Constrained $\mathcal{H}_\infty$-Optimization for Discrete-Time Control Systems

Thesis by

H. P. Rotstein

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1993

(Submitted July 2, 1992)

Acknowledgements

I would like to thank my advisor Thanasis Sideris for his guidance and support in this research; in spite of going through tough times, Thanasis managed to create a stimulating environment for my work and I really appreciate that. I would also like to thank Cacho Romagnoli and Alfredo Desges for guiding my first steps in control in Argentina.

Many people have contributed to my broad education, both at the Universidad Nacional del Sur, Bahía Blanca, Argentina and at Caltech. I thank John Doyle, Manfred Morari and their students for sharing their views and research in classes, seminars, preprints and informal discussions. I thank Al, Bobby, Carolyn, John, Jorge, Matt, Pete and Way Min for making the lab a friendly working place. I also thank Mario Sznaier, who is a nice uruguayo. I thank the people at the Departamento Once, Planta Piloto de Ingeniería Química, Alberto, Celeste, Stella, Jorge and Jose for valuable discussions in science and in how to beat hyperinflation (you cannot).

I would like to thank my mother Delia and late father Teddy, and my sister and brother Nora and Willy, for their love and support throughout my life. I thank also my uncle Quique Rotstein, whose love and care during these three years have been so helpful.

Finally, I have always read people thanking their families for their patience and support, but only now I clearly understand what these phrases mean. To my spouse Marcela and son Ariel, thank you very much.

To Marcela and Ariel, with all my love.

Constrained $\mathcal{H}_\infty$-Optimization for Discrete-Time Control Systems

H. P. Rotstein, Ph.D.
California Institute of Technology 1993

In order to formulate a problem in the $\mathcal{H}_\infty$-optimal control framework, all specifications have to be combined in a single $\mathcal{H}_\infty$-norm objective, by an appropriate selection of weighting functions. If some of the specifications have the form of hard time domain constraints, the task of finding weighting functions that achieve a satisfactory design can become arduous. In this thesis, a theory for constrained $\mathcal{H}_\infty$-control is presented, that can deal with the standard $\mathcal{H}_\infty$ objective *and* time domain constraints. Specifically, the following time domain constrained problem is solved: given a number $\gamma > 0$, and a set of fixed inputs $\{w^i\}$, find a controller such that the closed-loop transfer matrix has an $\mathcal{H}_\infty$-norm less than γ, and the time response y^i to the signal w^i belongs to some pre-specified set Ω^i for each i. Constraints are only imposed over a finite horizon, and this allows the formulation of a two step procedure. In the first step, the optimal way of clearing the constraints is found by computing a solution to a convex non differentiable problem. In the second, a standard unconstrained $\mathcal{H}_\infty$-problem is solved. The final controller results from putting together the solution to both subproblems.

The objective function for the minimization, and the solution to the whole problem are constructed using state-space formulas. The ellipsoid algorithm is argued to be a convenient procedure for performing the optimization since, if carefully implemented, it can deal with the two main characteristics of the problem, i.e., non-differentiability and large-scale. The validity of assuming constraints over a finite horizon is justified by presenting a procedure for computing a solution that gives an overall satisfactory behavior. For clarity of exposition, this thesis starts by discussing a very special instance of the problem, and then proceeds to give the solution to the general case. Also, a benchmark problem for robust control is solved to illustrate the applicability of the theory.

Table of Contents

List of Figures

Chapter 1

Introduction

One of the major recent advances in linear optimal control has been the development of the so-called $\mathcal{H}_\infty$-theory, which attempts to design a controller that achieves internal stability and minimizes a closed-loop $\mathcal{H}_\infty$-norm when wrapped around a system. In his 1981 paper, Zames discussed the shortcomings of classical LQG methods for dealing with uncertainty, and proposed the minimization of the $\mathcal{H}_\infty$ norm as the most natural robust control objective, thus initiating the interest in the field. The febrile activity that followed has produced valuable results, both from a theoretical and practical viewpoint. Among the former, one may cite the solution to the robust stabilization or the optimal rejection of norm bounded disturbance problems. Proof of the latter, is that algorithms for solving $\mathcal{H}_\infty$ designs are now standard in commercial control design packages, and numerous successful practical applications are being reported.

In spite of its success and popularity, $\mathcal{H}_\infty$ has also some disadvantages, that were recognized at the early stages of its development. First, the theory addresses the problem of minimizing the $\mathcal{H}_\infty$ norm of a single transfer matrix, whereas most control problems require either the simultaneous minimization of several norms or the minimization of a norm subject to some additional structure (e.g., the minimization of the structured singular value). Although it is usually possible to formulate any of these problems as one in the standard form, the conservatism introduced by this step can render the final result useless. Some iterative schemes to reduce this conservatism that seem to work well in practice have been proposed (e.g., Doyle's D-K iteration), but this is a current area of active research. A second disadvantage is that the control specifications must all be translated into a single $\mathcal{H}_\infty$ norm objective, by designing weighting functions that attempt to reflect the specifications; in the absence of a formal procedure, this step may turn out to be quite hard and time consuming. Critics of the paradigm have indeed claimed that the $\mathcal{H}_\infty$ approach to control just pushes the design problem one step backwards (from controller to weight design), but this is somewhat unfair, since a designer with a physical knowledge of the system that he or she wants to control, and a reasonable flexible set of specifications, will usually be able to come up with a satisfactory design. When some of the specifications are hard-bound constraints, then the criticism gains more credibility, since the trial and error process for designing weights might not produce a good design within a reasonable time frame. Moreover, the designer cannot assess, by using the $\mathcal{H}_\infty$ machinery, if a given constrained problem has a feasible solution or otherwise establish the limits of performance induced by his or her specifications. This drawback of $\mathcal{H}_\infty$ theory can in principle be circumvented by putting the analytic machinery aside and using numerical methods for determining the parameters of a controller that achieves the specifications. When the number of parameters is small and the problem relatively simple, then this naive scheme may produce good results, since in general simple correlations between the tunning parameters and the closed-loop behavior might be found. For larger or more complex problems, this is in general no longer true, and the computation of a controller would potentially require enormous computer power. And again, if the computer fails to produce a result, it is not possible to conclude that no solution exists.

1.1 Overview of Previous Works

Several design methods have been proposed for solving constrained control problems, in the face of the inadequacy of purely linear ones to deal with constraints. Examples of early efforts included the work of Truxal [70] and Horowitz [37] that attempted to design a closed-loop transfer function that satisfied a set of constraints formulated in the frequency domain. Their methods could not take time domain constraints exactly into account, and frequency domain translations were derived by using second order model approximations. Fegley and co-workers were probably the first ones to use numerical optimization to compute a closed-loop design for single input-single output discrete and continuous time systems, incorporating explicitly and exactly, several time domain constraints. A summary of the application of linear and quadratic programming to solve the resulting problems and further references can be found in [21].

In Fegley's approach, frequency domain specifications cannot be treated exactly, since only a finite number of constraints can be considered for each problem. When algorithms for nondifferentiable optimization became available, Polak, Mayne and co-workers were able to incorporate frequency domain constraints into the set of admissible specifications [58], and therefore solve in principle a very general instance of the constrained design problem. However, their method, as well as all the ones cited so far, are "local," in the sense that the final results cannot be guaranteed to be globally optimal due to the presence of non convex constraints and/or objective functions. This major limitation can be lifted by using the Youla Parametrization Lemma [25], that parametrizes all feasible closed-loop transfer function as an affine function of a stable transfer function. The affine dependence implies that convex constraints on the closed-loop translate into convex constraints on the design parameter, and stability of the former is equivalent to stability of the latter. Gustafson and Desoer [30] explored the use of the parametrization on constrained design, although they did not duel with the numerics of the problem. Polak and Salcudean [59] used the Youla parametrization to formulate the design problem as a constrained convex optimization problem, describing a fairly large number of convex specifications, both in the time and frequency domain. The constraints are nondifferentiable, but they have a nice continuity property (i.e., Lipschitz continuity) that makes it possible to apply the nondifferentiable optimization algorithms in [57]. Constraints are also semi-infinite, because for each frequency or time instant there corresponds a constraint; this issue was addressed by solving a sequence of finite optimization problems of increasing dimensions. The author feels that, although the whole approach may work well for relatively small problems (i.e., less than 50 variables and a couple of hundred of constraints), it will perform very poorly for larger ones, due to the inherent computational complexity of the algorithms employed.

The problem considered by Boyd and co-workers [10] was similar to the one in [59], but they concentrated on developing a program to translate discrete-time control problems into optimization programs, solvable by one of the available general nonlinear programming packages. This is achieved by formulating the specifications as constraints. The semi-infinity issue is addressed by considering a finite horizon and sampling the constraints at a finite number of frequencies. In the book [7], the approach is further refined be proposing some algorithms for convex programming that are able to deal with nondifferentiable constraints.

Although the idea of straightforwardly reducing constrained controller design into an optimization problem is appealing, the resulting minimization may be too hard even for state of the art numerical algorithms for convex programming. A different route was followed in the paper by Helton and Sideris [38]. In this approach, the iterative scheme designed by Lawson to solve the $\mathcal{H}_\infty$ optimal control problem without time domain constraints, was modified so as to accommodate linear time domain constraints. The resulting program is a combination of Lawson's algorithm steps with quadratic programming steps. If the algorithm converges, then it minimizes the infinity norm of a scalar transfer function subject to time domain constraints. Although this algorithm is intuitively appealing, its convergence properties are hard to ascertain as it seems to become ill-defined when the problem becomes nondifferentiable. Moreover, it was designed for scalar objective functions and it is not immediate how to generalize it to a more complex problem.

1.2 General Considerations, Thesis Outline and Notation

The purpose of this paper is to incorporate time domain constraints into the $\mathcal{H}_\infty$ design framework for discrete-time, linear time invariant systems. In principle, the approach is restricted to time domain constraints imposed only over a finite horizon. This additional assumption, allows the formulation of a two step procedure, by which first the constraints are optimally cleared, and then a standard unconstrained $\mathcal{H}_\infty$ problem is solved. The overall solution is found by putting together the solution to each sub-problem. The advantage of this scheme is that the problem of optimally clearing the constraints in an $\mathcal{H}_\infty$-norm sense can be formulated as a finite dimensional optimization program with a convex objective function; hence, the whole problem will be convex if the time domain constraints are. The subdivision of the problem is done following the idea proposed in [38], but unlike the route taken there, the state-space approach to $\mathcal{H}_\infty$-control is subsequently employed to find a solution. Considering constraints over a finite horizon may seem to be reasonable, since stability of the closed loop system implies a decay in the time domain responses, and one usually expects the system to be "dead" after a certain number of samples. It turns out that this intuitive idea can actually fail, since a good behavior after the horizon can conflict with the constrained $\mathcal{H}_\infty$ optimality criteria. However, taking the horizon long enough can produce the desired effect, if suboptimal solutions are taken instead of optimal ones.

There are two main reasons that make the treatment of discrete over continuous time systems desirable, besides the mathematical convenience. First, due to the use of digital computers and microprocessor based control systems, many control problems originate in discrete time; even if this in not true for a specific application, continuous-time system can always be approximated as close as desired by a discrete-time one. Second, the majority of the practical control problems involve a continuous-time system controlled by a digital device. Although one can always design a continuous-time controller and then discretize it for implementation, this requires an approximation and hence a potential loss in performance. If instead, the hybrid closed-loop system is considered directly, then it has recently been shown that the problem can be reformulated as a special discrete-time one.

The thesis may be divided into four different parts and the conclusions. The first part consists of Chapter 2, that is devoted to a fairly complete study of the unconstrained $\mathcal{H}_\infty$-control problem. The chapter is mostly technical, but is needed in order to solve the general instance of the constrained problem. It contains results that are either new or more general than the ones in the literature, and are required in subsequent developments. The chapter is independent from the rest of the material, and can therefore be skipped on a first reading and used only as as a convenient reference. The second part consists of Chapters 3 and 4, and includes the solution of a simple instance of the time domain constrained $\mathcal{H}_\infty$ problem. In spite of its simplicity, the problem captures the main ideas and properties of the solution, without the technical difficulties of the general case. In Chapter 3, the problem is transformed into a finite dimensional optimization problem, by exploiting the fact that constraints are only imposed over a finite horizon . Then, a state-space procedure is worked out for computing the objective function, a proof of convexity is given, and a property of the function is discussed for the limiting case with non binding constraints. The latter issue has interesting consequences for the numerical properties of the problem.

The chapter also contains some bounds over the order of the solution and a discussion on the overall behavior of the time response. The main idea in [38] is that if the horizon is taken long enough then satisfaction of the constraints over the horizon implies a nice behavior from zero to infinity. It turns out that this may not be the case (specially if the optimal solution is pursued), since letting the horizon length trend to infinity does not necessarily imply that the limiting solution will satisfy the constraints. However, it is shown that this difficulty can be easily overcome by taking suboptimal solutions; in particular, a scheme for computing one of such solutions that has guaranteed overall behavior if the horizon is taken long enough is given. In Chapter 4, the numerical aspects of the problem are treated. First descent algorithms for nondifferentiable optimizations are considered and subsequently discarded, due to their seeming inability to deal with large-scale problems. Then, a non descent method called the Ellipsoid Algorithm is briefly described, and an implementation that is able to cope with a large number of variables and constraints is discussed. The chapter also contains a textbook example to illustrate the theory.

The third part of the thesis consists of Chapter 5, where the general instance of the constrained problem is considered. First, a solution to the so-called "one-block" problem is presented, that follows closely the treatment of the simple case in Chapter 3. Then, this result is used in conjunction with the material in Chapter 2 to give the solution to the general problem. The problem is again shown to be equivalent to a finite convex minimization, and a simple formula for the objective function is given. In fact, several alternative optimization schemes are proposed, and their relative numerical merits are investigated. Finally, a state-space realization for a solution is given and the textbook example in Chapter 4 is continued to show the applicability of the methodology.

The last part of the thesis consists of Chapter 6, and is dedicated to a case-study. Although a textbook example is included in Chapters 4 and 5, the application of the theory and algorithms to a well-known benchmark problem in robust control is both interesting and enlightening. In particular, the tradeoff between stability robustness and time domain performance are investigated. The thesis is closed with some concluding remarks and current research lines in Chapter 7.

Notation

The notation in this paper is standard. $\overline{\sigma}$ and ρ denote the largest eigenvalue and spectral radius respectively. Let G be a matrix valued function. L_∞ denotes the Lebesgue space of all functions defined on the unit circle such that $\|G\|_\infty \doteq sup\{\overline{\sigma}(G(z))\ s.t.\ |z| = 1\} < \infty$. By $\mathcal{H}_\infty$ ($\mathcal{H}_\infty^\sim$) we denote the Hardy space of stable (antistable) functions $G \in L_\infty$, i.e., G with all its poles inside (outside) the unit disk. Therefore, z^{-1} represents the unit delay operator. $\mathcal{H}_2$ denotes the Hardy space of complex-valued norm square integrable functions on the unit circle with vanishing negative Fourier coefficients; i.e., if $G \in \mathcal{H}_2$ then $H(z) = \sum_{i=0}^{\infty} H_i z^{-i}$, and $\|H\|_2^2 = trace(\sum_{i=0}^{\infty} H_i' H_i)$. This norm has the alternative "frequency-domain" characterization $\|H\|_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{i=1}^{n}\sigma_i(G(e^{j\theta})G(e^{-j\theta})^t)d\theta$. Therefore $G \in \mathcal{H}_\infty \Rightarrow G \in \mathcal{H}_2$ and

$$1/\sqrt{n}\|G\|_2 \leq \|G\|_\infty. \qquad (1.2-1)$$

A real-rational transfer function may be written as

$$G(z) = G_p(z) + G_i(z), \qquad (1.2-2)$$

where $G_p(z)$ is proper and $G_i(z)$ is a polynomial in z. Note that this notation does not include the dimensions involved, but dimensions will be denoted with supra indices whenever necessary. For instance $\mathcal{H}_\infty^{m\times n}$ denotes a stable matrix transfer function with m outputs and n inputs. Let $\mathcal{RH}_\infty$ denote the set of stable, proper, real rational transfer functions with a state-space realization, i.e., $G(z) \in \mathcal{RH}_\infty$ may be written as

$$G(z) = D + C(zI - A)^{-1}B = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right). \qquad (1.2-3)$$

See Chapter 2 for more details on several representations of discrete-time systems. Particularly useful for subsequent development are the following definitions. Let

$$A_f^n = \begin{pmatrix} 0 & I_l & 0 & \cdots & 0 \\ 0 & 0 & I_l & \cdots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \cdots & 0 & I_l \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad (1.2-4)$$

$E_1^n \doteq [I_l\ 0\ \cdots 0]^t$ and $E_n \doteq [0\ \cdots 0\ I_l]^t$, then the function $z^{-n}I_l$ has a realization:

$$z^{-n}I_l = \left(\begin{array}{c|c} A_f^n & E_n \\ \hline E_1^t & 0 \end{array}\right) \qquad (1.2-5)$$

and if $G(z)$ is an FIR, then

$$G(z) = \sum_{i=0}^{n-1} G_i z^{-i} = \left(\begin{array}{c|c} A_f^{n-1} & E_{n-1} \\ \hline [G_{n-1} \cdots G_1] & G_0 \end{array}\right) = \left(\begin{array}{c|c} A_f^{n-1} & [G_1 \cdots G_{n-1}] \\ \hline E_1^t & G_0 \end{array}\right). \qquad (1.2-6)$$

A_f^n and E_i^n will be denoted A_f and E_i respectively whenever the block-dimension is clear from the context. If $G(z) \in \mathcal{RH}_\infty$, $G(z)$ can be expanded as:

$$G(z) = D + \sum_{i=0}^{\infty} CA^iBz^{-(i+1)}. \qquad (1.2-7)$$

For notational convenience, we will sometimes write $G_0 = D$ and $G_i = CA^{i-1}B$, $i = 1, 2, \cdots$, and define $\mathbf{G_n} = \begin{bmatrix} G_0 & \cdots & G_{n-1} \end{bmatrix}$. It will be assumed in the following that all transfer functions are in $\mathcal{RL}_\infty$. Define

$$G^\sim(z) \doteq G(1/z)^t; \qquad (1.2-8)$$

then if $det(A) \neq 0$, and $G(z)$ is given by 1.2-3,

$$G^{\sim} = \left(\begin{array}{c|c} A^{-t} & A^{-t}C^{t} \\ \hline -B^{-t}A^{-t} & D^{t} - B^{t}A^{-t}C^{t} \end{array} \right). \qquad (1.2-9)$$

Note that if $det(A) = 0$, then $G^{\sim}$ has a polynomial part.

Chapter 2

Unconstrained $\mathcal{H}_\infty$ Optimal Control

Although time domain constraints constitute the main topic in this thesis, the first long chapter will be devoted to study some general properties of discrete-time (DT) systems and the standard DT $\mathcal{H}_\infty$-control problem. Some of the results presented here are not new, in the sense that alternative expressions may be found in the literature. However, the problems that arise in constrained $\mathcal{H}_\infty$-control do not satisfy the conditions that are usually assumed elsewhere. As it turns out, these conditions are not due to any inherent property of discrete-time, but to the form that is used to represent them. Specifically, a particular Rosenbrock System Matrix form is argued to be the most natural way of representing real rational DT systems, as opposed to state-space. The usage of this form both allows the lifting of one of the most critical assumptions and the simplification of algebraic manipulations. Finally, the treatment in this chapter sheds additional light on the structure of DT $\mathcal{H}_\infty$-control. For example, a link will be displayed between the pure algebraic manipulations involved in solving the Nehari problem to be defined below, and the classical one-step dilation construction from the operator-theoretic approach to the problem.

2.1 Introduction

It is well known that most theoretical results for continuous-time systems have a direct extension to the discrete-time counterpart. This is specially true for $\mathcal{H}_\infty$-Control Theory since, as noted by Glover in his seminal 1984 paper [28], the results for DT can be obtained from those of continuous time via a bilinear transformation. Such a procedure often gives rise to complicated expressions that obscure the structure of the solutions, and it is obvious that a more direct route, i.e., a purely discrete time derivation, is desirable. Additionally, the fact that most of the results in the field have first been worked out for continuous-time and then translated into DT, has created some confusion onto what are the conditions that are needed in order to guarantee the existence of solutions to different problems. Although these objections may seem to be only technical details if one is just interested in the standard $\mathcal{H}_\infty$ theory, they become of paramount importance for some of the constructions in this thesis.

Consider, for instance, the condition frequently imposed in the literature, that systems should not have poles at zero. This assumption certainly does look like a technical detail, since zero is an interior point of the stability region and therefore should not merit any special consideration. Indeed, the sole reason for this assumption is that it guarantees state-space representation to be closed under conjugation; i.e., if a system has a state-space representation, then its conjugate will also have a state-space representation if and only if the system has no poles at zero. The condition therefore refers to the way that the system is represented rather than reflecting any inherent property. The reason that is usually offered to validate the assumption, namely that the condition will hold generically for any physically motivated system, is actually weak. For instance, the condition may fail when the feed-through terms from the control inputs to the controlled outputs or from the disturbances to the measurements are not full rank. As discussed below, the latter case may occur when the system comes from sampling a continuous time system. As for the present thesis, poles at zero are introduced as an artifice to solve the constrained $\mathcal{H}_\infty$ problem, and hence a theory that includes this case is necessary in order to get most of the results.

The remainder of this chapter can be divided into three parts. The first part consists of a discussion on some polynomial system matrix representations and in particular of the Rosenbrock System Matrix form, which is the next simplest way of representing a system, with the additional advantage of being closed under conjugation. The second part, comprising of sections 2.3 through 2.5 is a compilation of results needed for subsequent developments. Although these sections are mostly technicals, certain topics of independent interest (like the already mentioned above fact concerning sample-data systems) are also discussed. The last part, that includes sections 2.6 through 2.8, contains a discussion of the standard $\mathcal{H}_\infty$ problem in a way suitable for use on future chapters.

2.2 RSM Representation

In this section the Polynomial System Matrix (PSM) representation is reviewed. In particular, a special case of the Rosenbrock System Matrix form [61] is presented that yields the natural representation for a discrete, finitely dimensional, linear time invariant system when poles at zero (or infinity) are not ruled out. The reason for this is that the form preserves the ease of computation of the state-space form, and is also closed with respect to conjugation.

Let $P(z)$ be a transfer function; then $P(z)$ may be represented by a *polynomial system matrix* (PSM) [67] as a set of equations of the form

$$\mathbf{P}(\mathbf{z})\xi = 0,$$

where $\mathbf{P}(\mathbf{z})$ is a matrix which entries are polynomial in z, and $\mathbf{P}(\mathbf{z})$ and ξ are partitioned as

$$\begin{aligned} \mathbf{P}(\mathbf{z}) &= \begin{bmatrix} p_x(z) & p_u(z) & p_y(z) \end{bmatrix} \\ \xi &= \begin{bmatrix} x^t & u^t & y^t \end{bmatrix}^t . \end{aligned}$$

Here u, y, x are vectors of inputs, outputs and states respectively. If $\mathbf{P}(\mathbf{z})$ is a PSM for $P(z)$, then write $P(z) = \mathbf{P}(\mathbf{z})$. A PSM is said to be in *Rosenbrock system-matrix* form if it has the form

$$\mathbf{P}(\mathbf{z}) = \begin{pmatrix} -T(z) & U(z) & 0 \\ V(z) & W(z) & -I \end{pmatrix},$$

in which case $P(z) = V(z)T(z)^{-1}U(z) + W(z)$. A PSM is said to be in *descriptory* form (DF) if it has the form

$$\mathbf{P}(\mathbf{z}) = \begin{pmatrix} -Ez + A & B & 0 \\ C & D & -I \end{pmatrix},$$

in which case $P(z) = C(zE - A)^{-1}B + D$. A PSM is said to be in *state-space* form (SSF) if it has the form

$$\mathbf{P}(\mathbf{z}) = \begin{pmatrix} -Iz + A & B & 0 \\ C & D & -I \end{pmatrix},$$

in which case $P(z) = C(zI - A)^{-1}B + D$ denoted $P(z) = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$. Finally, a PSM is said to be in *discrete normal* form (DNF) if it has the form

$$\mathbf{P}(\mathbf{z}) = \begin{pmatrix} -Ez + A & B & 0 \\ zF + C & D & -I \end{pmatrix},$$

in which case $P(z) = (zF + C)(zE - A)^{-1}B + D$. Although the DNF is not standard, it is arguably the most natural way of representing a DT finite-dimensional system, as will hopefully be shown in the remainder of the section.

A useful property of PSM representation is that the transfer function between u and y is not changed if $\mathbf{P}(\mathbf{z})$ is multiplied on the right by a matrix of the form

$$R = \begin{pmatrix} R_x & R_u & R_y \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}$$

with R_x unimodular. Also, multiplication of $\mathbf{P}(\mathbf{z})$ on the left by *any* compatible dimensioned unimodular polynomial matrix $L(x)$ leaves the transfer functions from u to y unchanged.

To illustrate the convenience of the DNF representation and show some useful operations on transfer matrices in DNF, a number of results are collected in what follows.

Property 2.1 *Let $P(z)$ have a DNF:*

$$P(z) = \begin{pmatrix} -Ez + A & B & 0 \\ C & D & -I \end{pmatrix}.$$

Then $P(z)^{\sim} \doteq P(1/z)^t$ has the DNF:

$$P(z) = \begin{pmatrix} -E^t + zA^t & C^t & 0 \\ zB^t & D^t & -I \end{pmatrix}.$$

Note that, strictly speaking, DNF's are not closed under conjugation since a first order polynomial would alse be needed in the "12" entry. However, the form is general enough to capture all cases of interest.

Property 2.2 *Let $P(z)$ have a SSF:*

$$P(z) = \begin{pmatrix} -zI + A & B & 0 \\ C & D & -I \end{pmatrix}.$$

Let X and Y be two square matrices of appropriate dimensions, and let

$$P_1(z) = \begin{pmatrix} -YXz + YAX & YB & 0 \\ CX & D & -I \end{pmatrix}.$$

Then $P(z) = P_1(z)$ if X and Y are nonsingular or if $P(z)$ is stable and X, Y solve the equations

$$\begin{aligned} Y &= A^tYA + C^tC \\ X &= AXA^t + BB^t. \end{aligned}$$

Proof: the first part (i.e., X and Y non-singular) follows from general properties of PSM forms. The second part follows after recognizing that X and Y are the controllability and observability grammians respectively of $P(z)$ and hence that their null space corresponds to uncontrollable or unobservable modes.

□

To illustrate the addition and multiplication of DNF, the following result will be proved.

Theorem 2.1 *Let P have a minimal state-space realization, $P = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$, with $\rho(A) < 1$. Then $P^{\sim}P = I$ if and only if*

$$\begin{aligned} D^tC + B^tYA &= 0 \\ D^tD + B^tYB &= I, \end{aligned}$$

where Y denotes the observability grammian, solving the Lyapunov equation $Y = A^tYA + C^tC$.

