restriction to the diagonal multiplier class without introducing conservatism. In case of complex uncertainties as appearing in SSV-theory we will be able to easily derive similar insights in a considerable more general setting in the sections to follow.

Let us summarize our findings in the following general principle.

(Absence of relaxation gap). If the Lagrange multiplier corresponding the concrete S-procedure multiplier inequality has rank one then there is not relaxation gap.

It is the main purpose of this section to illustrate this principle by means of obtaining a novel result on the absence of a relaxation gap for (implicitly described) full-block multipliers. A similar derivation in a somewhat more general setting including full complex uncertainty blocks as they appear in robust control will follow in Section

Chapter 7

Analysis of input-output behavior

The main concern in control theory is to study how signals are processed by dynamical systems and how this processing can be influenced to achieve a certain desired behavior. For that purpose one has to specify the signals (time series, trajectories) of interest. This is done by deciding on the set of values which the signals can take (such as $\mathbb{R}^n$) and on the time set on which they are defined (such as the full time axis $\mathbb{R}$, the half axis $[0, \infty)$ or the corresponding discrete time versions $\mathbb{Z}$ and $\mathbb{N}$). A dynamical system is then nothing but a mapping that assigns to a certain input signal some output signal. Very often, such a mapping is defined by a differential equation with a fixed initial condition or by an integral equation, which amounts to considering systems or mappings with a specific explicit description or representation.

The first purpose of this chapter is to discuss stability properties of systems in a rather abstract and general setting. In a second step we turn to specialized system representations and investigate in how far one can arrive at refined results which are amenable to computational techniques.

We stress that it is somewhat restrictive to consider a system as a mapping of signals, thus fixing what is considered as an input signal or output signal. It is not too difficult to extend our results to the more general and elegant setting of viewing a system as a subset of signals or, in the modern language, as a behavior [**?**]. In particular concerning robust stability issues, this point of view has been adopted as well in the older literature in which systems have been defined as relations rather than mappings [**?**, **?**, **?**, **?**].

7.1 Basic Notions

Let us now start the concrete ingredients that are required to develop the general theory. The basis for defining signal spaces is formed by L^n, the set of all time-functions or signals $x : [0, \infty) \to \mathbb{R}^n$ which are Lebesgue-measurable. Without bothering too much about the exact definition one should recall that all continuous or piece-wise continuous signals are contained in L^n. The book by Rudin [**?**] is an excellent source for a precise exposition of all the mathematical concepts in this chapter. From an engineering perspective it is helpful to consult [**?**] in which some of the advanced constructions are based on an elementary yet insightful footing.

For any $x \in L^n$ one can calculate the energy integral

$$\|x\|_2 := \sqrt{\int_0^\infty \|x(t)\|^2\, dt}$$

which is either finite or infinite. If we collect all signals with a finite value we arrive at

$$L_2^n := \{x \in L^n : \ \|x\|_2 < \infty\}.$$

It can be shown that L_2^n is a linear space, that $\|.\|_2$ is a norm on L_2^n, and that L_2^n is complete or a Banach space. The quantity $\|x\|_2$ is often called L_2-norm or **energy** of the signal x. In the sequel if the number of components n of the underlying signal is understood from the context or irrelevant, we simply write L_2 instead of L_2^n.

Actually L_2^n admits an additional structure. Indeed, let us define the bilinear form

$$\langle x, y\rangle = \int_0^\infty x(t)^T y(t)\, dt$$

on $L_2^n \times L_2^n$. Bilinearity just means that $\langle ., y\rangle$ is linear for each $y \in L_2^n$ and $\langle x, .\rangle$ is linear for each $x \in L_2^n$. It is not difficult to see that $\langle ., .\rangle$ defines an inner product on L_2^n. It is obvious that the above defined norm and the inner product are related in a standard fashion by $\|x\|_2^2 = \langle x, x\rangle$. As a complete inner product space, L_2^n is in fact a Hilbert space.

For any $x \in L_2^n$ one can calculate the Fourier transform $\widehat{x}$ which is a function mapping the imaginary axis $\mathbb{C}^0$ into $\mathbb{C}^n$ such that

$$\int_{-\infty}^\infty \widehat{x}(i\omega)^* \widehat{x}(i\omega)\, d\omega \ \text{ is finite.}$$

Indeed, a fundamental results in the theory of the Fourier transformation on L_2-spaces, the so-called Parseval theorem, states that

$$\int_0^\infty x(t)^T y(t)\, dt = \frac{1}{2\pi}\int_{-\infty}^\infty \widehat{x}(i\omega)^* \widehat{y}(i\omega)\, d\omega.$$

Recall that the Fourier transform $\widehat{x}$ has in fact a unique continuation into $\mathbb{C}^0 \cup \mathbb{C}^+$ that is analytic in $\mathbb{C}^+$. Hence, $\widehat{x}$ is not just an element of $L_2(\mathbb{C}^0, \mathbb{C}^n)$ but even of the subspace $H_2(\mathbb{C}^+, \mathbb{C}^n)$, one of the Hardy spaces. Moreover, one has $L_2(\mathbb{C}^0, \mathbb{C}^n) = H_2(\mathbb{C}^-, \mathbb{C}^n) + H_2(\mathbb{C}^+, \mathbb{C}^n)$, the sum on the right is direct, and the two spaces are orthogonal to each other. This corresponds via the Payley-Wiener theorem to the orthogonal direct sum decomposition $L_2(\mathbb{R}, \mathbb{C}^n) = L_2((-\infty, 0], \mathbb{C}^n) + L_2([0, \infty), \mathbb{C}^n)$. These facts which can be found with more details e.g. in [**?**, **?**] are mentioned only for clarification and they are not alluded to in the sequel.

Roughly speaking, stability of systems is related to the property that it maps any input signal of bounded energy into an output signal which also has bounded energy. Since it is desired to deal with unstable systems, we cannot confine ourselves to signals with finite energy. This motivates to introduce a larger class of signals that have finite energy over finite intervals only.

For that purpose it is convenient to introduce for each $T \geq 0$ the so-called **truncation operator** P_T: It assigns to any signal $x \in L^n$ the signal $P_T x$ that is identical to x on $[0, T]$ and that vanishes identically on (T, ∞):

$$P_T : L^n \to L^n,\ (P_T x)(t) := \begin{cases} x(t) & \text{for} \quad t \in [0, T], \\ 0 & \text{for} \quad t \in (T, \infty). \end{cases}$$

Note that L^n is a linear space and that P_T is a linear operator on that space with the property $P_T P_T = P_T$. Hence P_T is a projection.

Now it is straightforward to define the so-called extended space L^n_{2e}. It just consists of all signals $x \in L^n$ such that $P_T x$ has finite energy for all $T \geq 0$:

$$L^n_{2e} := \{x \in L^n :\ P_T x \in L^n_2 \ \text{ for all }\ T \geq 0\}.$$

Hence any $x \in L^n_{2e}$ has the property that

$$\|P_T x\|_2 = \int_0^T \|x(t)\|^2\, dt$$

is finite for every T. We observe that $\|P_T x\|_2$ does not decrease if T increases. Therefore, $\|P_T x\|_2$ viewed as a function of T either stays bounded for $T \to \infty$ and then converges, or it is unbounded and then it diverges to ∞. For any $x \in L^n_{2e}$ we can hence conclude that $\|P_T x\|_2$ is bounded for $T \to \infty$ iff x is contained in L^n_2. Moreover,

$$x \in L^n_2 \ \text{ implies } \ \|x\|_2 = \lim_{T \to \infty} \|P_T x\|_2.$$

Example 7.1 The signal defined by $x(t) = e^t$ is contained in L_{2e} but not in L_2. The signal defined by $x(0) = 0$ and $x(t) = 1/t$ for $t > 0$ is not contained in L_{2e}. In general, since continuous or piece-wise continuous signals are bounded on $[0, T]$ for every $T \geq 0$, they are all contained in L^n_{2e}. ■

A **dynamical system** S is a mapping

$$S : U_e \to L^l_{2e} \ \text{ with } \ U_e \subset L^k_{2e}.$$

Note that the domain of definition U_e is not necessarily a subspace and that, in practical applications, it is typically smaller than L^k_{2e} since it imposes some additional regularity or boundedness properties on its signals. The system S is **causal** if its domain of definition satisfies

$$P_T U_e \subset U_e \ \text{ for all } \ T \geq 0$$

and if

$$P_T S(u) = P_T S(P_T u) \text{ for all } T \geq 0,\ u \in U_e.$$

It is easily seen that $P_T S = P_T S P_T$ is equivalent to the following more intuitive fact: If u_1 and u_2 are two input signals that are identical on $[0, T]$, $P_T u_1 = P_T u_2$, then Su_1 and Su_2 are also identical on the same time-interval, $P_T S(u_1) = P_T S(u_2)$. In other words, the future values of the inputs do not have any effect on the past output values. This mathematical definition matches the intuitive notion of causality.

Our main interest in this abstract setting is to characterize invertibility and stability of a system. Among the many possibilities to define system stability, those based on gain or incremental gain have turned out to be of prominent importance.

Definition 7.2 The L_2**-gain** of the system S defined on U_e is given as

$$\|S\| := \sup\left\{\frac{\|P_T S(u)\|_2}{\|P_T u\|_2} \mid u \in U_e,\ T \geq 0,\ \|P_T u\|_2 \neq 0\right\} = \\ = \inf\{\gamma \in \mathbb{R} \mid \forall\, u \in U_e,\ T \geq 0 :\ \|P_T S(u)\|_2 \leq \gamma \|P_T u\|_2\}.$$

If $\|S\| < \infty$, S is said to have finite L_2-gain.

The reader is invited to prove that the two different ways of expressing $\|S\|$ do indeed coincide. If S has finite L_2-gain we infer for $\gamma := \|S\|$ that

$$\|P_T S(u)\|_2 \leq \gamma \|P_T u\|_2 \text{ for all } T \geq 0,\ u \in U_e. \tag{7.1.1}$$

In particular, this implies

$$\|S(u)\| \leq \gamma \|u\|_2 \text{ for all } u \in U_e \cap L_2^k. \tag{7.1.2}$$

Hence, if the input has finite energy, then the output has finite energy and the output energy is bounded by a constant times the input energy. If S is actually *causal*, the converse is true as well and property (7.1.2) implies (7.1.1). Indeed, for $u \in U_e$ we infer $P_T u \in U_e \cap L_2^k$ such that we can apply (7.1.2) to conclude $\|S(P_T u)\|_2 \leq \gamma \|P_T u\|_2$. Using causality, it remains to observe $\|P_T S(u)\|_2 = \|P_T S(P_T u)\|_2 \leq \|S(P_T u)\|_2$ to end up with (7.1.1).

We conclude for causal S that

$$\|S\| = \sup_{u \in U_e \cap L_2^k,\ \|u\|_2 > 0} \frac{\|S(u)\|_2}{\|u\|_2} = \inf\{\gamma \in \mathbb{R} \mid \forall\, u \in U_e \cap L_2^k :\ \|S(u)\|_2 \leq \gamma \|u\|_2\}.$$

Hence the L_2-gain of S is the worst amplification of the system if measuring the size of the input- and output-signals by their energy.

Example 7.3 Consider the linear time-invariant system

$$\dot{x} = Ax + Bu,\quad y = Cx + Du,\quad x(0) = 0 \tag{7.1.3}$$

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times k}$, $C \in \mathbb{R}^{l\times n}$. A standard fact in the theory of differential equations reveals that any $u \in L_{2e}^k$ leads to a unique response $y \in L_{2e}^l$. It is not difficult to prove that the output has finite energy for all $u \in L_2^k$ if and only if the corresponding transfer matrix $C(sI-A)^{-1}B$ has all its poles in the open left-half plane. If (A, B) is stabilizable and (A, C) is detectable, this is equivalent to A begin Hurwitz. If (7.1.3) map L_2^k into L_2^l, it has finite L_2-gain which is know to equal the H_∞-norm of the transfer matrix $C(sI-A)^{-1}B+D$.

With the continuous function $\phi : \mathbb{R} \to \mathbb{R}$ let us define the static nonlinear system

$$S(u)(t) := \phi(u(t)) \;\text{ for }\; t \geq 0. \tag{7.1.4}$$

For S mapping L_{2e} into L_{2e}, one either requires additional properties on ϕ or one has to restrict the domain of definition. If U_e denotes all continuous signals in L_{2e}, we conclude that S maps U_e into $U_e \subset L_{2e}$, simply due to the fact that $t \to \phi(u(t))$ is continuous for $u \in U_e$. However, S is not guaranteed to have finite L_2-gain. Just think of $\phi(x) = \sqrt{x}$ and the signal $u(t) = 1/(t+1)$, $t \geq 0$. In general, if ϕ satisfies

$$|\phi(x)| \leq \alpha |x| \;\text{ for all }\; x \in \mathbb{R},$$

the reader is invited to show that $S : L_{2e} \to L_{2e}$ is causal and has finite L_2-gain which is bounded by α. ∎

For nonlinear systems it is often more relevant to compare the distance of two different input signals u_1, u_2 with the distance of the corresponding output signals. In engineering terms this amounts to investigating the effect of input variations onto the system's output, whereas in mathematically terms it is related to continuity properties. From an abundance of possible definitions, the most often used version compares the energy of the input increment $u_1 - u_2$ with the energy of the output increment $S(u_1) - S(u_2)$ which leads to the notion of incremental L_2-gain.

Definition 7.4 The **incremental L_2-gain** of the system $S : U_e \to L_{2e}^l$ is defined as

$$\begin{aligned} \|S\|_i &:= \sup\left\{ \frac{\|P_T S(u_1) - P_T S(u_2)\|_2}{\|P_T u_1 - P_T u_2\|_2} \mid u_1, u_2 \in U_e,\ T \geq 0,\ \|P_T u_1 - P_T u_2\|_2 \neq 0 \right\} = \\ &= \inf\{\gamma \in \mathbb{R} \mid \forall\, u_1, u_2 \in U_e, T \geq 0 :\ \|P_T S(u_1) - P_T S(u_2)\|_2 \leq \gamma \|P_T u_1 - P_T u_2\|_2\}. \end{aligned}$$

If $\|S\|_i < \infty$, S is said to have finite incremental L_2-gain.

Similarly as for the L_2-gain, the system S has finite incremental L_2-gain if there exists a real $\gamma > 0$ such that

$$\|P_T S(u_1) - P_T S(u_2)\|_2 \leq \gamma \|P_T u_1 - P_T u_2\|_2 \;\text{ for all }\; T \geq 0,\ u_1, u_2 \in U_e. \tag{7.1.5}$$

This reveals

$$\|S(u_1) - S(u_2)\|_2 \leq \gamma \|u_1 - u_2\|_2 \;\text{ for all }\; u_1, u_2 \in U_e \cap L_2^k. \tag{7.1.6}$$

If S is causal, it can again be easily seen that (7.1.6) implies (7.1.5), and that

$$\|S\|_i = \sup_{u_1,u_2\in U_e\cap L_2^k,\ \|u_1-u_2\|_2>0} \frac{\|S(u_1)-S(u_2)\|_2}{\|u_1-u_2\|_2} =$$
$$= \inf\{\gamma\in\mathbb{R} \mid \forall\, u_1,u_2\in U_e\cap L_2^k:\ \|S(u_1)-S(u_2)\|_2 \le \gamma\|u_1-u_2\|_2\}.$$

Example 7.5 If the function ϕ in Example 7.3 satisfies the Lipschitz condition

$$|\phi(x_1)-\phi(x_2)| \le \alpha|x_1-x_2| \ \text{ for all }\ x_1,x_2\in\mathbb{R},$$

the mapping defined by (7.1.4) has finite incremental L_2-gain bounded by α. Note that one cannot guarantee that it has finite L_2-gain, as seen with $\phi(x)=1+|x|$ and the signal $u(t)=0$ which maps into $\phi(u(t))=1$ not contained in L_2. ∎

It just follows from the definitions that

$$\|S\| \le \|S\|_i \ \text{ if }\ S(0)=0.$$

Hence systems mapping zero into zero do have finite L_2-gain if they have finite incremental L_2-gain. The converse is, in general, not true and the reader is asked to construct a suitable counterexample. It is obvious that

$$\|S\| = \|S\|_i \ \text{ if } S \text{ is linear.}$$

Hence linear systems have finite L_2-gain iff they have finite incremental L_2-gain.

For reasons of clarity we have opted to confine ourselves to a few concepts in the pretty specific L_2-setting which sets the stage for various modifications or extensions that have been suggested in the literature. Let us hint at a few of these generalizations in various respects. All notions are as easily introduced for L_p- and L_{pe}-spaces with $1\le p\le\infty$. The time-axis can be chosen as all nonnegative integers to investigate discrete-time systems. A mixture of continuous- and discrete-time allows to consider hybrid systems or systems with jumps. Moreover, the value set of the signals can be taken to be an arbitrary normed space, thus including infinite dimensional systems. Finally, the presented stability concepts are important samples out of a multitude of other possibilities. For example, stability of S if often defined to just require S to map $U_e\cap L_2$ into L_2 without necessarily having finite L_2-gain. On the other hand in various applications it is important to qualify in a more refined fashion how $\|P_T S(u)\|_2$ is related to $\|P_T u\|_2$ for varying T. For example, one could work with $\|P_T S(u)\|_2 \le \gamma(\|P_T u\|_2)$ to hold for $T\ge 0$, $u\in U_e$, and for some function $\gamma:[0,\infty)\to[0,\infty)$ in a specific class. We have confined ourselves to *linear* functions γ, but one could as well take the class of *affine* or *monotonic* functions, or classes of even more specific type with desired tangency and growth properties for $T\to 0$ and $T\to\infty$ respectively, some of which can be found e.g. in [**?**, **?**]. Many of the abstract results to be presented could be formulated with general designer chosen stability properties that only need to obey certain (technical) axiomatic hypotheses.

7.2 Robust Input-Output Stability

In general, robustness analysis is nothing but an investigation of the sensitivity of a interesting system property against (possibly large) perturbations that are known to belong to an a priori specified class which determines their structure and size. In general one can argue that robustness questions form center stage in most natural sciences, in engineering and in mathematics. In control one is typically interested in how a desired property of a system interconnection, such as e.g. stability or performance, is affected by perturbations in the system's components that are caused e.g. by modeling inaccuracies or by failures. The purpose of this section is to develop tools for guaranteeing the stability of an interconnection of linear time-invariant systems against rather general classes of system perturbations which are also called uncertainties.

7.2.1 A Specific Feedback Interconnection

As we have clarified in Section **??** robust stability and performance questions can be handled on the basis of the setup in Figure 7.1 in quite some generality. Here, M is a fixed nominal model and Δ describes an uncertainty which is allowed to vary in a certain class $\boldsymbol{\Delta}$. Both the nominal system and the uncertainty are interconnected via feedback as suggested in the diagram, and as analytically expressed by the relations

$$\begin{pmatrix} w \\ M(w) \end{pmatrix} - \begin{pmatrix} \Delta(z) \\ z \end{pmatrix} = \begin{pmatrix} w_{\text{in}} \\ z_{\text{in}} \end{pmatrix}. \tag{7.2.1}$$

The signals w_{in} and z_{in} are external inputs or disturbances and w, z constitute the corresponding response of the interconnection.

The two most basic properties of such feedback interconnections are well-posedness and stability. Well-posedness means that for all external signals w_{in} and z_{in} there exists a unique response w and z which satisfies (7.2.1). We can conclude that the relations (7.2.1) actually define a mapping $(w_{\text{in}}, z_{\text{in}}) \to (w, z)$. Stability is then just related to whether this mapping has finite gain or not.

After these intuitive explanations we intend to introduce the precise mathematical framework chosen for our exposition. We start by fixing the hypothesis on the component systems.