Proof: Using DNF representation,

$$P(z) = \begin{pmatrix} -zI + A & B & 0 \\ C & D & -I \end{pmatrix}$$

and

$$P(z)^{\sim} = \begin{pmatrix} -I + zA^t & C^t & 0 \\ zB^t & D^t & -I \end{pmatrix}.$$

A simple computation shows that:

$$P(z)^{\sim}P(z) = \begin{pmatrix} -zI + A & 0 & B & 0 \\ C^tC & -I + zA^t & C^tD & 0 \\ D^tC & zB^t & D^tD & -I \end{pmatrix}.$$

Multiplying this system matrix by $T = \begin{pmatrix} I & 0 & 0 \\ A^tX & I & 0 \\ B^tX & 0 & I \end{pmatrix}$ on the left leaves the transfer function between u_1 and y_2 unchanged:

$$P(z)^{\sim}P(z) = \begin{pmatrix} -zI + A & 0 & B & 0 \\ C^tC - zA^tX + A^tXA & -I + zA^t & C^tD + A^tXB & 0 \\ -zB^tX + B^tXA + D^tC & zB^t & D^tD + B^tXB & -I \end{pmatrix}.$$

Now multiplying by $T_1 = \begin{pmatrix} I & 0 & 0 & 0 \\ Y & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix}$ on the right:

$$P(z)^{\sim}P(z) = \begin{pmatrix} -zI + A & 0 & B & 0 \\ C^tC + A^tXA - X & -I + zA^t & C^tD + A^tXB & 0 \\ B^tXA + D^tC & zB^t & D^tD + B^tXB & -I \end{pmatrix}.$$

Let Y be the observability grammian for $P(z)$). Then:

$$P(z)^{\sim}P(z) = \begin{pmatrix} -zI + A & 0 & B & 0 \\ 0 & -I + zA^t & C^tD + A^tYB & 0 \\ (C^tD + A^tYB)^t & zB^t & D^tD + B^tYB & -I \end{pmatrix}.$$

Therefore $P(z)^{\sim}P(z) = I$ if and only if

$$\begin{aligned} C^tD + A^tYB &= 0 \\ D^tD + B^tYB &= I, \end{aligned}$$

thus proving the theorem.

□

Corollary 2.1 *$P^{\sim}P = I$ if and only if*

$$\begin{aligned} DB^t + CXA^t &= 0 \\ DD^t + CXC^t &= I \end{aligned}$$

where X denotes the observability grammian, solving the Lyapunov equation $X = AXA^t + BB^t$.

□

The section is closed with a simple property that is required bellow.

Lemma 2.1 *Let G, E_L, $E_R \in \mathcal{RH}_\infty$ have state-space realizations:*

$$G = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array}\right) \quad E_L = \left(\begin{array}{c|c} A_L & B_L \\ \hline C_L & D_L \end{array}\right) \quad E_R = \left(\begin{array}{c|c} A_R & B_R \\ \hline C_R & D_R \end{array}\right).$$

Then the stable part of $\begin{bmatrix} I & 0 \\ 0 & E_L^{\sim} \end{bmatrix} G \begin{bmatrix} I & 0 \\ 0 & E_R^{\sim} \end{bmatrix}$ has a state-space realization of the form $\left(\begin{array}{c|cc} A & B_1 & \hat{B}_2 \\ \hline C_1 & D_{11} & \hat{D}_{12} \\ \hat{C}_2 & \hat{D}_{21} & \hat{D}_{22} \end{array}\right)$, for some matrices $\hat{B}_2, \hat{C}_2, \hat{D}_{12}, \hat{D}_{21}$ and $\hat{D}_{22}$.

Proof: It follows from simple matrix manipulations.

□

2.3 The Discrete-time Riccati Equation

The algebraic Riccati Equation plays a central role in both the continuous and the discrete time $\mathcal{H}_\infty$ theories. In this section, the Discrete-Time Algebraic Riccati equation (DARE) will be reviewed, with a greater generality than the one usually found in the optimal control literature. Following [71], [2] the solution to the Riccati equation will be obtained by using the "stable" deflating subspace of a matrix pencil.

Let $A \in R^{n\times n}$, $B^{n\times m}$ and R, Q, T be real matrices of appropriate dimensions, with R and Q symmetric. Define the ordered pair of $2n+m$ by $2n+m$ matrices:

$$S \doteq \left(\begin{bmatrix} A & 0 & B \\ -Q & I & -T \\ T^t & 0 & R \end{bmatrix}, \begin{bmatrix} I & 0 & 0 \\ 0 & A^t & 0 \\ 0 & -B^t & 0 \end{bmatrix} \right) \doteq (S_1,\ S_2) \qquad (2.3-1)$$

associated with the pencil $\lambda S_2 - S_1$. It will be assumed in what follows that the pencil is regular, i.e., $\det(\lambda S_2 - S_1) \not\equiv 0$; it turns out that, for the case of interest, this is without loss of generality. Let λ_1 be such that $\det(\lambda_1 S_2 - S_1) = 0$ and x be such that $(\lambda_1 S_2 - S_1)x = 0$. Then λ_1 and x are called a generalized eigenvalue and eigenvector of $\lambda S_2 - S_1$ respectively. With an abuse of notation, λ_1 and x will be called an eigenvalue and an eigenvector of S respectively. By the structure of S_2, the pencil has at least m generalized eigenvalues at ∞. This follows form the fact that the pencil $S_2 - \lambda S_1$ has at least m eigenvalues at zero. When R is invertible, 2.3-1 is equivalent to:

$$S^r = \left(\begin{bmatrix} A - BR^{-1}T^t & 0 \\ TR^{-1}T^t - Q & I \end{bmatrix}, \begin{bmatrix} I & BR^{-1}B^t \\ 0 & (A - BR^{-1}T^t)^t \end{bmatrix} \right). \qquad (2.3-2)$$

The two pencils have the same finite generalized eigenvalues, and the corresponding deflating subspaces are closely related (see [71] for details). The pencil S^r is the one usually considered in the literature [39]. Suppose that $\lambda S_2^r - S_1^r$ has no eigenvalues on $\partial\mathcal{D}$. Then, since the pencil is symplectic, it must have n generalized eigenvalues on $\mathcal{D}$ [56]. Not surprisingly, a similar property holds for S. To see this, compare the functions $det(S_1 - \lambda S_2)$ and $det(\mu S_1 - S_2)$:

$$\det \begin{bmatrix} \mu A - I & 0 & \mu B \\ -\mu Q & \mu I - A^t & -\mu T \\ \mu T^t & B^t & \mu R \end{bmatrix} = \det \left(\begin{bmatrix} \mu A - I & 0 & B \\ -\mu Q & \mu I - A^t & -T \\ \mu T^t & B^t & R \end{bmatrix} \right) \mu^m \tag{2.3-3}$$

$$\begin{aligned} \det \left(\begin{bmatrix} A - \lambda I & 0 & B \\ -Q & I - \lambda A^t & -T \\ T^t & \lambda B^t & R \end{bmatrix} \right) &= \det \left(\begin{bmatrix} A^t - \lambda I & -Q & T \\ 0 & I - \lambda A & \lambda B \\ B^t & -T^t & R \end{bmatrix} \right) \\ &= \det \left(\begin{bmatrix} I - \lambda A & 0 & \lambda B \\ -Q & A^t - \lambda I & T \\ -T^t & B^t & R \end{bmatrix} \right) \\ &= \det \left(\begin{bmatrix} \frac{1}{\lambda} I - A & 0 & B \\ -Q & A^t - \lambda I & T \\ -T^t & B^t & R \end{bmatrix} \right) \lambda^n \\ &= \det \left(\begin{bmatrix} \lambda A - I & 0 & B \\ -\lambda Q & \lambda I - A^t & -T \\ \lambda T^t & B^t & R \end{bmatrix} \right) \end{aligned} \tag{2.3-4}$$

From 2.3-3, S has m infinite generalized eigenvalues. From 2.3-4, S has $2n$ additional generalized eigenvalues and comparing 2.3-3 and 2.3-4, λ is a generalized eigenvalue of S if and only if $\mu = \lambda^{-1}$ is also a generalized eigenvalue, with the convention that ∞ is the reciprocal of 0. In particular, if S has no eigenvalues on $\partial\mathcal{D}$ then it has n eigenvalues on $\mathcal{D}$.

Consider the n-dimensional deflating subspace $\mathcal{X}$ corresponding to these generalized eigenvalues. Finding a basis for this subspace and partitioning it conformally with the partition in S,

$$\mathcal{X} = \operatorname{Im} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \qquad (2.3-5)$$

where X_1, X_2 and X_3 are real matrices of appropriate dimensions. If X_1 is nonsingular, or equivalently, if the two subspaces

$$\mathcal{X}, \quad \operatorname{Im} \begin{bmatrix} 0 & 0 \\ I & 0 \\ 0 & I \end{bmatrix} \qquad (2.3-6)$$

are complementary, set $X = X_2 X_1^{-1}$ and $F = X_3 X_1^{-1}$. It is easy to see [25] that X and F are uniquely determined by S. Therefore, define the function *Ric* by $(X, F) = Ric(S)$. The domain of *Ric*, denoted $dom(Ric)$ will consist of pairs S, such that the pencil has no generalized eigenvalues on $\partial\mathcal{D}$[1] and the subspaces in 2.3-6 are complementary. Before proceeding, note (following [51]) that a state feedback does not affect the solvability of the problem. To be specific, recall that two pencils S and S_1 are called strictly equivalent (SE) if $D_L S D_R = S_1$, with D_L, D_R constant square non singular matrices. In particular, SE pencils share the same eigenvalues. Let $D_L = \begin{bmatrix} I & 0 & 0 \\ 0 & I & K^t \\ 0 & 0 & I \end{bmatrix}$, $D_R = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -K & I & 0 \end{bmatrix}$. Then S in 2.3-1 is SE with the pencil:

$$\hat{S} = \left(\begin{bmatrix} A - BK & 0 & B \\ K^t T^t + TK - Q - K^t R K & I & -(T - K^t R) \\ T^t - RK & 0 & R \end{bmatrix}, \begin{bmatrix} I & 0 & 0 \\ 0 & (A - BK)^t & 0 \\ 0 & -B^t & 0 \end{bmatrix} \right). \qquad (2.3-7)$$

Lemma 2.2 *Suppose that $S \in dom(Ric)$ and $(X, F) = Ric(S)$. Then $\hat{S} \in dom(Ric)$ and $(X, F + KB) = Ric(\hat{S})$.*

---

[1]In particular this implies that the pencil is regular

□

The following lemma generalizes Lemma 2.1 in [39].

Lemma 2.3 *Suppose that* $S \in$ *dom(Ric) and* $(X, F) = $ *Ric(S). Then.*

a) $X = X^t$

b) (X, F) *satisfies the equations*

$$X = A^tXA - F^t(B^tXB + R)F + Q$$

$$(B^tXA + T^t) + (B^tXB + R)F = 0.$$

c) The matrix $A + BF$ *is stable*

Proof: Since $S \in dom(Ric)$, there exist $[X_1^t\ X_2^t\ X_3^t]^t \in R^{2n+m\times n}$ with full column rank, and a matrix M_x with $\rho(M_x) < 1$, such that

$$\begin{bmatrix} A & 0 & B \\ -Q & I & -T \\ T^t & 0 & R \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & A^t & 0 \\ 0 & -B^t & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} M_x. \qquad (2.3-8)$$

The proof of part a follows along the lines of [39]. From 2.3-8

$$\begin{aligned} AX_1 + BX_3 &= X_1M_x \\ QX_1 + X_2 - TX_3 &= A^tX_2M_x \\ T^tX_1 + RX_3 &= -B^tX_2M_x \end{aligned}$$

Therefore:

$$\begin{aligned} X_1^tX_2 - X_2^tX_1 &= X_1^tTX_3 - X_3^tT^tX_1 + X_1A^tX_2M_x - M_x^tX_2^tAX_1 \\ &= X_3^tB^tX_2M_x - M_x^tX_2^tBX_3 + X_1^tA^tX_2M_x - M_x^tX_2^tAX_1 \\ &= (X_1^tA^t + X_3^tB^t)X_2M_x - M_x^tX_2^t(AX_1 + BX_3) \\ &= M_x^t(X_1^tX_2 - X_2^tX_1)M_x; \end{aligned}$$

this is a Lyapunov equation with M_x stable, and hence $X_1^tX_2 = X_2^tX_1$, which implies part a. Now 2.3-8 and the definitions of X and F, immediately imply parts b and c.

□

Lemma 2.4 *Assume that A is stable. Then S has no eigenvalues on $\partial\mathcal{D}$ if and only if $R - T^t(A-\lambda I)^{-1}B + \lambda B^t(I-\lambda A^t)^{-1}T - \lambda B^t(I-\lambda A^t)^{-1}Q(A-\lambda I)^{-1}B$ is non singular for each λ such that $|\lambda| = 1$.*

Proof: The stability of A implies that both $A - \lambda I$ and $I - \lambda A^t$ are invertible for $|\lambda| = 1$. The proof follows from row operations on $S_1 - \lambda S_2$.

□

Lemma 2.5 *Suppose that $R = D^tD$, $Q = C^tC$, $T = C^tD$ and assume that $G \doteq \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$ has no zeros on the unit circle and the pair (A, B) is stabilizable. Then S has no eigenvalues on $\partial\mathcal{D}$.*

Proof: assume first that A is stable. Then, from the previous lemma, S has no eigenvalues on $\partial\mathcal{D}$ if and only if $[D^t + \lambda B^t(I-\lambda A^t)^{-1}C^t][D + C(\lambda I - A)^{-1}B]$ is non singular for $\lambda \in \partial\mathcal{D}$. But this is equivalent to G not having zeros on $\partial\mathcal{D}$. If A is not stable, take K such that $A_K = A - BK$ is stable (the stabilizability condition guarantees that at least one such K exists). The resulting pencil has $Q = C_K^tC_K$ and $T = C_K^tD$, with $C_K = C - DK$. Since zeros are invariant with respect to state feedback, the proof is completed.

□

Lemma 2.6 *Suppose that $R = D^tD$ with D full column rank, $Q = C^tC$ and $T = C^tD$ and assume that $G \doteq \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$ has no zeros on the unit circle and the pair (A, B) is stabilizable. Then $S \in dom(Ric)$ and (X, F) satisfies $(B^tXA + T^t) + (B^tXB + R)F = 0$ and $X = (A+BF)^tX(A+BF) + (C+DF)^t(C+DF)$. Moreover, if $(X, F) = Ric(S)$, then $X \geq 0$.*

Proof: Let now X_1, X_2, X_3 be such that

$$\begin{bmatrix} A & 0 & B \\ -C^tC & I & -C^tD \\ D^tC & 0 & D^tD \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & A^t & 0 \\ 0 & -B^t & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} M_x \qquad (2.3-9)$$

where M_x is such that $\rho(M_x) < 1$, and $[X_1^t\ X_2^t\ X_3^t]'$ is full column rank. From 2.3-9,

$$AX_1 + BX_3 = X_1 M_x \qquad (2.3\text{-}10)$$

$$-C^tCX_1 + X_2 - C^tDX_3 = A^tX_2M_x \qquad (2.3\text{-}11)$$

$$D^tCX_1 + D^tDX_3 = -B^tX_2M_x. \qquad (2.3\text{-}12)$$

Multiplying 2.3-11 on the left by X_1:

$$X_1^tX_2 = X_1^tC^tCX_1 + X_1^tC^tDX_3 + X_1^tA^tX_2M_x.$$

Replacing 2.3-12 in 2.3-10 multiplied on the left by $M_x^tX_2^t$:

$$M_x^tX_2^tAX_1 - X_1^tC^tDX_3 - X_3^tD^tDX_3 = M_x^tX_2^tX_1M_x.$$

The last two equations now give

$$X_1^tX_2 = (CX_1 + DX_3)^t(CX_1 + DX_3) + M_x^tX_2^tX_1M_x. \qquad (2.3-13)$$

This is a discrete Lyapunov equation with M_x stable, and hence $X_1^tX_2$ is symmetric and positive semi-definite. Let z be such that $X_1z = 0$. Then 2.3-13 implies that $DX_3z = 0$. Since D is full column rank, this implies that $X_3z = 0$. Together with 2.3-10, this implies that $X_1M_xz = 0$, and hence the kernel of X_1 is M_x-invariant. Therefore w.l.o.g. take z such that $M_xz = \lambda z$, and $|\lambda| < 1$. From 2.3-11 and 2.3-12, this implies that $X_2z = \lambda A^tz$ and $\lambda B^tX_2z = 0$. By the full column rank assumption of $[X_1^t\ X_2^t\ X_3^t]^t$, $X_2z \neq 0$, and hence $\lambda \neq 0$ and $1/\lambda$ is an eigenvalue of A. Since (A, B) is stabilizable, this implies that $z = 0$ and hence X_1 is invertible. Multiplying 2.3-13 on the left by X_1^{-t}, on the right by X_1^{-1} and using 2.3-10,

$$X = (C + DF)^t(C + DF) + (A + BF)^tX(A + BF) \qquad (2.3-14)$$

and hence X is positive semi-definite.

□

The full-rank condition on D is sufficient but it is not necessary and S may belong to *dom(Ric)* even if this condition fails to hold.

2.4 Youla Parametrization Revisited

Consider the problem shown in Fig. 1. The plant $P(z)$ is assumed to be of dimension n, linear and time invariant. The input-output relations are:

$$\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix}. \qquad (2.4-1)$$

The closed-loop system resulting from the control law $u = Ky$ is $z = \mathcal{F}_l(P, K) \doteq P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}$. It is well known [25] that various robust control problems may be formulated as the minimization of $\|\mathcal{F}_l(P, K)\|_\infty$ with K varying over the set $\mathcal{K}$ of all internally stabilizing controllers. Suppose that $P(z)$ has the state-space realization:

$$P(z) = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array}\right) \qquad (2.4-2)$$

and assume that (A, B_2) is stabilizable and (C_2, A) is detectable. Then for each $K \in \mathcal{K}$ there exist a $Q \in \mathcal{H}_\infty$ such that

$$\mathcal{F}_l(P, K) = T_{11} - T_{12}QT_{21}, \qquad (2.4-3)$$

where

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$$

$$= \left(\begin{array}{cc|cc} A+B_2F & -B_2F & B_1 & B_2R_b^{-1} \\ 0 & A+HC_2 & B_1+HD_{21} & 0 \\ \hline C_1+D_{12}F & -D_{12}F & D_{11} & D_{12}R_b^{-1} \\ 0 & R_c^{-1}C_2 & R_c^{-1}D_{21} & 0 \end{array}\right), \qquad (2.4\text{-}4)$$

with F and H chosen so that $A+B_2F$ and $A+HC_2$ are stable. An important observation is that under some mild assumptions T_{12} and T_{21} may be chosen to be inner and co-inner respectively, and that there exist $T_{12\perp}$ and $T_{21\perp}$ such that $\begin{bmatrix} T_{12} \\ T_{12\perp} \end{bmatrix}$ and $[T_{21}\ T_{21\perp}]$ are unitary. This fact was derived for continuous time systems in [19] and for discrete time in [36] under the additional assumption that D_{12} and D_{21} were full column and row rank respectively. Although these conditions are sensible for continuous-time (see the discussion in [29]), they are not needed for discrete-time, as the following example shows. Consider the transfer function $P(z)$ given by:

$$P(z) = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \frac{1}{z-.5}\begin{bmatrix} z+.5 & z+.5 \\ 1 & 1 \end{bmatrix}. \qquad (2.4-5)$$

Suppose that one wants to find a controller $K(z)$ such that $\mathcal{F}_l(P,K) \doteq P_{11}+P_{12}K(I-P_{22}K)^{-1}P_{21}$ is internally stable and the infinity norm $\mathcal{F}_l(P,K)$ is as small as possible. Given that P_{22} is stable, all internally stabilizing controllers can be parametrized as $K(Q) = Q/(1+P_{22}Q) = \frac{(z-.5)Q}{(z-.5)+Q}$, where Q is any stable transfer function. Therefore, after some algebra, $\mathcal{F}_l(P,K(Q)) = \frac{z+.5}{z-.5} + \frac{z+.5}{z-.5}\frac{1}{z-.5}Q = \frac{z+.5}{z-.5} - z^{-1}\hat{Q}$. Here $\hat{Q} = -\frac{z+.5}{z-.5}\frac{z}{z-.5}Q$, and the mapping $Q \rightarrow \hat{Q}$ is stable with stable inverse. Therefore

$$\inf_{K\ stabilizing} \|\mathcal{F}_l(P,K)\|_\infty = \inf_{\hat{Q}\ stable} \|\frac{z+.5}{z-.5} - z^{-1}\hat{Q}\|_\infty = \inf_{\hat{Q}\ stable} \|z\frac{z+.5}{z-.5} - \hat{Q}\| =$$
$$= \inf_{\hat{Q}\ stable} \|\frac{1+.5z}{z(1-.5z)} - \hat{Q}^\sim\|_\infty,$$

where $\hat{Q}(z)^\sim \doteq Q(1/z)^t$, since taking the conjugate preserves the $\mathcal{H}_\infty$ norm. By calling $G \doteq \frac{1+.5z}{z(1-.5z)}$, the optimization problem has been transformed into the one-block problem with poles at zero: $\inf_{\hat{Q}\ stable} \|G - \hat{Q}^{\sim}\|_\infty$.

The full rank condition is not only unnecessary but can also be too restrictive for practical applications. Suppose, for instance, that $P(z)$ comes from sampling a continuous time system. In this case, the disturbances must be low-pass filtered before being sampled in order to prevent aliasing, and hence the corresponding feed-through term must be necessarily zero. In fact, the system could not in general be internally stabilized without this assumption (see [3]). Therefore, the condition will not be assumed in what follows.

Let T_{12} have a DNF:

$$T_{12} = \begin{pmatrix} -zY + YA_F & YB_2R_b^{-1} & 0 \\ C_F & D_{12}R_b^{-1} & -I \end{pmatrix}, \tag{2.4 – 6}$$

where $A_F = A + B_2F$, $C_F = C_1 + D_{12}F$, and consider the pencil

$$S = \left(\begin{bmatrix} A & 0 & B_2 \\ -C_1^tC_1 & I & -C_1^tD_{12} \\ D_{12}^tC_1 & 0 & D_{12}^tD_{12} \end{bmatrix}, \begin{bmatrix} I & 0 & 0 \\ 0 & A^t & 0 \\ 0 & -B^t & 0 \end{bmatrix} \right). \tag{2.4 – 7}$$

Assume that $S \in dom(Ric)$ and let $(Y, F) = Ric(S)$. Then, from Lemma 2.6,

$$B_2^tYA_F + D_{12}^tC_F = 0$$

$$Y = A_F^tYA_F + C_F^tC_F.$$

Let $R_b = (D_{12}^tD_{12} + B_2^tYB_2)^{1/2}$ and $R_{b\perp} = (D_{12\perp}^tD_{12\perp} + B_\perp YB_\perp)^{1/2}$, where $D_{12\perp}$, $B_\perp$ solve the system:

$$\begin{bmatrix} D_{12}^t & B_2^tY \\ C_F^t & A_F^t \end{bmatrix} \begin{bmatrix} D_{12\perp} \\ B_\perp \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

with $\begin{bmatrix} D_{12\perp} \\ B_\perp \end{bmatrix}$ full column rank and as many columns as the number of rows of C_F minus the number of columns of B_2.

Lemma 2.7 *Let T_{12} be as in 2.4-4, and assume that S in 2.4-7 belongs to dom(Ric). Consider the system $[T_{12}\ T_{12\perp}]$ with a DNF:*

$$\begin{bmatrix} T_{12\perp} & T_{12} \end{bmatrix} = \begin{pmatrix} -zY + YA_F & YB_\perp R_{b\perp}^{-1} & YB_2 R_b^{-1} \\ C_F & D_{12\perp} R_{b\perp}^{-1} & D_{12} R_b^{-1} \end{pmatrix}. \qquad (2.4-8)$$

Then $[T_{12}\ T_{12\perp}]$ is unitary.

Proof: With the previous definitions and after some algebra, it is possible to show that $[T_{12}\ T_{12\perp}]^{\sim}[T_{12}\ T_{12\perp}] = I$. Then the proof follows by construction, because the system is square.

□

Remark: the above realization is well defined even if R_b or $R_{b\perp}$ are singular.

A similar result holds for T_{21}. Let

$$S_1 = \left(\begin{bmatrix} A^t & 0 & C_2^t \\ -B_1 B_1^t & I & -B_1 D_{21}^t \\ D_{21} B_1^t & 0 & D_{21} D_{21}^t \end{bmatrix}, \begin{bmatrix} I & 0 & 0 \\ 0 & A & 0 \\ 0 & C_2 & 0 \end{bmatrix} \right). \qquad (2.4-9)$$

If $S_1 \in dom(Ric)$, let $(X, H) = Ric(S_1)$, $A_H = A + HC_2$, $B_H = B_1 + HD_{21}$ and Let $R_c \doteq (D_{21} D_{21}^t + C_2 X C_2^t)^{1/2}$. Moreover, let $C_\perp$, $D_{21\perp}$ solve the system:

$$\begin{bmatrix} C_\perp & D_{21\perp} \end{bmatrix} \begin{bmatrix} XA_H^t & XC_2^t \\ B_H^t & D_{21}^t \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}, \qquad (2.4-10)$$

with $\begin{bmatrix} C_\perp & D_{21\perp} \end{bmatrix}$ full row rank and as many rows as the number of columns of B_1 minus the number of rows of C_2, and $R_{c\perp} = (D_{21\perp} D_{21\perp}^t + C_\perp X C_\perp^t)^{1/2}$.

Lemma 2.8 *Let*

$$\begin{bmatrix} T_{21\perp} \\ T_{21} \end{bmatrix} = \begin{pmatrix} -zX + A_H X & B_H & 0 & 0 \\ R_{c\perp}^{-1} C_\perp X & R_{c\perp}^{-1} D_{21\perp} & -I & 0 \\ R_c^{-1} C_2 X & R_c^{-1} D_{21} & 0 & -I \end{pmatrix} \qquad (2.4-11)$$

where $A_H = A + HC$, $B_H = B_1 + HD_{21}$. Then 2.4-11 is unitary.

Since unitary transfer functions do not change the $\mathcal{H}_\infty$ norm, the following well known result holds:

$$\begin{aligned}
\|T_{11} - T_{12}QT_{21}\|_\infty &= \left\| [T_{12\perp}\; T_{12}]^\sim T_{11} \begin{bmatrix} T_{21\perp} \\ T_{21} \end{bmatrix}^\sim - \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix} \right\|_\infty \\
&= \left\| R - \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix} \right\|_\infty ,
\end{aligned} \tag{2.4-12}$$

where

$$R^\sim \doteq [T_{12\perp}\; T_{12}]^\sim T_{11} \begin{bmatrix} T_{21\perp} \\ T_{21} \end{bmatrix}^\sim \tag{2.4 - 13}$$

Procceding as in [19] it is possible to get the following realization for $G \doteq R^\sim$:

$$\begin{aligned}
G &= \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \\
&= \begin{pmatrix} -zX + A_H X & B_H & -A_H X F^t & 0 & 0 \\ R_{c\perp}^{-1} C_\perp X & R_{c\perp}^{-1} D_{21\perp} & -R_{c\perp}^{-1} C_\perp X F^t & 0 & -I \\ R_c^{-1} C_2 X & R_c^{-1} D_{21} & -R_c^{-1} C_2 X F^t & -I & 0 \end{pmatrix} \times \\
&\quad \begin{pmatrix} -zY + YA_F & YB_{2\perp} R_{b\perp}^{-1} & YB_2 R_b^{-1} & 0 & 0 \\ D_{11}^t C_F + B_1^t Y A_F & D_\alpha & D_\beta & -I & 0 \\ 0 & 0 & R_b & 0 & -I \end{pmatrix} ,
\end{aligned}$$

with $D_\alpha = (D_{11}^t D_{12\perp} + B_1^t Y B_{2\perp}) R_{b\perp}^{-1}$, $D_\beta = (D_{11}^t D_{12} + B_1^t Y B_2) R_b^{-1}$. Moreover, due to their structure, it is possible to reduce each one of the terms above to state-space form [67] and hence G always has a representation of the form:

$$G = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array} \right) \tag{2.4-14}$$

On the other hand, R is not guaranteed to have a SSF, since G may have some poles at zero.

From 2.4-12, the $\mathcal{H}_\infty$-Optimal Control Problem may be reformulated as the approximation problem:

Given $G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \in \mathcal{RH}_\infty$, *and a number* γ, *find all* $Q \in \mathcal{RH}_\infty$ *that achieve*

$$\left\| \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & Q^{\sim} \end{bmatrix} \right\|_\infty \leq \gamma. \qquad (2.4-15)$$

γ_o is said to be optimal if it is the infimum number for which 2.4-15 has a feasible solution.

When all the G_{ij}'s are non zero, 2.4-15 is called a "Four Block" problem [19]. When G_{11}, G_{12} and G_{21} are all simultaneously zero, the problem is called "One Block." It is well known that, while the latter has an easily computable solution in terms of the data, no simple closed-form solution exists to the former. However, given a constant γ it is possible to give necessary and sufficient conditions for the existence of $Q \in \mathcal{RH}_\infty$ that makes the norm in 2.4-12 less than or equal to γ.

2.5 Further Preliminary Results

2.5.1 Contractive Expansions

Let $T \in \mathcal{R}^{n \times m}$, and let $\|T\| = \overline{\sigma}(T)$. The matrix is said to be a contraction (strict contraction) if $\|T\| \leq 1 (< 1)$. If T is a contraction (strict contraction) then it is convenient to define the positive semi-definite (positive definite) matrix $\Delta_T \doteq (I - T^t T)^{1/2}$.

Theorem 2.2 *Let T be the matrix*

$$T = \begin{bmatrix} A & B \\ C & E \end{bmatrix} \qquad (2.5-1)$$

where A, B and C are matrices of appropriate dimension. Then T is a contraction if and only if $[A\ B]$ *and* $\begin{bmatrix} A \\ C \end{bmatrix}$ *are contractions and E is a matrix of the form*

$$E = -Z_1 A^t Z_2 + \Delta_{Z_1^t} U \Delta_{Z_2}, \tag{2.5-2}$$

with U an arbitrary contraction and Z_1 *and* Z_2 *are contractions that satisfy*

$$\begin{aligned} C &= Z_1 \Delta_A \\ B &= \Delta_{A^t} Z_2 \end{aligned} . \tag{2.5-3}$$

Suppose that instead of being E a real matrix, it is a matrix transfer function. Then Theorem 2.2 can be extended as follows.

Corollary 2.2 *Let T be the matrix*

$$T = \begin{bmatrix} A & B \\ C & E \end{bmatrix} \tag{2.5-4}$$

where A, B and C are matrices of appropriate dimension and E is a transfer matrix. Then T is a stable contraction (i.e., $T \in \mathcal{H}_\infty$ *and* $\|T\|_\infty \leq 1$*) if and only if* $[A\ B]$ *and* $\begin{bmatrix} A \\ C \end{bmatrix}$ *are contractions and E is a transfer matrix of the form 2.5-2 with U a stable contraction.*

Corollary 2.3 *With the notation of the theorem, the following equalities hold:*

$$I - \begin{bmatrix} A \\ C \end{bmatrix} \begin{bmatrix} A^t & C^t \end{bmatrix} = \begin{bmatrix} \Delta_{A^t} & 0 \\ -Z_1 A^t & \Delta_{Z_1^t} \end{bmatrix} \begin{bmatrix} \Delta_{A^t} & 0 \\ -Z_1 A^t & \Delta_{Z_1^t} \end{bmatrix}^t$$

$$I - \begin{bmatrix} A^t \\ B^t \end{bmatrix} \begin{bmatrix} A & B \end{bmatrix} = \begin{bmatrix} \Delta_A & 0 \\ -A Z_2 & \Delta_{Z_2} \end{bmatrix} \begin{bmatrix} \Delta_A & 0 \\ -A Z_2 & \Delta_{Z_2} \end{bmatrix}^t .$$

The theorem is the main result in [17]. The more general result given by the first corollary is taken from [24]. The second corollary is also from [24], and follows after some minor calculations.

2.5.2 All-pass Embedding

Theorem 2.3 *Let $G(z)$ be an pxm transfer function with a minimal state-space realization*

$$G(z) = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right),$$

and such that $\max_{\omega\in[0,\pi]} \overline{\sigma}(G(e^{j\omega})) < 1$. *Then there exist*

$$\begin{aligned} D_a &= \begin{matrix} & m & n & \\ & \left(\begin{matrix} D & D_{12} \\ D_{21} & D_{22} \end{matrix}\right) & & \begin{matrix} n \\ m \end{matrix} \end{matrix} \\ B_a &= (B \; B_2) \\ C_a &= \left(\begin{matrix} C \\ C_2 \end{matrix}\right), \end{aligned}$$

such that $G_a \doteq \left(\begin{array}{c|cc} A & B & B_2 \\ \hline C & D & D_{12} \\ C_2 & D_{21} & D_{22} \end{array}\right)$ *is an all-pass function, i.e.,*

$$G_a^{\sim} G_a = G_a G_a^{\sim} = I \qquad (2.5-5)$$

<u>Proof:</u> This result is the discrete time version of theorem 5.2 in [28], and can be proved from the latter by means of the bilinear transformation.

□

2.5.3 Computation of a Spectral Factor

In this section, the state-space formulae for a spectral factorization problem are reviewed. A more complete treatment and further references (although for the continuous time case) may be found in [25]. The first lemma is essentially taken from [28].

Lemma 2.9 *Let* $G = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \in \mathcal{RH}_\infty$ *with* p *rows and* m *columns. Then:*

a) Assume $p \leq m$ *and* G *controllable. Then* $GG^\sim = I$ *if and only if*

$$\begin{aligned} DB^t + CXA^t &= 0 \\ DD^t + CXC^t &= I \\ BB^t + AXA^t &= X. \end{aligned}$$

In this case, G *is said to be "co-inner."*

b) Assume $p \geq m$ *and* G *observable. Then* $G^\sim G = I$ *if and only if*

$$\begin{aligned} D^tC + B^tYA &= 0 \\ D^tD + B^tYB &= I \\ C^tC + A^tYA &= Y. \end{aligned}$$

In this case, G *is said to be "inner."*

Proof: See [36] or [63] for a more general proof.

□

Note that X and Y are the controllability and observability grammians of G. Next assume that $\|G\|_\infty < 1$ and $p \geq m$. From Lemma 2.3, there exist B_1 and D_1 such that $[G\ G_1] = \left(\begin{array}{c|cc} A & B & B_2 \\ \hline C & D & D_{12} \end{array}\right)$ is co-inner, i.e., G_1 is such that $G_1 G_1^\sim = I - GG^\sim$. The spectral factorization problem consists of computing G_1 such that G_1 is outer (i.e., stable with a stable inverse). Applying the previous lemma,

$$\begin{aligned} DB^t + D_1B_1^t + CXA^t &= 0 \\ DD^t + D_1D_1^t + CXC^t &= I \\ BB^t + B_1B_1^t + AXA^t &= X. \end{aligned} \qquad (2.5-6)$$

In particular, $D_1D_1^t = I - DD^t - CXC^t$, and the fact that $\|G\|_\infty < 1$ implies that $D_1D_1^t > 0$. Then, from 2.5-6,

$$X = AXA^t + (AXC^t + BD^t)(I - DD^t - CXC^t)^{-1}(CXA^t + DB^t) + BB^t. \quad (2.5-7)$$

Comparing 2.5-6 and 2.5-7 with Lemmas 2.3 and 2.6, $(X, -D_1^{-t}B_1^t) = Ric(S_X)$, where

$$S_X = \left(\begin{bmatrix} A^t & 0 & C^t \\ -BB^t & I & -BD^t \\ DB^t & 0 & DD^t - I \end{bmatrix}, \begin{bmatrix} I & 0 & 0 \\ 0 & A & 0 \\ 0 & -C & 0 \end{bmatrix} \right). \quad (2.5-8)$$

By Lemma 2.3, $A - B_1D_1^{-1}C$ is stable as required.

A similar result holds if $p \le m$. Let $\begin{bmatrix} G \\ G_1 \end{bmatrix} = \left[\begin{array}{c|c} A & B \\ \hline C & D \\ C_1 & D_1 \end{array}\right]$. Then C_1 and D_1 verify:

$$\begin{aligned} D^tC + D_1^tC_1 + B^tYA &= 0 \\ D^tD + D_1^tD_1 + B^tYB &= I \\ C^tC + C_1^tC_1 + A^tYA &= Y, \end{aligned} \quad (2.5-9)$$

and therefore $(Y, -D_1^{-1}C_1) = Ric(S_Y)$, where

$$S_Y = \left(\begin{bmatrix} A & 0 & B \\ -C^tC & I & -D^tC \\ C^tD & 0 & D^tD - I \end{bmatrix}, \begin{bmatrix} I & 0 & 0 \\ 0 & A^t & 0 \\ 0 & -B^t & 0 \end{bmatrix} \right), \quad (2.5-10)$$

and $A - BD_1^{-1}C_1$ is stable.

2.6 The One-Block Problem: Computation of a Solution

The one-block problem arises when, in 2.4-15, G_{11}, G_{12} and G_{21} are identically zero and $G_{22} = G$. Although the literature on the subject is abundant (see [25] and the references therein), the discrete-time case has received considerable less attention than the continuous-time one. Among the pure discrete-time derivations, Ball and Ran [5] parametrized all suboptimal solutions using the Ball-Helton interpolation theory under the additional assumption that the original system had no poles at zero. Moreover, the generator of all solutions falls short from having a simple dependence on the data of the problem, as compared with the continuous time counterpart. The exclusion of poles at zero also appeared in [36] and implicitly in [32].

In this section, a solution to the one-block problem that includes the case of having poles at the origin is worked out following the continuous time procedure in [45]. This route was also attempted in [32], but the results there are incomplete; in particular, poles at zero are implicitly excluded since their starting system is anti-stable and still has a state-space representation. Both [22] and [33] were also able to lift the additional assumption about the location of the poles, by using operator theoretic formulas (namely, the former used the Schur representation for the Commutant Lifting Theorem). The approach taken here is more convenient for studying the constrained $\mathcal{H}_\infty$ problem and, in contrast with [22], also holds for the optimal case. Also, the result highlights the relationship between the final formulas obtained through algebraic manipulations, and the one-step dilation procedure [1] which is central to the operator theory of the problem.

Consider the system $G(z)$ with a state-space representation:

$$G(z) = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \tag{2.6 - 1}$$

and assume that $\rho(A) < 1$. The problem considered in this section is the following:

given γ *find a* $Q(z) \in RH_\infty$ *such that* $\|G(z) - Q^{\sim}(z)\|_\infty \leq \gamma$.

This is called the suboptimal approximation problem. From Nehari's theorem, such a Q exists if and only if $\rho = \|G(z)\|_H < \gamma$ where $\|\cdot\|_H$ denotes the Hankel norm. Also the optimal approximation problem, i.e., finding a $Q(z) \in \mathcal{RH}_\infty$ such that

$\|G(z) - Q(z)^{\sim}\|_\infty = \rho$, is of interest, but this problem may be addressed by simply taking the limit $\gamma \downarrow \rho$ on the suboptimal solution.

To motivate the solution, consider the weighted $\mathcal{H}_2$ optimization problem:

$$\min_{Q \in \mathcal{RH}_\infty} \|[G - Q^{\sim}]W\|_2, \qquad (2.6-2)$$

where $G \in \mathcal{RH}_\infty$ and $W \in \mathcal{RH}_\infty^{\sim}$ such that $W^{-1} \in \mathcal{RH}_\infty^{\sim}$ are given. In view of the inner product structure of $\mathcal{L}_2$, 2.6-2 has an optimal solution given by $Q_{opt}^2 = [GW]_+ W^{-1}$, where $[\cdot]_+$ denotes the projection on $\mathcal{H}_\infty^{\sim}$. The central idea in the Wiener-Hopf approach to $\mathcal{H}_\infty$ control [64] followed here, is that it is possible to choose W such that $\min_{Q \in \mathcal{RH}_\infty} \|[G - Q^{\sim}]W\|_2 = \min_{Q \in \mathcal{RH}_\infty} \|G - Q^{\sim}\|_\infty$. With this motivation, consider a $W(z) \in RH_\infty^{\sim}$ with $W(z)^{-1} \in RH_\infty^{\sim}$ to be specified latter. Then

$$\begin{aligned} GW &= [GW]_- + [GW]_+ \\ GW &= ([GW]_- + D_1) + ([GW]_+ - D_1) \end{aligned} \qquad (2.6\text{-}3)$$

with D_1 a constant matrix also to be specified latter. Define

$$\begin{aligned} V_s &\doteq [GW]_- + D_1 \\ V_u &\doteq [GW]_+ - D_1, \end{aligned}$$

and let $E = G - Q^{\sim}$. Then $G = E + Q^{\sim}$, and

$$E \doteq V_s W^{-1} \qquad (2.6\text{-}4)$$

$$Q^{\sim} \doteq V_u W^{-1}. \qquad (2.6\text{-}5)$$

The free variables are now the system $W(z)$ and D_1. Note that $W(z)$ is anti-stable and hence need not have a state-space representation (i.e., it can have a polynomial term). Therefore assume that it has the following DNF:

$$W(z) = \begin{pmatrix} zA_w^t - I & C_w^t & 0 \\ zB_w^t & I & -I \end{pmatrix}$$

where A_w, B_w, C_w and I are the corresponding A, B, C and D matrices of a state-space representation of $W(z)^{\sim}$. Now impose the condition that [28]

$$E(z)^{\sim}E(z) = \gamma^2 I. \qquad (2.6-6)$$

From 2.6-4 this holds if and only if

$$V_s(z)^{\sim}V_s(z) = \gamma^2 W(z)^{\sim}W(z).$$

In the following construction, this condition is shown to be verified if $W(z)$ and D_1 are chosen appropriately. First $V_s(z)$ and $V_u(z)$ are computed as follows.

$$\begin{aligned} GW &= \begin{pmatrix} -zI+A & zBB_w^t & B & 0 \\ 0 & zA_w^t - I & C_w^t & 0 \\ C & zDB_w^t & D & -I \end{pmatrix} \\ &= \begin{pmatrix} -zI+A & zBB_w^t + zAX_1A_w^t - AX_1 & B + AX_1C_w^t & 0 \\ 0 & zA_w^t - I & C_w^t & 0 \\ C & z(DB_w^t + CX_1A_w^t) - CX_1 & D + CX_1C_w^t & -I \end{pmatrix} \\ &= \begin{pmatrix} -zI+A & z(BB_w^t + AX_1A_w^t - X_1) & B + AX_1C_w^t & 0 \\ 0 & zA_w^t - I & C_w^t & 0 \\ C & z(DB_w^t + CX_1A_w^t) & D + CX_1C_w^t & -I \end{pmatrix}. \end{aligned}$$

Taking X_1 so that

$$X_1 = AX_1A_w^t + BB_w^t, \qquad (2.6-7)$$

and introducing D_1,

$$GW = \begin{pmatrix} -zI+A & B + AX_1C_w^t & 0 \\ C & E_1 + CX_1C_w^t & -I \end{pmatrix} + \begin{pmatrix} zA_w^t - I & C_w^t & 0 \\ z(DB_w^t + CX_1A_w^t) & D_1 & -I \end{pmatrix}, \qquad (2.6-8)$$

where $E_1 \doteq D - D_1$. By comparison with 2.6-3, the first and second terms in 2.6-8 give a DNF of V_s and V_u respectively. Using similar manipulations, the following

DNF representation may be computed for $W^{\sim}W$ and $V_s^{\sim}V_s$:

$$W^{\sim}W = \begin{pmatrix} -zI + Aw & B_w + A_wX_2C_w^t & 0 \\ C_w & 0 & I \end{pmatrix} \tag{2.6-9}$$

$$+ \begin{pmatrix} -zI + Aw & B_w + A_wX_2C_w^t & 0 \\ C_w & 0 & I \end{pmatrix}^{\sim} (I + C_wX_2C_w^t), \tag{2.6-10}$$

where X_2 satisfies the Lyapunov equation

$$X_2 = A_wX_2A_w^t + B_wB_w^t. \tag{2.6 – 11}$$

Similarly

$$V_s^{\sim}V_s = \begin{pmatrix} -zI + A & B + AX_1C_w & 0 \\ E_1C^t + C_wX_1X_3 + BX_3A & 0 & -I \end{pmatrix} \tag{2.6-12}$$

$$+ \begin{pmatrix} -zI + A & B + AX_1C_w & 0 \\ E_1C^t + C_wX_1X_3 + BX_3A & 0 & -I \end{pmatrix}^{\sim}$$

$$+(E_1^t + C_wX_1C^t)(E_1^t + C_wX_1C^t)^t + (B^t + C_wX_1A^t)(B^t + C_wX_1A^t)^t,$$

where X_3 satisfies the Lyapunov equation

$$X_3 = A^tX_3A + C^tC. \tag{2.6 – 13}$$

That is, $X_3 = Y$, where Y denotes the observability grammian of the realization for $G(z)$. Condition 2.6-6 is satisfied if

$$A = A_w \tag{2.6 – 14}$$

$$B^t + C_w^tX_1A^t = B_w^t + C_wX_2A^t \tag{2.6 – 15}$$

$$\gamma^2C_w = E_1^tC + C_wX_1L_o + B^tL_oA \tag{2.6 – 16}$$

$$(E_1^t+C_wX_1C^t)(E_1^t+C_wX_1C^t)^t+(B^t+C_wX_1A^t)L_o(B^t+C_wX_1A^t)^t = \gamma^2(I+C_wX_2C_w^t). \tag{2.6 – 17}$$

Note that if $B = B_w$, then $X_1 = X_2 = X$, where X denotes the controllability grammian for the realization of $G(z)$, and 2.6-15 holds. Moreover, from 2.6-16

$$C_w = (E_1^t C + B^t Y A)(\gamma^2 I - XY)^{-1}, \qquad (2.6-18)$$

where the inverse exists if $\gamma > \|\Gamma_G\| = \rho(XY)$. Equation 2.6-17 is the main difference between the continuous and the discrete time derivation of the suboptimal solution. In the continuous time case one gets $D_{1c}^t D_{1c} = \gamma^2 I$ and hence D_{1c} may be chosen as γU_o where U_o is any unitary matrix of appropriate dimension (see [45]). In the discrete time case it is not obvious that a D_1 that solves 2.6-17 exists at all, and some lengthy matrix manipulation are needed to clarify that point. The final outcome is quite interesting, as it relates the pure state-space calculations with some deep results in interpolation theory.

Replacing 2.6-18 in 2.6-17 and after some calculations:

$$\begin{aligned} &E_1^t(I + CNXC^t)E_1 + B^tYANXC^tE_1 + E_1CNXA^tYB = \\ &= \gamma^2 I - B^tYB - B^tYANXA^tYB, \end{aligned}$$

where $N \doteq (\gamma^2 I - XY)^{-1}$. Let

$$\begin{aligned} M_1 &= (I + CNXC^t)^{\frac{1}{2}} \\ M_2^t &= M_1^{-1}CNXA^tYB \\ M_3 &= \gamma^2(I + B^tYNB)^{-1}. \end{aligned} \qquad (2.6-19)$$

Then 2.6-16,2.6-17 hold if and only if

$$(E_1^t M_1 + M_2)(E_1^t M_1 + M_2)^t = M_3. \qquad (2.6-20)$$

Since M_1 and M_3 are positive definite, it is now clear that a solution E_1 to 2.6-20 (and hence to 2.6-17) exists, and that any solution is given by

$$E_1 = M_1^{-1}[\gamma U(I + B^tYNB)^{-\frac{1}{2}} - M_2^t], \qquad (2.6-21)$$

where U is an arbitrary unitary matrix of appropriate dimension.

Summarizing,

$$W(z) = \begin{pmatrix} zA^t - I & N^t(C^tE_1 + A^tYB) & 0 \\ zB^t & I & -I \end{pmatrix} \qquad (2.6\text{-}22)$$

$$V_s(z) = \begin{pmatrix} -zI + A & B + ANX(C^tE_1 + A^tYB) & 0 \\ C & E_1 + CNX(C^tE_1 + A^tYB) & -I \end{pmatrix}$$

$$V_u(z) = \begin{pmatrix} zA^t - I & N^t(C^tD_1 + A^tYB) & 0 \\ z(DB^t + CXA^t) & D_1 & -I \end{pmatrix}.$$

From 2.6-22

$$W(z)^{-1} = \begin{pmatrix} z(A^t - C_w^tB_w^t) - I & C_w^t & 0 \\ -zB_w^t & I & -I \end{pmatrix},$$

and since $Q(z)^\sim = V_u(z)W(z)^{-1}$,

$$Q(z)^\sim = \begin{pmatrix} z(A^t - C_w^tB^t) - I & C_w^t & 0 \\ z(CXA^t + E_1B^t) & D_1 & -I \end{pmatrix}.$$

Note that W^{-1} is anti-stable if and only if $Q^\sim$ is anti-stable, and that Q has the SSF:

$$Q(z) = \left(\begin{array}{c|c} A - BC_w & AXC^t + BE_1^t \\ \hline C_w & D_1^t \end{array}\right).$$

Let

$$A_Q \doteq A - BC_w = A - B(D_1^tC + B^tYA)N \qquad (2.6-23)$$

and

$$B_Q \doteq AXC^t + BE_1^t. \qquad (2.6-24)$$

Stability of $Q(z)$ is established in the following lemma.

Lemma 2.10 *For any E_1 computed as a solution to 2.6-20, the matrix A_Q is asymptotically stable.*

Proof: patient computations show that the following Lyapunov equation holds:

$$N^{-1}X = A_Q N^{-1} X A_Q^t + B_Q B_Q^t \qquad (2.6-25)$$

where $N^{-1}X$ is symmetric and positive definite. Therefore, from standard stability theory, all the eigenvalues of A_Q are inside the closed unit disk. Moreover, if A_Q has an eigenvalue on the unit circle, it must be an uncontrollable mode of the pair (A_Q, B_Q). Now suppose that $e^{j\omega}$, $\omega \in [0, 2\pi]$, is such an eigenvalue. Then by the PBH test there exists a vector x such that $x^t A_Q = e^{j\omega} x^t$ and $x^t B_Q = 0$. But

$$\begin{aligned} A_Q &= A - BE_1^t CN - BB^t YAN \\ &= A - B_Q CN + (AXY - XYA)N \\ &= (\gamma^2 A - AXY + AXY - XYA)N - B_Q CN \\ &= N^{-1}AN - B_Q CN \end{aligned}$$

and hence $x^t B_Q = 0$ implies $x^t N^{-1} AN = e^{j\omega} x^t$, which contradicts $\rho(A) < 1$. Therefore $\rho(A_Q) < 1$.

□

The following result has been established.

Theorem 2.4 *Let $G(z) \in \mathcal{RH}_\infty$ have the minimal state-space representation 2.6-1, and assume that $\rho = \|G(z)\|_H < \gamma$. Let $Q(z) \in \mathcal{RH}_\infty$ have a SSF:*

$$Q = \left(\begin{array}{c|c} A_Q & B_Q \\ \hline C_Q & D_Q \end{array} \right) \qquad (2.6-26)$$

with

$$\begin{aligned} A_Q &= A - BC_Q \\ B_Q &= AXC^t + BE_1^t \\ C_Q &= (E_1^t C + B^t YA)N \\ D_Q &= D^t - E_1^t, \end{aligned}$$

where $N = (\gamma^2 I - XY)^{-1}$, *and for any unitary matrix* U,

$$E_1 = -(I + CNXC^t)^{-1} CNXA^tYB + \gamma(I + CNXC^t)^{-1/2} U (I + B^tYNB)^{-1/2}.$$

Then $\|G(z) - Q(z)^\sim\|_\infty = \gamma$ *and* $(G(z) - Q(z)^\sim)^\sim (G(z) - Q(z)^\sim) = \gamma^2 I.$

□

The following corollary allows the derivation of an optimal solution, i.e., the case $\gamma = \rho$.

Corollary 2.4 *In Theorem 2.4,* C_Q *can be taken as the solution of*

$$C_Q(\gamma^2 I - XY) = E_1^t C + B^tYA \qquad (2.6-27)$$

and

$$E_1 = -Z_1 X^{1/2} A^t Y^{1/2} Z_2 + \gamma(I - Z_1 Z_1^t)^{1/2} W (I - Z_2^t Z_2)^{1/2}, \qquad (2.6-28)$$

where Z_1 *and* Z_2 *satisfy:*

$$CX^{1/2} = Z_1(\gamma^2 I - X^{1/2}A^tYAX^{1/2})^{1/2} \qquad (2.6\text{-}29)$$

$$Y^{1/2}B = (\gamma^2 I - Y^{1/2}AXA^tY^{1/2})^{1/2} Z_2, \qquad (2.6\text{-}30)$$

and are such that $\overline{\sigma}(Z_1) \leq 1$, $\overline{\sigma}(Z_2) \leq 1$.

Proof: note that $\gamma > \rho$ implies that, for instance, $X^{1/2}A^tYAX^{1/2} < \gamma^2 I$. Therefore one can compute Z_1 and Z_2 from 2.6-29, 2.6-30 and replace in 2.6-28. Moreover, $xYx = xA^tYAx + xC^tCx < \gamma^2 I$ implies $\overline{\sigma}(Z_1) < 1$ and similarly $\overline{\sigma}(Z_2)$.