Assumption 7.6 For each $\Delta \in \boldsymbol{\Delta}$, the mappings

$$M : W_e \to L_{2e}^l \ \text{ and } \ \Delta : Z_e \to L_{2e}^k$$

with $W_e \subset L_{2e}^k$ and $Z_e \subset L_{2e}^l$ are causal and of finite L_2-gain. Moreover, $\boldsymbol{\Delta}$ is star-shaped with star center 0:

$$\Delta \in \boldsymbol{\Delta} \ \text{ implies } \ \tau\Delta \in \boldsymbol{\Delta} \ \text{ for all } \ \tau \in [0, 1].$$

■

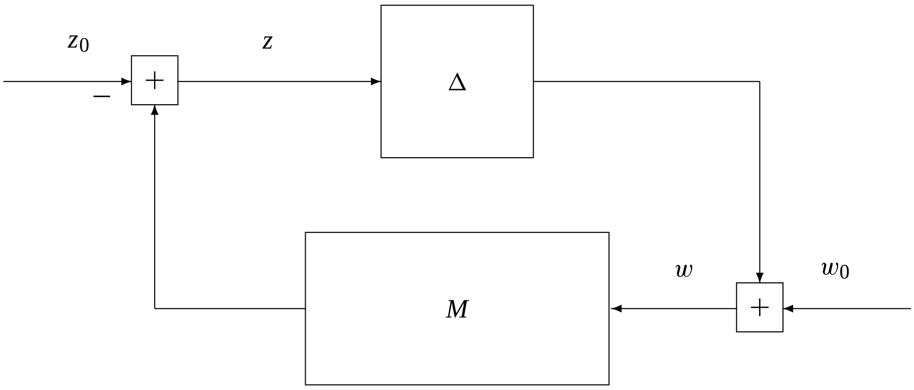


Figure 7.1: Uncertainty Feedback Configuration

For each $\Delta \in \mathbf{\Delta}$ we note $\{\tau\Delta \mid \tau \in [0,1]\}$ just defines a line segment in the set of all causal and stable systems connecting 0 with Δ. This justifies the terminology that $\mathbf{\Delta}$ is star-shaped with center 0 since any $\Delta \in \mathbf{\Delta}$ can be connected to 0 with such a line-segment that is fully contained in $\mathbf{\Delta}$. We infer that $0 \in \mathbf{\Delta}$ which is consistent with viewing Δ as an uncertainty where $\Delta = 0$ is related to the unperturbed or nominal interconnection described by

$$\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} w_{\text{in}} \\ M(w_{\text{in}}) - z_{\text{in}} \end{pmatrix}. \tag{7.2.2}$$

Just for convenience of notations we define the interconnection mapping

$$\mathcal{I}_M(\Delta) \begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} I & -\Delta \\ M & -I \end{pmatrix} \begin{pmatrix} w \\ z \end{pmatrix} := \begin{pmatrix} w \\ M(w) \end{pmatrix} - \begin{pmatrix} \Delta(z) \\ z \end{pmatrix} \tag{7.2.3}$$

Note that $\mathcal{I}_M(\Delta)$ depends both on the component systems M, Δ as well as on the specific interconnection under investigation. The asymmetric notation should stress that M is fixed whereas Δ is allowed to vary in the class $\mathbf{\Delta}$.

Now we are read to define well-posedness.

Definition 7.7 The mapping $\mathcal{I}_M(\Delta)$ is **well-posed** with respect to the external signal set $W_{\text{in,e}} \times Z_{\text{in,e}} \subset L_{2e}^k \times L_{2e}^l$ if for each $(w_{\text{in}}, z_{\text{in}}) \in W_{\text{in,e}} \times Z_{\text{in,e}}$ there exists a unique $(w, z) \in W_e \times Z_e$ satisfying

$$\mathcal{I}_M(\Delta) \begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} w_{\text{in}} \\ z_{\text{in}} \end{pmatrix}$$

such that the correspondingly defined map

$$W_{\text{in,e}} \times Z_{\text{in,e}} \ni \begin{pmatrix} w_{\text{in}} \\ z_{\text{in}} \end{pmatrix} \to \begin{pmatrix} w \\ z \end{pmatrix} \in W_e \times Z_e \tag{7.2.4}$$

is causal.

Mathematically, well-posedness hence just means that

$$\mathcal{I}_M(\Delta) : W_{\text{res,e}} \times Z_{\text{res,e}} \to W_{\text{in,e}} \times Z_{\text{in,e}} \text{ has a causal inverse } \mathcal{I}_M(\Delta)^{-1}$$

with some set $W_{\text{res,e}} \times Z_{\text{res,e}} \subset W_e \times Z_e$. Typically, $W_e \times Z_e$ and $W_{\text{in,e}} \times Z_{\text{in,e}}$ are explicitly given user-specified subsets of L_{2e} defined through additional regularity or boundedness properties on the corresponding signals. It is then often pretty simple to rely on standard existence and uniqueness results (for differential equations or integral equations) to guarantee well-posedness as illustrated in examples below. In contrast it is not required to explicitly describe the (generally Δ-dependent) set $W_{\text{res,e}} \times Z_{\text{res,e}}$ which collects all responses of the interconnection to disturbances in $W_{\text{in,e}} \times Z_{\text{in,e}}$.

For well-posedness it is clearly required to choose an input signal set with $P_T(W_{\text{in,e}} \times Z_{\text{in,e}}) \subset W_{\text{in,e}} \times Z_{\text{in,e}}$ for all $T \geq 0$. In addition the unperturbed interconnection mapping $I_M(0)$ is required to be well-posed. In view (7.2.2) this is guaranteed if the input signal sets satisfy $W_{\text{in,e}} \subset W_e$ and $M(W_{\text{in,e}}) - Z_{\text{in,e}} \subset Z_e$. This brings us to the following standing hypothesis on the interconnection, including some abbreviations in order to simplify the notation.

Assumption 7.8 Well-posedness of $I_M(\Delta)$ is always understood with respect to the disturbance input set $W_{\text{in,e}} \times Z_{\text{in,e}}$ satisfying $P_T(W_{\text{in,e}} \times Z_{\text{in,e}}) \subset W_{\text{in,e}} \times Z_{\text{in,e}}$ for all $T \geq 0$ and

$$W_{\text{in,e}} \subset W_e \text{ and } M(W_{\text{in,e}}) - Z_{\text{in,e}} \subset Z_e.$$

In addition the following standard abbreviations are employed:

$$W \times Z := (W_e \times Z_e) \cap L_2^{k+l} \text{ and } W_{\text{in}} \times Z_{\text{in}} = (W_{\text{in,e}} \times Z_{\text{in,e}}) \cap L_2^{k+l}.$$

■

Example 7.9 Consider $M : L_{2e}^k \to L_{2e}^l$ defined by

$$\dot{x} = Ax + Bw, \quad z = Cx$$

with zero initial condition and A Hurwitz. Let $\mathbf{\Delta}$ denote the class of all mappings

$$\Delta(z)(t) := \phi(z(t)) \text{ for } t \geq 0$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and satisfies $|\phi(x)| \leq |x|$ for all $x \in \mathbb{R}$. It is obvious that $|\phi(x)| \leq |x|$ for all $x \in \mathbb{R}$ implies $|\tau\phi(x)| \leq |x|$ for all $x \in \mathbb{R}$ and $\tau \in [0, 1]$. This shows that $\mathbf{\Delta}$ is star-shaped with star-center zero, and the remaining facts in Assumption 7.6 follow from Example 7.3.

The interconnection is described by

$$\dot{x} = Ax + Bw, \quad w_{\text{in}} = w - \phi(z), \quad z_{\text{in}} = Cx - z \tag{7.2.5}$$

and defines the mapping $\mathcal{I}_M(\Delta)$. To investigate well-posedness it is convenient to rewrite these relations into the non-linear initial value problem

$$\dot{x}(t) = Ax(t) + B\phi(Cx(t) - z_{\text{in}}(t)) + Bw_{\text{in}}(t), \; x(0) = 0 \tag{7.2.6}$$

with the interconnection response as output:

$$w(t) = \phi(Cx(t) - z_{\rm in}(t)) + w_{\rm in}(t), \quad z(t) = Cx(t) - z_{\rm in}(t).$$

Let us restrict the external disturbance input signals $(w_{\rm in}, z_{\rm in})$ to those which are piece-wise right-continuous thus defining the subspace $W_{\rm in,e} \times Z_{\rm in,e}$ of L_{2e}^{k+l}. Then standard elementary results in differential equation theory guarantee the existence of a unique continuously differentiable solution of the initial value problem (7.2.6) for each $(w_{\rm in}, z_{\rm in}) \in W_{\rm in,e} \times Z_{\rm in,e}$, and it is clear that $(w_{\rm in}, z_{\rm in}) \to (w, z)$ is causal. Through introducing convenient regularity properties on the input signals it turns out to be simple to establish well-posedness. Of course one can allow for more irregular disturbance inputs by relying on the more advanced theory of Carathéodory for results on differential equations. Our example illustrates that this more technical framework can be avoided without loosing practical relevance. ∎

In addition to well-posedness, it is of interest to investigate stability of the feedback interconnection. Among the various possible choices, we confine ourselves to $\mathcal{I}_M(\Delta)^{-1}$ having finite L_2-gain or finite incremental L_2-gain. If these properties hold for *all* $\Delta \in \mathbf{\Delta}$, the interconnection will be called robustly stable. Uniform robust stability then just means that the (incremental) L_2-gain of $\mathcal{I}_M(\Delta)^{-1}$ is uniformly bounded in $\Delta \in \mathbf{\Delta}$, i.e., the bound does not depend on $\Delta \in \mathbf{\Delta}$.

In this book we provide sufficient conditions for the following precise stability properties:

- Under the hypothesis that $\mathcal{I}_M(\Delta)$ is well-posed, guarantee that there exists a constant c with $\|\mathcal{I}_M(\Delta)^{-1}\| \leq c$ for all $\Delta \in \mathbf{\Delta}$. In particular, $\mathcal{I}_M(\Delta)^{-1}$ then has finite L_2-gain for all $\Delta \in \mathbf{\Delta}$.
- If M and all systems in $\mathbf{\Delta}$ have finite incremental L_2-gain, characterize that $\mathcal{I}_M(\Delta)$ is well-posed and that there exists a c with $\|\mathcal{I}_M(\Delta)^{-1}\|_i \leq c$ for all $\Delta \in \mathbf{\Delta}$. In particular, any $\mathcal{I}_M(\Delta)^{-1}$ has a finite incremental L_2-gain.

It is important to observe the fundamental difference in these two characterizations: In the first scenario one has to *assume* well-posedness and one *concludes* stability, whereas in the second *both* well-posedness and stability are *concluded*.

For systems described by differential stability is often related to the state converging to zero for $t \to \infty$. In continuing Example 7.9 let us now illustrate how to guarantee this attractiveness of the solutions of (nonlinear) differential by input-output stability.

Example 7.10 Let us consider the uncertain non-linear system

$$\dot{x} = Ax + B\phi(Cx), \quad x(t_0) = x_0 \tag{7.2.7}$$

with $t_0 > 0$, with A being Hurwitz and with the nonlinearity being continuously differentiable and satisfying $|\phi(x)| \leq |x|$. We are interested in the question of whether all trajectory of (7.2.7) converge to zero for $t \to \infty$. This holds true if $\mathcal{I}_M(\Delta)^{-1}$ as defined in Example 7.9 has a finite L_2-gain. Why? Since A is stable, it suffices to prove the desired properties for x_0 in the controllable subspace of (A, B). Given any such x_0, we can find $w_{\rm in}(.)$ on $[0, t_0)$ which steers the state-trajectory of $\dot{x} = Ax + Bw_{\rm in}$ from $x(0) = 0$ to $x(t_0) = x_0$. Define $z_{\rm in}(t) := Cx(t)$ on $[0, t_0)$, and concatenate these external inputs with $w_{\rm in}(t) = 0$, $z_{\rm in}(t) = 0$ on $[t_0, \infty)$, implying $(w_{\rm in}, z_{\rm in}) \in L_2^{k+l}$. At this point we exploit stability to infer that $(w, z) = \mathcal{I}_M(\Delta)^{-1}(w_{\rm in}, z_{\rm in})$ is also contained in L_2^{k+l}. Therefore (7.2.5) reveal that both x and $\dot{x}$ are in L_2 which implies $\lim_{t\to\infty} x(t) = 0$. It is obvious by the concatenation construction that the responses of (7.2.6) and (7.2.7) coincide on $[t_0, \infty)$ which proves the desired stability property.

In this construction it is clear that the energy of $w_{\rm in}$ and hence also of $z_{\rm in}$ can be bounded in terms of $\|x_0\|$. If $\mathcal{I}_M(\Delta)^{-1}(w_{\rm in}, z_{\rm in})$ has finite L_2-gain, we infer that the energy of the responses of w and z are also bounded in terms of $\|x_0\|$. Due to $\dot{x} = Ax + Bw$, the same is true for $\|x\|_2$ and even for $\|x\|_\infty$. These arguments reveal how to deduce Lyapunov stability. Uninformity with respect to the uncertainty is guaranteed by a uniform bound on the L_2-gain of $\mathcal{I}_M(\Delta)^{-1}(w_{\rm in}, z_{\rm in})$. ∎

Before we embark on a careful exposition of the corresponding two fundamental theorems which are meant to summarize a variety of diverse results in the literature, we include a section with a preparatory discussion meant to clarify the main ideas and to technically simplify the proofs.

7.2.2 An Elementary Auxiliary Result

The well-posedness and stability results to be derived are based on separation. The main ideas are easily demonstrated in terms of mappings M and Δ defined through matrix multiplication. Well-posedness is then related to whether

$$\begin{pmatrix} I & -\Delta \\ M & -I \end{pmatrix} \text{ is non-singular} \tag{7.2.8}$$

and stability translates into the existence of a bound c for the inverse's spectral norm

$$\left\| \begin{pmatrix} I & -\Delta \\ M & -I \end{pmatrix}^{-1} \right\| \leq c. \tag{7.2.9}$$

Let us define the graph subspaces

$$U = \operatorname{im} \begin{pmatrix} I \\ M \end{pmatrix} \text{ and } V = \operatorname{im} \begin{pmatrix} \Delta \\ I \end{pmatrix}$$

of the mappings M and Δ. It is obvious that (7.2.8) equivalently translates into the property that U and V only intersect in zero:

$$U \cap V = \{0\}. \tag{7.2.10}$$

Let us now prove that (7.2.9) can be similarly characterized by U and V actually keeping nonzero distance. Indeed

$$\left\| \begin{pmatrix} I & -\Delta \\ M & -I \end{pmatrix}^{-1} \begin{pmatrix} w_{\rm in} \\ w_{\rm in} \end{pmatrix} \right\| \leq c \left\| \begin{pmatrix} w_{\rm in} \\ z_{\rm in} \end{pmatrix} \right\| \text{ for all } (w_{\rm in}, z_{\rm in})$$

is equivalent to

$$\left\| \begin{pmatrix} w \\ z \end{pmatrix} \right\|^2 \leq c^2 \left\| \begin{pmatrix} I & -\Delta \\ M & -I \end{pmatrix} \begin{pmatrix} w \\ z \end{pmatrix} \right\|^2 = \text{ for all } (w, z). \tag{7.2.11}$$

If we exploit the evident inequality

$$\left\| \begin{pmatrix} w \\ z \end{pmatrix} \right\|^2 \leq \left\| \begin{pmatrix} I \\ M \end{pmatrix} w \right\|^2 + \left\| \begin{pmatrix} \Delta \\ I \end{pmatrix} z \right\|^2 \leq \underbrace{\max\{1 + \|M\|^2, 1 + \|\Delta\|^2\}}_{\gamma^2} \left\| \begin{pmatrix} w \\ z \end{pmatrix} \right\|^2$$

we observe that (7.2.11) *implies*

$$\left\| \begin{pmatrix} I \\ M \end{pmatrix} w \right\|^2 + \left\| \begin{pmatrix} \Delta \\ I \end{pmatrix} z \right\|^2 \leq (c\gamma)^2 \left\| \begin{pmatrix} I \\ M \end{pmatrix} w - \begin{pmatrix} \Delta \\ I \end{pmatrix} z \right\|^2 \text{ for all } (w, z)$$

and that (7.2.11) *is implied by*

$$\left\| \begin{pmatrix} I \\ M \end{pmatrix} w \right\|^2 + \left\| \begin{pmatrix} \Delta \\ I \end{pmatrix} z \right\|^2 \leq c^2 \left\| \begin{pmatrix} I \\ M \end{pmatrix} w - \begin{pmatrix} \Delta \\ I \end{pmatrix} z \right\|^2 \text{ for all } (w, z).$$

In terms of the subspaces U and V we conclude that (7.2.9) is guaranteed if

$$\|u\|^2 + \|v\|^2 \leq c^2 \|u - v\|^2 \text{ for all } u \in U, \; v \in V. \tag{7.2.12}$$

(Conversely, (7.2.9) implies (7.2.12) for c replaced by $c\gamma$ but this will be not exploited in the sequel.) Clearly (7.2.12) strengthens (7.2.10) in that it requires all vectors $u \in U$ and $v \in V$ of length 1 to keep at least the distance $\sqrt{2}/c$. Readers familiar with the standard gap-metric for subspaces will recognize that this just comes down to requiring U and V having positive distance.

It is not exaggerated to state that the verifiable characterization of whether two sets are separated is at the heart of huge variety of mathematical problems. For example in convex analysis sets are separated by trying to locate them into opposite half-spaces. This can be alternatively expressed as trying to located either set into the zero sub-level and sup-level set of an affine function. As the reader can immediately verify, it is impossible to separate subspaces in this fashion. This motivates to try separation in exactly the same vein by *quadratic* functions instead. For the Euclidean subspaces U and V quadratic separation then means to find a symmetric P such that $u^T P u \leq 0$ and $v^T P v \geq 0$ for all $u \in U$, $v \in V$. Strict quadratic separation is defined as the existence of $\epsilon > 0$ for which

$$u^T P u \leq -\epsilon \|u\|^2 \text{ and } v^T P v \geq 0 \text{ for all } u \in U, \; v \in V. \tag{7.2.13}$$

It is the main purpose of this section to prove for a considerably more general situation that (7.2.13) does allow to compute a positive c for which (7.2.12) is satisfied. In the exercises the reader is asked to prove the converse for the specific matrix problem discussed so far.

Let us turn to an abstraction. We assume U and V to be subsets (not necessarily subspaces) of a normed linear space X. Let us first fix what we mean by referring to a quadratic mapping.

Definition 7.11 $\Sigma : X \to \mathbb{R}$ is said to be **quadratic** if

$$\Sigma(x) = \langle x, x \rangle$$

holds for some mapping $\langle ., . \rangle : X \times X \to \mathbb{R}$ that is additive in both arguments and bounded. $\langle ., . \rangle$ is additive if

$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle, \quad \langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$$

and it is bounded if there exists a $\sigma > 0$ with

$$|\langle x, y \rangle| \leq \sigma \|x\| \|y\|$$

for all $x, y, z \in X$.

After all these preparations we now turn to the very simple quadratic separation result which will form the basis for proof of our main results in the next two sections.

Lemma 7.12 *Suppose that $\Sigma : X \to \mathbb{R}$ is quadratic, and that there exists an $\epsilon > 0$ for which*

$$\Sigma(u) \leq -\epsilon \|u\|^2 \text{ and } \Sigma(v) \geq 0 \text{ for all } u \in U,\ v \in V. \tag{7.2.14}$$

Then there exists a positive real number c (which only depends on ϵ and Σ) such that

$$\|u\|^2 + \|v\|^2 \leq c\|u - v\|^2 \text{ for all } u \in U,\ v \in V. \tag{7.2.15}$$

Proof. Let us choose α with $0 < \alpha < \epsilon$. We have

$$\Sigma(x + u) - \Sigma(u) - \alpha\|u\|^2 = \langle x, x \rangle + \langle x, u \rangle + \langle u, x \rangle - \alpha\|u\|^2 \leq$$
$$\leq \sigma\|x\|^2 + 2\sigma\|x\|\|u\| - \alpha\|x\|^2 \leq \delta\|x\|^2$$

where $\delta > 0$ is chosen as $\max\{\sigma + 2\sigma t - \alpha t^2 :\ t \in \mathbb{R}\}$ such that the last inequality follows from $\sigma + 2\sigma\frac{\|u\|}{\|x\|} - \alpha\frac{\|u\|^2}{\|x\|^2} \leq \delta$ for $\|x\| \neq 0$. We can conclude

$$\Sigma(v) - \Sigma(u) - \alpha\|u\|^2 \leq \delta\|u - v\|^2.$$

We now exploit (7.2.14) to infer

$$(\epsilon - \alpha)\|u\|^2 \leq \delta\|u - v\|^2.$$

Recall that $\epsilon - \alpha > 0$ which motivates the choice of α. If we observe that $\|v\|^2 = \|v - u + u\|^2 \leq 2\|u - v\|^2 + 2\|u\|^2$ we end up with

$$\|u\|^2 + \|v\|^2 \leq 3\|u\|^2 + 2\|u - v\|^2 \leq \underbrace{\left(\frac{3\delta}{\epsilon - \alpha} + 2\right)}_{c}\|u - v\|^2.$$

■

Remark. Note that it is not really essential that Σ is quadratic. In fact, we only exploited that

$$\frac{\Sigma(v) - \Sigma(u) - \alpha\|u\|^2}{\|v-u\|^2} \text{ is bounded on } \{(u,v) : \ u \in U, \ v \in V, \ u \neq v\}$$

for some $0 < \alpha < \epsilon$. The concept of quadratic continuity [**?**] (meaning that boundedness is guaranteed for *all* $\alpha > 0$) has been introduced just for this purpose. Our preference for the more specific version is motivated by the robust stability tests that are amenable to computational techniques as discussed in later sections. Far more general results about topological separation to guarantee robust stability are discussed in the literature, and the interested reader is referred to [**?**, **?**, **?**]

7.2.3 An Abstract Stability Characterization

Let us now continue the discussion in Section 7.2.1 with our first fundamental result on guaranteeing stability under the assumption of well-posedness. More concretely we assume that $\mathcal{I}_M(\Delta)^{-1}$ defined on $W_{\mathrm{in,e}} \times Z_{\mathrm{in,e}}$ exists and is causal for all $\Delta \in \mathbf{\Delta}$. The main point is to formulate sufficient conditions for this inverse to admit a uniform bound on its L_2-gain. In view of our preparations it is not difficult to provide an abstract criterion on the basis of Lemma 7.12.