□

Corollary 2.5 *The expressions for* C_Q *and* E_1 *in Corollary 2.4 remain valid for the case* $\rho = \gamma$. *Indeed, 2.6-26 provides a non-minimal realization for an optimal solution, and this realization has at least as many non-controllable modes as the multiplicity of* ρ *as an eigenvalue of* XY

Proof: The proof of this corollary is based on the contractive completion of a matrix. Consider the matrix:

$$\hat{D}(E_1) \doteq \begin{bmatrix} Y^{1/2}AX^{1/2} & Y^{1/2}B \\ CX^{1/2} & E_1 \end{bmatrix}. \qquad (2.6-31)$$

Then since $\overline{\sigma}([Y^{1/2}AX^{1/2}\ Y^{1/2}B]) = \overline{\sigma}\left(\begin{bmatrix} Y^{1/2}AX^{1/2} \\ CX^{1/2} \end{bmatrix}\right) = \rho$, E_1 may be chosen so that $\overline{\sigma}(\hat{D}(E_1)) = \rho$. Moreover all E_1 that achieve this are of the form

$$E_1 \ = \ -Z_1X^{1/2}A^tY^{1/2}Z_2 + \rho(I - Z_1Z_1^t)^{1/2}U(I - Z_2^tZ_2)^{1/2},$$

where Z_1 and Z_2 are contractions that satisfy 2.6-29, 2.6-30 for $\gamma = \rho$. To prove that C_Q satisfying 2.6-27 exists, let ν be a unitary vector such that $X^{1/2}YX^{1/2}\nu = \rho^2\nu$, and compute

$$\hat{D}(E_1)^tD(E_1) = \begin{bmatrix} X^{1/2}YX^{1/2} & X^{1/2}(A^tYB + C^tE_1) \\ (B^tYA + E_1^tC)X^{1/2} & B^tYB + E_1^tE_1 \end{bmatrix}.$$

Since

$$\hat{D}(E_1)^tD(E_1)\begin{bmatrix} \nu \\ 0 \end{bmatrix} = \begin{bmatrix} X^{1/2}YX^{1/2}\nu \\ (B^tYA + E_1^tC)X^{1/2}\nu \end{bmatrix} = \begin{bmatrix} \rho^2\nu \\ (B^tYA + E_1^tC)X^{1/2}\nu \end{bmatrix},$$

$(B^tYA + E_1^tC)X^{1/2}\nu = 0$. But $X^{1/2}YX^{1/2}\nu = \rho^2\nu \Rightarrow (\rho^2I - XY)(X^{1/2}\nu) = 0$, and therefore $\mathrm{Ker}(\rho^2I - XY) \subset \mathrm{Ker}(B^tYA + E_1^tC)$. By elementary linear algebra, C_Q solving 2.6-27 exists. To complete the proof, note that from 2.6-25 the controllability grammian of $Q^{\sim}(z)$ is $(\gamma^2I - XY)X$, so that the null space of the grammian has the same dimension as the multiplicity of the largest eigenvalue of XY.

□

Remark: Corollary 2.5 shows that if there exist a E_1 such that the norm of the matrix in 2.6-31 remains less than or equal to γ, then there exist a solution. Given that $D_Q = D^t - D_1^t$, and D_Q is the first term in the Taylor's expansion of Q, this can be rephrased as follows. Let Q be expanded as $Q(z) = \sum_{i=0}^{\infty} Q_i z^{-i}$ Then if it is possible to compute the first Q_0, then it is possible to compute the remainder Q_i's and hence a stable Q that solves the suboptimal approximation problem. This was shown in [63] to be equivalent to the classical one-step extension method introduced by Adamjan, Arov and Krein [1,60].

The last result in the section is an alternative realization for Q, that follows from previous manipulations.

Corollary 2.6 *$Q(z)$ has the alternative state-space representation:*

$$Q(z) = \left(\begin{array}{c|c} A - N(AXC^t + BE_1^t)C & N(AXC^t + BE_1^t) \\ \hline E_1^t C + B^t Y A & D_1^t \end{array}\right). \qquad (2.6-32)$$

□

Realization 2.6-32 may again be modified to hold also for the optimal case.

2.7 The One-Block Problem: Parametrization of All Solutions

The approach taken in [28] to parametrize all solutions for the one-block problem is to construct an augmented all-pass error system, and then connect a contraction around the augmented system to generate all solutions. A similar approach may be pursued using the formulas derived in the previous section. First some results from [28] are briefly reviewed.

2.7.1 Linear Fractional Transformations

Consider the interconnection structure of Fig. 1. The transfer function between w and z is called a *linear fractional transformation (LFT) of K with coefficient matrix G* denoted

$$\mathcal{F}_l(G, K) = G_{11} + G_{12}K(I - G_{22}K)^{-1}G_{21}, \qquad (2.7-1)$$

where

$$G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}, \quad G_{11} : p_1 \times m_1 \,, \; G_{22} : p_2 \times m_2.$$

If K is connected around the upper loop, the resulting transfer function is denoted $\mathcal{F}_u(G,K) = G_{22} + G_{21}K(I - G_{11}K)^{-1}G_{12}$. An important property of LFT's is that any interconnection of LFT's is again an LFT. For example, given G_1 and G_2, they can be interconnected as in Fig. 2.2, and then

$$\mathcal{F}_l[G_1, \mathcal{F}_l(G_2, K)] = \mathcal{F}_l(\mathcal{F}_l(G_1, G_2), K).$$

To make the notation $\mathcal{F}_l(G_1, G_2)$ unambiguous, the sizes of u_1 and y_1 should be specified together with G_1 and G_2, but this additional information will always be apparent from the context. Let

$$G_1 = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & 0 \end{array}\right) \qquad G_2 = \begin{bmatrix} \hat{D}_{11} & \hat{D}_{12} \\ \hat{D}_{21} & 0 \end{bmatrix}.$$

Then the following formula is not hard to derive:

$$\mathcal{F}_l(G_1, G_2) = \left(\begin{array}{c|cc} A + B_2\hat{D}_{11}C_2 & B_1 + B_2\hat{D}_{11}D_{21} & B_2\hat{D}_{12} \\ \hline C_1 + D_{12}\hat{D}_{11}C_2 & D_{11} + D_{12}\hat{D}_{11}D_{21} & D_{12}\hat{D}_{12} \\ \hat{D}_{12}C_2 & \hat{D}_{21}D_{21} & 0 \end{array}\right) \qquad (2.7-2)$$

This formula is a clearly a special case of the generic composition of LFT's.

The following results are the discrete-time versions of Theorem 2.3, Corollary 2.6 and Lemma 2.7 in [34] (see also [28]).

Theorem 2.5 *Let* $det(I - P_{22}K)(\infty) \neq 0$ *and assume that* $rank(P_{21}(e^{j\omega})) = p_2 \; \forall \omega \in [0, \pi]$ *with* $P^{\sim}P = I$. *Then* $\|\mathcal{F}_l(P,K)\|_\infty < 1$ *if and only if* $\|K\|_\infty < 1$.

□

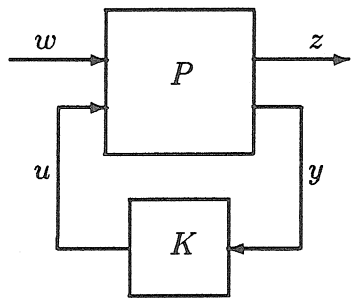

Figure 2.1: Linear Fractional Interconnection

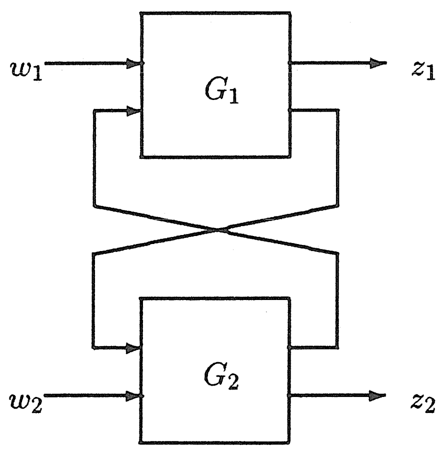

Figure 2.2: Composition of LFT's

Lemma 2.11 *Let P have the state-space representation*

$$P = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array} \right),$$

where $\mathrm{rank}\, D_{12} = m_2$, $\mathrm{rank}\, D_{21} = p_2$, $B_2 = B_{20} D_{12}$, $C_2 = D_{21} C_{20}$, *and let K have a minimal realization. Suppose that P has exactly k poles outside the unit disk, that* $\rho(A - B_{20} C_1) < 1$, $\rho(A - B_1 C_{20}) < 1$ *and let* $K \in RH_\infty$ *be such that* $\|KP_{22}\|_\infty < 1$. *Then* $\mathcal{F}_l(P, K)$ *has exactly k poles outside the unit disk.*

□

Lemma 2.12 *Let P and K be rational transfer function matrices, and let* $G = \mathcal{F}_l(P, K)$. *Then if P and G are proper,* $\det P(\infty) \neq 0$, $\det\left(P + \begin{bmatrix} G & 0 \\ 0 & 0 \end{bmatrix}\right)(\infty) \neq 0$ *and* P_{12} *and* P_{21} *are square and invertible for almost all z,* $|z| = 1$. *Then K is proper and* $K = \mathcal{F}_u(P^{-1}, G)$.

□

Parametrization of all Suboptimal Solutions

Consider first the suboptimal case $\gamma > \rho$. The first step in the construction is to augment G(z) as:

$$G_a = \begin{pmatrix} G & 0 \\ 0 & 0 \end{pmatrix}. \qquad (2.7-3)$$

A natural state-space representation for $G_a(z)$ is $G_a(z) = \left(\begin{array}{c|c} A & B_a \\ \hline C_a & D_a \end{array} \right)$, with $B_a = [B\ 0]$, $C_a = \begin{bmatrix} C \\ 0 \end{bmatrix}$ and $D_a = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$. Replacing in 2.6-19:

$$M_{a1} = \begin{bmatrix} M_1 & 0 \\ 0 & I \end{bmatrix}$$

$$M_{a2} = \begin{bmatrix} M_2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$M_{a3} = \begin{bmatrix} M_3 & 0 \\ 0 & \gamma^2 I \end{bmatrix}.$$

Partitioning the corresponding arbitrary unitary matrix U_a as $U_a = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$, 2.6-21 yields:

$$\begin{bmatrix} E_{11} & -D_{12} \\ -D_{21} & -D_{22} \end{bmatrix} = \begin{bmatrix} M_1^{-1}(\gamma U_{11} M_3^{\frac{1}{2}} - M_2^t) & \gamma M_1^{-1} U_{12} \\ \gamma U_{21} M_3^{\frac{1}{2}} & \gamma U_{22} \end{bmatrix}.$$

For reasons to be clarified below, D_{12} and D_{21} must be non-singular. Equivalently, U_{12} and U_{21} must be invertible. In particular, if $U_{11} = U_{22} = 0$ and $U_{12} = U_{21} = I$, then

$$\begin{bmatrix} E_{11} & -D_{12} \\ -D_{21} & -D_{22} \end{bmatrix} = \begin{bmatrix} -M_1^{-1} M_2^t & \gamma M_1^{-1} \\ \gamma M_3^{\frac{1}{2}} & 0 \end{bmatrix}.$$

From 2.6-26, the state-space representation for Q is:

$$Q_a(z) = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \left(\begin{array}{c|cc} A_Q & B_{Q1} & B_{Q2} \\ \hline C_{Q1} & D_{Q11} & D_{Q12} \\ C_{Q2} & D_{Q21} & D_{Q22} \end{array}\right)$$

$$= \left(\begin{array}{c|cc} A - B(B^tYA + E_{11}^tC)N & (AXC^t + BE_{11}^t) & -BD_{21}^t \\ \hline (E_{11}^tC + B^tYA)N & D_{11}^t & D_{21}^t \\ -D_{12}^tCN & D_{12}^t & D_{22}^t \end{array}\right). \quad (2.7\text{-}4)$$

The following properties of 2.7-4 are instrumental in proving the main result of the section:

$$A_Q - B_{Q1} D_{Q12}^{-1} C_{Q2} = NAN^{-1} \qquad (2.7-5)$$

and

$$A_Q - B_{Q2} D_{Q21}^{-1} C_{Q1} = A; \qquad (2.7-6)$$

therefore both Q_{12} and Q_{21} have a stable inverse. But then they have no zero on the unit circle and in particular they are full rank on the unit circle.

Theorem 2.6 *Given $G(z) \in RH_\infty$ with minimal state-space realization 2.6-1 and a number $\gamma > \|\Gamma_G(z)\|$. Then all $Q(z) \in RH_\infty$ such that*

$$\|G(z) - Q(z)^\sim\|_\infty \leq \gamma \qquad (2.7-7)$$

are given by

$$Q(z) = \mathcal{F}_l(Q_a(z), \Phi(z)), \ \Phi(z) \in RH_\infty, \ \|\Phi(z)\| < 1/\gamma, \qquad (2.7-8)$$

with $Q_a(z)$ defined as in 2.7-4.

Proof: Assume without loss of generality that $\gamma = 1$ (otherwise the problem can be scaled so that this assumption holds). Let $G(z)^\sim - Q(z) = \mathcal{F}_l(G_a(z)^\sim - Q_a(z), \Phi(z))$. Then, since $(G_a - Q_a^\sim)(G_a^\sim - Q_a) = I$ and, from 2.7-6, $(G_a - Q_a^\sim)_{21} = Q_{a12}^\sim$ has full rank on the unit circle, then Theorem 2.5 implies that $\|G(z)^\sim - Q(z)\|_\infty \leq 1$ if and only if $\|\Phi(z)\|_\infty < 1$. Hence, any $\Phi(z)$ in the conditions of the theorem will yield a $Q(z)$ that satisfies 2.7-7. To prove the converse, suppose that $Q(z) \in \mathcal{RH}_\infty$ is such that 2.7-7 holds. Consider the equation $Q = \mathcal{F}_l(Q_a, \Phi)$. The invertibility of D_{12} and D_{21}, and

$$Q_a(\infty) = \begin{pmatrix} D^t - D_{11}^t & D_{21}^t \\ D_{12}^t & 0 \end{pmatrix},$$

imply both $det\,(Q_a(\infty)) \neq 0$ and $det\left(Q_a(\infty) + \begin{bmatrix} Q(\infty) & 0 \\ 0 & 0 \end{bmatrix}\right) \neq 0$. Moreover, from 2.7-5 and 2.7-6, Q_{a12} and Q_{a21} are invertible for all z, $|z| = 1$. Therefore, the hypothesis in Lemma *3* hold and there exists a proper Φ such that $\mathcal{F}_l(Q_a(z), \Phi(z)) = Q(z)$. From 2.7-6 $Q_{21}(z)$ has full rank on the unit circle and hence theorem 2.5 implies that $\|\Phi(z)\| < 1$. Finally Lemma 2.12 and 2.7-5, 2.7-6 imply that $\Phi(z)$ has to be stable, and the proof is completed.

□

Corollary 2.7 *Q_a defined by 2.7-4 has the alternative realization:*

$$Q_a = \left(\begin{array}{c|cc} AM & AMNXC_1^t & -\gamma NB(I+B^tYNB)^{-\frac{1}{2}} \\ \hline B^tYAM & B^tYAMNXC^t + D^t & \gamma(I+B^tYNB)^{-\frac{1}{2}} \\ \gamma(I+CNXC^t)^{-\frac{1}{2}}C & -\gamma(I+CNXC^t)^{-\frac{1}{2}} & 0 \end{array} \right) \tag{2.7-9}$$

with $M \doteq (I + NXC^tC)^{-1}$.

□

The central solution

The central solution (i.e., the one obtained by setting Φ to zero) has some interesting properties which are easy to derive from its realization.

Theorem 2.7 *Consider the central solution Q_c with a state-space realization*

$$Q_c = \left(\begin{array}{c|c} A(I+NXC^tC)^{-1} & A(I+NXC^tC)^{-1}NXC^t \\ \hline B^tYA(I+NXC^tC)^{-1} & B^tYA(I+NXC^tC)^{-1}NXC^t + D^t \end{array} \right). \tag{2.7-10}$$

Then:

a) If A is invertible, then $Q_c^{\sim}$ has a state-space realization

$$Q_c^{\sim} = \left(\begin{array}{c|c} A^{-t}(I+C^tCNX) & YB \\ \hline CNX & D \end{array} \right). \tag{2.7-11}$$

b) If the realization 2.7-10 for Q_c has some poles at zero (i.e., A_Q is singular), then the poles are non controllable. Therefore $Q_c^{\sim}$ always has a state-space representation.

c) Assume that $(\gamma^2 I - XA^tYA)$ is nonsingular (note that this is generically true). Then realization 2.7-10 may be expressed as

$$Q_c = \left(\begin{array}{c|c} A(\gamma^2 I - XA^tYA)^{-1}N^{-1} & A(\gamma^2 I - XA^tYA)^{-1}XC^t \\ \hline B^tYA(\gamma^2 I - XA^tYA)^{-1}N^{-1} & B^tYA(\gamma^2 I - XA^tYA)^{-1}XC^t \end{array} \right), \tag{2.7-12}$$

that makes sense even for the case $\gamma = \rho$.

Proof: the proof is straightforward.

□

Parametrization of all Optimal Solutions

The parametrization of all optimal solutions to the approximation problem may be obtained from the suboptimal case, using Corollaries 2.4 and 2.5. Let $Z_{1a} \doteq \begin{bmatrix} Z_1 \\ 0 \end{bmatrix}$ and $Z_{2a} \doteq [Z_2 \; 0]$, with Z_1 and Z_2 defined in corollary 2.5, and let W_a be as above. Then

$$\begin{aligned} E_a &= \begin{bmatrix} E_{11} & -D_{12} \\ -D_{21} & -D_{22} \end{bmatrix} \\ &= \begin{bmatrix} -Z_1 X^{1/2} A^t Y^{1/2} Z_2 & 0 \\ 0 & 0 \end{bmatrix} + \\ &\qquad \gamma \begin{bmatrix} (I - Z_1 Z_1^t)^{1/2} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} (I - Z_2^t Z_2)^{1/2} & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

and therefore

$$\begin{aligned} E_{11} &= -Z_1 X^{1/2} A^t Y^{1/2} Z_2 \\ -D_{12} &= \gamma (I - Z_1 Z_1^t)^{1/2} \\ -D_{21} &= \gamma (I - Z_2^t Z_2)^{1/2} \\ -D_{22} &= 0. \end{aligned}$$

Theorem 2.8 *(Parametrization of all Solutions)*

Let $G(z) \in \mathcal{RH}_\infty$ with minimal state-space realization 2.6-1 be such that $\|G(z)\|_H = \rho$, and let $\gamma \geq \rho$. Let C_{Q1}, C_{Q2} solve the equations

$$\begin{aligned} C_{Q1}(\gamma^2 I - XY) &= E_{11}^t C + B^t Y A \\ C_{Q2}(\gamma^2 I - XY) &= (I - Z_1 Z_1^t)^{1/2} C, \end{aligned}$$

where $E_{11} = -Z_1 X^{\frac{1}{2}} A^t Y^{\frac{1}{2}} Z_2$ and Z_1, Z_2 are contractions that satisfy:

$$\begin{aligned} C X^{1/2} &= Z_1(\gamma^2 I - X^{1/2} A^t Y A X^{1/2})^{1/2} \\ Y^{1/2} B &= (\gamma^2 I - Y^{1/2} A X A^t Y^{1/2})^{1/2} Z_2. \end{aligned}$$

Finally, let

$$Q_a = \left(\begin{array}{c|cc} A - BC_{Q1} & AXC^t + BE_{11}^t & B(I - Z_2^t Z_2)^{1/2} \\ \hline C_{Q1} & D^t - E_{11}^t & -(I - Z_2^t Z_2)^{1/2} \\ C_{Q2} & -(I - Z_1 Z_1^t)^{1/2} & 0 \end{array} \right) \in \mathcal{RH}_\infty. \qquad (2.7-13)$$

Then $\|G(z) - Q(z)^\sim\|_\infty \leq \gamma$ if and only if $Q(z) = \mathcal{F}_l(Q_a(z), \Phi(z))$ for some $\Phi(z) \in \mathcal{RH}_\infty$, $\|\Phi(z)\|_\infty \leq \gamma$.

□

Theorem 2.9 *(Alternative Parametrization)*

Let $G(z) \in \mathcal{RH}_\infty$ with minimal state-space realization 2.6-1 be such that $\|G(z)\|_H = \rho$, and let $\gamma \geq \rho$. Then $\|G(z) - Q(z)^\sim\|_\infty \leq \gamma$ if and only if

$$Q = \left(\begin{array}{c|c} A_Q & B_Q \\ \hline C_Q & D_Q \end{array} \right),$$

with

$$\begin{aligned} A_Q &= A - BC_Q \\ B_Q &= AXC^t + BE_1(z)^t \\ C_Q(\gamma^2 I - XY) &= E_1(z)^t C + B^t Y A \\ D_Q &= D^t - E_1(z)^t \end{aligned}$$

and E_1 is a stable transfer function such that the matrix function

$$\hat{D}(E_1(z)) = \begin{bmatrix} Y^{1/2}AX^{1/2} & Y^{1/2}B \\ CX^{1/2} & E_1(z) \end{bmatrix}$$

has $\|\hat{D}(E_1)\|_\infty \leq \gamma$.

Proof: This result follows from Theorem 2.8 after parametrizing all stable E_1 that give $\overline{\sigma}(\hat{D}(E_1)) \leq \gamma$.

□

This last parametrization will be the basic tool for proving necessary and sufficient conditions for both the unconstrained and the constrained four-block $\mathcal{H}_\infty$ control problems.

2.8 The Four-Block Problem

Although the literature on the discrete-time $\mathcal{H}_\infty$ control problem is now abundant [39,69,27,48,50], all recent works do not require the solution to a four-block problem. For example, [39] is the discrete-time counterpart of [18], and gives necessary and sufficient conditions for the existence of controllers that achieve an $\mathcal{H}_\infty$-norm bound, and a parametrization of all such controllers directly in terms of the data for the problem.

In the next two sections, the discrete-time counterpart to [45] or [35] is worked out since, as shown in the next chapter, the formulation of the approximation problem is needed to transform time domain constrained $\mathcal{H}_\infty$ control into a tractable optimization problem. The approach has the further advantage of clarifying the role of two standard assumptions on the existing discrete time theory, namely the full column and rank conditions imposed on the feed-through terms from the control inputs to the controlled outputs or from the disturbances to the measurements. It turns out that these conditions are not necessary and this is a relevant fact since, as stated before, the condition may be too stringent in practice.

2.8.1 Necessary Condition

Solving a four-block instance of the optimal approximation problem is essentially harder than solving the one-block case; for instance, while the optimal value of the latter can be easily found by computing the Hankel norm, no simple closed-form expression exists for the former. However, if instead of the optimal solution one is contented with finding a solution to Problem 2.4-15 for a fixed γ or establish that none exists, then it is possible to reduce the general instance of the problem to the one-block case; this is the route followed in this section. Set again $\gamma = 1$, and assume in what follows that:

$$\max \left\{ \|G_{11} \; G_{12}\|_\infty, \left\| \begin{matrix} G_{11} \\ G_{21} \end{matrix} \right\|_\infty \right\} < 1. \tag{2.8 - 1}$$

This is again without loss of generality if one is only interested in strictly suboptimal solutions.

Let $G \in \mathcal{RH}_\infty$, with

$$G = \begin{matrix} & \begin{matrix} m_1 & m_2 \end{matrix} & \\ & \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} & \begin{matrix} p_1 \\ p_2 \end{matrix} \end{matrix} .$$

The objective is to find necessary and sufficient conditions for the existence of $Q \in \mathcal{RH}_\infty$ such that

$$\left\| \begin{matrix} G_{11} & G_{12} \\ G_{21} & G_{22} - Q^\sim \end{matrix} \right\| \leq 1, \tag{2.8 - 2}$$

and parametrize all such solutions. Suppose that there exist $Q \in \mathcal{RH}_\infty$ such that

2.8-2 holds, and let $Q^{\sim} = \left(\begin{array}{c|c} A_{\tilde{Q}} & B_{\tilde{Q}} \\ \hline C_{\tilde{Q}} & D_{\tilde{Q}} \end{array}\right)$. Define:

$$
\begin{aligned}
E &= \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} - Q^{\sim} \end{bmatrix} \\
&= \left(\begin{array}{cc|cc} A & 0 & B_1 & B_2 \\ 0 & A_{\tilde{Q}} & 0 & B_{\tilde{Q}} \\ \hline C_1 & 0 & D_{11} & D_{12} \\ C_2 & -C_{\tilde{Q}} & D_{21} & -D_{\tilde{Q}} \end{array}\right).
\end{aligned}
$$

Then, by Theorem 2.3, there exist matrices $\hat{B}_i$, $\hat{C}_i$, $B_{\tilde{Q}i}$, $C_{\tilde{Q}i}$ and $\hat{D}_{ij}$ such that E_{aa} is all-pass, where

$$
\begin{aligned}
E_{aa} &= \begin{pmatrix} G_{11} & G_{12} & E_{13} & E_{14} \\ G_{21} & G_{22} - Q^{\sim} & E_{23} & E_{24} \\ E_{31} & E_{32} & E_{33} & E_{34} \\ E_{41} & E_{42} & E_{43} & E_{44} \end{pmatrix} \\
&= \left(\begin{array}{cc|cccc} & & m_1 & m_2 & p_1 & p_2 \\ A & 0 & B_1 & B_2 & \hat{B}_3 & \hat{B}_4 \\ 0 & A_{\tilde{Q}} & 0 & B_{\tilde{Q}} & B_{\tilde{Q}3} & B_{\tilde{Q}4} \\ \hline C_1 & 0 & D_{11} & D_{12} & \hat{D}_{13} & \hat{D}_{14} \\ C_2 & -C_{\tilde{Q}} & D_{21} & -D_{\tilde{Q}} & \hat{D}_{23} & \hat{D}_{24} \\ \hat{C}_3 & -C_{\tilde{Q}3} & \hat{D}_{31} & \hat{D}_{32} & \hat{D}_{33} & \hat{D}_{34} \\ \hat{C}_4 & -C_{\tilde{Q}4} & \hat{D}_{41} & \hat{D}_{42} & \hat{D}_{43} & \hat{D}_{44} \end{array}\right) \begin{array}{c} \\ \\ \\ p_1 \\ p_2 \\ m_1 \\ m_2 \end{array}.
\end{aligned}
\tag{2.8-3}
$$

Remark: In 2.8-3 it has been implicitly assumed that $Q^{\sim}$ has a state-space representation. As discussed in [63], this assumption may fail to hold. Although this loss of generality can be removed using the representation in Section 2.2, doing so would obscure the derivations. Moreover, the results remain valid even when the assumption is violated, since all calculations are done in terms of a state-space representation of Q.