Theorem 7.13 *Suppose that for all $\Delta \in \mathbf{\Delta}$ the map $\mathcal{I}_M(\Delta)$ is well-posed, and $(w,z) = \mathcal{I}_M(\Delta)^{-1}(w_{\mathrm{in}}, z_{\mathrm{in}})$ implies $w_{\mathrm{in}} + \delta\Delta(z) \in W_{\mathrm{in,e}}$ for all $\delta \in [0,1]$. Let $\Sigma : L_2^{k+l} \to \mathbb{R}$ be quadratic, and suppose*

$$\Sigma\begin{pmatrix} \Delta(z) \\ z \end{pmatrix} \geq 0 \ \text{ for all } \ z \in Z, \quad \Delta \in \mathbf{\Delta}. \tag{7.2.16}$$

If there exists an $\epsilon > 0$ with

$$\Sigma\begin{pmatrix} w \\ M(w) \end{pmatrix} \leq -\epsilon\|w\|_2^2 \ \text{ for all } \ w \in W, \tag{7.2.17}$$

there exists a constant c such that

$$\|\mathcal{I}_M(\Delta)^{-1}\| \leq c \ \text{ for all } \ \Delta \in \mathbf{\Delta}.$$

Proof. Fix an arbitrary $\Delta \in \mathbf{\Delta}$ and recall that $\tau\Delta \in \mathbf{\Delta}$ for all $\tau \in [0,1]$. Our first goal is to show that there exists $c > 0$ such that

$$\|x\|_2 \leq c\|\mathcal{I}_M(\tau\Delta)(x)\|_2 \ \text{ for all } \ x \in W \times Z, \quad \tau \in [0,1]. \tag{7.2.18}$$

The construction will reveal that c does neither depend on τ nor on Δ. Property (7.2.18) is proved by applying Lemma 7.12 for the sets

$$U := \left\{\begin{pmatrix} w \\ M(w) \end{pmatrix} : \ w \in W\right\} \ \text{ and } \ V := \left\{\begin{pmatrix} \tau\Delta(z) \\ z \end{pmatrix} : \ z \in Z\right\}.$$

Due to (7.2.16) and $\tau\Delta \in \mathbf{\Delta}$ we certainly have $\Sigma(v) \geq 0$ for all $v \in V$. Moreover, for all $u \in U$ we infer with $u = (w, M(w))$ that

$$\Sigma(u) = \Sigma\begin{pmatrix} w \\ M(w) \end{pmatrix} \leq -\epsilon\|w\|^2 \leq -\frac{\epsilon}{1+\|M\|_2^2}\left\|\begin{pmatrix} w \\ M(w) \end{pmatrix}\right\|_2 = -\frac{\epsilon}{1+\|M\|_2^2}\|u\|^2$$

due to $\|(w, M(w))\|_2^2 \leq (1+\|M\|_2^2)\|w\|_2^2$. By Lemma 7.12 there exists $c > 0$ with

$$\left\|\begin{pmatrix} w \\ M(w) \end{pmatrix}\right\|_2^2 + \left\|\begin{pmatrix} \tau\Delta(z) \\ z \end{pmatrix}\right\|_2^2 \leq c^2\left\|\begin{pmatrix} w \\ M(w) \end{pmatrix} - \begin{pmatrix} \tau\Delta(z) \\ z \end{pmatrix}\right\|_2^2$$

for all $w \in W$ and $z \in Z$. Since the left-hand side bounds $\|(w, z)\|_2^2$ from above, we infer $\|x\|_2 \leq c\|\mathcal{I}_M(\tau\Delta)(x)\|_2$ for all $x \in W \times Z$. By construction the constant c only depends on $\epsilon > 0$, $\|M\|$, Σ which guarantees (7.2.18) uniformly in $\Delta \in \mathbf{\Delta}$.

If we knew that $\mathcal{I}_M(\tau\Delta)^{-1}$ mapped $W_{\text{in}} \times Z_{\text{in}}$ into $W \times Z$ (and not just into $W_e \times Z_e$ as guaranteed by hypothesis), the proof would be finished since then (7.2.18) just implies that the gain of $\mathcal{I}_M(\tau\Delta)^{-1}$ is bounded by c for all $\tau \in [0, 1]$, and hence also for $\tau = 1$. To finish the proof it thus suffices to show that

$$\mathcal{I}_M(\tau\Delta)^{-1}(W_{\text{in}} \times Z_{\text{in}}) \subset W \times Z \tag{7.2.19}$$

for all $\tau \in [0, 1]$. Let us assume that this is true for any $\tau_0 \in [0, 1]$. For $\tau > \tau_0$ take $y = (w_{\text{in}}, z_{\text{in}}) \in W_{\text{in}} \times Z_{\text{in}}$ and set $x_\tau = \mathcal{I}_M(\tau\Delta)^{-1}(y) \in W_e \times Z_e$. With $x_\tau = (w_\tau, z_\tau)$ we have

$$\mathcal{I}_M(\tau_0\Delta)(x_\tau) = [\mathcal{I}_M(\tau_0\Delta)(x_\tau) - \mathcal{I}_M(\tau\Delta)(x_\tau)] + \mathcal{I}_M(\tau\Delta)(x_\tau) = \begin{pmatrix} (\tau-\tau_0)\Delta(z_\tau) \\ 0 \end{pmatrix} + \begin{pmatrix} w_{\text{in}} \\ z_{\text{in}} \end{pmatrix}.$$

We now exploit the extra hypothesis (applied for $\tau\Delta(z_\tau)$ and $\delta = (\tau - \tau_0)/\tau \in [0, 1]$) to infer that $(\tau-\tau_0)\Delta(z_\tau) + w_{\text{in}} \in W_{\text{in,e}}$. Hence the right-hand side is contained in the domain of definition of $\mathcal{I}_M(\tau_0\Delta)^{-1}$ such that

$$x_\tau = \mathcal{I}_M(\tau_0\Delta)^{-1}\left[\begin{pmatrix} (\tau-\tau_0)\Delta(z_\tau) \\ 0 \end{pmatrix} + y\right].$$

Since the L_2-gain of $\mathcal{I}_M(\tau_0\Delta)^{-1}$ is bounded by c we obtain

$$\|P_T x_\tau\|_2 \leq c\|P_T(\tau-\tau_0)\Delta(z_\tau)\| + c\|P_T y\|_2 \leq c|\tau-\tau_0|\,\|\Delta\|\|P_T x_\tau\|_2 + c\|P_T y\|_2.$$

Together with $\|P_T y\|_2 \leq \|y\|_2$ we arrive at

$$(1 - c|\tau-\tau_0|\,\|\Delta\|)\ \|P_T x_\tau\|_2 \leq c\|y\|_2.$$

For all $\tau \in [\tau_0, 1]$ with $1 - c|\tau-\tau_0|\,\|\Delta\| < 1$, we can hence infer that $\|P_T x_\tau\|_2$ is bounded for all $T \geq 0$ which actually implies $x_\tau \in L_2^{k+l}$ and hence $x_\tau \in W \times Z$.

We have thus proved: If (7.2.19) holds for $\tau_0 \in [0, 1]$, the it also holds automatically for all $\tau \in [0, 1] \cap [\tau_0, \tau_0 + \delta_0]$ if $0 < \delta_0 < 1/(c\|\Delta\|)$. Starting with $\tau_0 = 0$, we conclude by induction that (7.2.19) holds for all $\tau \in [0, 1] \cap [0, k\delta_0]$ for $k = 0, 1, 2, \ldots$ and thus, with sufficiently large $k > 0$, indeed for all $\tau \in [0, 1]$. ■

Example 7.14 Let us continue Example 7.9. We have seen in Example 7.10 that $\boldsymbol{\Delta}$ is star-shaped with center zero. Well-posedness has been verified, where we observe that the interconnection response (w, z) is piece-wise right-continuous for $(w_{\text{in}}, z_{\text{in}}) \in W_{\text{in,e}} \times Z_{\text{in,e}}$. We infer $\Delta(z) \in W_{\text{in,e}}$ and thus $w_{\text{in}} + \delta\Delta(z) \in W_{\text{in,e}}$ for all $\delta \in [0, 1]$ just because $W_{\text{in,e}}$ is a linear space. Moreover, the Lipschitz condition $|\phi(z)| \leq |z|$ is equivalent to the sector condition

$$0 \leq z\phi(z) \leq z^2 \ \text{ for all } \ z \in \mathbb{R}$$

which can be written as

$$\begin{pmatrix} \phi(z) \\ z \end{pmatrix}^T \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}}_{P} \begin{pmatrix} \phi(z) \\ z \end{pmatrix} \geq 0 \ \text{ for all } \ z \in \mathbb{R}.$$

This motivates to define the quadratic function

$$\Sigma(w, z) := \int_0^\infty \begin{pmatrix} w(t) \\ z(t) \end{pmatrix}^T P \begin{pmatrix} w(t) \\ z(t) \end{pmatrix} dt$$

on L_2^2 to conclude that (7.2.16) is indeed satisfied. Now (7.2.17) simply amounts to requiring that M is strictly dissipative with respect to the quadratic supply rate defined with P, and this dissipativity guarantees that $\mathcal{I}_M(\Delta)^{-1}$ has uniformly bounded L_2-gain. If M is described by a linear time-invariant system its dissipativity with respect to P can be numerically verified by translation into a feasibility test for a suitably defined linear matrix inequality.

The example demonstrates how to actually apply Theorem 7.13 in concrete situations, and how dissipativity will result in computational LMI-tests for robust stability. It also illustrates how to verify the rather abstract hypotheses and that they often considerably simplify in concrete situations. We will explicitly formulate such specializations in later sections.

Remarks.

- Since we allow for general domains of definitions that are just subsets of L_2 or L_{2e}, we require the extra condition that $w_{\text{in}} + \delta\Delta(z) = (1-\delta)w_{\text{in}} + \delta w \in W_{\text{in,e}}$ for all $\delta \in [0, 1]$ and all interconnection responses $(w, z) = \mathcal{I}_M(\Delta)^{-1}(w_{\text{in}}, z_{\text{in}})$ with $(w_{\text{in}}, z_{\text{in}}) \in W_{\text{in}} \times Z_{\text{in}}$. This technical property is often simple to verify. For example it holds if W_e equals $W_{\text{in,e}}$ and both are convex, or if $\Delta(Z) \subset W_{\text{in,e}}$ and $W_{\text{in,e}}$ is a subspace of L_{2e}^k. In subsequent discussions M will be taken linear such that it is natural to assume that its domain of definition W_e is a subspace of L_{2e}^k and to take $W_{\text{in,e}} = W_e$.
- A careful analysis of the proof reveals that (7.2.16)-(7.2.17) has to be only guaranteed for all interconnection responses $(w, z) = \mathcal{I}_M(\Delta)^{-1}(w_{\text{in}}, z_{\text{in}})$ with $(w_{\text{in}}, z_{\text{in}}) \in W_{\text{in}} \times Z_{\text{in}}$. This insight might simplify the actual verification of the two quadratic inequalities.
- The proof of the theorem proceeds via a homotopy argument: The L_2-gain of $\mathcal{I}_M(0)^{-1}$ is finite. Then one proves that $\mathcal{I}_M(\Delta)^{-1}$ has finite L_2-gain by showing that the gain $\|\mathcal{I}_M(\tau\Delta)^{-1}\|_2$

is bounded by c which avoids that it blows up for τ moving from zero to one. It is hence not difficult to obtain similar stability results if replacing the line segment $[0, 1] \ni \tau \to \tau\Delta$ by any continuous curve $\gamma : [0, 1] \to \mathbf{\Delta}$ connecting 0 and Δ as $\gamma(0) = 0$, $\gamma(1) = \Delta$. Instead of being star-shaped, it suffices to require that $\mathbf{\Delta}$ contains 0 and is path-wise connected. In view of our detailed proofs the advanced reader should be in the position to generalize our arguments to more advanced stability question that are not covered directly by Theorem 7.13.

- One can easily avoid homotopy arguments and directly guarantee that $\mathcal{I}_M(\Delta)^{-1}$ has finite L_2-gain for a *fixed* $\Delta \in \mathbf{\Delta}$ with the quadratic inequalities

$$\Sigma\begin{pmatrix} P_T\Delta(z) \\ P_T z \end{pmatrix} \geq 0 \;\text{ for all }\; z \in Z_e,\; T \geq 0$$

and

$$\Sigma\begin{pmatrix} P_T w \\ P_T M(w) \end{pmatrix} \leq -\epsilon\|P_T w\|^2 \;\text{ for all }\; w \in W_e,\; T \geq 0$$

for some $\epsilon > 0$. It even suffices to ask these properties only for interconnection responses $(w, z) = \mathcal{I}_M(\Delta)^{-1}(w_{\text{in}}, z_{\text{in}})$ with $(w_{\text{in}}, z_{\text{in}}) \in W_{\text{in}} \times Z_{\text{in}}$. Due to Lemma 7.12 the corresponding proof is so simple that it can be left as an exercise.

7.2.4 A Characterization of Well-Posedness and Stability

The goal of this section is to get rid of the hypothesis that $\mathcal{I}_M(\Delta)$ has a causal inverse. The price to be paid is the requirement that all mappings have finite incremental L_2-gain, and to replace the crucial positivity and negativity conditions in (7.2.16) and (7.2.17) by their incremental versions. The main result will guarantee a uninform bound on the incremental L_2-gain of $\mathcal{I}_M(\Delta)^{-1}$. Technically, we will apply Banach's fixed point theorem for which we require completeness of $W \times Z$ as a subset of L_2^{k+l}.

Theorem 7.15 *Suppose there exists a $\delta_0 \in (0, 1]$ such that $W_{\text{in}} + \delta\Delta(Z) \subset W_{\text{in,e}}$ for all $\delta \in [0, \delta_0]$ and all $\Delta \in \mathbf{\Delta}$. Let all $\Delta \in \mathbf{\Delta}$ have finite incremental L_2-gain, and let $\Sigma : L_2^{k+l} \to \mathbb{R}$ be quadratic such that that*

$$\Sigma\begin{pmatrix} \Delta(z_1) - \Delta(z_2) \\ z_1 - z_2 \end{pmatrix} \geq 0 \;\text{ for all }\; z_1, z_2 \in Z. \tag{7.2.20}$$

Moreover, let M have finite incremental L_2-gain and let there exist an $\epsilon > 0$ such that

$$\Sigma\begin{pmatrix} w_1 - w_2 \\ M(w_1) - M(w_2) \end{pmatrix} \leq -\epsilon\|w_1 - w_2\|_2^2 \;\text{ for all }\; w_1, w_2 \in W. \tag{7.2.21}$$

If $W \times Z$ be a closed subset of L_2^{k+l} then $\mathcal{I}_M(\Delta)$ has a causal inverse and there exists a constant $c > 0$ with

$$\|\mathcal{I}_M(\Delta)^{-1}\|_{2i} \leq c \;\text{ for all }\; \Delta \in \mathbf{\Delta}.$$

Proof. Let us fix $\Delta \in \mathbf{\Delta}$.

First step. Similarly as in the proof of Theorem 7.15, the first step is to show the existence of some $c > 0$ with

$$\|x_1 - x_2\|_2^2 \leq c^2 \|\mathcal{I}_M(\tau\Delta)(x_1) - \mathcal{I}_M(\tau\Delta)(x_2)\|_2^2 \ \text{ for all } \ x_1, x_2 \in W \times Z, \ \tau \in [0,1] \tag{7.2.22}$$

where c is independent from $\Delta \in \mathbf{\Delta}$. For that purpose we now choose

$$U := \left\{ \begin{pmatrix} w_1 - w_2 \\ M(w_1) - M(w_2) \end{pmatrix} : \ w_1, w_2 \in W \right\}, \ V := \left\{ \begin{pmatrix} \tau\Delta(z_1) - \tau\Delta(z_2) \\ z_1 - z_2 \end{pmatrix} : \ z_1, z_2 \in Z \right\}.$$

For $v \in V$ we have $\Sigma(v) \geq 0$ by (7.2.20). For $u \in U$ we conclude with $u = (w_1 - w_2, M(w_1) - M(w_2))$ from (7.2.21) that

$$\Sigma(u) \leq -\epsilon\|w_1 - w_2\|^2 \leq -\frac{\epsilon}{1 + \|M\|_{2i}^2} \left\| \begin{pmatrix} w_1 - w_2 \\ M(w_1) - M(w_2) \end{pmatrix} \right\|_2 = -\frac{\epsilon}{1 + \|M\|_{2i}^2} \|u\|^2$$

due to $\|(w_1 - w_2, M(w_1) - M(w_2))\|_2^2 \leq (1 + \|M\|_i^2)\|w_1 - w_2\|_2^2$. Hence Lemma 7.12 implies the existence of $c > 0$ (only depending on ϵ, $\|M\|_i$, Σ) such that

$$\left\| \begin{pmatrix} w_1 - w_2 \\ M(w_1) - M(w_2) \end{pmatrix} \right\|_2^2 + \left\| \begin{pmatrix} \tau\Delta(z_1) - \tau\Delta(z_2) \\ z_1 - z_2 \end{pmatrix} \right\|_2^2 \leq$$
$$\leq c^2 \left\| \begin{pmatrix} w_1 - w_2 \\ M(w_1) - M(w_2) \end{pmatrix} - \begin{pmatrix} \tau\Delta(z_1) - \tau\Delta(z_2) \\ z_1 - z_2 \end{pmatrix} \right\|_2^2$$

for all $(w_j, z_j) \in W \times Z$. Since the left-hand side bounds $\|w_1 - w_2\|_2^2 + \|z_1 - z_2\|_2^2$ from above, we arrive at (7.2.22).

Second step. Suppose that, for $\tau_0 \in [0,1]$, $\mathcal{I}_M(\tau_0\Delta)$ has an inverse defined on $W_{\text{in}} \times Z_{\text{in}}$. Let us then prove that

$$\mathcal{I}_M(\tau\Delta) \ \text{ has an inverse defined on } W_{\text{in}} \times Z_{\text{in}} \text{ for all } \ \tau \in [0,1] \cap [\tau_0, \tau_0 + \tilde{\delta}_0] \tag{7.2.23}$$

with $0 < \tilde{\delta}_0 < \min\{\delta_0, (c\|\Delta\|_i)^{-1}\}$. Let us first observe that (7.2.22) together with the existence of the inverse imply

$$\|\mathcal{I}_M(\tau_0\Delta)^{-1}(y_1) - \mathcal{I}_M(\tau_0\Delta)^{-1}(y_2)\|_2^2 \leq c^2\|y_1 - y_2\|_2^2 \ \text{ for all } \ y_1, y_2 \in W_{\text{in}} \times Z_{\text{in}}. \tag{7.2.24}$$

To verify that $\mathcal{I}_M(\tau\Delta)$ has an inverse amounts to checking that for all $y \in W_{\text{in}} \times Z_{\text{in}}$ there is a unique $x \in W \times Z$ with $\mathcal{I}_M(\tau\Delta)(x) = y$. This relation can be rewritten to $\mathcal{I}_M(\tau_0\Delta)(x) = y - \mathcal{I}_M(\tau\Delta)(x) + \mathcal{I}_M(\tau_0\Delta)(x)$ or to the fixed-point equation

$$x = \mathcal{I}_M(\tau_0\Delta)^{-1}(\, y - \mathcal{I}_M(\tau\Delta)(x) + \mathcal{I}_M(\tau_0\Delta)(x)\,) =: F(x). \tag{7.2.25}$$

Hence we need to guarantee the existence of a unique $x \in W \times Z$ with $F(x) = x$, and we simply apply Banach's fixed-point theorem. Recall that $W \times Z$ is complete. With $y = (w_{\text{in}}, z_{\text{in}})$, $x = (w, z)$ we have

$$y - \mathcal{I}_M(\tau\Delta)(x) + \mathcal{I}_M(\tau_0\Delta)(x) = \begin{pmatrix} w_{\text{in}} + (\tau - \tau_0)\Delta(z) \\ z_{\text{in}} \end{pmatrix}. \tag{7.2.26}$$

For $\tau \in [\tau_0, \tau_0 + \delta_0]$, we can exploit the hypothesis to infer that $W_{\text{in}} + (\tau - \tau_0)\Delta(Z) \subset W_{\text{in,e}}$ and thus (**??**) is contained in $W_{\text{in,e}} \times Z_{\text{in}}$; since $\Delta(Z) \subset L_2^k$ it is even contained in $W_{\text{in}} \times Z_{\text{in}}$. This reveals that F maps $W \times Z$ into $W \times Z$. Finally, using (7.2.24) and with $x_j = (w_j, z_j)$ we conclude

$$\begin{aligned}
&\|F(x_1) - F(x_2)\|_2^2 \leq \\
&\leq c^2 \| \, [y - \mathcal{I}_M(\tau\Delta)(x_1) + \mathcal{I}_M(\tau_0\Delta)(x_1)] - [y - \mathcal{I}_M(\tau\Delta)(x_2) + \mathcal{I}_M(\tau_0\Delta)(x_2)] \, \|_2^2 \leq \\
&\leq c^2 \| \, [\mathcal{I}_M(\tau_0\Delta)(x_1) - \mathcal{I}_M(\tau\Delta)(x_1)] - [\mathcal{I}_M(\tau_0\Delta)(x_2) - \mathcal{I}_M(\tau\Delta)(x_2)] \, \|_2^2 \leq \\
&\leq c^2 \left\| \begin{pmatrix} (\tau - \tau_0)\Delta(z_1) \\ 0 \end{pmatrix} - \begin{pmatrix} (\tau - \tau_0)\Delta(z_2) \\ 0 \end{pmatrix} \right\|_2^2 \leq \\
&\leq c^2 |\tau - \tau_0|^2 \|\Delta(z_1) - \Delta(z_2)\|_2^2 \leq c^2 |\tau - \tau_0|^2 \|\Delta\|_i^2 \|z_1 - z_2\|_2^2 \leq c^2 |\tau - \tau_0|^2 \|\Delta\|_i^2 \|x_1 - x_2\|_2^2.
\end{aligned}$$

For $\tau \in [\tau_0, \tau_0 + \tilde{\delta}_0]$ we have $c^2 |\tau - \tau_0|^2 \|\Delta\|_i^2 < 1$ such that F is a strict contraction. All this guarantees that F has a unique fixed-point in $W \times Z$.

Third step. Obviously, $\mathcal{I}_M(\tau\Delta)$ does have an inverse for $\tau = 0$. Hence this mapping has an inverse for $\tau \in [0, 1] \cap [0, \tilde{\delta}_0]$ and, by induction, for $\tau \in [0, 1] \cap [0, k\tilde{\delta}_0]$ with $k = 0, 1, 2, ...$ For sufficiently large k we conclude that $\mathcal{I}_M(\Delta)$ has an inverse. By (7.2.23) it is clear that the incremental gain of this inverse is bounded by c. Then it is a simple exercise to prove that the causal mapping $\mathcal{I}_M(\Delta)$ has a causal inverse $\mathcal{I}_M(\Delta)^{-1}$ defined on $W_{\text{in,e}} \times Z_{\text{in,e}}$ with the same bound c on its incremental L_2-gain. ■

Remarks.

- The technical property $W_{\text{in}} + \delta\Delta(Z) \subset W_{\text{in,e}}$ for $\delta \in [0, \delta_0]$ is certainly true if Δ maps Z into $W_{\text{in,e}}$ and $W_{\text{in,e}}$ is a linear space. For linear M defined on a linear space W_e and the choice $W_{\text{in,e}} = W_e$, the uncertainties have to map Z_e into W_e. We stress that (7.2.21) and (7.2.17) are then obviously equivalent.

- Even if dealing with nonlinear uncertainties Δ, they often are guaranteed to have the property $\Delta(0) = 0$. Then we infer that $\mathcal{I}_M(\Delta)(0) = 0$ such that the same must hold for its inverse. Therefore
$$\|\mathcal{I}_M(\Delta)^{-1}\|_2 \leq \|\mathcal{I}_M(\Delta)^{-1}\|_{2i}$$
and Theorem 7.15 guarantees a bound on the L_2-gain of the inverse as well.

- If the uncertainties $\Delta \in \mathbf{\Delta}$ are linear and defined on the linear space Z_e, (7.2.20) is equivalent (7.2.16), and $\mathcal{I}_M(\Delta)$ as well as its inverse (if existing) are linear. Since uncertainties could be defined by infinite-dimensional systems (partial differential equations) even in this simplified situation Theorem 7.15 might be extremely helpful in guaranteeing well-posedness.

- We have made use of a rather straightforward application of the global version of Banach's fixed-point theorem. As variants on could work with well-established local versions thereof. We finally also stress for Theorem 7.15 that it might not directly be applicable to specific practical problems, but the techniques of proof might be successfully modified along the lines of the rather the detailed exposition of the basic ingredients.

However, one can also change the viewpoint: Given a quadratic mapping Σ, *define* the class of uncertainties $\boldsymbol{\Delta}$ as those that satisfy (7.2.16). Then all systems that have the property (7.2.17) cannot be destabilized by this class of uncertainties. Classical small-gain and passivity theorems fall in this class as will be discussed Section **??**.

Chapter 8

Robust controller synthesis

8.1 Robust controller design

So far we have presented techniques to design controllers for nominal stability and nominal performance. Previous chapters have been devoted to a thorough discussion of how to analyze, for a fixed stabilizing controller, robust stability or robust performance. For time-invariant or time-varying parametric uncertainties, we have seen direct tests formulated as searching for constant or parameter-dependent quadratic Lyapunov functions. For much larger classes of uncertainties, we have derived tests in terms of integral quadratic constraints (IQC's) that involve additional variables which have been called scalings or multipliers.

Typically, only those IQC tests with a class of multipliers that admit a state-space description as discussed in Sections **??**-**??** of Chapter 4 are amenable to a systematic output-feedback controller design procedure which is a reminiscent of the D/K-iteration in μ-theory. This will be the first subject of this chapter.

In a second section we consider as a particular information structure the robust state-feedback design problem. We will reveal that the search for static state-feedback gains which achieve robust performance can be transformed into a convex optimization problem.

The discussion is confined to the quadratic performance problem since most results can be extended in a pretty straightforward fashion to the other specifications considered in these notes.

8.1.1 Robust output-feedback controller design

If characterizing robust performance by an IQC, the goal in robust design is to find a controller *and* a multiplier such that, for the closed-loop system, the corresponding IQC test is satisfied. Hence, the multiplier appears as an extra unknown what makes the problem hard if not impossible to solve.

However, if the multiplier is held fixed, searching for a controller amounts to a nominal design problem that can be approached with the techniques described earlier. If the controller is held fixed, the analysis techniques presented in Chapter 7 can be used to find a suitable multiplier. Hence, instead of trying to search for a controller and a multiplier commonly, one iterates between the search for a controller with fixed multiplier and the search for a multiplier with fixed controller. This procedure is known from μ-theory as scalings/controller iteration or D/K iteration.

To be more concrete, we consider the specific *example* of achieving robust quadratic performance against time-varying parametric uncertainties as discussed in Section **??**.

The uncontrolled unperturbed system is described by (4.1.1). We assume that $w_1 \to z_1$ is the uncertainty channel and the uncontrolled uncertain system is described by including

$$w_1(t) = \Delta(t) z_1(t)$$

where $\Delta(.)$ varies in the set of continuous curves satisfying

$$\Delta(t) \in \mathbf{\Delta}_c := \text{conv}\{\Delta_1, ..., \Delta_N\} \text{ for all } t \geq 0.$$

We assume (w.l.o.g.) that

$$0 \in \text{conv}\{\Delta_1, ..., \Delta_N\}.$$

The performance channel is assumed to be given by $w_2 \to z_2$, and the performance index

$$P_p = \begin{pmatrix} Q_p & S_p \\ S_p^T & R_p \end{pmatrix}, \quad R_p \geq 0 \quad \text{with the inverse} \quad \tilde{P}_p^{-1} = \begin{pmatrix} \tilde{Q}_p & \tilde{S}_p \\ \tilde{S}_p^T & \tilde{R}_p \end{pmatrix}, \quad \tilde{Q}_p \leq 0$$

is used to define the quadratic performance specification

$$\int_0^\infty \begin{pmatrix} w_2(t) \\ z_2(t) \end{pmatrix}^T P_p \begin{pmatrix} w_2(t) \\ z_2(t) \end{pmatrix} dt \leq -\epsilon \|w_2\|_2^2.$$

The goal is to design a controller that achieves robust stability and robust quadratic performance. We can guarantee both properties by finding a controller, a Lyapunov matrix $\mathcal{X}$, and a multiplier

$$P = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}, \quad Q < 0, \quad \begin{pmatrix} \Delta_j \\ I \end{pmatrix}^T \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \begin{pmatrix} \Delta_j \\ I \end{pmatrix} > 0 \text{ for all } j = 1, \ldots, N \tag{8.1.1}$$

that satisfy the inequalities

$$\mathcal{X} > 0, \quad \left(\begin{array}{ccc} I & 0 & 0 \\ \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_1 & \mathcal{X}\mathcal{B}_2 \\ \hline 0 & I & 0 \\ \mathcal{C}_1 & \mathcal{D}_1 & \mathcal{D}_{12} \\ \hline 0 & 0 & I \\ \mathcal{C}_2 & \mathcal{D}_{21} & \mathcal{D}_2 \end{array}\right)^T \left(\begin{array}{cc|cc|cc} 0 & I & 0 & 0 & 0 & 0 \\ I & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & Q & S & 0 & 0 \\ 0 & 0 & S^T & R & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & Q_p & S_p \\ 0 & 0 & 0 & 0 & S_p^T & R_p \end{array}\right) \left(\begin{array}{ccc} I & 0 & 0 \\ \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_1 & \mathcal{X}\mathcal{B}_2 \\ \hline 0 & I & 0 \\ \mathcal{C}_1 & \mathcal{D}_1 & \mathcal{D}_{12} \\ \hline 0 & 0 & I \\ \mathcal{C}_2 & \mathcal{D}_{21} & \mathcal{D}_2 \end{array}\right) < 0.$$

(Recall that the condition on the left-upper block of P can be relaxed in particular cases what could reduce the conservatism of the test.)

If we apply the controller parameter transformation of Chapter **??**, we arrive at the synthesis matrix inequalities

$$\boldsymbol{X}(v) > 0, \quad \left(\begin{array}{c} * \\ * \\ \hline * \\ * \\ \hline * \\ * \end{array}\right)^T \left(\begin{array}{cc|cc|cc} 0 & I & 0 & 0 & 0 & 0 \\ I & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & Q & S & 0 & 0 \\ 0 & 0 & S^T & R & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & Q_p & S_p \\ 0 & 0 & 0 & 0 & S_p^T & R_p \end{array}\right) \left(\begin{array}{ccc} I & 0 & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_1(v) & \boldsymbol{B}_2(v) \\ \hline 0 & I & 0 \\ \boldsymbol{C}_1(v) & \boldsymbol{D}_1(v) & \boldsymbol{D}_{12}(v) \\ \hline 0 & 0 & I \\ \boldsymbol{C}_2(v) & \boldsymbol{D}_{21}(v) & \boldsymbol{D}_2(v) \end{array}\right) < 0.$$

Unfortunately, there is no obvious way how to render these synthesis inequalities convex in *all* variables v, Q, S, R.

This is the reason why we consider, instead, the problem with a scaled uncertainty

$$w_1(t) = [r\Delta(t)]z_1(t), \quad \Delta(t) \in \boldsymbol{\Delta}_c \tag{8.1.2}$$

where the scaling factor is contained in the interval $[0, 1]$. Due to

$$\begin{pmatrix} r\Delta \\ I \end{pmatrix}^T \begin{pmatrix} Q & rS \\ rS^T & r^2R \end{pmatrix} \begin{pmatrix} r\Delta \\ I \end{pmatrix} = r^2 \begin{pmatrix} \Delta \\ I \end{pmatrix}^T \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \begin{pmatrix} \Delta \\ I \end{pmatrix},$$

we conclude that the corresponding analysis or synthesis are given by (8.1.1) and

$$\mathcal{X} > 0, \quad \left(\begin{array}{ccc} I & 0 & 0 \\ \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_1 & \mathcal{X}\mathcal{B}_2 \\ \hline 0 & I & 0 \\ \mathcal{C}_1 & \mathcal{D}_1 & \mathcal{D}_{12} \\ \hline 0 & 0 & I \\ \mathcal{C}_2 & \mathcal{D}_{21} & \mathcal{D}_2 \end{array}\right)^T \left(\begin{array}{cc|cc|cc} 0 & I & 0 & 0 & 0 & 0 \\ I & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & Q & rS & 0 & 0 \\ 0 & 0 & rS^T & r^2R & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & Q_p & S_p \\ 0 & 0 & 0 & 0 & S_p^T & R_p \end{array}\right) \left(\begin{array}{ccc} I & 0 & 0 \\ \mathcal{X}\mathcal{A} & \mathcal{X}\mathcal{B}_1 & \mathcal{X}\mathcal{B}_2 \\ \hline 0 & I & 0 \\ \mathcal{C}_1 & \mathcal{D}_1 & \mathcal{D}_{12} \\ \hline 0 & 0 & I \\ \mathcal{C}_2 & \mathcal{D}_{21} & \mathcal{D}_2 \end{array}\right) < 0 \tag{8.1.3}$$

or

$$
\boldsymbol{X}(v) > 0, \quad
\begin{pmatrix} * \\ * \\ \hline * \\ * \\ \hline * \\ * \end{pmatrix}^T
\left(\begin{array}{cc|cc|cc}
0 & I & 0 & 0 & 0 & 0 \\
I & 0 & 0 & 0 & 0 & 0 \\ \hline
0 & 0 & Q & rS & 0 & 0 \\
0 & 0 & rS^T & r^2R & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & Q_p & S_p \\
0 & 0 & 0 & 0 & S_p^T & R_p
\end{array}\right)
\left(\begin{array}{ccc}
I & 0 & 0 \\
\boldsymbol{A}(v) & \boldsymbol{B}_1(v) & \boldsymbol{B}_2(v) \\ \hline
0 & I & 0 \\
\boldsymbol{C}_1(v) & \boldsymbol{D}_1(v) & \boldsymbol{D}_{12}(v) \\ \hline
0 & 0 & I \\
\boldsymbol{C}_2(v) & \boldsymbol{D}_{21}(v) & \boldsymbol{D}_2(v)
\end{array}\right) < 0. \tag{8.1.4}
$$

For $r = 0$, we hence have to solve the nominal quadratic performance synthesis inequalities. If they are not solvable, the robust quadratic performance synthesis problem is not solvable either and we can stop. If they are solvable, the idea is to try to increase, keeping the synthesis inequalities feasible, the parameter r from zero to one. Increasing r is achieved by alternatingly maximizing r over v satisfying (8.1.4) (for fixed P) and by varying $\mathcal{X}$ and P in (8.1.3) (for a fixed controller).

The maximization of r proceeds along the following steps:

Initialization. Perform a nominal quadratic performance design by solving (8.1.4) for $r = 0$. Proceed if these inequalities are feasible and compute a corresponding controller.

After this initial phase, the iteration is started. The $j-1$-st step of the iteration leads to a controller, a Lyapunov matrix $\mathcal{X}$, and a multiplier P that satisfy the inequalities (8.1.1) and (8.1.3) for the parameter $r = r_{j-1}$. Then it proceeds as follows:

First step: Fix the controller and maximize r by varying the Lyapunov matrix $\mathcal{X}$ and the scaling such that such that (8.1.1) and (8.1.3) hold. The maximal radius is denoted as $\widehat{r}_j$ and it satisfies $r_{j-1} \leq \widehat{r}_j$.

Second step: Fix the resulting scaling P and find the largest r by varying the variables v in (8.1.4). The obtained maximum r_j clearly satisfies $\widehat{r}_j \leq r_j$.

The iteration defines a sequence of radii

$$r_1 \leq r_2 \leq r_3 \leq \cdots$$

and a corresponding controller that guarantee robust stability and robust quadratic performance for all uncertainties (8.1.2) with radius $r = r_j$.

If we are in the lucky situation that there is an index for which $r_j \geq 1$, the corresponding controller is robustly performing for all uncertainties with values in $\boldsymbol{\Delta}_c$ as desired, and we are done. However, if $r_j < 1$ for all indices, we cannot guarantee robust performance for $r = 1$, but we still have a guarantee of robust performance for $r = r_j$!

Before entering a brief discussion of this procedure, let us include the following remarks on the start-up and on the computations. If the nominal performance synthesis problem has a solution, the LMI's (8.1.1)-(8.1.3) do have a solution $\mathcal{X}$ and P for the resulting controller and for some - possibly small - $r > 0$; this just follows by continuity. Hence the iteration does not get stuck after the first step. Secondly, for a fixed r, the first step of the iteration amounts to solving an analysis problem, and finding a solution v of (8.1.4) can be converted to an LMI problem. Therefore, the maximization of r can be performed by bisection.

Even if the inequalities (8.1.1)-(8.1.4) are solvable for $r = 1$, it can happen the the limit of r_j is smaller than one. As a remedy, one could consider another parameter to maximize, or one could modify the iteration scheme that has been sketched above. For example, it is possible to take the fine structure of the involved functions into account and to suggest other variable combinations that render the resulting iteration steps convex. Unfortunately, one cannot give general recommendations for modifications which guarantee success.

Remark. It should be noted that the controller/multiplier iteration can be extended to all robust performance tests that are based on families of dynamic IQC's which are described by real rational multipliers. Technically, one just requires a parametrization of the multipliers such that the corresponding analysis test (for a fixed controller) and the controller synthesis (for a fixed multiplier) both reduce to solving standard LMI problems.

8.1.2 Robust state-feedback controller design

For the same set-up as in the previous section we consider the corresponding synthesis problem if the state of the underlying system is measurable. According to our discussion in Section 4.6, the resulting synthesis inequalities read as

$$Q < 0,\ \begin{pmatrix} \Delta_j \\ I \end{pmatrix}^T \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \begin{pmatrix} \Delta_j \\ I \end{pmatrix} > 0 \ \text{ for all } \ j = 1, \ldots, N$$

and

$$Y > 0,\ \left(\begin{array}{c} * \\ * \\ \hline * \\ * \\ \hline * \\ * \end{array}\right)^T \left(\begin{array}{cc|cc|cc} 0 & I & 0 & 0 & 0 & 0 \\ I & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & Q & S & 0 & 0 \\ 0 & 0 & S^T & R & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & Q_p & S_p \\ 0 & 0 & 0 & 0 & S_p^T & R_p \end{array}\right) \left(\begin{array}{ccc} I & 0 & 0 \\ AY + BM & B_1 & B_2 \\ \hline 0 & I & 0 \\ C_1Y + E_1M & D_1 & D_{12} \\ \hline 0 & 0 & I \\ C_2Y + E_1M & D_{21} & D_2 \end{array}\right) < 0$$

in the variables Y, M, Q, S, R.