By the all-pass character of E_{aa},

$$[G_{11}\ G_{12}]\begin{bmatrix} G_{11}^{\sim} \\ G_{12}^{\sim} \end{bmatrix} + [E_{13}\ E_{14}]\begin{bmatrix} E_{13}^{\sim} E_{14}^{\sim} \end{bmatrix} = I. \qquad (2.8-4)$$

Consider now an inner-outer factorization for $[E_{13}\ E_{14}]$, i.e., $[E_{13}\ E_{14}] = G_{13}[E_{13}^{i}\ E_{14}^{i}]$, with G_{13} outer (i.e., stable with a stable inverse) and $[E_{13}^{i}\ E_{14}^{i}]$ inner, i.e.,

$$[E_{13}^{i}\ E_{14}^{i}][E_{13}^{i}\ E_{14}^{i}]^{\sim} = I.$$

Similarly

$$[G_{11}^{\sim}\ G_{21}^{\sim}]\begin{bmatrix} G_{11} \\ G_{21} \end{bmatrix} + [E_{31}^{\sim} E_{41}^{\sim}]\begin{bmatrix} E_{31} \\ E_{41} \end{bmatrix} = I, \qquad (2.8-5)$$

and factorize $\begin{bmatrix} E_{31} \\ E_{41} \end{bmatrix} = \begin{bmatrix} E_{31}^{i} \\ E_{41}^{i} \end{bmatrix} G_{31}$, with G_{31} outer and $\begin{bmatrix} E_{31}^{i} \\ E_{41}^{i} \end{bmatrix}^{\sim} \begin{bmatrix} E_{31}^{i} \\ E_{41}^{i} \end{bmatrix} = I.$

Define:

$$\begin{aligned} E_a &\doteq \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & E_{31}^{i}{}^{\sim} & E_{41}^{i}{}^{\sim} \end{bmatrix} E_{aa} \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & E_{13}^{i}{}^{\sim} \\ 0 & 0 & E_{14}^{i}{}^{\sim} \end{bmatrix} \\ &= \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} - Q^{\sim} & \hat{E}_{23} \\ G_{31} & \hat{E}_{32} & \hat{E}_{33} \end{bmatrix}. \end{aligned}$$

Given that $\|E_{aa}\| = 1$ and the matrices multiplying on the left and right have by construction unitary norm, it follows that $\|E_a\| \leq 1$. Therefore, Nehari's Theorem [25] implies that if G_a denotes the stable part of E_a then $\|G_a\|_H \leq 1$. Moreover, Lemma 2.1 implies that G_a has a realization of the form:

$$G_a = \left(\begin{array}{c|ccc} A & B_1 & B_2 & B_3 \\ \hline C_1 & D_{11} & D_{12} & D_{13} \\ C_2 & D_{21} & E_{22} & E_{23} \\ C_3 & D_{31} & E_{32} & E_{33} \end{array}\right). \tag{2.8-6}$$

Now note that since G_{13} and G_{31} are uniquely determined as the outer solutions to the spectral factorization problems

$$G_{13}G_{13}^{\sim} = I - G_{11}G_{11}^{\sim} - G_{12}G_{12}^{\sim}$$

and

$$G_{31}^{\sim}G_{31} = I - G_{11}^{\sim}G_{11} - G_{21}^{\sim}G_{21}.$$

From Section 2.5.3, it is possible to compute the unknown terms B_3, C_3, D_{13} and D_{31} *independently* of Q. Then 2.5-6 and 2.5-9 imply that:

$$\begin{array}{rcl} D_{11}B_1^t + D_{12}B_2^t + D_{13}B_3^t + C_1XA^t & = & 0 \\ D_{11}D_{11}^t + D_{12}D_{12}^t + D_{13}D_{13}^t + C_1XC_1^t & = & I \\ D_{11}^tC_1 + D_{21}^tC_2 + D_{31}^tC_3 + B_1^tYA & = & 0 \\ D_{11}^tD_{11} + D_{21}^tD_{21} + D_{31}^tD_{31} + B_1^tYB_1 & = & I \end{array}, \tag{2.8-7}$$

and therefore

$$D_{13} \doteq (I - D_{11}D_{11}^t - D_{12}D_{12}^t - C_1XC_1^t)^{1/2} \tag{2.8-8}$$

$$B_3 \doteq -(AXC_1^t + B_1D_{11}^t + B_2D_{12}^t)D_{13}^{-t} \tag{2.8-9}$$

$$D_{31} \doteq (I - D_{11}^tD_{11} - D_{21}^tD_{21} - B_1^tYB_1)^{1/2} \tag{2.8-10}$$

$$C_3 \doteq -D_{31}^{-t}(B_1^tYA + D_{11}^tC_1 + D_{21}^tC_2). \tag{2.8-11}$$

Moreover, X and Y satisfy:

$$X = AXA^t + B_1B_1^t + B_2B_2^t + B_3B_3^t \tag{2.8-12}$$

$$Y = A^tYA + C_1^tC_1 + C_2^tC_2 + C_3^tC_3. \tag{2.8-13}$$

From Section 2.5.3, it follows that $(X, -D_{13}^{-t}B_3^t) = ric(S_X)$ and $(Y, -D_{31}^{-1}C_3) = ric(S_Y)$, where:

$$S_X = \left(\begin{bmatrix} A^t & 0 & C_1^t \\ -(B_1B_1^t + B_2B_2^t) & I & -(B_1D_{11}^t + B_2D_{12}^t) \\ D_{11}B_1^t + D_{12}B_2^t & 0 & D_{11}D_{11}^t + D_{12}D_{12}^t - I \end{bmatrix}, \begin{bmatrix} I & 0 & 0 \\ 0 & A & 0 \\ 0 & -C_1 & 0 \end{bmatrix} \right) \quad (2.8\text{-}14)$$

$$S_Y = \left(\begin{bmatrix} A & 0 & B_1 \\ -(C_1^tC_1 + C_2^tC_2) & I & -(D_{11}^tC_1 + D_{21}^tC_2) \\ C_1^tD_{11} + C_2^tD_{21} & 0 & D_{11}^tD_{11} + D_{21}^tD_{21} - I \end{bmatrix}, \begin{bmatrix} I & 0 & 0 \\ 0 & A^t & 0 \\ 0 & -B_1^t & 0 \end{bmatrix} \right). \quad (2.8\text{-}15)$$

In particular, X and Y are positive definite, and $A - B_3D_{13}^{-1}C_1$ and $A - B_1D_{31}^{-1}C_1$ are stable. Note that Assumption 2.8-1 implies that all the terms are well defined; in particular $D_{13}D_{13}^t > 0$ and $D_{31}^tD_{31} > 0$ or equivalently:

$$(I - C_1XC_1^t - D_{11}D_{11}^t - D_{12}D_{12}^t)^{-1} \; > \; 0 \quad (2.8\text{-}16)$$

$$(I - B_1^tYB_1 - D_{11}^tD_{11} - D_{21}^tD_{21})^{-1} \; > \; 0. \quad (2.8\text{-}17)$$

For future reference, x and y denote the positive square roots of X and Y respectively. The result obtained so far can be summarized as follows.

Lemma 2.13 *(Necessary Condition) Let* $G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$, *have a state-space realization:*

$$G = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & 0 \end{array} \right) \in \mathcal{RH}_\infty$$

and assume that 2.8-1 is satisfied. If there exists $Q \in \mathcal{RH}_\infty$ *such that 2.8-2 holds, then* S_X, $S_Y \in dom(Ric)$, *and their associated solutions* X *and* Y *respectively are such that* $\rho(XY) \leq 1$.

Proof: The fact that $S_X,\ S_Y \in dom(Ric)$ was established in Section 2.5.3. From 2.8-12 and 2.8-13, X and Y are the controllability and observability grammians of G_a respectively. By Theorem 2.4, the existence of a solution then implies that $\rho(XY) \leq 1$, since the Hankel norm needs to be less than or equal to one.

□

2.8.2 Sufficient Condition and Parametrization of the Solutions

The fact that the conditions discussed in the lemma are necessary is quite natural; what is surprising is that they are also sufficient. To see this, assume that the conditions hold, construct B_3, D_{13}, C_3 and D_{31} as above, and let

$$G_a = \left(\begin{array}{c|ccc} A & B_1 & B_2 & B_3 \\ \hline C_1 & D_{11} & D_{12} & D_{13} \\ C_2 & D_{21} & 0 & 0 \\ C_3 & D_{31} & 0 & 0 \end{array}\right). \qquad (2.8-1)$$

By construction, equations 2.8-7, 2.8-12 and 2.8-13 hold. The condition $\rho(XY) \leq 1$ implies that there exists $Q_a(z) \in \mathcal{RH}_\infty$ such that $\|\mathcal{E}_a(z)\|_\infty = \|G_a(z) - Q_a(z)^\sim\|_\infty \leq 1$, and hence it only needs to be shown that there exists a solution of the form:

$$Q_a(z) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & Q_{22} & Q_{23} \\ 0 & Q_{32} & Q_{33} \end{pmatrix}; \qquad (2.8-2)$$

$Q = Q_{22}$ then gives a solution to the original problem. Moreover, a parametrization of all solutions of the form 2.8-2 yields a parametrization of all solutions to the four-block problem. Let $B = [B_1\ B_2\ B_3]$, $C = [C_1^t\ C_2^t\ C_3^t]^t$, $D = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & 0 & D_{23} \\ D_{31} & D_{32} & D_{33} \end{bmatrix}$ and

$N = (I - XY)^{-1}$. By Theorem 2.9, $Q_a = \mathcal{F}_l(Q_{aa}, E^t)$, where

$$Q_{aa} = \left(\begin{array}{c|cc} A - NAXC^tC & NAXC^t & -NB \\ \hline B^tYA & D^t & I \\ C & -I & 0 \end{array}\right) \qquad (2.8\text{-}3)$$

and E is such that

$$\begin{aligned} D_a(z) &= \begin{bmatrix} yAx & yB \\ Cx & E(z) \end{bmatrix} \\ &= \begin{bmatrix} yAx & yB_1 & yB_2 & yB_3 \\ C_1x & E_{11}(z) & E_{12}(z) & E_{13}(z) \\ C_2x & E_{21}(z) & E_{22}(z) & E_{33}(z) \\ C_3x & E_{31}(z) & E_{32}(z) & E_{33}(z) \end{bmatrix} \end{aligned}$$

is a contraction. From 2.8-3, $Q(\infty)^t = D - E(\infty)$. Therefore, if $Q(z)$ is to be of the form 2.8-2, then

$$E(\infty) = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & 0 & D_{23} \\ D_{31} & D_{32} & D_{33} \end{bmatrix} \qquad (2.8-4)$$

and

$$D_a(\infty) = \begin{bmatrix} yAx & yB_1 & yB_2 & yB_3 \\ C_1x & D_{11} & D_{12} & D_{13} \\ C_2x & D_{21} & 0 & D_{33} \\ C_3x & D_{31} & D_{32} & D_{33} \end{bmatrix}. \qquad (2.8-5)$$

Since $D_a(z)$ has to be a contraction, and

$$\begin{bmatrix} C_1x & D_{11} & D_{12} & D_{13} \end{bmatrix} \begin{bmatrix} xC_1^t \\ D_{11}^t \\ D_{12}^t \\ D_{13}^t \end{bmatrix} = I$$

$$\begin{bmatrix} B_1^t y & D_{11}^t & D_{21}^t & D_{31}^t \end{bmatrix} \begin{bmatrix} yB_1 \\ D_{11} \\ D_{21} \\ D_{31} \end{bmatrix} = I,$$

then necessarily,

$$\begin{aligned} \begin{bmatrix} E_{11}(z) & E_{12}(z) & E_{13}(z) \end{bmatrix} &= \begin{bmatrix} D_{11} & D_{12} & D_{13} \end{bmatrix} \\ \begin{bmatrix} E_{11}(z)^t & E_{21}(z)^t & E_{31}(z)^t \end{bmatrix} &= \begin{bmatrix} D_{11}^t & D_{21}^t & D_{31}^t \end{bmatrix}. \end{aligned}$$

To see this, let $H(z) \doteq [C_1x\ E_{11}(z)\ E_{12}(z)\ E_{13}(z)]$. Since $H \in \mathcal{H}_\infty$, $H = \sum_{i=0}^{\infty} H_i z^{-i}$, where $H_0 = [C_1x\ D_{11}\ D_{12}\ D_{13}]$, and $H_i = [0\ E_{11}^i\ E_{12}^i\ E_{13}^i]$. Using 1.2-1,

$$\begin{aligned} 1 &\geq \|H\|_\infty^2 \geq (1/n)\|H\|_2^2 \\ &= (1/n)\text{trace}\left(\sum_{i=0}^{\infty} H_i^t H_i\right) = (1/n)\left[\text{trace}(H_0^t H_0) + \text{trace}\left(\sum_{i=1}^{\infty} H_i^t H_i\right)\right] \\ &= 1 + \text{trace}\left(\sum_{i=1}^{\infty} H_i^t H_i\right)/n, \end{aligned}$$

impying $H_i = 0$ for $i \geq 1$.

Theorem 2.10 *All solutions $Q_a(z) \in \mathcal{RH}_\infty$ of $\|G_a - Q_a^{\sim}\| \leq 1$ of the form*

$$Q_a(z) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & Q_{22}(z) & Q_{23}(z) \\ 0 & Q_{32}(z) & Q_{33}(z) \end{bmatrix}$$

are of the form $Q_a = \mathcal{F}_l(Q_{aa}, E)$, where Q_{aa} is as in 2.8-3 and

$$E(z) = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & E_{22}(z) & E_{23}(z) \\ D_{31} & E_{32}(z) & E_{33}(z) \end{bmatrix}$$

is stable and makes the matrix

$$D_a(z) = \begin{bmatrix} yAx & yB \\ Cx & E(z) \end{bmatrix} \qquad (2.8-6)$$

a contraction. All such solutions actually satisfy $\|G_a - Q_a^{\sim}\|_\infty = 1$.

Proof: from the previous discussion, the condition is necessary. To establish sufficiency, write $E(z) = \mathcal{F}_l(E_a, \begin{bmatrix} E_{22}(z) & E_{23}(z) \\ E_{32}(z) & E_{33}(z) \end{bmatrix})$, where

$$E_a \doteq \left[\begin{array}{ccc|cc} D_{11} & D_{12} & D_{13} & 0 & 0 \\ D_{21} & 0 & 0 & I & 0 \\ D_{31} & 0 & 0 & 0 & I \\ \hline 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \end{array}\right].$$

Then, from the discussion on LFT's in Section 2.7.1,

$$\mathcal{F}_l(Q_{aa}, E^t) = \mathcal{F}_l\left(\hat{Q}_{aa}, \begin{bmatrix} E_{22} & E_{23} \\ E_{32} & E_{33} \end{bmatrix}^t\right),$$

with $\hat{Q}_{aa} = \mathcal{F}_l(Q_{aa}, E_a^t)$; using Formula 2.7-2

$$\hat{Q}_{aa} = \left(\begin{array}{c|cc} A - NAXC^tC - NB\hat{D}^t & NAXC^t + NB\hat{D}^t & -NB\begin{bmatrix} 0 & 0 \\ I & 0 \\ 0 & I \end{bmatrix} \\ \hline B^tYA + \hat{D}^tC & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ I & 0 \\ 0 & I \end{bmatrix} \\ \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}\begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} & \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} & 0 \end{array}\right),$$

where $\hat{D} = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & 0 & 0 \\ D_{31} & 0 & 0 \end{bmatrix}$. Then, using 2.8-7:

$$N(AXC^t + B\hat{D}^t) = N\begin{bmatrix} 0 & AXC_2^t + B_1D_{21}^t & AXC_3^t + B_1D_{31}^t \end{bmatrix}$$

$$B^tYA + \hat{D}^tC \;=\; \begin{bmatrix} 0 \\ B_2^tYA + D_{12}^tC_1 \\ B_3^tYA + D_{13}^tC_1 \end{bmatrix}.$$

This implies that Q_{aa12} and Q_{aa21} have the structure $Q_{aa12}^t = Q_{aa21} = \begin{bmatrix} 0 & * & * \end{bmatrix}$; then, from 2.7-1,

$$\mathcal{F}_l\left(\hat{Q}_{aa}(z)\begin{bmatrix} E_{22}(z) & E_{23}(z) \\ E_{32}(z) & E_{33}(z) \end{bmatrix}^t\right) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & Q_{22}(z) & Q_{23}(z) \\ 0 & Q_{32}(z) & Q_{33}(z) \end{bmatrix}$$

as required. Finally, note that by construction $\left\| \begin{matrix} G_{11} & G_{12} & G_{13} \end{matrix} \right\|_\infty = \left\| \begin{matrix} G_{11} \\ G_{21} \\ G_{31} \end{matrix} \right\|_\infty = 1$ and therefore, for all such Q_a, $\|G_a - Q_a^{\sim}\|_\infty = 1$.

□

Necessary and sufficient conditions can now be derived using Lemma 2.13 and Theorem 2.10.

Theorem 2.11 *Let $G \in \mathcal{RH}_\infty$ be as in Lemma 2.13 and assume that the additional condition 2.8-1 is satisfied. Then the following two conditions are equivalent:*

a) There exist a $Q(z) \in \mathcal{RH}_\infty$ such that $\left\| \begin{matrix} G_{11} & G_{12} \\ G_{21} & G_{22} - Q^{\sim} \end{matrix} \right\| \leq 1$.

b) The discrete time algebraic Riccati equations associated with the pencils 2.8-14 and 2.8-15 have positive semi-definite solutions X and Y respectively, such that $\rho(XY) \leq 1$.

Proof: $a \Rightarrow b$ was shown in Lemma 2.13. For the converse, note that:

$$\|yAx \;\; yB\|^2 = \rho\left([yAx\; yB]\begin{bmatrix} xA^ty \\ B^ty \end{bmatrix}\right) = \rho(yXy) \leq 1$$

and similarly $\left\| \begin{array}{c} yAx \\ Cx \end{array} \right\| \leq 1$. Therefore, in Theorem 2.10 at least one E that makes D_a in 2.8-6 into a contraction exists. Taking $Q = \begin{bmatrix} 0 & I & 0 \end{bmatrix} Q_a \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix}$ completes the proof.

□

The following result is a useful corollary of the theorem.

Corollary 2.8 *There exist a* $Q(z) \in \mathcal{RH}_\infty$ *such that* $\left\| \begin{array}{cc} G_{11} & G_{12} \\ G_{21} & G_{22} - Q^\sim \end{array} \right\| \leq 1$ *if and only if there exist a* $Q_a(z) \in \mathcal{RH}_\infty$ *such that* $\|G_a - Q_a^\sim\|_\infty = 1$.

In the parametrization in Theorem 2.10, the free parameter has dimensions $(m_2 + p_1 \times p_2 + m_1)$, larger than those of Q, but it is possible to re-parametrize the solutions in terms of a contraction with the same dimensions as Q. Consider $D_a(z)$ partitioned as:

$$\begin{aligned} D_a(z) &= \left[\begin{array}{cc|cc} yAx & yB_1 & yB_2 & yB_3 \\ C_1x & D_{11} & D_{12} & D_{13} \\ \hline C_2x & D_{21} & E_{22} & E_{23} \\ C_3x & D_{31} & E_{32} & E_{33} \end{array} \right] \\ &= \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{E}(z) \end{bmatrix}. \end{aligned}$$

From Theorem 2.2 the set of transfer functions $\mathcal{E}(z)$ that make $D_a(z)$ into a stable contraction are parametrized by:

$$\mathcal{E}(z) = -Z_1 \mathcal{A}^t Z_2 + \Delta_{Z_1^t} U_a(z) \Delta_{Z_2}, \qquad (2.8-7)$$

where $U_a(z)$ is an $(m_2+p_1 \times p_2+m_1)$ stable contraction, and Z_1, Z_2 are contractions that satisfy:

$$\begin{aligned} \mathcal{C} &= Z_1 \Delta_{\mathcal{A}} \\ \mathcal{B} &= \Delta_{\mathcal{A}^t} Z_2 \end{aligned} . \tag{2.8-8}$$

Note that, from Corollary 2.3,

$$I - \begin{bmatrix} yAx & yB_1 \\ C_1x & D_{11} \\ C_2x & D_{21} \\ C_3x & D_{31} \end{bmatrix} \begin{bmatrix} xA^ty & xC_1^t & xC_2^t & xC_3^t \\ B_1^ty & D_{11}^t & D_{21}^t & D_{31}^t \end{bmatrix} = \begin{bmatrix} \Delta_{\mathcal{A}^t} & 0 \\ -Z_1\mathcal{A}^t & \Delta_{Z_1^t} \end{bmatrix} \begin{bmatrix} \Delta_{\mathcal{A}^t} & -\mathcal{A}Z_1^t \\ 0 & \Delta_{Z_1^t} \end{bmatrix}.$$

Multiplying this expression on the right by $\begin{bmatrix} yB_1 \\ D_{11} \\ D_{21} \\ D_{31} \end{bmatrix}$ and using 2.8-7,

$$\Delta_{Z_1^t} \begin{bmatrix} D_{21} \\ D_{31} \end{bmatrix} = 0. \tag{2.8-9}$$

Similarly,

$$\begin{bmatrix} D_{12} & D_{13} \end{bmatrix} \Delta_{Z_2} = 0. \tag{2.8-10}$$

Since D_{13} and D_{31} have full column and row rank respectively, $\Delta_{Z_1^t}$ and Δ_{Z_2} have a singular value decomposition:

$$\begin{aligned} \Delta_{Z_1^t} &= U_1 \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} V_1^t \\ \Delta_{Z_2} &= U_2 \begin{bmatrix} \Sigma_2 & 0 \\ 0 & 0 \end{bmatrix} V_2^t. \end{aligned}$$

Define $W_{12} \doteq U_1 \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix}$, $W_{21} \doteq \begin{bmatrix} \Sigma_2 & 0 \end{bmatrix} V_2^t$.

Theorem 2.12 *With the previous definitions, all solutions* $Q \in \mathcal{RH}_\infty$ *of* $\|G_a - Q_a^\sim\| \leq 1$ *of the form*

$$Q_a = \begin{bmatrix} 0 & 0 & 0 \\ 0 & Q_{22} & Q_{23} \\ 0 & Q_{32} & Q_{33} \end{bmatrix},$$

are obtained as $Q_a = \mathcal{F}_l(J, U)$*, where* $U \in \mathcal{RH}_\infty$ *is an* $m_2 \times p_2$ *contraction, and*

$$J = \left(\begin{array}{c|cc} A - N(AXC^t + BE_c^t)C & N(AXC^t + BE_c^t) \;\; -N\begin{bmatrix} B_2 & B_3 \end{bmatrix} W_{21}^t \\ \hline B^tYA + E_c^tC & \begin{bmatrix} 0 & 0 \\ 0 & Z_2^t A Z_1^t \end{bmatrix} & \begin{bmatrix} 0 \\ W_{21}^t \end{bmatrix} \\ W_{12}^t \begin{bmatrix} C_2 \\ C_3 \end{bmatrix} & [0 \; W_{12}^t] & 0 \end{array}\right) \qquad (2.8-11)$$

with

$$E_c = \begin{bmatrix} D_{11} & [D_{12} \; D_{13}] \\ \begin{bmatrix} D_{21} \\ D_{31} \end{bmatrix} & -Z_1 A^t Z_2 \end{bmatrix}. \qquad (2.8-12)$$

Proof: From the previous argument, E makes 2.8-6 a stable contraction if and only if $E = \mathcal{F}_l(\hat{E}_a, U)$, with U an $m_2 \times p_2$ stable contraction and

$$\hat{E}_a = \begin{bmatrix} D_{11} & D_{12} \; D_{13} & 0 \\ \begin{matrix} D_{21} \\ D_{31} \end{matrix} & -Z_1 A^t Z_2 & W_{12} \\ 0 & W_{12} & 0 \end{bmatrix}.$$

Now cascade 2.8-3 with $\hat{E}_a^t$ to complete the proof.

□

Chapter 3

Constrained $\mathcal{H}_\infty$ Control: The SISO Case

3.1 Introduction

In this chapter, the problem of designing an internally stabilizing controller for a single input-single output system that solves a tracking problem with maximal robustness is discussed in detail. It is well known that, for pure $\mathcal{H}_\infty$ control, the problem of maximal robustness has little relevance by itself, since one is usually interested in other specifications besides robust stability. Even if that were the case, the resulting controller may turn out to be useless from a practical viewpoint, or even worst, the whole problem may be ill-posed [72]. In spite of these shortcomings, the problem is treated here because it captures all the ideas involved in the solution of the constrained $\mathcal{H}_\infty$ problem. At the same time, the notation is simplified by the fact that the problem is SISO, and the technical difficulty of the underlying $\mathcal{H}_\infty$ theory is reduced, because it involves the solution of a one-block problem.

The chapter is essentially self-contained, since very little of the previous chapter is used throughout. This will hopefully allow the reader to proceed through the derivations without the need to look for further clarifications of the details. Once the ideas on this chapter are comprehended, then it should be easier to understand the general instance of the problem, that is technically more involved.

3.2 Problem Formulation

Let $p(z)$ be the nominal plant, $k(z)$ be the controller to be designed and $t(z) \doteq p(z)k(z)[1 + p(z)k(z)]^{-1}$ be the complementary sensitivity transfer function which represents the command response transfer function in the closed loop system of $p(z)$ and $k(z)$. As $t(z)$ must be stable it has an expansion

$$t(z) = \sum_{i=0}^{\infty} t_i z^{-i}. \tag{3.2 - 1}$$

Let $\mathbf{t_n} = [\ t_0 \quad \cdots \quad t_{n-1}\ ]$. The problem which studied in this chapter is the following.

> **Time Domain Constrained $\mathcal{H}_\infty$ Problem:** Design a controller k such that convex constraints on $\mathbf{t_n}$ are satisfied and $\|t\|_\infty$ is minimized.

This problem can be motivated as follows. Suppose that it is desired that the response of the closed loop system resulting from given test inputs (for example a step function) stays within certain bounds for the first n clock times. It is easy to see that this requirement results in linear constraints on $\mathbf{t_n}$. The solution of the problem above will give a controller that satisfies the given time domain input-output specifications, and achieves optimal robustness with respect to multiplicative perturbations of the plant model. Note that no constraints are imposed over the time-domain response once the finite horizon has been cleared. It is not clear at this point if an overall time response can be guaranteed, and this issue will be discussed in a later section of this chapter.