In this form these inequalities are *not convex*. However, we can apply the Dualization Lemma (Section 4.5.1) to arrive at the equivalent inequalities

$$\tilde{R} > 0,\ \begin{pmatrix} I \\ -\Delta_j^T \end{pmatrix}^T \begin{pmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{pmatrix} \begin{pmatrix} I \\ -\Delta_j^T \end{pmatrix} < 0 \ \text{ for all } \ j = 1, \ldots, N$$

and $Y > 0$,

$$
* \left(\begin{array}{cc|cc|cc}
0 & I & 0 & 0 & 0 & 0 \\
I & 0 & 0 & 0 & 0 & 0 \\ \hline
0 & 0 & \tilde{Q} & \tilde{S} & 0 & 0 \\
0 & 0 & \tilde{S}^T & \tilde{R} & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & \tilde{Q}_p & \tilde{S}_p \\
0 & 0 & 0 & 0 & \tilde{S}_p^T & \tilde{R}_p
\end{array}\right)
\left(\begin{array}{ccc}
-(AY + BM)^T & -(C_1 Y + E_1 M)^T & -(C_2 Y + E_2 M)^T \\
I & 0 & 0 \\
-B_1^T & -D_1^T & -D_{21}^T \\ \hline
0 & I & 0 \\
-B_2^T & -D_{12}^T & -D_2^T \\ \hline
0 & 0 & I
\end{array}\right) > 0
$$

in the variables Y, M, $\tilde{Q}$, $\tilde{S}$, $\tilde{R}$. It turns out that these dual inequalities are all *affine* in the unknowns. Testing feasibility hence amounts to solving a standard LMI problem. If the LMI's are feasible, a robust static state-feedback gain is given by $\mathcal{D} = MY^{-1}$. This is one of the very few lucky instances in the world of designing robust controllers!

8.1.3 Affine parameter dependence

Let us finally consider the system

$$
\begin{pmatrix} \dot{x} \\ z \\ y \end{pmatrix} = \left(\begin{array}{c|cc}
A(\Delta(t)) & B_1(\Delta(t)) & B(\Delta(t)) \\ \hline
C_1(\Delta(t)) & D(\Delta(t)) & E(\Delta(t)) \\
C(\Delta(t)) & F(\Delta(t)) & 0
\end{array}\right) \begin{pmatrix} x \\ w \\ u \end{pmatrix}, \quad \Delta(t) \in \mathrm{conv}\{\Delta_1, ..., \Delta_N\}
$$

where the describing matrices depend *affinely* on the time-varying parameters. If designing output-feedback controllers, there is no systematic alternative to pulling out the uncertainties and applying the scalings techniques as in Section 8.1.1.

For robust state-feedback design there is an alternative without scalings. One just needs to directly solve the system of LMI's

$$
Y > 0, \quad \left(\begin{array}{c} * \\ * \\ \hline * \\ * \end{array}\right)^T
\left(\begin{array}{cc|cc}
0 & I & 0 & 0 \\
I & 0 & 0 & 0 \\ \hline
0 & 0 & Q_p & S_p \\
0 & 0 & S_p^T & R_p
\end{array}\right)
\left(\begin{array}{cc}
I & 0 \\
A(\Delta_j)Y + B(\Delta_j)M & B_1(\Delta_j) \\ \hline
0 & I \\
C_1(\Delta_j)Y + E(\Delta_j)M & D(\Delta_j)
\end{array}\right) < 0, \ j = 1, \ldots, N \tag{8.1.5}
$$

in the variables Y and M.

For the controller gain $D_c = MY^{-1}$ we obtain

$$
Y > 0, \quad \left(\begin{array}{c} * \\ * \\ \hline * \\ * \end{array}\right)^T
\left(\begin{array}{cc|cc}
0 & I & 0 & 0 \\
I & 0 & 0 & 0 \\ \hline
0 & 0 & Q_p & S_p \\
0 & 0 & S_p^T & R_p
\end{array}\right)
\left(\begin{array}{cc}
I & 0 \\
(A(\Delta_j) + B(\Delta_j)D_c)Y & B_1(\Delta_j) \\ \hline
0 & I \\
(C_1(\Delta_j) + E(\Delta_j)D_c)Y & D(\Delta_j)
\end{array}\right) < 0, \ j = 1, \ldots, N
$$

A convexity argument leads to

$$Y > 0, \quad \begin{pmatrix} * \\ * \\ \hline * \\ * \end{pmatrix}^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & Q_p & S_p \\ 0 & 0 & S_p^T & R_p \end{array}\right) \begin{pmatrix} I & 0 \\ (A(\Delta(t)) + B(\Delta(t))D_c)Y & B_1(\Delta(t)) \\ \hline 0 & I \\ (C_1(\Delta(t)) + E(\Delta(t))D_c)Y & D(\Delta(t)) \end{pmatrix} < 0$$

for all parameter curves $\Delta(t) \in \text{conv}\{\Delta_1, \ldots, \Delta_N\}$, and we can perform a congruence transformation as in Section 4.6 to get

$$\mathcal{X} > 0, \quad \begin{pmatrix} * \\ * \\ \hline * \\ * \end{pmatrix}^T \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \\ \hline 0 & 0 & Q_p & S_p \\ 0 & 0 & S_p^T & R_p \end{array}\right) \begin{pmatrix} I & 0 \\ \mathcal{X}(A(\Delta(t)) + B(\Delta(t))D_c) & \mathcal{X}B_1(\Delta(t)) \\ \hline 0 & I \\ (C_1(\Delta(t)) + E(\Delta(t))D_c) & D(\Delta(t)) \end{pmatrix} < 0.$$

These two inequalities imply, in turn, robust exponential stability and robust quadratic performance for the controlled system as seen in Section **??**.

We have proved that it suffices to directly solve the LMI's (8.1.5) to compute a robust static state-feedback controller. Hence, if the system's parameter dependence is affine, we have found two equivalent sets of synthesis inequalities that differ in the number of the involved variables and in the sizes of the LMI's that are involved. In practice, the correct choice is dictated by whatever system can be solved faster, more efficiently, or numerically more reliably.

Remark. Here is the reason why it is possible to directly solve the robust performance problem by state-feedback without scalings, and why this technique does, unfortunately, not extend to output-feedback control: The linearizing controller parameter transformation for state-feedback problems *does not involve the matrices that describe the open-loop system,* whereas that for that for ouptut-feedback problems indeed depends on the matrices A, B, C of the open-loop system description.

Let us conclude this chapter by stressing, again, that these techniques find straightforward extensions to other performance specifications. As an exercise, the reader is asked to work out the details of the corresponding results for the robust H_2-synthesis problem by state- or output-feedback.

8.2 Exercises

Exercise 1

This is an exercise on robust control. To reduce the complexity of programming, we consider a non-dynamic system only.

Suppose you have given the algebraic uncertain system

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ \hline z \\ \hline y_1 \\ y_2 \end{pmatrix} = \left(\begin{array}{cccc|c|cc} 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0.5 & 0 & 0.5 & 0 & 1 & 0 & 1 \\ 2a & 0 & a & 0 & 1 & 0 & 0 \\ 0 & -2a & 0 & -a & 1 & 0 & 0 \\ \hline 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ \hline w \\ \hline u_1 \\ u_2 \end{pmatrix},$$

with a time-varying uncertainty

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix} = \begin{pmatrix} \delta_1(t) & & & 0 \\ & \delta_1(t) & & \\ & & \delta_2(t) & \\ 0 & & & \delta_2(t) \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}, \quad |\delta_1(t)| \leq 0.7, \quad |\delta_2(t)| \leq 0.7.$$

As the performance measure we choose the L_2-gain of the channel $w \to z$.

(a) For the uncontrolled system and for each $a \in [0, 1]$, find the minimal robust L_2-gain level of the channel $w \to z$ by applying the robust performance analysis test in Chapter 3 with the following class of scalings $P = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}$:

- P is as in μ-theory: Q, S, R are block-diagonal, $Q < 0$, R is related to Q (how?), and S is skew-symmetric.
- P is general with $Q < 0$.
- P is general with $Q_1 < 0$, $Q_2 < 0$, where Q_j denote the blocks $Q(1 : 2, 1 : 2)$ and $Q(3 : 4, 3 : 4)$ in Matlab notation.

Draw plots of the corresponding optimal values versus the parameter a and comment!

(b) For $a = 0.9$, apply the controller

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

and perform the analysis test with the largest class of scalings for $k \in [-1, 1]$. Plot the resulting optimal value over k and comment.

(c) Perform a controller/scaling iteration to minimize the optimal values for the controller structures

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & k_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} k_1 & k_{12} \\ k_{21} & k_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Start from gain zero and plot the optimal values that can are reached in each step of the iteration to reveal how they decrease. Comment on the convergence.

(d) With the last full controller from the previous exercise for a performance level that is close to the limit, redo the analysis of the first part. Plot the curves and comment.

Chapter 9

Linear parameterically varying systems

Linear parameterically varying (LPV) systems are linear systems whose describing matrices depend on a time-varying parameter such that both the parameter itself and its rate of variation are known to be contained in pre-specified sets.

In robust control, the goal is to find *one fixed* controller that achieves robust stability and robust performance for all possible parameter variations, irrespective of which specific parameter curve does indeed perturb the system.

Instead, in LPV control, it is assumed that the parameter (and, possibly, its rate of variation), although not known a priori, is (are) on-line measurable. Hence the actual parameter value (and its derivative) can be used as extra information to control the system - the controller will turn out to depend on the parameter as well. We will actually choose also an LPV structure for the controller to be designed.

We would like to stress the decisive distinction to the control of time-varying systems: In the standard techniques to controlling time-varying systems, the model description is assumed to be known a priori over the whole time interval $[0, \infty)$. In LPV control, the model is assumed to be known, at time instant t, only over the interval $[0, t]$.

The techniques we would like to develop closely resemble those for robust control we have investigated earlier. It is possible to apply them

- to control certain classes of nonlinear systems
- to provide a systematic procedure for gain-scheduling

with guarantees for stability and performance.

Before we explore these applications in more detail we would like to start presenting the available problem setups and solution techniques to LPV control.

9.1 General Parameter Dependence

Suppose that $\boldsymbol{\delta}_c, \dot{\boldsymbol{\delta}}_c \subset \mathbb{R}^m$ are two parameter sets such that

$$\boldsymbol{\delta}_c \times \dot{\boldsymbol{\delta}}_c \text{ is compact,}$$

and that the matrix valued function

$$\left(\begin{array}{c|cc} A(p) & B_p(p) & B(p) \\ \hline C_p(p) & D_p(p) & E(p) \\ C(p) & F(p) & 0 \end{array}\right) \text{ is continuous in } p \in \boldsymbol{\delta}_c. \tag{9.1.1}$$

Consider the *Linear Parameterically Varying (LPV) system* that is described as

$$\left(\begin{array}{c} \dot{x} \\ \hline z_p \\ y \end{array}\right) = \left(\begin{array}{c|cc} A(\delta(t)) & B_p(\delta(t)) & B(\delta(t)) \\ \hline C_p(\delta(t)) & D_p(\delta(t)) & E(\delta(t)) \\ C(\delta(t)) & F(\delta(t)) & 0 \end{array}\right) \left(\begin{array}{c} x \\ \hline w_p \\ u \end{array}\right), \quad \delta(t) \in \boldsymbol{\delta}_c, \quad \dot{\delta}(t) \in \dot{\boldsymbol{\delta}}_c. \tag{9.1.2}$$

We actually mean the family of systems that is obtained if letting $\delta(.)$ vary in the set of continuously differentiable parameter curves

$$\delta : [0, \infty) \to \mathbb{R}^m \text{ with } \delta(t) \in \boldsymbol{\delta}_c, \quad \dot{\delta}(t) \in \dot{\boldsymbol{\delta}}_c \text{ for all } t \geq 0.$$

The signals admit the same interpretations as in Chapter 4: u is the control input, y is the measured output available for control, and $w_p \to z_p$ denotes the performance channel.

In *LPV control*, it is assumed that the parameter $\delta(t)$ is on-line measurable. Hence the actual value of $\delta(t)$ can be taken as extra information for the controller to achieve the desired design goal.

In view of the specific structure of the system description, we assume that the controller admits a similar structure. In fact, an *LPV controller* is defined by functions

$$\left(\begin{array}{cc} A_c(p) & B_c(p) \\ C_c(p) & D_c(p) \end{array}\right) \text{ that are continuous in } p \in \boldsymbol{\delta}_c \tag{9.1.3}$$

as

$$\left(\begin{array}{c} \dot{x}_c \\ u \end{array}\right) = \left(\begin{array}{cc} A_c(\delta(t)) & B_c(\delta(t)) \\ C_c(\delta(t)) & D_c(\delta(t)) \end{array}\right) \left(\begin{array}{c} x_c \\ y \end{array}\right)$$

with the following interpretation: It evolves according to linear dynamics that are defined at time-instant t via the actually measured value of $\delta(t)$.

Note that a robust controller would be simply defined with a constant matrix

$$\left(\begin{array}{cc} A_c & B_c \\ C_c & D_c \end{array}\right)$$

that does not depend on δ what clarifies the difference between robust controllers and LPV controllers.

The controlled system admits the description

$$\begin{pmatrix} \dot{\xi} \\ z_p \end{pmatrix} = \begin{pmatrix} \mathcal{A}(\delta(t)) & \mathcal{B}(\delta(t)) \\ \mathcal{C}(\delta(t)) & \mathcal{D}(\delta(t)) \end{pmatrix} \begin{pmatrix} \xi \\ w_p \end{pmatrix}, \quad \delta(t) \in \boldsymbol{\delta}_c, \quad \dot{\delta}(t) \in \dot{\boldsymbol{\delta}}_c \tag{9.1.4}$$

where the function

$$\begin{pmatrix} \mathcal{A}(p) & \mathcal{B}(p) \\ \mathcal{C}(p) & \mathcal{D}(p) \end{pmatrix} \text{ is continuous in } \ p \in \boldsymbol{\delta}_c$$

and given as

$$\left(\begin{array}{cc|c} A(p) + B(p)D_c(p)C(p) & B(p)C_c(p) & B_p(p) + B(p)D_c(p)F(p) \\ B_c(p)C(p) & A_c(p) & B_c(p)F(p) \\ \hline C_p(p) + E(p)D_c(p)C(p) & E(p)C_c(p) & D_p(p) + E(p)D_c(p)F(p) \end{array}\right)$$

or

$$\left(\begin{array}{cc|c} A(p) & 0 & B_p(p) \\ 0 & 0 & 0 \\ \hline C_p(p) & 0 & D_p(p) \end{array}\right) + \left(\begin{array}{cc} 0 & B(p) \\ I & 0 \\ \hline 0 & E(p) \end{array}\right) \begin{pmatrix} A_c(p) & B_c(p) \\ C_c(p) & D_c(p) \end{pmatrix} \left(\begin{array}{cc|c} 0 & I & 0 \\ C(p) & 0 & F(p) \end{array}\right).$$

To evaluate performance, we concentrate again on the quadratic specification

$$\int_0^\infty \begin{pmatrix} w(t) \\ z(t) \end{pmatrix}^T P_p \begin{pmatrix} w(t) \\ z(t) \end{pmatrix} dt \leq -\epsilon \|w\|_2^2 \tag{9.1.5}$$

with an index

$$P_p = \begin{pmatrix} Q_p & S_p \\ S_p^T & R_p \end{pmatrix}, \quad R_p \geq 0 \quad \text{that has the inverse} \quad \tilde{P}_p^{-1} = \begin{pmatrix} \tilde{Q}_p & \tilde{S}_p \\ \tilde{S}_p^T & \tilde{R}_p \end{pmatrix}, \quad \tilde{Q}_p \leq 0.$$

In order to abbreviate the formulation of the analysis result we introduce the following differential operator.

Definition 9.1 If $X : \boldsymbol{\delta}_c \ni p \to X(p) \in \mathbb{R}^{n \times n}$ is continuously differentiable, the continuous mapping

$$\partial X : \boldsymbol{\delta}_c \times \dot{\boldsymbol{\delta}}_c \to \mathbb{R}^{n \times n} \ \text{ is defined as } \ \partial X(p, q) := \sum_{j=1}^m \frac{\partial X}{\partial p_j}(p) q_j.$$

Note that this definition is simply motivated by the fact that, along any continuously differentiable parameter curve $\delta(.)$, we have

$$\frac{d}{dt} X(\delta(t)) = \sum_{j=1}^m \frac{\partial X}{\partial p_j}(\delta(t)) \dot{\delta}_j(t) = \partial X(\delta(t), \dot{\delta}(t)). \tag{9.1.6}$$

(We carefully wrote down the definitions and relations, and one should read all this correctly. X and ∂X are functions of the parameters $p \in \boldsymbol{\delta}_c$ and $q \in \dot{\boldsymbol{\delta}}_c$ respectively. In the definition of ∂X, no time-trajectories are involved. The definition of ∂X is just tailored to obtain the property (9.1.6) if plugging in a function of time.)

In view of the former discussion, the following analysis result comes as no surprise.

Theorem 9.2 *Suppose there exists a continuously differentiable $\mathcal{X}(p)$ defined for $p \in \boldsymbol{\delta}_c$ such that for all $p \in \boldsymbol{\delta}_c$ and $q \in \dot{\boldsymbol{\delta}}_c$ one has*

$$\mathcal{X}(p) > 0, \quad \begin{pmatrix} \partial\mathcal{X}(p,q) + \mathcal{A}(p)^T\mathcal{X}(p) + \mathcal{X}(p)\mathcal{A}(p) & \mathcal{X}(p)\mathcal{B}(p) \\ \mathcal{B}(p)^T\mathcal{X}(p) & 0 \end{pmatrix} + \\ + \begin{pmatrix} 0 & I \\ \mathcal{C}(p) & \mathcal{D}(p) \end{pmatrix}^T P_p \begin{pmatrix} 0 & I \\ \mathcal{C}(p) & \mathcal{D}(p) \end{pmatrix} < 0. \quad (9.1.7)$$

Then there exists an $\epsilon > 0$ such that, for each parameter curve with $\delta(t) \in \boldsymbol{\delta}_c$ and $\dot{\delta}(t) \in \dot{\boldsymbol{\delta}}_c$, the system (9.1.4) *is exponentially stable and satisfies* (9.1.5) *if the initial condition is zero and if $w_p \in L_2$.*

In view of our preparations the proof is a simple exercise that is left to the reader.

We can now use the same procedure as for LTI systems to arrive at the corresponding synthesis result. It is just required to obey that all the matrices are actually functions of $p \in \boldsymbol{\delta}_c$ or of $(p,q) \in \boldsymbol{\delta}_c \times \dot{\boldsymbol{\delta}}_c$. If partitioning

$$\mathcal{X} = \begin{pmatrix} X & U \\ U^T & * \end{pmatrix}, \quad \mathcal{X}^{-1} = \begin{pmatrix} Y & V \\ V^T & * \end{pmatrix},$$

we can again assume w.l.o.g. that U, V have full row rank. (Note that this requires the compactness hypothesis on $\boldsymbol{\delta}_c$ and $\dot{\boldsymbol{\delta}}_c$. Why?) With

$$\mathcal{Y} = \begin{pmatrix} Y & I \\ V^T & 0 \end{pmatrix} \text{ and } \mathcal{Z} = \begin{pmatrix} I & 0 \\ X & U \end{pmatrix}$$

we obtain the identities

$$\mathcal{Y}^T\mathcal{X} = \mathcal{Z} \text{ and } I - XY = UV^T.$$

If we apply the differential operator ∂ to the first functional identity, we arrive at $(\partial\mathcal{Y})^T\mathcal{X} + \mathcal{Y}^T(\partial\mathcal{X}) = \partial\mathcal{Z}$. (Do the simple calculations. Note that ∂ is not the usual differentiation such that you cannot apply the standard product rule.) If we right-multiply $\mathcal{Y}$, this leads to

$$\mathcal{Y}^T(\partial\mathcal{X})\mathcal{Y} = (\partial\mathcal{Z})\mathcal{Y} - (\partial\mathcal{Y})^T\mathcal{Z}^T = \begin{pmatrix} 0 & 0 \\ \partial X & \partial U \end{pmatrix}\begin{pmatrix} Y & I \\ V^T & 0 \end{pmatrix} - \begin{pmatrix} \partial Y & \partial V \\ 0 & 0 \end{pmatrix}\begin{pmatrix} I & X \\ 0 & U^T \end{pmatrix}$$

and hence to

$$\mathcal{Y}^T(\partial\mathcal{X})\mathcal{Y} = \begin{pmatrix} -\partial Y & -(\partial Y)X - (\partial V)U^T \\ (\partial X)Y + (\partial U)V^T & \partial X \end{pmatrix}.$$

If we introduce the transformed controller parameters

$$\begin{pmatrix} K & L \\ M & N \end{pmatrix} = \begin{pmatrix} U & XB \\ 0 & I \end{pmatrix} \begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \begin{pmatrix} V^T & 0 \\ CY & I \end{pmatrix} + \begin{pmatrix} XAY & 0 \\ 0 & 0 \end{pmatrix} + \\ + \begin{pmatrix} (\partial X)Y + (\partial U)V^T & 0 \\ 0 & 0 \end{pmatrix},$$

a brief calculation reveals that

$$\mathcal{Y}^T(\partial \mathcal{X} + \mathcal{A}^T \mathcal{X} + \mathcal{X}\mathcal{A})\mathcal{Y} = \begin{pmatrix} -\partial Y + \mathrm{sym}\,(AY + BM) & (A + BNC) + K^T \\ (A + BNC)^T + K & \partial X + \mathrm{sym}\,(AX + LC) \end{pmatrix}$$

$$\mathcal{Y}^T \mathcal{X}\mathcal{B} = \begin{pmatrix} B_p + BNF \\ XB_p + LF \end{pmatrix}, \quad \mathcal{C}\mathcal{Y} = \begin{pmatrix} C_pY + EM & C_p + ENC \end{pmatrix}, \quad \mathcal{D} = D_p + ENF$$

where we used again the abbreviation $\mathrm{sym}\,(M) = M + M^T$. If compared to a parameter independent Lyapunov function, we have modified the transformation to K by $(\partial X)Y + (\partial U)V^T$ in order to eliminate this extra term that appears from the congruence transformation of $\partial \mathcal{X}$. If $\mathcal{X}$ is does not depend on p, $\partial \mathcal{X}$ vanishes identically and the original transformation is recovered.

We observe that L, M, N are functions of $p \in \boldsymbol{\delta}_c$ only, whereas K also depends on $q \in \dot{\boldsymbol{\delta}}_c$. In fact, this function has the structure

$$K(p, q) = K_0(p) + \sum_{i=1}^{m} K_i(p) q_i \tag{9.1.8}$$

(why?) and, hence, it is fully described by specifying

$$K_i(p), \quad i = 0, 1, \ldots, m$$

that depend, as well, on $p \in \boldsymbol{\delta}_c$ only.

Literally as in Theorem 4.3 one can now prove the following synthesis result for LPV systems.

Theorem 9.3 *If there exists an LPV controller defined by* (9.1.3) *and a continuously differentiable* $\mathcal{X}(.)$ *defined for* $p \in \boldsymbol{\delta}_c$ *that satisfy* (9.1.7)*, then there exist continuously differentiable functions* X, Y *and continuous functions* K_i, L, M, N *defined on* $\boldsymbol{\delta}_c$ *such that, with* K *given by* (9.1.8)*, the inequalities*

$$\begin{pmatrix} Y & I \\ I & X \end{pmatrix} > 0 \tag{9.1.9}$$

and

$$\left(\begin{array}{cc|c} -\partial Y + \mathrm{sym}\,(AY + BM) & (A + BNC) + K^T & B_p + BNF \\ (A + BNC)^T + K & \partial X + \mathrm{sym}\,(AX + LC) & XB_p + LF \\ \hline (B_p + BNF)^T & (XB_p + LF)^T & 0 \end{array}\right) + \\ + \begin{pmatrix} * \\ * \end{pmatrix}^T P_p \left(\begin{array}{cc|c} 0 & 0 & I \\ C_pY + EM & C_p + ENC & D_p + ENF \end{array}\right) < 0 \tag{9.1.10}$$

hold on $\boldsymbol{\delta}_c \times \dot{\boldsymbol{\delta}}_c$. *Conversely, suppose the continuously differentiable* X, Y *and the continuous* K_i, *defining* K *as in* (9.1.8), L, M, N *satisfy these synthesis inequalities. Then one can factorize* $I - XY = UV^T$ *with continuously differentiable square and nonsingular* U, V, *and*

$$\mathcal{X} = \begin{pmatrix} Y & V \\ I & 0 \end{pmatrix}^{-1} \begin{pmatrix} I & 0 \\ X & U \end{pmatrix} \tag{9.1.11}$$

$$\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} = \begin{pmatrix} U & XB \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} K - XAY - [(\partial X)Y + (\partial U)V^T] & L \\ M & N \end{pmatrix} \begin{pmatrix} V^T & 0 \\ CY & I \end{pmatrix}^{-1} \tag{9.1.12}$$

render the analysis inequalities (9.1.7) *satisfied.*

Remark. Note that the formula (9.1.12) just emerges from the modified controller parameter transformation. We observe that the matrices B_c, C_c, D_c are functions of $p \in \boldsymbol{\delta}_c$ only. Due to the dependence of K on q and due to the extra term $U^{-1}[(\partial X)Y + (\partial U)V^T]V^{-T}$ in the formula for A_c, this latter matrix is a function that depends both on $p \in \boldsymbol{\delta}_c$ and $q \in \dot{\boldsymbol{\delta}}_c$. It has the same structure as K and can be written as

$$A_c(p, q) = A_0(p) + \sum_{i=1}^{m} A_i(p)q_i.$$

A straightforward calculation reveals that

$$A_i = U^{-1}[K_i V^{-T} - \frac{\partial X}{\partial p_i} Y V^{-T} - \frac{\partial U}{\partial p_i}], \quad i = 1, \ldots, m.$$

Hence, to implement this controller, one indeed requires not only to measure $\delta(t)$ but also its rate of variation $\dot{\delta}(t)$. However, one could possibly exploit the freedom in choosing U and V to render $A_i = 0$ such that A_c does not depend on q any more. Recall that U and V need to be related by $I - XY = UV^T$; hence let us choose

$$V^T := U^{-1}(I - XY).$$

This leads to

$$A_i = U^{-1}[(K_i - \frac{\partial X}{\partial p_i} Y)(I - XY)^{-1} U - \frac{\partial U}{\partial p_i}], \quad i = 1, \ldots, m.$$

Therefore, U should be chosen as a nonsingular solution of the system of first order partial differential equations

$$\frac{\partial U}{\partial p_i}(p) = [K_i(p) - \frac{\partial X}{\partial p_i}(p)Y(p)](I - X(p)Y(p))^{-1}U(p), \quad j = 1, \ldots, m.$$

This leads to $A_i = 0$ such that the implementation of the LPV controller does not require any on-line measurements of the rate of the parameter variations. First order partial differential equations can be solved by the method of characteristics [10]. We cannot go into further details at this point.

In order to construct a controller that solves the LPV problem, one has to verify the solvability of the synthesis inequalities in the unknown *functions* X, Y, K_i, L, M, N, and for designing a controller, one has to find functions that solve them.

However, standard algorithms do not allow to solve functional inequalities directly. Hence we need to include a discussion of how to reduce these functional inequalities to finitely many LMI's in real variables.

First step. Since $q \in \dot{\boldsymbol{\delta}}_c$ enters the inequality (9.1.10) affinely, we can replace the set $\dot{\boldsymbol{\delta}}_c$, if convex, by its extreme points. Let us make the, in practice non-restrictive, assumption that this set has finitely many generators:

$$\dot{\boldsymbol{\delta}}_c = \text{conv}\{\dot{\delta}^1, \ldots, \dot{\delta}^k\}.$$

Solving (9.1.9)-(9.1.10) over $(p, q) \in \boldsymbol{\delta}_c \times \dot{\boldsymbol{\delta}}_c$ is equivalent to solving (9.1.9)-(9.1.10) for

$$p \in \boldsymbol{\delta}_c, \quad q \in \{\dot{\delta}^1, \ldots, \dot{\delta}^k\}. \tag{9.1.13}$$

Second step. Instead of searching over the set of all continuous functions, we restrict the search to a finite dimensional subspace thereof, as is standard in Ritz-Galerkin techniques. Let us hence choose basis functions

$$f_1(.), \ldots, f_l(.) \ \text{ that are continuously differentiable on } \ \boldsymbol{\delta}_c.$$

Then all the functions to be found are assumed to belong to the subspace spanned by the functions f_j. This leads to the Ansatz

$$X(p) = \sum_{j=1}^{l} X_j f_j(p), \quad Y(p) = \sum_{j=1}^{l} Y_j f_j(p)$$

$$K_i(p) = \sum_{j=1}^{l} K_j^i f_j(p), \quad i = 0, 1, \ldots, m,$$

$$L(p) = \sum_{j=1}^{l} L_j f_j(p), \quad M(p) = \sum_{j=1}^{l} M_j f_j(p), \quad N(p) = \sum_{j=1}^{l} N_j f_j(p).$$

We observe

$$\partial X(p, q) = \sum_{j=1}^{l} X_j \, \partial f_j(p, q), \quad \partial Y(p, q) = \sum_{j=1}^{l} Y_j \, \partial f_j(p, q).$$

If we plug these formulas into the inequalities (9.1.9)-(9.1.10), we observe that all the coefficient matrices enter affinely. After this substitution, (9.1.9)-(9.1.10) turns out to be a family of linear matrix inequalities in the

$$\text{matrix variables } \ X_j, Y_j, K_j^i, L_j, M_j, N_j$$

that is parameterized by (9.1.13). The variables of this system of LMI's are now real numbers; however, since the parameter p still varies in the infinite set $\boldsymbol{\delta}_c$, we have to solve infinitely many LMI's. This is, in fact, a so-called semi-infinite (not infinite dimensional as often claimed) convex optimization problem.

Third step. To reduce the semi-infinite system of LMI's to finitely many LMI's, the presently chosen route is to just fix a *finite subset*

$$\boldsymbol{\delta}_{\text{finite}} \subset \boldsymbol{\delta}_c$$

and solve the LMI system in those points only. Hence the resulting family of LMI's is parameterized by

$$p \in \boldsymbol{\delta}_{\text{finite}} \text{ and } q \in \{\dot{\delta}^1, \ldots, \dot{\delta}^k\}.$$

We end up with a finite family of linear matrix inequalities in real valued unknowns that can be solved by standard algorithms. Since a systematic choice of points $\boldsymbol{\delta}_{\text{finite}}$ is obtained by gridding the parameter set, this last step is often called the gridding phase, and the whole procedure is said to be a gridding technique.

Remark on the second step. Due to Weierstraß' approximation theorem, one can choose a sequence of functions $f_1, f_2, \ldots$ on $\boldsymbol{\delta}_c$ such that the union of the subspaces

$$\mathcal{S}_\nu = \text{span}\{f_1, \ldots, f_\nu\}$$

is *dense* in the set of all continuously differentiable mappings on $\boldsymbol{\delta}_c$ with respect to the norm

$$\|f\| = \max\{|f(p)| \mid p \in \boldsymbol{\delta}_c\} + \sum_{j=1}^{m} \max\{|\frac{\partial f}{\partial p_j}(p)| \mid p \in \boldsymbol{\delta}_c\}.$$

This implies that, given any continuously differentiable g on $\boldsymbol{\delta}_c$ and any accuracy level $\epsilon > 0$, one can find an index ν_0 such that there exists an $f \in \mathcal{S}_{\nu_0}$ for which

$$\forall\, p \in \boldsymbol{\delta}_c,\ q \in \dot{\boldsymbol{\delta}}_c:\ \ |g(p) - f(p)| \leq \epsilon,\ \ |\partial g(p,q) - \partial f(p,q)| \leq \epsilon.$$

(Provide the details.) Functions in the subspace $\mathcal{S}_\nu$ hence approximate any function g and its image ∂g under the differential operator ∂ up to arbitrary accuracy, if the index ν is chosen sufficiently large.

Therefore, if (9.1.9)-(9.1.10) viewed as functional inequalities do have a solution, then they have a solution if restricting the search over the finite dimensional subspace $\mathcal{S}_\nu$ for sufficiently large ν, i.e., if incorporating sufficiently many basis functions. However, the number of basis functions determines the number of variables in the resulting LMI problem. To keep the number of unknowns small requires an efficient choice of the basis functions what is, in theory and practice, a difficult problem for which one can hardly give any general recipes.

Remark on the third step. By compactness of $\boldsymbol{\delta}_c$ and continuity of all functions, solving the LMI's for $p \in \boldsymbol{\delta}_c$ or for $p \in \boldsymbol{\delta}_{\text{finite}}$ is equivalent if only the points are chosen sufficiently dense. A measure of density is the infimal ϵ such that the balls of radius ϵ around each of the finitely many points in $\boldsymbol{\delta}_{\text{finite}}$ already cover $\boldsymbol{\delta}_c$:

$$\boldsymbol{\delta}_c \subset \bigcup_{p_0 \in \boldsymbol{\delta}_{\text{finite}}} \{u \mid \|p - p_0\| \leq \epsilon\}.$$

If the data functions describing the system are also differentiable in δ, one can apply the mean value theorem to provide explicit estimates of the accuracy of the required approximation. Again, however, it is important to observe that the number of LMI's to solve depends on the number of grid-points; hence one has to keep this number small in order to avoid large LMI's.

Remark on extensions. Only slight adaptations are required to treat all the other performance specifications (such as bounds on the L_2-gain and on the analogue of the H_2-norm or generalized H_2-norm for time-varying systems) as well as the corresponding mixed problems as discussed in Chapter **??** in full generality. Note also that, for single-objective problems, the techniques to eliminate parameters literally apply; there is no need go into the details. In particular for solving gain-scheduling problems, it is important to observe that one can as well let the performance index depend on the measured parameter without any additional difficulty. As a designer, one can hence ask *different performance properties* in *different parameter ranges* what has considerable relevance in practical controller design.

Remark on robust LPV control. As another important extension we mention *robust LPV design*. It might happen that some parameters are indeed on-line measurable, whereas others have to be considered as unknown perturbations with which the controller cannot be scheduled. Again, it is straightforward to extend the robustness design techniques that have been presented in Chapter **??** from LTI systems and controllers to LPV systems and controllers. This even allows to include dynamic uncertainties if using IQC's to capture their properties. Note that the scalings that appear in such techniques constitute extra problem variables. In many circumstances it causes no extra technical difficulties to let these scalings also depend on the scheduling parameter what reduces the conservatism.

9.2 Affine Parameter Dependence

Suppose that the matrices (9.1.1) describing the system are *affine* functions on the set

$$\boldsymbol{\delta}_c = \mathrm{conv}\{\delta^1, \ldots, \delta^k\}.$$

In that case we intend to search, as well, for an LPV controller that is defined with *affine* functions (9.1.3). Note that the describing matrices for the cosed-loop system are also *affine* in the parameter if

$$\begin{pmatrix} B \\ E \end{pmatrix} \text{ and } \begin{pmatrix} C & F \end{pmatrix} \text{ are parameter independent}$$

what is assumed from now on. Finally, we let $\mathcal{X}$ in Theorem 9.2 be *constant*.

Since $R_p \geq 0$, we infer that (9.1.7) is satisfied if and only if it holds for the generators $p = \delta^j$ of the set $\boldsymbol{\delta}_c$. Therefore, the analysis inequalities reduce to the finite set of LMI's

$$\mathcal{X} > 0, \ \begin{pmatrix} \mathcal{A}(\delta^j)^T\mathcal{X} + \mathcal{X}\mathcal{A}(\delta^j) & \mathcal{X}\mathcal{B}(\delta^j) \\ \mathcal{B}(\delta^j)^T\mathcal{X} & 0 \end{pmatrix} +$$
$$+ \begin{pmatrix} 0 & I \\ \mathcal{C}(\delta^j) & \mathcal{D}(\delta^j) \end{pmatrix}^T P_p \begin{pmatrix} 0 & I \\ \mathcal{C}(\delta^j) & \mathcal{D}(\delta^j) \end{pmatrix} < 0 \ \text{ for all } \ j = 1, \ldots, k.$$

Under the present structural assumptions, the affine functions $\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}$ are transformed into affine functions $\begin{pmatrix} K & L \\ M & N \end{pmatrix}$ under the controller parameter transformation as considered in the previous section.

Then the synthesis inequalities (9.1.9)-(9.1.10) whose variables are the constant X and Y and the affine functions $\begin{pmatrix} K & L \\ M & N \end{pmatrix}$ turn out to be *affine* in the parameter p. This implies for the synthesis inequalities that we can replace the search over $\boldsymbol{\delta}_c$ without loss of generality by the search over the generators δ^j of this set. Therefore, solving the design problem amounts to testing whether the LMI's

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

and

$$\left(\begin{array}{cc|c} \operatorname{sym}\left(A(\delta^j)Y + BM(\delta^j)\right) & * & * \\ (A(\delta^j) + BN(\delta^j)C)^T + K(\delta^j) & \operatorname{sym}\left(A(\delta^j)X + L(\delta^j)C\right) & * \\ \hline (B_p(\delta^j) + BN(\delta^j)F)^T & (XB_p(\delta^j) + L(\delta^j)F)^T & 0 \end{array}\right) +$$
$$+ \begin{pmatrix} * \\ * \end{pmatrix}^T P_p \left(\begin{array}{cc|c} 0 & 0 & I \\ C_p(\delta^j)Y + EM(\delta^j) & C_p(\delta^j) + EN(\delta^j)C & D_p(\delta^j) + EN(\delta^j)F \end{array}\right) < 0$$

for $j = 1, \ldots, k$ admit a solution.

Since affine, the functions K, L, M, N are parameterized as

$$\begin{pmatrix} K(p) & L(p) \\ M(p) & N(p) \end{pmatrix} = \begin{pmatrix} K_0 & L_0 \\ M_0 & N_0 \end{pmatrix} + \sum_{i=1}^{m} \begin{pmatrix} K_i & L_i \\ M_i & N_i \end{pmatrix} p_i$$

with real matrices K_i, L_i, M_i, N_i. Hence, the synthesis inequalities form genuine linear matrix inequalities that can be solved by standard algorithms.

9.3 LFT System Description

Similarly as for our discussion of robust controller design, let us assume in this section that the LPV system is described as and LTI system

$$\begin{pmatrix} \dot{x} \\ \hline z_u \\ z_p \\ y \end{pmatrix} = \left(\begin{array}{c|ccc} A & B_u & B_p & B \\ \hline C_u & D_{uu} & D_{up} & E_u \\ C_p & D_{pu} & D_{pp} & E_p \\ C & F_u & F_p & 0 \end{array}\right) \begin{pmatrix} x \\ \hline w_u \\ w_p \\ u \end{pmatrix} \tag{9.3.1}$$

in wich the parameter enters via the uncertainty channel $w_u \to z_u$ as

$$w_u(t) = \Delta(t) z_u(t), \quad \Delta(t) \in \mathbf{\Delta}_c. \tag{9.3.2}$$

The size and the structure of the possible parameter values $\Delta(t)$ is captured by the convex set

$$\mathbf{\Delta}_c := \mathrm{conv}\{\Delta_1, ..., \Delta_N\}$$

whose generators Δ_j are given explicitly. We assume w.l.o.g. that $0 \in \mathbf{\Delta}_c$. As before, we concentrate on the quadratic performance specification with index P_p imposed on the performance channel $w_p \to z_p$.

Adjusted to the structure of (9.3.1)-(9.3.2), we assume that the measured parameter curve enters the controller also in a linear fractional fashion. Therefore, we assume that the to-be-designed LPV controller is defined by scheduling the LTI system

$$\dot{x}_c = A_c x_c + B_c \begin{pmatrix} y \\ w_c \end{pmatrix}, \quad \begin{pmatrix} u \\ z_c \end{pmatrix} = C_c x_c + D_c \begin{pmatrix} y \\ w_c \end{pmatrix} \tag{9.3.3}$$

with the actual parameter curve entering as

$$w_c(t) = \Delta_c(\Delta(t)) z_c(t). \tag{9.3.4}$$

The LPV controller is hence parameterized through the matrices A_c, B_c, C_c, D_c, and through a possibly non-linear matrix-valued scheduling function

$$\Delta_c(\Delta) \in \mathbb{R}^{n_r \times n_c} \text{ defined on } \mathbf{\Delta}_c.$$

Figure 9.1 illustrates this configuration.

The goal is to construct an LPV controller such that, for all admissible parameter curves, the controlled system is exponentially stable and, the quadratic performance specification with index P_p for the channel $w_p \to z_p$ is satisfied.