From Chapter 2, the set of all admissible (i.e., resulting from internally stabilizing controllers) closed loop transfer functions $t(z)$ can be parametrized in terms of a stable transfer function $q(z)$ (the free "parameter") as:

$$t(z) = u(z) - v(z)q(z), \qquad (3.2-2)$$

where u, v are stable and v can be selected to be inner, i.e such that $v(z^{-1})v(z) = 1$ for $|z| = 1$. Let $u(z) = \sum_{i=0}^{\infty} u_i z^{-i}$, $v(z) = \sum_{i=0}^{\infty} v_i z^{i}$, and $q(z) = \sum_{i=0}^{\infty} q_i z^{-i}$. Let $\mathbf{u_n} \doteq [u_0 \ \cdots \ u_{n-1}]$, $\mathbf{v_n} \doteq [v_0 \ \cdots \ v_{n-1}]$ and

$$\mathbf{q_n} = [q_0 \cdots q_{n-1}]. \qquad (3.2-3)$$

Then it is easy to see that:

$$t_i = u_i - \sum_{j=0}^{i} v_{(i-j)} q_j. \qquad (3.2-4)$$

In matrix notation:

$$\mathbf{t_n} = \mathbf{u_n} - \mathbf{q_n} V \qquad (3.2-5)$$

where

$$V = \begin{bmatrix} v_0 & v_1 & \cdots & v_{n-1} \\ 0 & v_0 & \cdots & v_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & v_0 \end{bmatrix}. \qquad (3.2-6)$$

Equation (3.2-5) implies that convex (linear) constraints on t_n translate into convex (linear) constraints on $\mathbf{q_n}$. From the discussion in Section 2.4, the minimization of $\|t\|_\infty$ is equivalent to the minimization of $\|g - q^\sim\|_\infty$ where $g \doteq u^\sim v \in \mathcal{RH}_\infty$. The proposed problem can then be formulated as

$$\inf_{\substack{q \in \mathcal{H}_\infty \\ \mathbf{q_n} \in \Omega}} \|g - q^\sim\|_\infty, \qquad (3.2-7)$$

where $\Omega \doteq \{\mathbf{q_n} \in R^n \text{s.t.} lb \leq \mathbf{q_n} V \leq ub\}$. More generally, if $w^i(z)$ denotes the z-transform of an input signal, then $y^i(z) = [w^i(z)u(z)] - [w^i(z)v(z)]q(z)$, and the constraints determine a set $\Omega^i = \{\mathbf{q_n} \in R^n \text{s.t.} lb^i \leq \mathbf{q_n} V^i \leq ub^i\}$. The pair (w^i, Ω^i) is called an *input-output time domain constraint.*

3.3 Problem Transformation and Solution

In the previous section, the problem of robust performance (with performance specified in the time domain) was formulated as the constrained $\mathcal{H}_\infty$ optimization problem 3.2-7. For any z such that $|z| = 1$,

$$\begin{aligned} |g(z) - \tilde{q}(z)| &= |g(z) - \sum_{i=0}^{\infty} q_i z^i| = |g(z) - \sum_{i=0}^{n-1} q_i z^i - \sum_{i=n}^{\infty} q_i z^i| \\ &= |z^{-n}[g(z) - \sum_{i=0}^{n-1} q_i z^i] - q_t(z)^{\sim}| \end{aligned}$$

where $q_t(z)^{\sim} \doteq \sum_{i=0}^{\infty} q_{i+n} z^i$. Let

$$g(z; \mathbf{q_n}) \doteq z^{-n}[g(z) - \sum_{i=0}^{n-1} q_i z^i] \qquad (3.3-1)$$

and define

$$\varphi(\mathbf{q_n}) \doteq \min_{q_t(z)\in\mathcal{H}_\infty} \|g(z; \mathbf{q_n}) - q_t(z)^{\sim}\|_\infty. \qquad (3.3-2)$$

Then

$$\inf_{q\in\mathcal{H}_\infty,\ \mathbf{q_n}\in\Omega} \|g - q\|_\infty = \min_{\mathbf{q_n}\in\Omega} \varphi(\mathbf{q_n}) \qquad (3.3-3)$$

The following result is now obvious.

Theorem 3.1 *With the previous definitions, the optimization problem 3.2-7 is equivalent to the program*

$$\min_{\mathbf{q_n}\in\Omega} \varphi(\mathbf{q_n}) \qquad (3.3-4)$$

in the sense that if $\mathbf{q_n^*} = [q_0^* \ \cdots \ q_{n-1}^*]$ *solves 3.3-4 and* $q_t^*(z)^{\sim}$ *is the best antistable approximation to* $g(z; \mathbf{q_n^*})$ *then* $q^*(z) \doteq \sum_{i=1}^{n-1} h_i^* z^{-i} + z^{-n} q_t^*(z)$ *solves problem (3.2-7), and conversely if* $h^*(z) = \sum_{i=0}^{\infty} q_i^* z^{-i}$ *solves 3.2-7, then* $\mathbf{q_n^*} = [q_0^*, \ \cdots \ q_{n-1}^*]$ *solves 3.3-4.*

□

For each vector of parameters $\mathbf{q_n}$, the value of φ and $q_t(z)$ may be computed as the solution to an unconstrained Nehari extension problem. Then Theorem 3.1 shows how to transform (3.2-7) into a finite dimensional optimization problem, namely the minimization of $\varphi(\mathbf{q_n})$ subject to some constraints. The following theorem shows that $\varphi(\cdot)$ is a "nice" function.

Theorem 3.2 *Let $\varphi(\mathbf{q_n})$ be as defined in 3.3-2. Then $\varphi(.)$ is convex.*

Proof: Suppose $\mathbf{q_n^1}$, $\mathbf{q_n^2}$ are given. Then there exist q_t^1, $q_t^2 \in \mathcal{RH}_\infty$ such that

$$\begin{aligned}\varphi(\mathbf{q_n^1}) &= \|g(z;\mathbf{q_n^1}) - q_t^{1\sim}(z)\|_\infty \\ \varphi(\mathbf{q_n^2}) &= \|g(z;\mathbf{q_n^2}) - q_t^{2\sim}(z)\|_\infty.\end{aligned}$$

For $0 \leq \lambda \leq 1$, let $\hat{q}_t(z) \doteq \lambda q_t^1(z) + (1-\lambda)q_t^2(z)$. Then

$$\begin{aligned}\varphi(\lambda\mathbf{q_n^1} + (1-\lambda)\mathbf{q_n^2}) &= \min_{q_t \in \mathcal{RH}_\infty} \|z^{-n}[g(z) - \sum_{i=0}^{n-1}(\lambda q_i^1 + (1-\lambda)q_i^2)z^i] - q_t^\sim(z)\|_\infty \\ &\leq \|z^{-n}[g(z) - \sum_{i=0}^{n-1}(\lambda q_i^1 + (1-\lambda)q_i^2)z^i] - \hat{q}_t^\sim(z)\|_\infty \\ &\leq \lambda\|g(z;\mathbf{q_n^1}) - q_t^{1\sim}(z)\|_\infty + (1-\lambda)\|g(z;\mathbf{q_n^2}) - q_t^{2\sim}(z)\|_\infty.\end{aligned}$$

As this holds for any $\mathbf{q_n^1}$, $\mathbf{q_n^2}$ the function is convex.

□

Since φ is convex, then it is well known that there exist effective algorithms for computing its global minimum, whenever the constrained set is convex. However, convexity is not enough to guarantee the existence of *efficient* algorithms, because the evaluation of the function by the procedure suggested above, involving the solution of two discrete Lyapunov equations, becomes unreasonably expensive for a large horizon. This topic is addressed in the next section, where state-space formulas for computing $\varphi(\mathbf{q_n})$ are presented.

3.4 State-Space Computation of $\varphi(\mathbf{q_n})$

Let $g(z) \in L_\infty$, and let $\Gamma_g : l_2(-\infty, 0) \mapsto l_2(0, \infty)$ denote the associated Hankel operator with operator norm $\|\Gamma_g\|$. Then there exists a closest function $q \in \mathcal{H}_\infty$ to g such that $\|g - q^\sim\|_\infty = \|\Gamma_g\|$ [25]. In the special case under consideration, the Hankel norm is easy to compute. Let $\left(\begin{array}{c|c} A & b \\ \hline c & 0 \end{array}\right)$ be a minimal realization for $g(z) \in \mathcal{RH}_\infty$, and X, Y denote its controllability and observability grammians respectively. Then $\|\Gamma_g\|^2 = \max_i \lambda_i(XY)$. Recall that X and Y are symmetric positive definite matrices which satisfy the discrete Lyapunov equations

$$X = AXA^t + bb^t \tag{3.4-1}$$

$$Y = A^tYA + c^tc. \tag{3.4-2}$$

Note that in general the product XY need not be symmetric but it is easy to show that $\rho(XY) = \rho(X^{1/2}YX^{1/2})$. A closed-form expression for $\varphi(\cdot)$ can then be found by computing the grammians for

$$g(z; \mathbf{q_n}) = z^{-n}g(z) - \sum_{i=0}^{n-1} q_i z^{-n+i}.$$

<u>Remark</u>: the assumption that $g \in \mathcal{RH}_\infty$ is without loss of generality. If the starting $g \in \mathcal{L}_\infty$, then the anti-stable part of $g(z)$ can be subtracted first, and this amounts to a constant translation of the vector of variables $\mathbf{q_n}$. That is, if $\mathbf{g_n^u} = [g_0^u, \cdots, g_{n-1}^u]$ denotes the first n terms in the expansion of the anti-stable part of $g(z)$, define $\hat{\mathbf{q}}_\mathbf{n} \doteq \mathbf{q_n} - \mathbf{g_n^u}$ and then:

$$\begin{aligned} \inf_{q \in \mathcal{H}_\infty^\sim} \|g(z) \; &- \; \sum_{i=0}^{n-1} q_i z^i - z^n q(z)\|_\infty = \\ &= \inf_{\hat{q} \in \mathcal{H}_\infty^\sim} \|g_s(z) - \sum_{i=0}^{n-1} \hat{q}_i z^i - z^n \hat{q}_t(z)\|_\infty. \end{aligned}$$

Let then $g(z) = \left(\begin{array}{c|c} A & b \\ \hline c & 0 \end{array}\right) \in \mathcal{RH}_\infty$. After some manipulations:

$$g(z;\mathbf{q_n}) = \left(\begin{array}{cc|c} A & be_1^t & 0 \\ 0 & A_f & e_n \\ \hline c & -\mathbf{q_n} & 0 \end{array}\right) = \left(\begin{array}{c|c} A_e & b_e \\ \hline c_e & 0 \end{array}\right) \qquad (3.4-3)$$

where $\mathbf{q_n} = [q_0\ q_1 \cdots q_{n-1}]$.

Lemma 3.1 *Assume that the realization for $g(z)$ is minimal. Then the realization for $g(z;\mathbf{q_n})$ in 3.4-3 is controllable, and it is observable if and only if*

$$\det\left(\begin{bmatrix} A & b \\ c & -q_o \end{bmatrix}\right) \neq 0$$

In particular, if A is invertible then 3.4-3 is observable if and only if $q_o + cA^{-1}b \neq 0$.

The lemma follows from a simple application of the PBH test. Note that the system is controllable and that by the minimality of $g(z)$ the poles of $g(z)$ are observable. Therefore, observability would be lost if and only if $g(z;\mathbf{q_n})$ has some zeros at 0.

Let X and Y denote the controllability and observability grammians of $g(z)$ respectively. The grammians of 3.4-3 solve the equations:

$$X_e = A_e X_e A_e^t + b_e b_e^t \qquad (3.4\text{-}4)$$

$$Y_e = A_e^t Y_e A_e + c_e^t c_e. \qquad (3.4\text{-}5)$$

Due to the special form of A_e and b_e, the following explicit formulas for X_e and Y_e can be obtained (see Appendix A for a derivation):

$$X_e = \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix} \qquad (3.4-6)$$

$$Y_e = \begin{bmatrix} Y & Y_{12}^0 - Y_{12}^l Q_n^t \\ Y_{12}^{0t} - Q_n Y_{12}^{lt} & Y_{22}^0 - Y_{22}^l Q_n^t - Q_n Y_{22}^{lt} + Q_n Q_n^t \end{bmatrix}, \qquad (3.4-7)$$

where:

$$Y_{12}^0 = [A^tYb\, A^{2t}Yb \cdots A^{nt}Yb] \quad (3.4\text{-}8)$$

$$Y_{12}^l = [c^t\, A^tc^t \cdots A^{(n-1)t}c^t] \quad (3.4\text{-}9)$$

$$Y_{22}^0 = \begin{bmatrix} b^tYb & b^tA^tYb & \cdots & b^tA^{(n-1)t}Yb \\ b^tYAb & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ b^tYA^{n-1}b & \cdots & \cdots & b^tYb \end{bmatrix} \quad (3.4\text{-}10)$$

$$Y_{22}^l = \begin{bmatrix} 0 & b^tc^t & \cdots & b^tA^{(n-2)t}c \\ \vdots & & \ddots & \vdots \\ \vdots & & \ddots & b^tc^t \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \quad (3.4\text{-}11)$$

and

$$Q_n = \begin{bmatrix} q_0 & q_1 & \cdots & q_{n-1} \\ 0 & q_0 & \cdots & q_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & q_0 \end{bmatrix}. \quad (3.4\text{-}12)$$

Using the notation $x = X^{1/2}$, $y = Y^{1/2}$, the following result holds.

Theorem 3.3 *With the previous notation, let*

$$W(\mathbf{q_n}) = \begin{bmatrix} xYx & xY_{12}^0 \\ Y_{12}^{0t}x & Y_{22}^0 \end{bmatrix} + \begin{bmatrix} 0 & -xY_{12}^lQ_n \\ -(Y_{12}^lQ_n)^tx & -Y_{22}^lQ_n - (Y_{22}^lQ_n)^t + Q_n^tQ_n \end{bmatrix}; \quad (3.4-13)$$

then $\varphi(\mathbf{q_n})^2 = \rho[W(\mathbf{q_n})]$.

□

Corollary 3.1 *Let*

$$W_1(\mathbf{q_n}) = \begin{bmatrix} yA^n x & yA^{n-1}b & \cdots & \cdots & yAb & yb \\ cA^{n-1}x & cA^{n-2}b & \cdots & \cdots & cb & -q_0 \\ cA^{n-2}x & cA^{n-3}b & \cdots & \cdots & -q_0 & -q_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ cAx & cb & -q_0 & \cdots & -q_{n-3} & -q_{n-2} \\ cx & -q_0 & -q_1 & \cdots & -q_{n-2} & -q_{n-1} \end{bmatrix}. \qquad (3.4-14)$$

Then

$$\varphi(\mathbf{q_n}) = \rho[W(\mathbf{q_n})]^{\frac{1}{2}} = \overline{\sigma}[W_1(Q_n)]. \qquad (3.4-15)$$

<u>Proof:</u> See Appendix A.

□

The corollary also gives a proof of the convexity of $\varphi(\cdot)$, since it expresses φ as the norm of a matrix depending linearly on the independent variables. Summarizing, a solution to the time domain constrained $\mathcal{H}_\infty$ problem can be found by the constrained minimization of the spectral radius of a matrix depending quadratically on the variables or the maximum singular value of a matrix depending linearly on the variables. It turns out that the singular value objective function can be again transformed into a spectral radius one. To see this, recall that if the realization for a scalar transfer function is balanced, then for some sign matrix R (i.e., a diagonal matrix with either 1 or -1 in its diagonal) the following equalities hold [65]:

$$\begin{aligned} c &= b^t R \\ RA &= A^t R \end{aligned}. \qquad (3.4-16)$$

Corollary 3.2 *Assume that the realization for g is balanced, and that y is also chosen diagonal. Let*

$$W_s(\mathrm{q_n}) = \begin{bmatrix} yRA^ny & yRA^{n-1}b & \cdots & \cdots & yRAb & yRb \\ b^tRA^{n-1}y & b^tRA^{n-2}b & \cdots & \cdots & b^tRb & -q_0 \\ b^tRA^{n-2}y & b^tRA^{n-3}b & \cdots & \cdots & -q_0 & -q_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b^tRAy & b^tRb & -q_0 & \cdots & -q_{n-3} & -q_{n-2} \\ b^tRy & -q_0 & -q_1 & \cdots & -q_{n-2} & -q_{n-1} \end{bmatrix}, \qquad (3.4-17)$$

where R is a sign matrix such that 3.4-16 holds. Then $\varphi(\mathrm{q_n}) = \rho(W_s)$

Proof: The norm of W_1 is preserved if the first block row is multiplied on the left by unitary matrix R. Since both R and y are diagonal they commute and using 3.4-16, $\overline{\sigma}(W_1) = \overline{\sigma}(W_s)$. But W_s is symmetric and hence the result follows.

□

3.5 Degree Bounds

In this section, bounds on the order of the optimal solution q of problem 3.2-7 and on the optimal controller k are given. Let

$$g(z; \mathrm{q_n}) = z^{-n}[g(z) - \sum_{i=0}^{n-1} q_i z^i] \qquad (3.5-1)$$

where $g(z) \in \mathcal{RH}_\infty$, $\mathrm{q_n} = [q_0, \cdots, q_{n-1}] \in R^n$ is a fixed vector, and $q_t \in \mathcal{RH}_\infty$ is such that:

$$\min_{q \in \mathcal{RH}_\infty} \|g(\cdot; \mathrm{q_n}) - q^\sim\|_\infty = \|g(\cdot; \mathrm{q_n}) - q_t^\sim\|_\infty; \qquad (3.5-2)$$

moreover, let

$$q(z) = \sum_{i=0}^{n-1} q_i z^{-i} + q_t(z) z^{-n}. \qquad (3.5-3)$$

Theorem 3.4 *With the previous notation, the order of q is bounded by $m+n-r_0$ where m is the order of $g(z)$ and r_0 is the multiplicity of the maximum Hankel singular value of $g(z;\mathrm{q_n})$ in 3.4-3.*

Proof: Let X_e, Y_e be the controllability and observability grammians of $g(z;\mathrm{q_n})$. Define $\alpha(z)$ and $\tilde{\beta}(z)$ as

$$\alpha \doteq \left(\begin{array}{c|c} A_e & \omega \\ \hline c_e & 0 \end{array}\right) \qquad \beta^{\sim} \doteq \left(\begin{array}{c|c} A_e & b_e \\ \hline \nu^t A_e & \nu^t b_e \end{array}\right), \tag{3.5 – 4}$$

where ν is an eigenvector associated with the largest eigenvalue ρ^2 of $Y_e X_e$, i.e., $Y_e X_e \nu = \rho^2 \nu$ and $\omega = \rho^{-1} X_e \nu$. Then q minimizing $\|g(\cdot;\mathrm{q_n}) - q_t^{\sim}\|_\infty$ is given by:

$$q_t^{\sim} = g(\cdot;\mathrm{q_n}) - \rho\frac{\alpha}{\beta}. \tag{3.5 – 5}$$

Note that $\beta^{\sim}$ is defined rather than β because the former has a polynomial term in z and thus no state-space realization. Replacing in 3.5-4,

$$\begin{aligned}
\alpha &= \left(\begin{array}{cc|c} A & be_1^t & \omega_1 \\ 0 & A_f & w_2 \\ \hline c & -\mathrm{q_n} & 0 \end{array}\right) \\
\tilde{\beta} &= \left(\begin{array}{cc|c} A & be_1^t & 0 \\ 0 & A_f & e_n \\ \hline \nu_1^t A & \nu_1^t b e_1^t + \nu_2^t A_f & \nu_2^t e_n \end{array}\right) \\
&= z^{-n}\left(\begin{array}{c|c} A & b \\ \hline \nu_1^t A & 0 \end{array}\right) + \left[\begin{array}{cc} \nu_1^t b & \nu_2^t \end{array}\right]\left[\begin{array}{c} z^{-n} \\ \vdots \\ 1 \end{array}\right],
\end{aligned}$$

with ω and ν partitioned according to the block structure of A_e. If $\nu_2^t = (\nu_2^1, \cdots, \nu_2^n)$, define $\nu_3^t \doteq (\nu_2^n, \cdots, \nu_2^1)$. Then:

$$z^{-n}\beta = \left(\begin{array}{cc|c} A^{-t} & 0 & \nu_1 \\ 0 & A_f & e_n \\ \hline -b^t A^{-t} & \nu_3^t & 0 \end{array}\right). \tag{3.5 – 6}$$

Since the multiplicity of the largest Hankel singular value of $g(z;\mathrm{q_n})$ is r_o, there exist r_o linearly independent eigenvectors ν^i associated with the largest eigenvalue of YX. Therefore, one can choose ν to have its last $r_o - 1$ entries zero. By the special form of X_e, this implies that the last $r_o - 1$ entries of ω are also zero, and therefore $\alpha(z)$ has at least $r_o - 1$ non controllable poles at zero while $z^{-n}\beta$ has at least $r_o - 1$ non observable poles at zero. More general, the number of uncontrollable poles of $\alpha(z)$ at zero is exactly the same as the number of unobservable poles of $z^{-n}\beta(z)$ at zero. From 3.5-5, $q_t(z)^{\sim} = g(z;\mathrm{q_n}) - \rho\frac{\alpha(z)}{z^n(z^{-n}\beta(z))}$ and hence $q(z)^{\sim} = g(z) - \rho\frac{\alpha(z)}{z^{-n}\beta(z)}$. In order to estimate the number of poles of $q(z)^{\sim}$, note that the poles of g and α must cancel out since they are both stable while q is anti-stable. In particular the poles of $\alpha(z)$ at zero cancel with those of $z^{-n}\beta(z)$ since it was established that the realizations for α and $z^{-n}\beta(z)$ have the same number of controllable and observable poles at zero. Therefore,

$$\#\ poles\ of\ \tilde{h}(z) = \#\ zeros\ of\ z^{-n}\beta(z) \leq \#\ poles\ of z^{-n}\beta(z) \qquad (3.5-7)$$

since $z^{-n}\beta$ is strictly proper. Now since at least $r_o - 1$ modes in the realization 3.5-6 of $z^{-n}\beta(z)$ at zero are non-observable,

$$\#\ poles\ of z^{-n}\beta(z) \leq m + n - r_o + 1 \qquad (3.5-8)$$

and the result follows from 3.5-7 and 3.5-8.

□

For a standard $\mathcal{H}_\infty$ optimization problem, the multiplicity of the largest Hankel singular value is generically one. On the contrary, for constrained $\mathcal{H}_\infty$ a multiplicity larger than one is likely to occur. The reason for this is that a constrained solution is computed by minimizing the largest singular value of a matrix; since singular values are nonnegative it is intuitively clear that the minimization of the largest one will tend to increase its multiplicity. Theorem 3.5 shows that this increase is dramatic when constraints are not binding.

Theorem 3.5 : *Assume that* $\mathbf{q}_n^* = [h_0^*, \cdots, h_{n-1}^*]$ *solves the unconstrained optimization problem* $\min_{\mathbf{q}_n \in R^n} \overline{\sigma}[W_1(\mathbf{q}_n)]$, *with* $W_1(\mathbf{q}_n)$ *defined by 3.4-14. Then* $W_1(\mathbf{q}_n{}^*)$ *has at least* $n + r_u$ *singular values, say* $\sigma_1 \cdots \sigma_{n+r_u}$ *such that*

$$\sigma_i = \overline{\sigma}[W_1(\mathbf{q}_n^*)], \;\; i = 1, \;\cdots, n + r_u, \qquad (3.5-9)$$

where r_u *is the multiplicity of the largest Hankel singular value associated with* $g(z)$.

The proof of Theorem 3.5 is rather technical and relegated to Appendix B. Here an argument is given to motivate its validity. Let q^* be the solution to the unconstrained optimal problem with a minimal state-space realization.

$$q^* = \left(\begin{array}{c|c} A_q & b_q \\ \hline c_q & d_q \end{array}\right).$$

Then q^* has the non-minimal realization:

$$q^* = \left(\begin{array}{c|c} A_f & e_n \\ \hline [q_{n-1}^* \cdots q_1^*] & q_0 \end{array}\right) + z^{-n}\left(\begin{array}{c|c} A_q & A_q^{n-1}b_q \\ \hline c_q & q_n^* \end{array}\right). \qquad (3.5-10)$$

The order of q^* is $m - r_u$, where r_u denotes the multiplicity of the largest Hankel singular value of $g(z)$. On the other hand, by uniqueness of the solution to the Nehari extension problem in the SISO case and Theorem 1, q^* may be written as

$$q^*(z) = \sum_{i=0}^{n-1} q_i^* z^{-i} + z^{-n} q_t^{\sim}(z), \qquad (3.5-11)$$

where q_t solves

$$\inf_{q_t \in \mathcal{H}_\infty} \|g(\cdot; \mathbf{q}_n^*) - q_t^{\sim}\|_\infty. \qquad (3.5-12)$$

From Theorem 4, the order of q_t is bounded by $n + m - r_0$ and by comparing 3.5-10 and 3.5-11 the order of q_t is also bounded by $m - r_u$. Both bounds would coincide if the multiplicity r_o of the largest Hankel singular value of $g(\cdot; \mathbf{q}_n^*)$ were equal to $n + r_u$. Theorem 3.5 shows that this is indeed the case.

In the remaider of the section, a bound on the order of the optimal controller k is derived using an argument similar to the one used by Limebeer and Hung in [49] for the unconstrained problem.

Theorem 3.6 *Let p be a plant of order m. Then the order r_k of the $\mathcal{H}_\infty$-optimal controller with constraints on a horizon of length n is at most $n + m - r_0$.*

Proof: Consider the input-output transfer function

$$\begin{aligned} t(z) &= u(z) - v(z)h(z) \\ &= [u(z) - v(z)\sum_{i=0}^{n-1} q_i z^{-i}] - z^{-n}v(z)h^1(z). \end{aligned}$$

By a well known result [49] the order of $t(z)$ for $h^1(z)$ minimizing its infinity norm is not larger than the number of interpolation points, namely the number of unstable zeros of $v(z)$ plus n (since z^{-n} has n zeros at infinity), minus the multiplicity r_0 of the largest Hankel singular value of $[u(z) - v(z)\sum_{i=0}^{n-1} q_i z^{-i}]$. Since the degree of v is at most m, it follows that the degree of t is at most $n + m - r_0$. But $t = pk/(1 + pk)$, and hence

$$r_k \leq n + m - r_0 - m + c,$$

where c denotes the cancellations between p and k which occur as a result of closing the feedback loop. It is clear that c is bounded above by the order of the plant m, and then

$$r_k \leq n + m - r_0. \qquad (3.5-13)$$

□

Remark: Theorem 3.5 and some numerical experience suggest that in a typical problem, if the horizon length n is increased, then the multiplicity of r_o also increases. Theorem 3.6 then shows that the complexity of the solution increases only by $n - r_o$, which can be significantly smaller than n.

3.6 Behavior of the Overall Response

In the previous sections, attention was focused on shaping the first part of the time domain response by imposing constraints over a finite horizon; once the first n parameters are determined, the tail of the response is computed to minimize the infinity norm. Unfortunately, this procedure may generate an undesirable behavior after the time horizon considered. To illustrate this, consider in Fig. 1 the optimal response for the example treated in the next chapter, for a horizon of 40 samples. After clearing the constraints, the response has an unreasonable large peak, and then oscillates and attenuates with a definite pattern. It turns out that rather than being pathological, these oscillations are due to the character of the optimal solution. Specifically, recall from the previous section that when the variables of the problem are replaced by their optimal unconstrained values, the multiplicity of the largest Hankel singular value is $n + r$ in which case the order of the optimal controller drops to $m - r$. If a small perturbation on the vector of variables is introduced so that the largest singular value becomes unique, the order of the resulting optimal solution correspondingly increases to $m + n - r$. This variation on the order is associated with a pole-zero cancellation occurring in the limiting case. Because of the all-pass character of the closed-loop transfer function, poles and zeros are inside and outside the closed unit disk respectively, and hence the cancellations must occur on the unit circle. Suppose now that the constraints are such that the optimal constrained and unconstrained solution do not differ much; from the previous discussion, n poles and zeros should be distributed closed to the unit circle. The behavior is illustrated in Fig. 3.2, which shows the poles and zeros for the example cited above. Each pole closed to the unit circle is coupled with a zero, and when excited generates a poorly attenuated oscillation. During the first N samples, the resulting oscillations combine in such a way that the constraints are satisfied, but once the constraints are cleared, they produce large peaks in the time response.