The solution of this problem is approached with a simple trick. In fact, the controlled system can, alternatively, be obtained by scheduling the LTI system

$$\begin{pmatrix} \dot{x} \\ \hline z_u \\ z_c \\ \hline z_p \\ \hline y \\ w_c \end{pmatrix} = \left(\begin{array}{c|cc|c|cc} A & B_u & 0 & B_p & B & 0 \\ \hline C_u & D_{uu} & 0 & D_{up} & E_u & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{n_c} \\ \hline C_p & D_{pu} & 0 & D_{uu} & E_p & 0 \\ \hline C & F_u & 0 & F_p & 0 & 0 \\ 0 & 0 & I_{n_r} & 0 & 0 & 0 \end{array}\right) \begin{pmatrix} x \\ \hline w_u \\ w_c \\ \hline w_p \\ \hline u \\ z_c \end{pmatrix} \tag{9.3.5}$$

with the parameter as

$$\begin{pmatrix} w_1 \\ w_c \end{pmatrix} = \begin{pmatrix} \Delta(t) & 0 \\ 0 & \Delta_c(\Delta(t)) \end{pmatrix} \begin{pmatrix} z_1 \\ z_c \end{pmatrix}, \tag{9.3.6}$$

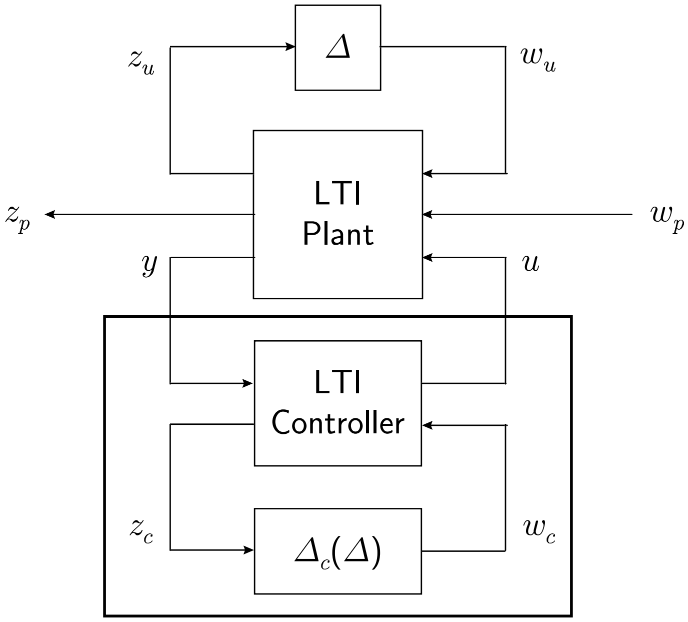


Figure 9.1: LPV system and LPV controller with LFT description

and then controlling this parameter dependent system with the LTI controller (9.3.3). Alternatively, we can interconnect the LTI system (9.3.5) with the LTI controller (9.3.3) to arrive at the LTI system

$$\begin{pmatrix} \dot{x} \\ \hline z_u \\ z_c \\ z_p \end{pmatrix} = \left(\begin{array}{c|ccc} \mathcal{A} & \mathcal{B}_u & \mathcal{B}_c & \mathcal{B}_p \\ \hline \mathcal{C}_u & \mathcal{D}_{uu} & \mathcal{D}_{uc} & \mathcal{D}_{up} \\ \mathcal{C}_c & \mathcal{D}_{cu} & \mathcal{D}_{cc} & \mathcal{D}_{cp} \\ \mathcal{C}_p & \mathcal{D}_{pu} & \mathcal{D}_{pc} & \mathcal{D}_{pp} \end{array}\right) \begin{pmatrix} x \\ \hline w_u \\ w_c \\ w_p \end{pmatrix}, \tag{9.3.7}$$

and then re-connect the parameter as (9.3.6). This latter interconnection order is illustrated in Figure 9.2.

Note that (9.3.5) is an extension of the original system (9.3.1) with an additional uncertainty channel $w_c \to z_c$ and with an additional control channel $z_c \to w_c$; the number n_r and n_c of the components of w_c and z_c dictate the size of the identity matrices I_{n_r} and I_{n_c} that are indicated by their respective indices.

Once the scheduling function $\Delta_c(\Delta)$ has been fixed, it turns out that (9.3.3) is a *robust controller* for the system (9.3.5) with uncertainty (9.3.6). The genuine robust control problem in which the parameter is not measured on-line would relate to the situation that $n_r = 0$ and $n_c = 0$ such that (9.3.5) and (9.3.1) are identical. In LPV control we have the extra freedom of being able to first extend the system as in (9.3.5) and design for this extended system a robust controller. It will turn out that this extra freedom will render the corresponding synthesis inequalities convex.

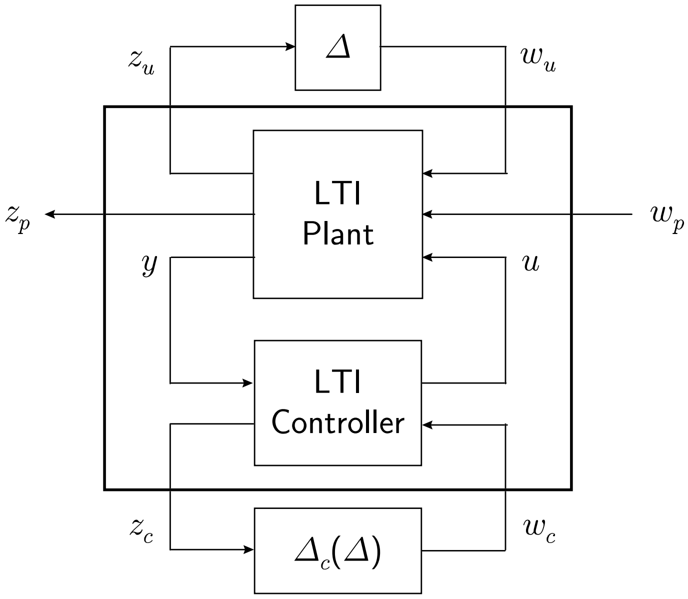


Figure 9.2: LPV system and LPV controller: alternative interpretation

Before we embark on a solution of the LPV problem, let us include some further comments on the corresponding genuine robust control problem. We have seen in section 8.1.1 that the search for a robust controller leads to the problem of having to solve the matrix inequalities

$$
\boldsymbol{X}(v) > 0, \quad \left(\begin{array}{c} * \\ * \\ \hline * \\ * \\ \hline * \\ * \end{array}\right)^T \left(\begin{array}{cc|cc|cc} 0 & I & 0 & 0 & 0 & 0 \\ I & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & Q & S & 0 & 0 \\ 0 & 0 & S^T & R & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & Q_p & S_p \\ 0 & 0 & 0 & 0 & S_p^T & R_p \end{array}\right) \left(\begin{array}{ccc} I & 0 & 0 \\ \boldsymbol{A}(v) & \boldsymbol{B}_u(v) & \boldsymbol{B}_p(v) \\ \hline 0 & I & 0 \\ \boldsymbol{C}_u(v) & \boldsymbol{D}_{uu}(v) & \boldsymbol{D}_{up}(v) \\ \hline 0 & 0 & I \\ \boldsymbol{C}_p(v) & \boldsymbol{D}_{pu}(v) & \boldsymbol{D}_{pp}(v) \end{array}\right) < 0
$$

$$
\begin{pmatrix} \Delta \\ I \end{pmatrix}^T \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \begin{pmatrix} \Delta \\ I \end{pmatrix} > 0 \text{ for all } \Delta \in \boldsymbol{\Delta}_c
$$

in the parameter v and in the multiplier $P = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}$.

Recall from our earlier discussion that one of the difficulties is a numerical tractable parameterization of the set of multipliers. This was the reason to introduce, at the expense of conservatism, the following subset of multipliers that admits a description in terms of finitely many LMI's:

$$
\boldsymbol{P} := \left\{ P = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \mid Q < 0, \begin{pmatrix} \Delta_j \\ I \end{pmatrix}^T P \begin{pmatrix} \Delta_j \\ I \end{pmatrix} > 0 \text{ for } j = 1, \ldots, N \right\}. \qquad (9.3.8)
$$

Even after confining the search to v and $P \in \boldsymbol{P}$, no technique is known how to solve the resulting still non-convex synthesis inequalities by standard algorithms.

In contrast to what we have seen for state-feedback design, the same is true of the dual inequalities that read as

$$\boldsymbol{X}(v) > 0, \quad \left(\begin{array}{c} * \\ * \\ \hline * \\ * \\ \hline * \\ * \end{array}\right)^T \left(\begin{array}{cc|cc|cc} 0 & X & 0 & 0 & 0 & 0 \\ X & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \tilde{Q} & \tilde{S} & 0 & 0 \\ X & 0 & \tilde{S}^T & \tilde{R} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \tilde{Q}_p & \tilde{S}_p \\ 0 & 0 & 0 & 0 & \tilde{S}_p^T & \tilde{R}_p \end{array}\right) \left(\begin{array}{ccc} -\boldsymbol{A}(v)^T & -\boldsymbol{C}_u(v)^T & -\boldsymbol{C}_p(v)^T \\ I & 0 & 0 \\ \hline -\boldsymbol{B}_u(v)^T & -\boldsymbol{D}_{uu}(v)^T & -\boldsymbol{D}_{pu}(v)^T \\ 0 & I & 0 \\ \hline -\boldsymbol{B}_p(v)^T & -\boldsymbol{D}_{up}(v)^T & -\boldsymbol{D}_{pp}(v)^T \\ 0 & 0 & I \end{array}\right) > 0$$

$$\begin{pmatrix} I \\ -\Delta^T \end{pmatrix}^T \begin{pmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{pmatrix} \begin{pmatrix} I \\ -\Delta^T \end{pmatrix} < 0 \text{ for all } \Delta \in \boldsymbol{\Delta}_c.$$

Again, even confining the search to the set of multipliers

$$\tilde{\boldsymbol{P}} := \left\{ \tilde{P} = \begin{pmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{pmatrix} \mid \tilde{R} > 0, \begin{pmatrix} I \\ -\Delta_j^T \end{pmatrix}^T \tilde{P} \begin{pmatrix} I \\ -\Delta_j^T \end{pmatrix} < 0 \text{ for } j = 1, \ldots, N \right\} \tag{9.3.9}$$

does not lead to a convex feasibility problem.

Since non-convexity is caused by the multiplication of functions that depend on v with the multipliers, one could be lead to the idea that it might help to eliminate as many of the variables that are involved in v as possible. We can indeed apply the technique exposed in Section 4.5.3 and eliminate K, L, M, N.

For that purpose one needs to compute basis matrices

$$\Phi = \begin{pmatrix} \Phi^1 \\ \Phi^2 \\ \Phi^3 \end{pmatrix} \text{ of } \ker \begin{pmatrix} B^T & E_u^T & E_p^T \end{pmatrix} \quad \text{and} \quad \Psi = \begin{pmatrix} \Psi^1 \\ \Psi^2 \\ \Psi^3 \end{pmatrix} \text{ of } \ker \begin{pmatrix} C & F_u & F_p \end{pmatrix}$$

respectively. After elimination, the synthesis inequalities read as

$$\begin{pmatrix} Y & I \\ I & X \end{pmatrix} > 0, \tag{9.3.10}$$

$$\Psi^T \left(\begin{array}{ccc} I & 0 & 0 \\ A & B_u & B_p \\ \hline 0 & I & 0 \\ C_u & D_{uu} & D_{up} \\ \hline 0 & 0 & I \\ C_p & D_{pu} & D_{pp} \end{array}\right)^T \left(\begin{array}{cc|cc|cc} 0 & X & 0 & 0 & 0 & 0 \\ X & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & Q & S & 0 & 0 \\ 0 & 0 & S^T & R & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & Q_p & S_p \\ 0 & 0 & 0 & 0 & S_p^T & R_p \end{array}\right) \left(\begin{array}{ccc} I & 0 & 0 \\ A & B_u & B_p \\ \hline 0 & I & 0 \\ C_u & D_{uu} & D_{up} \\ \hline 0 & 0 & I \\ C_p & D_{pu} & D_{pp} \end{array}\right) \Psi < 0, \tag{9.3.11}$$

$$\Phi^T \left(\begin{array}{ccc} -A^T & -C_u^T & -C_p^T \\ I & 0 & 0 \\ \hline -B_u^T & -D_{uu}^T & -D_{pu}^T \\ 0 & I & 0 \\ \hline -B_p^T & -D_{pu}^T & -D_{pp}^T \\ 0 & 0 & I \end{array}\right)^T \left(\begin{array}{cc|cc|cc} 0 & X & 0 & 0 & 0 & 0 \\ X & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \tilde{Q} & \tilde{S} & 0 & 0 \\ 0 & 0 & \tilde{S}^T & \tilde{R} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \tilde{Q}_p & \tilde{S}_p \\ 0 & 0 & 0 & 0 & \tilde{S}_p^T & \tilde{R}_p \end{array}\right) \left(\begin{array}{ccc} -A^T & -C_u^T & -C_p^T \\ I & 0 & 0 \\ \hline -B_u^T & -D_{uu}^T & -D_{pu}^T \\ 0 & I & 0 \\ \hline -B_p^T & -D_{pu}^T & -D_{pp}^T \\ 0 & 0 & I \end{array}\right) \Phi > 0 \tag{9.3.12}$$

in the variables X, Y, and in the multiplier P and $\tilde{P}$ that are coupled as

$$\tilde{P} = \begin{pmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{pmatrix} = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}^{-1} = P^{-1}. \tag{9.3.13}$$

Hence, after elimination, it turns out that the inequalities (9.3.10)-(9.3.12) are indeed affine in the unknowns X, Y, P and $\tilde{P}$. Unfortunately, non-convexity re-appears through the coupling (9.3.13) of the multipliers P and $\tilde{P}$.

Let us now turn back to the LPV problem where we allow, via the scheduling function $\Delta_c(\Delta)$ in the controller, extra freedom in the design process.

For guaranteeing stability and performance of the controlled system, we employ *extended* multipliers adjusted to the extended uncertainty structure (9.3.6) that are given as

$$P_e = \left(\begin{array}{c|c} Q_e & S_e \\ \hline S_e^T & R_e \end{array}\right) = \left(\begin{array}{cc|cc} Q & Q_{12} & S & S_{12} \\ Q_{21} & Q_{22} & S_{21} & S_{22} \\ \hline * & * & R & R_{12} \\ * & * & R_{21} & R_{22} \end{array}\right) \text{ with } Q_e < 0, \quad R_e > 0 \tag{9.3.14}$$

and that satisfy

$$\left(\begin{array}{cc} \Delta & 0 \\ 0 & \Delta_c(\Delta) \\ \hline I & 0 \\ 0 & I \end{array}\right) P_e \left(\begin{array}{cc} \Delta & 0 \\ 0 & \Delta_c(\Delta) \\ \hline I & 0 \\ 0 & I \end{array}\right) > 0 \text{ for all } \Delta \in \mathbf{\Delta}. \tag{9.3.15}$$

The corresponding dual multipliers $\tilde{P}_e = P_e^{-1}$ are partitioned similarly as

$$\tilde{P}_e = \left(\begin{array}{c|c} \tilde{Q}_e & \tilde{S}_e \\ \hline \tilde{S}_e^T & \tilde{R}_e \end{array}\right) = \left(\begin{array}{cc|cc} \tilde{Q} & \tilde{Q}_{12} & \tilde{S} & \tilde{S}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} & \tilde{S}_{21} & \tilde{S}_{22} \\ \hline * & * & \tilde{R} & \tilde{R}_{12} \\ * & * & \tilde{R}_{21} & \tilde{R}_{12} \end{array}\right) \text{ with } \tilde{Q}_e < 0, \quad \tilde{R}_e > 0 \tag{9.3.16}$$

and they satisfy

$$\left(\begin{array}{cc} I & 0 \\ 0 & I \\ \hline -\Delta^T & 0 \\ 0 & -\Delta_c(\Delta)^T \end{array}\right)^T P_e \left(\begin{array}{cc} I & 0 \\ 0 & I \\ \hline -\Delta^T & 0 \\ 0 & -\Delta_c(\Delta)^T \end{array}\right) > 0 \text{ for all } \Delta \in \mathbf{\Delta}.$$

As indicated by our notation, we observe that

$$\begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \in \boldsymbol{P} \ \text{ and } \ \begin{pmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{pmatrix} \in \tilde{\boldsymbol{P}}$$

for the corresponding sub-matrices of P_e and $\tilde{P}_e$ respectively.

If we recall the description (9.3.6)-(9.3.7) of the controlled LPV system, the desired exponential stability and quadratic performance property is satisfied if we can find a Lyapunov matrix $\mathcal{X}$ and an extended scaling P_e with (9.3.14)-(9.3.15) such that

$$\mathcal{X} > 0, \quad \left(\begin{array}{c} * \\ * \\ \hline * \\ * \\ * \\ * \\ \hline * \\ * \end{array}\right)^T \left(\begin{array}{cc|cccc|cc} 0 & \mathcal{X} & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathcal{X} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & Q & Q_{12} & S & S_{12} & 0 & 0 \\ 0 & 0 & Q_{21} & Q_{22} & S_{21} & S_{22} & 0 & 0 \\ 0 & 0 & * & * & R & R_{12} & 0 & 0 \\ 0 & 0 & * & * & R_{21} & R_{22} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & Q_p & S_p \\ 0 & 0 & 0 & 0 & 0 & 0 & S_p^T & R_p \end{array}\right) \left(\begin{array}{cccc} I & 0 & 0 & 0 \\ \mathcal{A} & \mathcal{B}_u & \mathcal{B}_c & \mathcal{B}_p \\ \hline 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ \mathcal{C}_u & \mathcal{D}_{uu} & \mathcal{D}_{uc} & \mathcal{D}_{up} \\ \mathcal{C}_c & \mathcal{D}_{cu} & \mathcal{D}_{cc} & \mathcal{D}_{cp} \\ \hline 0 & 0 & 0 & I \\ \mathcal{C}_p & \mathcal{D}_{pu} & \mathcal{D}_{pc} & \mathcal{D}_{pp} \end{array}\right) < 0. \tag{9.3.17}$$

We are now ready to formulate an LMI test for the existence of an LPV controller such that the controlled LPV system fulfills this latter analysis test.

Theorem 9.4 *The following statements are equivalent:*

(a) There exists a controller (9.3.3) *and a scheduling function* $\Delta_c(\Delta)$ *such that the controlled system as described by* (9.3.4)-(9.3.7) *admits a Lyapunov matrix* $\mathcal{X}$ *and a multiplier* (9.3.14)-(9.3.15) *that satisfy* (9.3.17).

(b) There exist X, Y *and multipliers* $P \in \boldsymbol{P}$, $\tilde{P} \in \tilde{\boldsymbol{P}}$ *that satisfy the linear matrix inequalities* (9.3.10)-(9.3.12).

Proof. Let us first prove $1 \Rightarrow 2$. We can apply the technique as described in Section 4.5.3 to eliminate the controller parameters in the inequality (9.3.17). According to Corollary 4.15, this leads to the coupling condition (4.5.24) and to the two synthesis inequalities (4.5.25)-(4.5.26). The whole point is to show that the latter two inequalities can indeed be simplified to (9.3.11)-(9.3.12). Let us illustrate this simplification for the first inequality only since a duality argument leads to the same conclusions for the second one.

With

$$\Psi_e = \begin{pmatrix} \Psi^1 \\ \Psi^2 \\ 0 \\ \Psi^3 \end{pmatrix} \ \text{ as a basis matrix of } \ \ker \begin{pmatrix} C & F_u & 0 & F_p \\ 0 & 0 & I_{n_r} & 0 \end{pmatrix},$$

the inequality that corresponds to (4.5.24) reads as

$$\Psi_e^T\left(\begin{array}{c} * \\ * \\ \hline * \\ * \\ * \\ * \\ \hline * \\ * \end{array}\right)^T\left(\begin{array}{cc|cccc|cc} 0 & X & 0 & 0 & 0 & 0 & 0 & 0 \\ X & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & Q & Q_{12} & S & S_{12} & 0 & 0 \\ 0 & 0 & Q_{21} & Q_{22} & S_{21} & S_{22} & 0 & 0 \\ 0 & 0 & * & * & R & R_{12} & 0 & 0 \\ 0 & 0 & * & * & R_{21} & R_{22} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & Q_p & S_p \\ 0 & 0 & 0 & 0 & 0 & 0 & S_p^T & R_p \end{array}\right)\left(\begin{array}{cccc} I & 0 & 0 & 0 \\ A & B_u & 0 & B_p \\ \hline 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ C_u & D_{uu} & 0 & D_{up} \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & I \\ C_p & D_{pu} & D_{pc} & D_{pp} \end{array}\right)\Psi_e < 0.$$

Due to the zero block in Ψ_e, it is obvious that this is the same as

$$\Psi^T\left(\begin{array}{c} * \\ * \\ \hline * \\ * \\ * \\ * \\ \hline * \\ * \end{array}\right)^T\left(\begin{array}{cc|cccc|cc} 0 & X & 0 & 0 & 0 & 0 & 0 & 0 \\ X & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & Q & Q_{12} & S & S_{12} & 0 & 0 \\ 0 & 0 & Q_{21} & Q_{22} & S_{21} & S_{22} & 0 & 0 \\ 0 & 0 & * & * & R & R_{12} & 0 & 0 \\ 0 & 0 & * & * & R_{21} & R_{22} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & Q_p & S_p \\ 0 & 0 & 0 & 0 & 0 & 0 & S_p^T & R_p \end{array}\right)\left(\begin{array}{ccc} I & 0 & 0 \\ A & B_u & B_p \\ \hline 0 & I & 0 \\ 0 & 0 & 0 \\ C_u & D_{uu} & D_{up} \\ 0 & 0 & 0 \\ \hline 0 & 0 & I \\ C_p & D_{pu} & D_{pp} \end{array}\right)\Psi < 0.$$

The two zero block rows in the outer factors allow to simplify this latter inequality to (9.3.11), what finishes the proof of $1 \Rightarrow 2$.