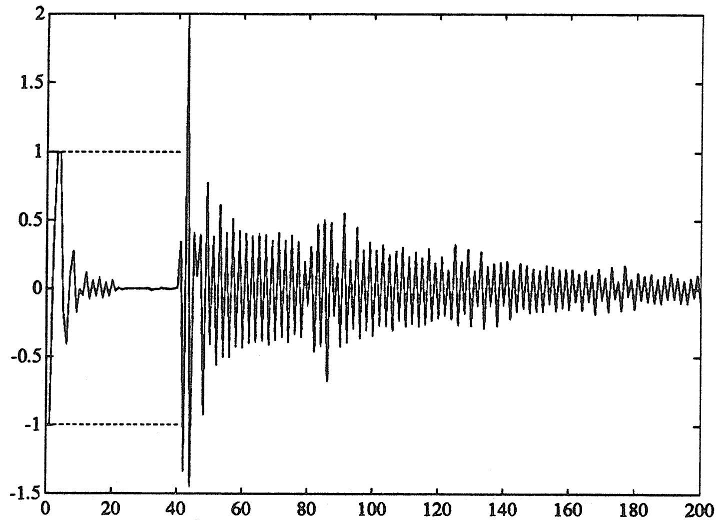

Figure 3.1: Time Response for Optimal Controller

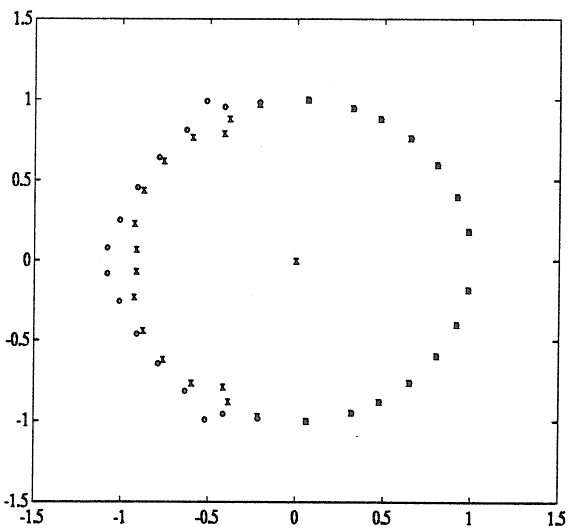

Figure 3.2: Optimal Closed-loop: Pole/Zero Configuration

The following argument gives a quantitative explanation for this phenomenon. Since the frequency response for the optimal approximation is flat, the 2 and the ∞ norm coincide, and hence

$$\begin{aligned}\|g(z) - \sum_{i=0}^{n-1} q_i z^i - z^n q_t(z)^{\sim}\|_{\infty}^2 &= \|g(z) - \sum_{i=0}^{n-1} q_i z^i - z^n q_t(z)\|_2^2 \\ &= \|g(z) - d_r\|_2^2 + \|d_r + \sum_{i=0}^{n-1} q_i z^i\|_2^2 + \|q_t(z)\|_2^2 \qquad (3.6\text{-}1)\end{aligned}$$

where the last equality follows from the l_2-orthogonality of the different terms. If the constraints are such that the 2-norm of the first n terms is small, then the norm of the tail needs to be large in order to satisfy 3.6-1. Moreover, since the absolute value of the variables after the horizon is limited by the steady-state constraint, the norm of the tail will tend to remain essentially constant or decrease slowly when the number of variables is increased. This implies that taking a larger horizon will not necessarily produce a better overall response; in fact, the existence of a solution for the limiting case $n \to \infty$ cannot be guaranteed.

The salient point of the argument above is the flat response character of optimal controllers. A natural way to circumvent this difficulty, is to replace the optimal controller by a suboptimal one that satisfies the time domain constraints but has a smaller bandwidth and hence, hopefully, better overall time response. Although the algorithm for the one-block problem in Section 2.5 can be used to compute a suboptimal solution, it is useful to combine the computation with the following scheme, which may be used to guarantee good overall behavior. First, introduce the change of variables $z \longrightarrow \rho z$, $\rho > 1$ which maps the unit disk into one of radius ρ. Then, compute a constrained $\mathcal{H}_\infty$ sub-optimal controller for the transformed plant using the formulas in Section 2.5. The "degree of sub-optimality," i.e., the number *upper bound on norm/optimal norm*, is usually chosen between 1.05 and 1.1. Finally transform the resulting controller back into the original z-plane by using the change of variables $z \longrightarrow z/\rho$. In this procedure, the time domain constraints should be scaled so that they are satisfied by the final closed-loop and not by the transformed one. The change of variables has the following consequences:

- By the Maximum Modulus Theorem, the norm of the transformed closed-loop gives a non-achievable upper bound over the actual norm.
- When transformed back into the original variables, the controller will neither be optimal nor all-pass; this follows from the previous observation and the fact that the change of variables does not preserve the symmetry between poles and zeros that characterizes a flat response behavior.
- All the poles of the closed-loop are placed inside a disk of radius $1/\rho$, thus inducing a decay ratio on the time response.

The pole-placement characteristic is essential in establishing the main result of this section since it can be used to guarantee an attenuation in the time domain response. Some further notation is needed before a formal statement. Let $y \in l^2$, with the $k-th$ sample denoted by y_k. For some fixed N, let $\mathbf{y_N} \doteq \begin{bmatrix} y_0 & y_1 & \cdots & y_{N-1} \end{bmatrix}$. Let $\Omega \in \mathcal{R}^N$ be a convex set and define $\hat{\Omega}^n \doteq \{\mathrm{y} \in l^2 \; s.t. \; \mathbf{y_N} \in \Omega, \; |y_k| < \epsilon \; n \geq k > N\}$, where ϵ is some small constant. Then $\hat{\Omega}^n$ corresponds to extending the horizon by n samples and a tolerance ϵ. Finally, let $t^n(z)$ be the closed loop response computed by the procedure above for some $\rho > 1$ and time domain constraints defined by $\hat{\Omega}^n$.

Theorem 3.7 *Assume that for some constant $K > 0$, $\|t^n(z)\|_\infty < K$ for each $n > N$. Then there exists M such that $t^m \in \hat{\Omega}^\infty$ for every $m > M$.*

<u>Proof</u>: It is well known that[1]

$$t_k^n = \frac{1}{2\pi j} \oint_C T^n(z) z^{(k-1)} dz$$

[1]The idea of using the convolution integral in order to get the needed bound was suggested by Prof. M. Sznaier

and since all the singularities are confined to a circle of radius $1/\rho$,

$$t_k^n = \frac{1}{2\pi}\int_{-\pi}^{\pi} T^n(e^{j\theta}/\rho)\rho^{-(k-1)}e^{j(k-1)\theta}d\theta.$$

This implies $|t_k^n| \leq K\rho^{-(k-1)}$ and hence by taking $M > -\log(\epsilon/K)/\log(\rho)+1$ the result follows.

□

This theorem relies on two facts: the existence of a uniform bound K (which is assured by the existence of a stable-closed loop satisfying the constraints), and the location of the poles inside a circle of radius strictly less than one. Note that the latter fails if no transformation is used. If the problem requires a nonzero steady-state value, a similar result holds.

A uniform bound can be computed as follows. Recall that $t = u - vh$, and assume that $A = \{\alpha_1, \alpha_2, \cdots, \alpha_l\}$ is the set of zeros of $v(z)$ outside the unit disk. Then, it is well known that t is a stable closed-loop transfer function if and only if $t(\alpha_i) = u(\alpha_i)$, $i = 1, \cdots, l$ (assuming that each α_i is of multiplicity one). Next compute (if possible) $t_{fir}(z) = \sum_{i=0}^{N-1} t_i z^{-i}$ such that $\mathbf{t}_{\mathbf{fir}}^{\mathbf{N}} \doteq [t_0\ t_1 \ldots t_{N-1}] \in \Omega$ and $t_{fir}(\alpha_i) = u(\alpha_i)$, $i = 1, \cdots, l$ and set $K = \|t_{fir}(z)\|_\infty$. If t_{fir} exists, then it may be computed by solving a linear programming problem whenever the constraints defining Ω are linear.

Theorem 3.7 concludes the discussion of the theoretical aspects in the solution of the simplified instance of the time domain constrained $\mathcal{H}_\infty$ control problem. The numerical problems involved in the actual computation of a solution are discussed in the next chapter.

3.7 Appendix A: Additional Proofs for Section 3.4

To prove formulas 3.4-6 and 3.4-7, start by partitioning X_e conformally with the block structure of A_e:

$$\begin{bmatrix} X_{e11} & X_{e12} \\ X_{e12}^t & X_{e22} \end{bmatrix} = \begin{bmatrix} A & be_1^t \\ 0 & A_f \end{bmatrix} \begin{bmatrix} X_{e11} & X_{e12} \\ X_{e12}^t & X_{e22} \end{bmatrix} \begin{bmatrix} A^t & 0 \\ e_1 b^t & A_f^t \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & e_n e_n^t \end{bmatrix}.$$

In particular:

$$X_{e22} = A_f^t X_{e22} A_f^t + e_n e_n^t$$

Multiplying by $e_n, \ e_{n-1}, \ \cdots, \ e_1$ on the right, one gets

$$\begin{aligned} X_{e22} e_n &= e_n \\ X_{e22} e_{n-1} &= e_{n-1} \\ &\vdots \\ X_{e22} e_1 &= e_1, \end{aligned}$$

giving $X_{e22} = I$. Using this fact in the equation for X_{e12}:

$$\begin{aligned} X_{e12} &= A X_{e12} A_f^t + be_1^t X_{e22} A_f^t \\ &= A X_{e12} A_f^t + be_1^t A_f^t \ = \ A X_{e12} A_f^t, \end{aligned}$$

implying $X_{e12} = 0$. Finally replacing X_{e22} and X_{e12} in the equation for X_{e11} above, $X_{e11} = X$ and 3.4-6 has been established. The computation of Y_e is slightly more laborious. Again start by partitioning Y_e to get:

$$\begin{bmatrix} Y_{e11} & Y_{e12} \\ Y_{e12}^t & Y_{e22} \end{bmatrix} = \begin{bmatrix} A^t & 0 \\ e_1 b^t & A_f^t \end{bmatrix} \begin{bmatrix} Y_{e11} & Y_{e12} \\ Y_{e12}^t & Y_{e22} \end{bmatrix} \begin{bmatrix} A & be_1^t \\ 0 & A_f \end{bmatrix} + \begin{bmatrix} c^t c & -c^t \mathbf{q_n} \\ \mathbf{q_n}^t c & \mathbf{q_n}^t \mathbf{q_n} \end{bmatrix}.$$

Then, $Y_{e11} = Y$, and

$$Y_{e12} = A^t Y b e_1^t + A^t Y_{e12} A_f - c^t \mathbf{q_n}$$

. Thus:

$$
\begin{aligned}
Y_{e12}e_1 &= A^tYb - c^tq_0 \\
Y_{e12}e_2 &= A^tY_{e12}e_1 - c^tq_1 = A^{2t}Yb - A^tc^tq_0 - c^tq_1 \\
&\vdots \\
Y_{e12}e_n &= A^tY_{e12}e_{n-1} - c^tq_{n-1} = A^{nt}Yb - A^{(n-1)t}c^tq_0 + \cdots + A^tYb - c^tq_{n-1}.
\end{aligned}
$$

In compact notation:

$$Y_{e12} = Y_{12}^0 - Y_{12}^l Q_n^t \qquad (3.7-1)$$

with Y_{12}^0, Y_{12}^l and Q_n defined by 3.4-8, 3.4-9 and 3.4-12 respectively. Finally,

$$Y_{e22} = e_1b^tYbe_1^t + A_f^tY_{e12}^tbe_1^t + e_1b^tY_{e12}A_f + A_f^tY_{e22}A_f + \mathbf{q}_\mathbf{n}^t\mathbf{q}_\mathbf{n}$$

. Proceeding as above and using 3.7-1,

$$Y_{e22} = Y_{22}^0 - Y_{22}^lQ_n^t - Q_nY_{22}^{lt} + Q_nQ_n^t \qquad (3.7-2)$$

with Y_{22}^0 and Y_{22}^l defined by 3.4-10 and 3.4-11 respectively, thus establishing 3.4-7

□

<u>Proof of Corollary 3.1</u>: Note that $W(\mathbf{q}_\mathbf{n})$ in 3.4-13 may be factorized as

$$W(\mathbf{q}_\mathbf{n}) = \begin{bmatrix} I & 0 & xY_{12}^l \\ 0 & I & Y_{22}^l - Q_n \end{bmatrix} \begin{bmatrix} & & 0 \\ & U & \\ & & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & I \\ Y_{12}^{lt}x & Y_{22}^{lt} - Q_n^t \end{bmatrix}, \qquad (3.7-3)$$

where

$$U = \begin{bmatrix} xYx - xY_{12}^lY_{12}^{lt}x & xY_{12}^0 - xY_{12}^lY_{22}^{lt} \\ Y_{12}^{0t}x - Y_{22}^lY_{12}^{lt}x & Y_{22}^0 - Y_{22}^lY_{22}^{lt} \end{bmatrix}. \qquad (3.7-4)$$

From the definitions 3.4-8 through 3.4-11 and using the Lyapunov equations 3.4-4, 3.4-5:

$$\begin{aligned}
Y^l_{12}Y^{lt}_{12} &= Y - A^{nt}YA \\
Y^0_{12} - Y^l_{12}Y^{lt}_{22} &= [A^{nt}YA^{n-1}b \cdots A^{nt}Yb] \\
Y^0_{22} - Y^l_{22}Y^{lt}_{22} &= \begin{bmatrix} b^tA^{(n-1)t}YA^{n-1}b & \cdots & bA^{(n-1)t}YAb & b^tA^{(n-1)t}Yb \\ \vdots & & \vdots & \vdots \\ b^tA^tYA^{n-1}b & \cdots & b^tA^tYAb & b^tA^tYb \\ b^tYA^{n-1}b & \cdots & b^tYAb & b^tYb \end{bmatrix}
\end{aligned}$$

and then:

$$U = \begin{bmatrix} xA^{nt}y \\ b^tA^{(n-1)t}y \\ b^tA^{(n-2)t}y \\ \vdots \\ b^ty \end{bmatrix} \begin{bmatrix} yA^nx & yA^{n-1}b & \cdots & yb) \end{bmatrix}. \qquad (3.7-5)$$

Using this factorization and 3.7-3, $W(\mathbf{q_n}) = W_1(\mathbf{q_n})^tW_1(\mathbf{q_n})$, with $W_1(\mathbf{q_n})$ as in 3.4-14.

3.8 Appendix B: Additional Proofs for Section 3.5

The purpose of the first part of this Appendix is to prove the claim, used in Theorem 4, about the best anti-stable approximation to a stable system in an $\mathcal{H}_\infty$ sense. Recall that the Hankel operator Γ_g corresponding to the stable system $g(z) = c(zI - A)^{-1}b$ has finite rank, and

$$\Gamma_g\beta_i = \sigma_i\alpha_i \qquad (3.8-1)$$

for $i = 1, \cdots, n$, where n is the McMillan degree of g, σ_i is a (Hankel) singular value, and $\begin{pmatrix} \alpha_i & \beta_i \end{pmatrix}$ are the corresponding Schmidt pairs. The fact that the best approximation $h^1(z) \in H_\infty^\sim$ to $g(z)$ is given by $h^1(z) = g(z) - \rho\frac{\alpha(z)}{\beta(z)}$, where $\rho \doteq \overline{\sigma}(\Gamma_g)$ was first established by Adamjan, Arov and Krein. In the continuous time case, state-space formulas for $\alpha(z)$ and $\beta(z)$ are given in [25]. If $g(z)$ has no poles at zero (i.e., the A matrix is invertible) then the discrete time counterparts may be deduced following the same reasoning. The case with non-invertible A matrix requires the use of the DNF. From $\tilde{\beta}(z) \doteq \beta(1/z)^t \sim \begin{pmatrix} -zI + A & b & 0 \\ \nu^t A & \nu^t b & -1 \end{pmatrix}$ it follows that

$$\beta \sim \begin{pmatrix} zA^t - I & \nu & 0 \\ b^t & 0 & -1 \end{pmatrix}$$

and hence

$$g \cdot \beta \sim \begin{pmatrix} zA^t - I & 0 & \nu & 0 \\ bb^t & -zI + A & 0 & 0 \\ 0 & c & 0 & -1 \end{pmatrix}. \qquad (3.8-2)$$

Now let χ denote the anti-stable system with a DNF

$$\chi \doteq \begin{pmatrix} zA^t - I & \nu & 0 \\ -cXA^t & 0 & -1 \end{pmatrix},$$

where X denotes the controllability grammian of g. Then

$$\chi + g \cdot \beta \sim \begin{pmatrix} zA^t - I & 0 & \nu & 0 \\ bb^t & -zI + A & 0 & 0 \\ -cXA^t & c & 0 & -1 \end{pmatrix}.$$

Multiplying first on the left and then on the right by appropriate matrices, and using the fact that $X = AXA^t + bb^t$,

$$\chi + g \cdot \beta \quad \sim \quad \begin{pmatrix} zA^t - I & 0 & \nu & 0 \\ bb^t - zXA^t + AXA^t & -zI + A & 0 & 0 \\ 0 & c & 0 & -1 \end{pmatrix}$$

$$\sim \begin{pmatrix} zA^t - I & 0 & \nu & 0 \\ 0 & -zI + A & X\nu & 0 \\ 0 & c & 0 & -1 \end{pmatrix}.$$

From $X\nu = \rho\omega$ and eliminating unobservable states:

$$\chi + g \cdot \beta \sim \rho \begin{pmatrix} -zI + A & w & 0 \\ c & 0 & -1 \end{pmatrix}. \qquad (3.8-3)$$

That is, $\chi + g \cdot \beta = \rho\alpha$. Taking the stable part on both sides, one gets $\Gamma_g \beta = \rho\alpha$ as desired.

Proof of Theorem 3.5 Assume that $dim(A) > 0$ (otherwise the result is trivial). The proof is by induction on n. Consider the matrix

$$W_2(\mathrm{q}_\mathrm{n}^*) = \begin{pmatrix} cx & -q_0 & -q_1 & \cdots & -q_{n-1} \\ cAx & cb & -q_0 & \cdots & -q_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ cA^{n-1}x & cA^{n-2}b & \cdots & cb & -q_0 \\ yA^n x & yA^{n-1}b & \cdots & \cdots & yb \end{pmatrix}. \qquad (3.8-4)$$

It is clear that W_2 has the same singular values as W_1. Consider first:

$$W_2(\mathrm{q}_1^*) = \begin{pmatrix} cx & -q_0 \\ yAx & yb \end{pmatrix}.$$

Note that:

$$[xc^t \; xA^t y] \begin{bmatrix} cx \\ yAx \end{bmatrix} = xYx.$$

Let $\{\nu_1, \cdots, \nu_{r_u}\}$ be a set of orthonormal eigenvectors of xYx associated with ρ^2, and define $\hat{\nu}_i \doteq \begin{bmatrix} \nu_i \\ o \end{bmatrix}$. Then

$$\|W_2(\mathrm{q_n}^*)\hat{\nu}_i\|_2^2 = \left\| \begin{bmatrix} cx\nu \\ yAx\nu \end{bmatrix} \right\|_2^2 = |\nu^t xYx\nu| = \rho^2.$$

Assume that for an eigenvector ω of yXy associated with ρ^2, $b^t y\omega \neq 0$, and let

$$\hat{\nu}_0 \doteq \begin{bmatrix} xA^t y \\ b^t y \end{bmatrix} \omega$$

with $\|\hat{\nu}_0\| = \rho$. Then:

$$W_2(\mathbf{q}_1^*)\hat{\nu}_0 = \begin{bmatrix} (cXA^t y - q_0 b^t y)\omega \\ yXy\omega \end{bmatrix} = \begin{bmatrix} (cXA^t y - q_0 b^t y)\omega \\ \rho^2 \omega \end{bmatrix}.$$

Therefore $\|W_2(\mathbf{q}_1^*)\hat{\nu}_0\|/\|\hat{\nu}_0\| \geq \rho$; but since $\overline{\sigma}(W_2) = \rho$, equality must hold. By assumption $b^t y\omega \neq 0$, implying that $\{\hat{\nu}_0 \ \hat{\nu}_1 \ \cdots \hat{\nu}_{r_u}\}$ is a set of linearly independent vectors such that $\|W_2(\mathbf{q_n}^*)\hat{\nu}_i\|/\|\nu_i\| = \rho$. Since $\overline{\sigma}[W_2(\mathbf{q_n}^*)] = \rho$, the multiplicity of the largest singular value is at least $r_u + 1$.

Next suppose that no ω exists such that $b^t y\omega \neq 0$, but $b^t A^t y\omega \neq 0$ for some ω. Define

$$\hat{\nu}_0 \doteq \begin{pmatrix} xA^{2t} y \\ b^t A^t y \end{pmatrix} \omega.$$

It is easy to check that $\|\hat{\nu}_0\| = \rho$, and

$$W_2(\mathbf{h}_1^*)\hat{\nu}_0 = \begin{bmatrix} (cXA^{2t} y - q_0 b^t A^t y)\omega \\ yXA^t y\omega \end{bmatrix}.$$

The claim is that $\|yXA^t y\omega\|^2 = \omega^t yAXYXA^t y\omega = \rho^4$. To see this, note that

$$\omega^t yXYXy\omega = \rho^4$$

and then from 3.4-1:

$$\omega^t yXYXy\omega = \omega^t y(AXA^t + bb^t)Y(AXA^t + bb^t)y\omega.$$

The claim follows from the fact that $b^t y\omega = 0$. If $b^t A^t y\omega = 0$ for every i then the same argument may be repeated for increasing powers of A until $b^t A^{jt} y\omega \neq 0$ for some ω and some j necessarily finite.

Now suppose that for $n-1$ the multiplicity is r_u+n-1. Let

$$W_2(\mathbf{q_n}^*) = \left[\begin{array}{cccc|c} cx & -q_0 & \cdots & -q_{n-2} & -q_{n-1} \\ cAx & cb & \cdots & -q_{n-3} & -q_{n-2} \\ \vdots & \vdots & \vdots & \vdots & \cdots \\ \hline cA^{n-1}x & cA^{n-2}b & \cdots & cb & -q_0 \\ yA^n x & yA^{n-1}b & \cdots & yAb & yb \end{array}\right] = \left[\begin{array}{cc} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{array}\right]. \quad (3.8-5)$$

Then

$$[Y_{11}^t \;\; Y_{21}^t]\left[\begin{array}{c} Y_{11} \\ Y_{21} \end{array}\right] = Y_{11}^t Y_{11} +$$

$$+\left[\begin{array}{cccc} xA^{(n-1)t}YA^{n-1}x & xA^{(n-1)t}YA^{n-2}b & \cdots & xA^{(n-1)t}Yb \\ b^tA^{(n-2)t}YA^{n-1}x & b^tA^{(n-2)t}YA^{n-2}b & \cdots & xA^{(n-2)t}Yb \\ \vdots & \vdots & \vdots & \vdots \\ b^tYA^{n-1}x & b^tYA^{n-2}b & \cdots & xYb \end{array}\right],$$

which implies $[Y_{11}^t \quad Y_{21}^t]\left[\begin{array}{c} Y_{11} \\ Y_{21} \end{array}\right] = W_2(\mathbf{q_{n-1}}^*)^t W_2(\mathbf{q_{n-1}}^*)$. By the inductive hypothesis, $W_2(\mathbf{q_n}^*)$ has $n+r_u-1$ singular vectors associated with ρ of the form $\hat{\nu}_i = \left(\begin{array}{c} \nu_i \\ 0 \end{array}\right)$, $i=1,\cdots,n+r_0-1$, where the ν_i's are singular vectors of $W_2(\mathbf{q_{n-1}}^*)$ associated with ρ. Assume that $b^t y\omega \neq 0$ and define

$$\hat{\nu}_0 = \left[\begin{array}{c} xA^{nt}y \\ b^tA^{(n-1)t}y \\ \vdots \\ b^tA^ty \\ b^ty \end{array}\right]\omega$$

such that $\|W_2(\mathbf{Q_n})\hat{\nu}_0\|_2/\|\hat{\nu}_0\|_2 = \rho$. Since by assumption $b^t y\omega \neq 0$, the set $\{\hat{\nu}_i, \;\; i = 0,\cdots,n+r_u-1\}$ is linearly independent. If $b^t y\omega = 0$ for every ω then the extra singular vector must be modified as in the first part of the proof to guarantee that the set is linearly independent.

□

Chapter 4

Numerical Optimization

In this chapter, the problem of computing the optimal constrained $\mathcal{H}_\infty$ norm is addressed. The two main characteristics that make the solution to the problem difficult, i.e., nondifferentiability and large scale, are recognized. Descent methods for nondifferentiable optimization are first considered together with the reasons that renders them inefficient for the problem. A particular non-descent method, namely the Ellipsoid Algorithm, is reviewed and its advantages over descent methods are discussed. The computation of the objective function is shown to be the critical step in the application of the ellipsoid algorithm and hence the topic is treated in detail. Finally, the calculation of a starting point for the algorithm is discussed and a "textbook" example of a time domain constrained $\mathcal{H}_\infty$ control problem is solved in detail.

4.1 Introduction

This chapter is devoted to study a numerically efficient algorithm for solving the time domain constrained $\mathcal{H}_\infty$ control problem considered in Chapter 3. The numerical problem that has to be solved is of the form:

$$\begin{aligned} \min \quad & \varphi(\mathbf{q_n}). \\ s.t. \quad & S\mathbf{q_n} \leq s \end{aligned} \qquad (4.1-1)$$

From Theorem 3.3 and its corollaries, $\varphi(\mathbf{q_n})$ may be alternatively computed as $\rho(W(\mathbf{q_n}))$, $\overline{\sigma}(W_1(\mathbf{q_n}))^2$ or $\overline{\sigma}(W_s(\mathbf{q_n}))^2$. Here $W(\cdot)$, $W_1(\cdot)$ are symmetric matrices, $W(\cdot)$, $W_1(\cdot)$, $W_s(\cdot) \in R^{r+n \times r+n}$, r denotes the number of states in the realization of g, and n denotes the length of the horizon. The number of variables of the problem is n, while the number of constraints is $2tn$, where t denotes the number of specified input-output time domain constraints. The factor of 2 arises because each sample is usually constrained to lie within a given upper and lower bound. If n is relatively large, then it is clear that 4.1-1 becomes a large-scale problem. This is the first main characteristic of the problem that needs to be taken into account when formulating an algorithm for computing its solution.