The constructive proof of $2 \Rightarrow 1$ is more involved and proceeds in three steps. Let us assume that we have computed solutions X, Y and $P \in \boldsymbol{P}$, $\tilde{P} \in \tilde{\boldsymbol{P}}$ with (9.3.10)-(9.3.12).

First step: Extension of Scalings. Since $P \in \boldsymbol{P}$ and $\tilde{P} \in \tilde{\boldsymbol{P}}$, let us recall that we have

$$\left(\begin{array}{c} \Delta \\ I \end{array}\right)^T P\left(\begin{array}{c} \Delta \\ I \end{array}\right) > 0 \text{ and } \left(\begin{array}{c} I \\ -\Delta^T \end{array}\right)^T \tilde{P}\left(\begin{array}{c} I \\ -\Delta^T \end{array}\right) < 0 \text{ for all } \Delta \in \boldsymbol{\Delta}. \tag{9.3.18}$$

Due to $0 \in \boldsymbol{\Delta}_c$, we get $R > 0$ and $\tilde{Q} < 0$. Hence we conclude for the diagonal blocks of P that $Q < 0$ and $R > 0$, and for the diagonal blocks of $\tilde{P}$ that $\tilde{Q} > 0$ and $\tilde{R} < 0$. If we introduce

$$Z = \left(\begin{array}{c} I \\ 0 \end{array}\right) \text{ and } \tilde{Z} = \left(\begin{array}{c} 0 \\ I \end{array}\right)$$

with the same row partition as P, these properties can be expressed as

$$Z^T P Z < 0,\ \tilde{Z}^T P \tilde{Z} > 0 \text{ and } Z^T \tilde{P} Z < 0,\ \tilde{Z}^T \tilde{P} \tilde{Z} > 0. \tag{9.3.19}$$

If we observe that $\text{im}(\tilde{Z})$ is the orthogonal complement of $\text{im}(Z)$, we can apply the Dualization Lemma to infer

$$\tilde{Z}^T P^{-1} \tilde{Z} > 0,\ Z^T P^{-1} Z < 0 \text{ and } \tilde{Z}^T \tilde{P}^{-1} \tilde{Z} > 0,\ Z^T \tilde{P}^{-1} Z < 0. \tag{9.3.20}$$

For the given P and $\tilde{P}$, we try to find an extension P_e with (9.3.14) such that the dual multiplier $\tilde{P}_e = P_e^{-1}$ is related to the given $\tilde{P}$ as in (9.3.16). After a suitable permutation, this amounts to finding an extension

$$\begin{pmatrix} P & T \\ T^T & T^T N T \end{pmatrix} \text{ with } \begin{pmatrix} \tilde{P} & * \\ * & * \end{pmatrix} = \begin{pmatrix} P & T \\ T^T & T^T N T \end{pmatrix}^{-1}, \tag{9.3.21}$$

where the specific parameterization of the new blocks in terms of a non-singular matrix T and some symmetric N will turn out convenient. Such an extension is very simple to obtain. However, we also need to obey the positivity/negativity constraints in (9.3.14) that amount to

$$\begin{pmatrix} Z & 0 \\ 0 & Z \end{pmatrix}^T \begin{pmatrix} P & T \\ T^T & T^T N T \end{pmatrix} \begin{pmatrix} Z & 0 \\ 0 & Z \end{pmatrix} < 0 \tag{9.3.22}$$

and

$$\begin{pmatrix} \tilde{Z} & 0 \\ 0 & \tilde{Z} \end{pmatrix}^T \begin{pmatrix} P & T \\ T^T & T^T N T \end{pmatrix} \begin{pmatrix} \tilde{Z} & 0 \\ 0 & \tilde{Z} \end{pmatrix} > 0. \tag{9.3.23}$$

We can assume w.l.o.g. (perturb, if necessary) that $P - \tilde{P}^{-1}$ is non-singular. Then we set

$$N = (P - \tilde{P}^{-1})^{-1}$$

and observe that (9.3.21) holds for any non-singular T.

The main goal is to adjust T to render (9.3.22)-(9.3.23) satisfied. We will in fact construct the sub-blocks $T_1 = TZ$ and $T_2 = T\tilde{Z}$ of $T = (T_1\ T_2)$. Due to (9.3.19), the conditions (9.3.22)-(9.3.23) read in terms of these blocks as (Schur)

$$T_1^T \left[N - Z(Z^T P Z)^{-1} Z^T\right] T_1 < 0 \text{ and } T_2^T \left[N - \tilde{Z}(\tilde{Z}^T P \tilde{Z})^{-1} \tilde{Z}^T\right] T_2 > 0. \tag{9.3.24}$$

If we denote by $n_+(S)$, $n_-(S)$ the number of positive, negative eigenvalues of the symmetric matrix S, we hence have to calculate $n_-(N - Z(Z^T P Z)^{-1} Z^T)$ and $n_+(N - \tilde{Z}(\tilde{Z}^T P \tilde{Z})^{-1} \tilde{Z}^T)$. Simple Schur complement arguments reveal that

$$\begin{aligned} n_-\begin{pmatrix} Z^T P Z & Z^T \\ Z & N \end{pmatrix} &= n_-(Z^T P Z) + n_-(N - Z(Z^T P Z)^{-1} Z^T) = \\ &= n_-(N) + n_-(Z^T P Z - Z^T N^{-1} Z) = n_-(N) + n_-(Z^T \tilde{P}^{-1} Z). \end{aligned}$$

Since $Z^T P Z$ and $Z^T \tilde{P}^{-1} Z$ have the same size and are both negative definite by (9.3.19) and (9.3.20), we conclude $n_-(Z^T P Z) = n_-(Z^T \tilde{P}^{-1} Z)$. This leads to

$$n_-(N - Z(Z^T P Z)^{-1} Z^T) = n_-(N).$$

Literally the same arguments will reveal

$$n_+(N - \tilde{Z}(\tilde{Z}^T P \tilde{Z})^{-1} \tilde{Z}^T) = n_+(N).$$

These two relations imply that there exist T_1, T_2 with $n_-(N)$, $n_+(N)$ columns that satisfy (9.3.24). Hence the matrix $T = (T_1\ T_2)$ has $n_+(N) + n_-(N)$ columns. Since the number of rows of T_1, T_2, Z, $\tilde{Z}$, N are all identical, T is actually a *square* matrix. We can assume w.l.o.g. - by perturbing T_1 or T_2 if necessary - that the square matrix T is non-singular.

This finishes the construction of the extended multiplier (9.3.14). Let us observe that the dimensions of Q_{22}/R_{22} equal the number of columns of T_1/T_2 which are, in turn, given by the integers $n_-(N)/n_+(N)$.

Second Step: Construction of the scheduling function. Let us fix Δ and let us apply the Elimination Lemma to (9.3.15) with $\Delta_c(\Delta)$ viewed as the unknown. We observe that the solvability conditions of the Elimination Lemma just amount to the two inequalities (9.3.18). We conclude that for any $\Delta \in \mathbf{\Delta}$ one can indeed compute a $\Delta_c(\Delta)$ which satisfies (9.3.15).

Due to the structural simplicity, we can even provide an explicit formula which shows that $\Delta_c(\Delta)$ can be selected to depend smoothly on Δ. Indeed, by a straightforward Schur-complement argument, (9.3.15) is equivalent to

$$\left(\begin{array}{cc|cc} U_{11} & U_{12} & (W_{11}+\Delta)^T & W_{21}^T \\ U_{21} & U_{22} & W_{12}^T & (W_{22}+\Delta_c(\Delta))^T \\ \hline W_{11}+\Delta & W_{12} & V_{11} & V_{12} \\ W_{21} & W_{22}+\Delta_c(\Delta) & V_{21} & V_{22} \end{array}\right) > 0$$

for $U = R_e - S_e^T Q_e^{-1} S_e > 0$, $V = -Q_e^{-1} > 0$, $W = Q_e^{-1} S_e$. Obviously this can be rewritten to

$$\left(\begin{array}{cc} U_{22} & * \\ W_{22}+\Delta_c(\Delta) & V_{22} \end{array}\right) - \left(\begin{array}{cc} U_{21} & W_{12}^T \\ W_{21} & V_{21} \end{array}\right) \left(\begin{array}{cc} U_{11} & (W_{11}+\Delta)^T \\ W_{11}+\Delta & V_{11} \end{array}\right)^{-1} \left(\begin{array}{cc} U_{12} & W_{21}^T \\ W_{12} & V_{12} \end{array}\right) > 0$$

in which $\Delta_c(\Delta)$ only appears in the off-diagonal position. Since we are sure that there does indeed exist a $\Delta_c(\Delta)$ that renders the inequality satisfied, the diagonal blocks must be positive definite. If we then choose $\Delta_c(\Delta)$ such that the off-diagonal block vanishes, we have found a solution of the inequality; this leads to the following explicit formula

$$\Delta_c(\Delta) = -W_{22} + \left(\begin{array}{cc} W_{21} & V_{21} \end{array}\right) \left(\begin{array}{cc} U_{11} & * \\ W_{11}+\Delta & V_{11} \end{array}\right)^{-1} \left(\begin{array}{c} U_{12} \\ W_{12} \end{array}\right)$$

for the scheduling function. We note that $\Delta_c(\Delta)$ has the dimension $n_-(N) \times n_+(N)$.

Third Step: LTI controller construction. After having constructed the scalings, the last step is to construct an LTI controller and Lyapunov matrix that render the inequality (9.3.17) satisfied. We are confronted with a standard nominal quadratic design problem of which we are sure that it admits a solution, and for which the controller construction proceed along the steps that have been intensively discussed in Chapter **??**. ■

We have shown that the LMI's that needed to be solved for designing an LPV controller are *identical* to those for designing a robust controller, with the only exception that the coupling condition

(9.3.13) drops out. Therefore, the search for X and Y and for the multipliers $P \in \boldsymbol{P}$ and $\tilde{P} \in \tilde{\boldsymbol{P}}$ to satisfy (9.3.10)-(9.3.12) amounts to testing the feasibility of standard LMI's. Moreover, the controller construction in the proof of Theorem 9.4 is constructive. Hence we conclude that we have found a full solution to the quadratic performance LPV control problem (including L_2-gain and dissipativity specifications) for full block scalings P_e that satisfy $Q_e < 0$. The more interesting general case without this still restrictive negativity hypotheses is dealt with in future work.

Remarks.

- The proof reveals that the scheduling function $\Delta_c(\Delta)$ has a many rows/colums as there are negative/positive eigenvalues of $P - \tilde{P}^{-1}$ (if assuming w.l.o.g. that the latter is non-singular.) If it happens that $P - \tilde{P}^{-1}$ is positive or negative definite, there is no need to schedule the controller at all; we obtain a controller that solves the robust quadratic performance problem.

- Previous approaches to the LPV problem [2,6,21,37] were based on $\Delta_c(\Delta) = \Delta$ such that the controller is scheduled with an identical copy of the parameters. These results were based on block-diagonal parameter matrices and multipliers that were as well assumed block-diagonal. The use of full block scalings [32] require the extension to a more general scheduling function that is - as seen a posteriori - a quadratic function of the parameter Δ.

- It is possible to extend the procedure to H_2-control and to the other performance specifications in these notes. However, this requires restrictive hypotheses on the system description. The extension to general mixed problems seems nontrivial and is open in its full generality.

9.4 A Sketch of Possible Applications

It is obvious how to apply robust or LPV control techniques in linear design: If the underlying system is affected, possibly in a nonlinear fashion, by some possibly time-varying parameter (such as varying resonance poles and alike), one could strive

- either for designing a robust controller if the actual parameter changes are not available as on-line information

- or for constructing an LPV controller if the parameter (and its rate of variation) can be measured on-line.

As such the presented techniques can be a useful extension to the nominal design specifications that have been considered previously.

In a brief final and informal discussion we would like to point out possible applications of robust and LPV control techniques to the control of nonlinear systems:

- They clearly apply if one can systematically embed a nonlinear system in a class of linear systems that admit an LPV parameterization.
- Even if it is required to perform a heuristic linearization step, they can improve classical gain-scheduling design schemes for nonlinear systems since they lead to a one-shot construction of a family of linear controllers.

9.4.1 From Nonlinear Systems to LPV Systems

In order to apply the techniques discussed in these notes to nonlinear systems, one uses variations of what is often called *global linearization*.

Consider a nonlinear system described by

$$\dot{x} = f(x) \tag{9.4.1}$$

where we assume that $f : \mathbb{R}^n \to \mathbb{R}^n$ is a smooth vector field.

If $f(0) = 0$, it is often possible to rewrite $f(x) = A(x)x$ with a smooth matrix valued mapping $A(.)$. If one can guarantee that the LPV system

$$\dot{x} = A(\delta(t))x$$

is exponentially stable, we can conclude that the nonlinear system

$$\dot{x} = A(x)x$$

has 0 as a globally exponentially stable equilibrium. Note that one can and should impose a priori bounds on the state-trajectories such as $x(t) \in M$ for some set M such that the stability of the LPV system only has to be assured for $\delta(t) \in M$; of course, one can then only conclude stability for trajectories of the nonlinear system that remain in M.

A slightly more general procedure allows to consider arbitrary system trajectories instead of equilibrium points (or constant trajectories) only. In fact, suppose $x_1(.)$ and $x_2(.)$ are two trajectories of (9.4.1). By the mean-value theorem, there exist

$$\eta_j(t) \in \operatorname{conv}\{x_1(t), x_2(t)\}$$

such that

$$\dot{x}_1(t) - \dot{x}_2(t) = f(x_1(t)) - f(x_2(t)) = \begin{pmatrix} \frac{\partial f_1}{\partial x}(\eta_1(t)) \\ \vdots \\ \frac{\partial f_n}{\partial x}(\eta_n(t)) \end{pmatrix} (x_1(t) - x_2(t)).$$

Therefore, the increment $\xi(t) = x_1(t) - x_2(t)$ satisfies the LPV system

$$\dot{\xi}(t) = A(\eta_1(t), \ldots, \eta_n(t))\xi(t)$$

with parameters $\eta_1, \ldots, \eta_n$. Once this LPV system is shown to be exponentially stable, one can conclude that $\xi(t) = x_1(t) - x_2(t)$ converges exponentially to zero for $t \to \infty$. If $x_2(.)$ is a nominal system trajectory (such as an equilibrium point or a given trajectory to be investigated), we can conclude that $x_1(t)$ approaches this nominal trajectory exponentially.

Finally, the following procedure is often referred to as *global linearization*. Let

$$\mathcal{F} \text{ be the closure of } \operatorname{conv}\{f_x(x) \mid x \in \mathbb{R}^n\}.$$

Clearly, $\mathcal{F}$ is a closed and convex subset of $\mathbb{R}^{n\times n}$. It is not difficult to see that any pair of trajectories $x_1(.)$, $x_2(.)$ of (9.4.1) satisfies the linear differential inclusion

$$\dot{x}_1(t) - \dot{x}_2(t) \in \mathcal{F}(x_1(t) - x_2(t)). \tag{9.4.2}$$

Proof. Fix any t and consider the closed convex set

$$\mathcal{F}[x_1(t) - x_2(t)] \subset \mathbb{R}^n.$$

Suppose this set is contained in the negative half-space defined by the vector $y \in \mathbb{R}^n$:

$$y^T \mathcal{F}[x_1(t) - x_2(t)] \le 0.$$

Due to the mean-value theorem, there exists a $\xi \in \operatorname{conv}\{x_1(t), x_2(t)\}$ with

$$y^T[\dot{x}_1(t) - \dot{x}_2(t)] = y^T[f(x_1(t)) - f(x_2(t))] = y^T f_x(\xi)[x_1(t) - x_2(t)].$$

Since $f_x(\xi) \in \mathcal{F}$, we infer

$$y^T[\dot{x}_1(t) - \dot{x}_2(t)] \le 0.$$

Hence $\dot{x}_1(t) - \dot{x}_2(t)$ is contained, as well, in the negative half-space defined by y. Since $\mathcal{F}$ is closed and convex, we can indeed infer (9.4.2) as desired. ■

To analyze the stability of the differential inclusion, one can cover the set $\mathcal{F}$ by the convex hull of finitely many matrices A_j and apply the techniques that have been presented in these notes.

Remarks. Of course, there are many other possibilities to embed nonlinear systems in a family of linear systems that depend on a time-varying parameter. Since there is no general recipe to transform a given problem to the LPV scenario, we have only sketched a few ideas. Although we concentrated on stability analysis, these ideas straightforwardly extend to various nominal or robust performance design problems what is a considerable advantage over other techniques for nonlinear systems. This is particularly important since, in practical problems, non-linearities are often highly structured and not all states enter non-linearly. For example, in a stabilization problem, one might arrive at a system

$$\dot{x} = A(y)x + B(y)u, \quad y = Cx$$

where u is the control input and y is the measured output that captures, as well, those states that enter the system non-linearly. We can use the LPV techniques to design a stabilizing LPV controller for this system. Since y is the scheduling variable, this controller will depend, in general, non-linearly on y; hence LPV control amounts to a systematic technique to design nonlinear controllers for nonlinear systems 'whose non-linearities can be measured'.

9.4.2 Gain-Scheduling

A typical engineering technique to attack design problems for nonlinear systems proceeds as follows: Linearize the system around a couple of operating points, design good linear controllers for each of these points, and then glue these linear controllers together to control the nonlinear system.

Although this scheme seems to work reasonably well in many practical circumstances, there are considerable drawbacks:

- There is no general recipe how to glue controllers together. It is hard to discriminate between several conceivable controller interpolation techniques.
- It is not clear how to design the linear controllers such that, after interpolation, the overall controlled system shows the desired performance.
- There are no guarantees whatsoever that the overall system is even stabilized, not to speak of guarantees for performance. Only through nonlinear simulations one can roughly assess that the chosen design scenario has been successful.

Based on LPV techniques, one can provide a recipe to systematically design a family of linear controllers that is scheduled on the operating point without the need for ad-hoc interpolation strategies. Moreover, one can provide, at least for the linearized family of systems, guarantees for stability and performance, even if the system undergoes rapid changes of the operating condition.

Again, we just look at the stabilization problem and observe that the extensions to include as well performance specifications are straightforward.

Suppose a nonlinear system

$$\dot{x} = a(x,u), \quad y = c(x,u) - r \tag{9.4.3}$$

has x as its state, u as its control, r as a reference input, and y as a tracking error output that is also the measured output. We assume that, for each reference input r, the system admits a unique equilibrium (operating condition)

$$0 = a(x_0(r), u_0(r)), \quad 0 = c(x_0(r), u_0(r)) - r$$

such that $x_0(.)$, $u_0(.)$ are smooth in r. (In general, one applies the implicit function theorem to guarantee the existence of such a parameterized family of equilibria under certain conditions. In practice, the calculation of these operating points is the first step to be done.)

The next step is to linearize the the system around each operating point to obtain

$$\dot{x} = f_x(x_0(r), u_0(r))x + f_u(x_0(r), u_0(r))u, \quad y = c_x(x_0(r), u_0(r))x + c_u(x_0(r), u_0(r))u - r.$$

This is indeed a family of linear systems that is parameterized by r.

In standard gain-scheduling, linear techniques are used to find, for each r, a good tracking controller for each of these systems, and the resulting controllers are then somehow interpolated.

At this point we can exploit the LPV techniques to systematically design an LPV controller that achieves good tracking for all reference trajectories in a certain class, even if these references vary quickly with time. This systematic approach directly leads to a family of linear systems, where the interpolation step is taken care of by the algorithm. Still, however, one has to confirm by nonlinear simulations that the resulting LPV controller works well for the original nonlinear system. Note that the Taylor linearization can sometimes be replaced by global linearization (as discussed in the previous section) what leads to a priori guarantees for the controlled nonlinear system.

Again, this was only a very brief sketch of ideas to apply LPV control in gain-scheduling, and we refer to [13] for a broader exposition of gain-scheduling in nonlinear control.