The eigenvalue minimization of symmetric matrices appears in many different areas of mathematics and engineering, and has received considerable attention in recent years. Indeed, the closely related problem of optimizing over positive semidefinite matrices has been shown to be equivalent to a many different convex programming problems [52]. The reason for the difficulty of the problem and the ongoing activity in this research area, is that the eigenvalues of a matrix function will not be differentiable at the points where they coalesce, even if the matrix function depends smoothly on the variables. The occurrence of non differentiable points during minimization cannot be discarded as pathological cases, since the optimization objective tends to squeeze the eigenvalues together and hence to increase the multiplicity. This intuitive argument is strengthened by Theorem 3.5, since the multiplicity in the unconstrained case is equal to the number of variables.

To illustrate the short-comings of nondifferentiability, consider the following algorithm for solving Problem 4.1-1 that attempts to exploit the information provided by the largest singular value *and* an associated singular vector.

Algorithm 4.1 *A Naive Procedure.*

Step 1. Let $\mathrm{q}_{\mathrm{n}}^{0}$ *be such that* $S\mathrm{q}_{\mathrm{n}}^{0} \leq s,\ k = 0.$

Step 2. Let $\sigma^k,\ u^k,\ v^k$ *be such that* $\sigma^k = \overline{\sigma}(M_1(\mathrm{q}_{\mathrm{n}}^{\mathrm{k}})),\ \|u^k\|,\ \|v^k\| = 1,$ $u^{kt}M_1(\mathrm{q}_{\mathrm{n}}^{\mathrm{k}})v^k = \sigma^k.$

Step 3. Let $\mathrm{q}_{\mathrm{n}}^{\mathrm{k+1}} = arg\{\min u^{kt}M_1(\mathrm{q}_{\mathrm{n}})v^k,\ s.t.\ S\mathrm{q}_{\mathrm{n}} \leq s\}.$

Step 4. Set $k = k+1$ *and goto* 2.

The appealing characteristic of the algorithm, is that the minimization problem in *step 3* is a linear programming problem (LP), for which efficient codes exist. Therefore, the optimization is reduced to solving a sequence of LP's, although the algorithm is certainly crude (in particular, one would add a trust region in *step 3* and a linear search between *steps 3* and *4*). Suppose now that at the k-th iteration, the largest singular value of $W_1(\mathrm{q}_{\mathrm{n}}^{\mathrm{k}})$ has multiplicity one. Then $\overline{\sigma}$ is differentiable and

$$\frac{\partial\overline{\sigma}(W_1(\mathrm{q}_{\mathrm{n}}^{\mathrm{k}}))}{\partial q_i^k} = u^{kt}W_{1i}v^k, \qquad (4.1-2)$$

where W_{1i} denotes the derivative of W_1 with respect to q_i. Then, it is easy to see that the cost vector for the problem in *Step 3* is the gradient of the function and therefore Algorithm 4.1 is just a gradient-based descent procedure. Unfortunately, if at $\mathrm{q}_{\mathrm{n}}^{\mathrm{k}}$ the multiplicity of the largest singular value is larger than one, then the gradient does *not* exist at that point, or equivalently, the cost vector is not uniquely defined since u^k and v^k may be chosen from a subspace with the same dimension as the multiplicity. As a consequence, the value of the objective function can no longer be guaranteed to decrease, and hence the whole idea of the algorithm breaks down. It is possible to circumvent this difficulty by considering the so-called "generalized gradient" or "subgradient," to be discussed next.

4.2 Nondifferentiable Optimization

Consider the nonlinear optimization problem:

$$\min f_0(x) \quad s.t. f_i(x) \leq 0 \quad i = 1, \cdots, m \tag{4.2-1}$$

where f_0 is a convex continuous function of x. When $f(x)$ is not continuously differentiable, then 4.2-1 is a problem of nondifferentiable optimization. These problems have received much attention during the past decades due to the wide range of applications that they encompass. The interested reader is referred to [68] for a comprehensive introduction and [43] for numerical considerations and an extensive bibliography up to 1985. Engineering problems are considered in the tutorial [57] which also includes the semi-infinite case.

Nondifferentiable problems have inherent poor analytic properties, making it hard to develop local models for the objective functions, to be used for computations. Algorithm 4.1 illustrates this fact, since it is seen to have a plausible behavior at those points where the gradient exists (note that in this case the gradient provides a good model for the local behavior of the function) but this plausibility breaks down together with differentiability. When Problem 4.2-1 is Locally Lipchitz (a property verified by most problems of practical interest) then Clarke's work provides a framework for unifying this nondifferentiable problem with standard differentiable optimization theory. In particular, it provides a way of establishing nondifferentiable analogues to the Kuhn-Tucker or F. John conditions for optimality, and to the Lagrange Multipiers. Unfortunately, the development of nondifferentiable algorithms has proven to be more arduous, and although there exist efficient algorithms for some specific problems, no general purpose nondifferentiable optimization algorithm is currently available. In particular, most "general purpose" algorithms in the literature cannot deal with large-scale programs and hence cannot be used to solve the problem of interest.

Let f_0 be a convex continuous differentiable function. Then, from classical analysis and for every x

$$f_0(x) \geq f_0(x_0) + \nabla f_0(x_0)^t (x - x_0),$$

where $\nabla f_0(x_0)$ denotes the gradient of f_0 at x_0. When f_0 is non differentiable, ∇f_0 may be substituted by the generalized gradient or sub-gradient defined as [13]:

$$\partial f_0(x_0) \doteq \{g \in R^n \;\; s.t. \;\; f_0(x) \geq f_0(x_0) + g^t(x - x_0)\}.$$

It is well known that ∂f_0 is a nonempty convex compact set which reduces to $\{\nabla f_0(x_0)\}$ whenever the function is differentiable at x_0. Moreover, the negative of the unique element of $\partial f_0(x_0)$ with minimum 2-norm is a descent direction (in fact, the first order steepest descent direction) for f_0 at x_0. Unfortunately, the generalized gradient can be discontinuous and hence ∂f_0 cannot be used to formulate algorithms for computing unconstrained minimums. An example of one such algorithm that produces a sequence converging to a non-optimal stationary point may be found in [75]. Instead, ∇f_0 should be replaced by the smeared sub gradient or ϵ-generalized gradient defined as:

$$\partial_\epsilon f_0(x_0) \doteq \{g \in R^n \;\; s.t. \;\; f_0(x) \geq f_0(x_0) + g^t(x - x_0) + \epsilon, \quad \epsilon \geq 0\}.$$

This set, that reduces to ∂f_0 when $\epsilon = 0$, collects the generalized gradient information for a whole neighborhood of x_0. In most algorithms based on the ϵ generalized gradient, ϵ is taken big enough at the beginning of the iterations and then it is gradually reduced to zero as the algorithm converges.

Generalized Gradient for the Largest Eigenvalue and Singular Value

Suppose that the maximum eigenvalue of the symmetric matrix function $M(x)$ has multiplicity t, with a corresponding orthonormal basis of eigenvalues $U(x)$. Then, defining $< B, C >= tr(B^t C)$, and $M_k = \partial M / \partial x_k$, the generalized gradient of $\overline{\lambda}(x)$

is the set:

$$\partial\overline{\lambda}(x) = \{g \in R^n \ \ s.\ t.\ \ g_k =< W, U(x)^t M_k(x) U(x) >,\ for\ some$$
$$W\ symmetric\ and\ s.t.\ W \geq 0,\ tr(W) = 1\}.$$

This formula for the generalized gradient is from [54], where a self-contained derivation is given. Moreover, using the formula and the generalized gradient algebra from [13]

$$\partial\overline{\sigma}(x) = \{g \in R^n \ \ s.\ t.\ \ g_k =< W, U(x)^t M_k(x) V(x) >,\ for\ some$$
$$W\ symmetric\ and\ s.t.\ W \geq 0, tr(W) = 1\}.$$

where $U(x)$ and $V(x)$ denote a set of orthonormal left and right singular vectors of $M(x)$ associated with $\overline{\sigma}$. In particular, taking $M = W_s$, with W_s defined by 3.4-14 and the notation of the previous section, it is not hard to get the following expression for the components of a vector $g \in \partial\varphi(\mathbf{q_n})$:

$$g_k = \sum_{i=0}^{n-k} u_{r+k+i} v_{n+k-i}, \tag{4.2 - 2}$$

where u and v denote a left and a right singular vector respectively of $M_1(\mathbf{q_n})$, associated with the largest singular value.

4.3 Descent Methods

Algorithms for nondifferentiable optimization can be roughly classified into two groups: non-descent and descent methods. In this section, descent methods are considered. Two such algorithms are briefly described: one designed for problems with linear constraints and a nondifferentiable objective function, and one specialized on eigenvalue minimization.

Descent methods attempt to construct a sequence of feasible points with monotonically decreasing cost, that will converge to the optimal solution if one exists.

They have an intuitive appeal, and were the first choice in attempting to solve the constrained minimization problem of interest; however, they had to be discarded for reasons to be explained below. Part of the appeal of descent methods, comes from the fact that they usually constitute a generalization of well established differential optimization algorithms. As stated in the previous section, on performing this extension it is necessary to replace the gradient by the ϵ-generalized gradient, and drive ϵ to zero as the optimal solution is approached. The lack of first (and hence higher) derivatives usually make these methods converge at most at linear rate, although quadratic convergence can be achieved in some cases. For instance, the algorithm in [53] to be discussed next, is claimed to have quadratic convergence if a certain parameter is chosen carefully. Descent methods are not new in the robust control literature; for instance, an upper bound over an appropriately defined notion of stability margin (i.e. the μ function) may be computed by the unconstrained minimization of a convex nondifferentiable function, and in [16] a descent algorithm is proposed for performing this task.

When formulating an algorithm for general nondifferentiable programming, it is usually not reasonable to assume that a complete description of the generalized gradient exists, since this may be expensive to compute or even not available. Due to this reason, most algorithms start by assuming that only one vector on the ϵ-generalized gradient is available and then proceed to construct a model for the generalized gradient using the information obtained in subsequent iterations. This gives rise to algorithms consisting of two kind of steps: a *real* step, in which a linear search along a descent direction is performed, and a *null* step, in which the model for the ϵ-generalized gradient is updated with no improvement on the value of the objective function. Algorithms of this type are available, both for unconstrained and constrained optimization. [43,44,58]. Consider, for instance, the algorithm for solving linearly constrained convex nondifferentiable problems in [44], which guarantees convergence to the optimal solution and requires the evaluation of the objective function and only one vector in the generalized gradient for each feasible point. At each iteration, the algorithm computes a descent direction, performs a linear search, and then updates the model for the generalized gradient. The descent direction is computed by solving a quadratic programming problem (QP), that has the original constraints of the problem plus a polyhedral approximation to the objective function obtained from previous iterations. The contribution of [44] over previous closely related works ([47], for instance) is that the number of constraints in the QP does not increase with each iteration, since the approximation is constructed with only a few previously calculated vectors in the generalized gradient and an aggregated subgradient [42] is computed as a convex combination of past vectors. Therefore the algorithm has bounded storage requirements, but if the number of constraints and variables is large, then the time required for solving the QP at each iteration becomes unreasonably long and hence the whole approach is ill suited for large-scale problems. Actually, most of the descent algorithms found in the literature use a QP for computing a descent direction, and are, for the same reasons, not adequate to deal with large-scale problems. In order to alleviate this difficulty, one may attempt to replace the QP by an LP for computing the descent direction, or attempt to update the solution to the problem from one iteration to the other using perturbation techniques. However, no algorithm that addresses the large-scale issue and preserves guaranteed convergence has currently been given.

The fact that a relatively simple description exists for the generalized gradient in the case where the objective function is the largest eigenvalue or singular value of a matrix, suggests that specialized methods might exist for solving these problems. This was explored by Overton and co-workers in a series of papers [26,53,54], using several results in nondifferentiable analysis. The basic idea of the approach is to formulate the necessary conditions for optimality (which is also sufficient in the case of interest since the functions involved are convex) and then iterate trying to satisfy this condition. The descent direction is computed based on the violation of the condition, in a way that resembles the procedure for moving off constraints associated with Lagrange multipliers with the wrong sign in differentiable optimization. The basic iteration of the algorithm involves the solution of a QP for computing the descent direction, and a linear search for updating the value of the function. This results in an overall Successive QP (SQP) algorithm, and it is claimed in [54] that, if the quadratic term in the objective functions of the QP's is chosen carefully, then quadratic convergence can be obtained. Unfortunately, the constraints of the problem include the ones of the original one plus $(3n - n + r(r+1)/2)$,where n denotes the number of variables and r denote the estimated multiplicity of the largest eigenvalue. Moreover, in [54] only equality constraints are considered, and therefore slack variables are needed in order to deal with the inequalities. The basic algorithm has been reported to perform satisfactorily for a small number of variables (i.e., $n \leq 40$), but it becomes very inefficient for larger problems. Given that for large-scale problems, the time required for solving a QP outweighs the benefits of quadratic convergence, an LP that preserves the key features of the SQP was considered in [54]. Moreover, the LP was replaced by a partial LP solver that works with a smaller number of constraints. The fact that the solution is forced to lie within a trust radius (responsible for 2n additional constraints), makes this approach reasonable. Although this modification reduces computation time, the procedure is no longer guaranteed to converge to the optimal value. Another critical step of the algorithm is the computation of the objective function and the characterization of the generalized gradient. Since only the largest eigenvalues are of interest, it is not efficient to use the standard QR algorithm for eigenvalue decomposition. This is specially true because the matrices on each iteration do not differ much (except, perhaps, for the first few) but this information cannot be incorporated into the QR calculations, since no effective updating scheme exists. In [54] several alternate schemes are discussed for the computation, but when the multiplicity of the largest eigenvalue or singular value is large, these schemes may perform poorly. A more detailed discussion on the eigenvalue computation is deferred to Section 4.5.

To summarize, general descent methods are inefficient for solving the time domain constrained $\mathcal{H}_\infty$ problem because the computation of the descent directions is too time consuming. Specialized algorithms are in principle more promising, but unfortunately, in the numerical experiments, they also perform poorly when the number of variables *and* constraints is large. Further work should be done to find an implementation of the latter algorithms tuned up for the requirements of the minimization problem of interest.

4.4 The Ellipsoid Algorithm

Having reviewed and discarded descent methods for solving the minimization problem of interest, the purpose now is to discuss an algorithm capable of solving it, even when a large number of variables and constraints are used, namely the Ellipsoid Algorithm (EA). The EA was chosen among other non-descent methods in the belief that it is the best suited to address simultaneously the two main characteristics of the problem, i.e., nondifferentiability and large-scale. The other popular algorithm for convex programming is Kelley's cutting plane algorithm [41,7], but it was discarded because it has, in principle, unbounded storage requirements. It is noteworthy to say that the distinction between descent and non-descent algorithms is vague; for instance Schor gave an interpretation of the EA as a variable-metric descent algorithm [68].

The EA was initially developed by the Russian mathematicians Iudin, Nemirovsky and Schor but it only began to draw attention when Khachiyan used it to give a polynomial time algorithm for LP. Some years later, when Karmarkar gave an ellipsoid-type algorithm which seemed to surpass Simplex as a tool for solving linear problems, this approach became a focus of intense research. In particular, and following the original motivation of Iudin, Nemirovsky and Schor, several improvements and a numerically stable implementation were given for solving convex nondifferentiable problems. In this section, a brief introduction to the EA is given; the interested reader is referred to [6] and [7] for a detailed treatment; application to convex nonlinear programming can be found in [20].

The Ellipsoid Algorithm constructs a sequence of ellipsoids E_0, E_1, $\cdots$ with decreasing volume, in such a way that if the first E_0 contains the optimal solution, then each one of the E_i also contains it. Since a positive lower bound over the rate by which the volume decreases is available, then the volume of E_0 determines a finite upper bound on the number of iterations required to confine the optimal solution within a given tolerance. On the $(k+1)$st iteration, the algorithm checks whether the center x_k of the ellipsoid E_k satisfies the constraints. If not, then one of the violated constraints is selected, say $a \cdot x_k \leq b$, and the ellipsoid of minimum volume that contains the half ellipsoid $\{x \in E_k \ / \ a \cdot x \leq a \cdot x_k\}$ is constructed. Note that, by linearity, all the points discarded also violate the selected constraint, implying that all feasible points contained in E_k are also contained in the new ellipsoid. If the point x_k is already feasible, then a vector g_k in the generalized gradient of the objective function at x_k is computed, and the ellipsoid of minimum volume that contains the half ellipsoid $\{x \in E_k \ / \ g_k^t x \leq g_k^t x_k\}$ is constructed. By the definition of the generalized gradient, the objective function evaluated at any of the discarded point is at least as large as the value at x_k; i.e., all points in E_k with costs less than the value of the objective function at x_k are preserved. Therefore, each new ellipsoid contains the minimizer of the objective function, assuming that it was contained in the original E_0.

The ellipsoid E_k, centered at x_k can be represented as

$$E_k = \{x \in R^n \;\; s.\,t.\; (x - x_k)^t H_k^{-1}(x - x_k) \leq 1\}, \qquad (4.4-1)$$

where H_k is a symmetric positive semidefinite matrix. Then, it is shown in [6] that for the next iteration

$$x_{k+1} = x_k - \tau(H_k a / \sqrt{a^t H_k a}) \qquad (4.4\text{-}2)$$

$$H_{k+1} = \delta(H_k - \sigma(H_k a (H_k a)^t / (a^t H_k a))), \qquad (4.4\text{-}3)$$

where

$$\tau = 1/(n+1), \quad \sigma = 2/(n+1), \quad \delta = n^2/(n^2 - 1)$$

for a "feasibility iteration," that is, an iteration with x_k unfeasible in which a is selected as one of the violated constraints. For a "function iteration" the same formulas hold with the vector g_k in the generalized gradient replacing a. Moreover, defining

$$u_k \doteq \min_{i=1,\cdots,k} f_0(x_i)$$

$$l_k \doteq \max_{i=1,\cdots,k} \left\{ f_0(x_i) - \sqrt{a_i^t H_i a_i} \right\},$$

then u_k and l_k give readily computable upper and lower bounds over the optimum, and then it is easy to get a stopping criterion.

Several implementation aspects should be taken into account to avoid numerical instability and accelerate convergence. If the starting H_0 is positive definite, the H_k's are positive definite for all values of k. However, due to roundoff errors the matrix may become singular or even indefinite, which, from 4.4-2, 4.4-3 is unacceptable. Fortunately, there is an easy way of circumventing this difficulty, by taking the factorization $H_k = J_k J_k^t$ and expressing the formulas in terms of J_k. In particular, the value of J_k can be updated directly, with no explicit computation of H_k required. See [6] for the formulas. Convergence can be accelerated by introducing the so-called "deep-cuts" in the algorithm. Suppose that, at the $k-th$ iteration, the algorithm is working on a feasibility iteration, i.e., the new ellipsoid is computed so that $ax \leq ax_k$, for each $x \in E_{k+1}$. But a feasible point should rather satisfy $ax \leq b$, for some b, and hence it is intuitively clear that an E_{k+1} that contains those points and has smaller volume may be used instead of the one given by 4.4-2, 4.4-3.

Note, also, that the constraints in the $\mathcal{H}_\infty$ problems usually come in pairs, i.e., for each constraint of the form $ax \leq b_1$ there is another one of the form $-ax \leq b_2$ (this is because the time response is usually specified to lie within upper and lower bounds) and then it is possible to use both constraints simultaneously to compute the new ellipsoid. This results in a further reduction of the volume of the updated ellipsoid. Let $\alpha = (ax_k - b_1)/\sqrt{aH_ka^t}$ and $\hat{\alpha} = (ax_k - b_2)/\sqrt{(aH_ka^t)}$, and suppose that $\alpha\hat{\alpha} < 1/n$ and $\alpha \leq -\hat{\alpha} \leq 1$. Then the update formulas are 4.4-2, 4.4-3 with:

$$\begin{aligned}
\sigma &= (1/(n+1))(n + (2/(\alpha - \hat{\alpha})^2)(1 - \alpha\hat{\alpha} - \rho/2)) \\
\tau &= ((\alpha - \hat{\alpha}))\sigma \\
\delta &= (n^2/(n^2-1))(1 - (\alpha^2 + \hat{\alpha}^2 - \rho/n)/2) \\
\rho &= \sqrt{4(1-\alpha^2)(1-\hat{\alpha}^2) + n^2(\hat{\alpha}^2 - \alpha^2)^2}.
\end{aligned}$$

These formulas were again taken from [6], as originally proposed by [66]. A similar deep-cut may be implemented for function iterations. In that case, let x^r be a point such that $f_0(x^r) \leq f_0(x_i)$ for all $0 \leq i \leq k$. Then, the optimal point lies in the subspace $g_k^t x \leq g_k^t x_k + (f_0(x^r) - f_0(x_k))$. Since the term on the right is less than zero, the resulting ellipsoid will again have a smaller volume.

Several other alternative deep-cuts have been proposed in the literature, but they usually require either a linear search or some other additional computation. However, from numerical experience [20] deep-cuts do not necessarily produce a substantial acceleration on the convergence rate for the EA as applied to general convex programming, and therefore only the simple ones described above were implemented.

In fact, as discussed in the following section, the speed of convergence is more related to the appropriate selection of the algorithm for computing the objective function.

4.5 Computation of the Objective Function

A nice property of the Ellipsoid algorithm is that it is relatively easy to obtain an upper bound for the number of iterations required to compute a solution that comes as close to the minimum as desired. The key reason for this is that, if $E_k \subset R^n$, then

$$vol(E_{k+1}) < e^{-1/2n} vol(E_k).$$

Let ϵ be given, and let K be such that $\varphi(\mathbf{q_n^K}) > \varphi(\mathbf{q_n^*}) + \epsilon$, $\varphi(\mathbf{q_n^{K+1}}) \leq \varphi(\mathbf{q_n^*}) + \epsilon$, where $\mathbf{q_n^*}$ denotes the optimal solution. Then, from *14.4.2* in [7],

$$e^{-K/2n} vol(E_0) \geq (\epsilon/G_n)^n \beta_n,$$

where $G_n \doteq \max\{\|g\| \text{s.t.} g \in \partial\varphi(\mathrm{q_n}), \mathrm{q_n} \in E_0\}$ and β_n denotes the volume of the unit ball in R^n. It is a standard result that $vol(E_0) = det(H_0)\beta_n$, and since from 4.2-2, $G_n \leq 1$, then

$$e^{-K/2n} \det(H_0)\beta_n \geq \epsilon^n \beta_n.$$

Let h denotes the algebraic mean of the initial axis lengths, i.e., $h \doteq (\Pi_{i=1}^n h_{ii})^{1/n}$. Then, taking the *log* in the expression above:

$$-k/2n + n \log h > n \log \epsilon,$$

or

$$K < 2n^2 log(h/\epsilon). \qquad (4.5-2)$$

The formula 4.5-1 used to derive the upper bound was actually proved in [7] for *unconstrained* optimization. However, it is not difficult to extend it to the constrained case if a violation ϵ of the constraints is tolerated, by redefining G_n to include the gradients of the constraints.

Given that the upper bound on the number of iterations is quadratic in the number of variables, it is clear that in order to get an efficient algorithm, the computation time per iteration has to be small. Consider first a feasibility iteration, and let c_f denote the flop count. From the previous chapter, the constraints are of the form

$$lb^i \leq \mathbf{q_n} V^i \leq ub^i, \quad i = 1, \cdots, t,$$

where t denotes the number of specified input-output time domain constraints, and V^i is an upper triangular Toeplitz matrix. Since the number of flops for such iteration step, is essentially given by the number of flops required for evaluating the constraints,

$$c_f \sim n^2 t.$$

On the other hand, function iterations are much more computationally demanding. Note first, that the evaluation of $W(\cdot)$ at each iteration requires $o(n^3)$ flops, while both $W_1(\cdot)$ and $W_s(\cdot)$ can be stored when the algorithm is initialized and require no further calculations (except for the update of the vector $\mathbf{q_n}$). Moreover, the computation of a vector in the generalized gradient depends linearly on n for $W_1(\cdot)$ and $W_s(\cdot)$, but requires $o(n^2)$ computations for $W(\cdot)$. The number of flops required for the computation of the eigenvalues of a symmetric matrix is smaller that the one required for an SVD computation, $\rho(W_s)$, and hence the best choice for the objective function is $\varphi(\mathbf{q_n}) = \rho(W_s(\mathbf{q_n}))$.

It is well known that the computation of the eigenvalue decomposition for a symmetric matrix requires $o(n^3)$ flops; with a cubic growth in the cost of each iteration, the optimization algorithm becomes very slow for problems with a large number of variables. Indeed, with more than 60 variables the behavior of the algorithm deteriorates rapidly, and hence an alternative algorithm for computing the largest eigenvalue was pursued. Following the reasoning in [54], section 6, the extended QR algorithm is inadequate for this task since *a)* only the largest eigenvalue are of interest and *b)* matrices will not differ much from iteration to iteration after the first few. In [54] subspace iterations were selected as the appropriate algorithm for computing the eigenvalues and eigenvectors, because the computation of the descent direction requires a complete description of the generalized gradient and hence of the subspace associated not only with the largest eigenvalue but also with every other eigenvalue in a neighborhood of the largest one. Subspace iterations generalize the idea of the power iteration, and are adequate to deal with problems that have a cluster of eigenvalues close or equal to the largest one and a good separation between eigenvalues otherwise. The separation of the eigenvalues is a critical issue, and has to be treated with great care [54]. In the time-domain constrained $\mathcal{H}_\infty$ problem, it is usually the case that a large number of eigenvalues have absolute values close to the maximum one, and no clear separation between "large" and "small" eigenvalues exists. The unconstrained case is the extreme example of this behavior, having $n+1$ eigenvalues equal to the maximum. The existence of large dimensional invariant subspaces associated with a cluster of large eigenvalues implies that subspace iteration may become too costly.

In the previous section, it was shown that the EA requires, for each function iteration, one vector in the generalized gradient. This is equivalent to having a good estimate of the largest eigenvalue, since an eigenvector may be then recovered using an inverse iteration. From this eigenvector, a vector in the generalized gradient may be computed using Formula 4.1-2. The vector-Lanczos iteration [55,14,31] appears to be the best choice for this purpose because of its fast convergence rate, even in the unfavorable case of matrices with a cluster of largest eigenvalues. In contrast, the power method [31] is not useful for the present problem because it converges quickly if the ratio $|\lambda_2/\lambda_1|$ is sufficiently smaller than 1, where λ_1 and λ_2 denote the largest and second largest eigenvalues (in absolute value). Lanczos iteration consists of two steps: *a)* computation of a reduced tridiagonal form, and *b)* computation of the largest eigenvalue and associated eigenvector for the reduced form. The following is an algorithm for reducing to tridiagonal form, taken from [31].

Algorithm 4.2 *Reduction of W to Tridiagonal Form.*

Step 1. Let $w = 0$; $\beta_0 = 1$, $j = 0$ *and* v *be a random unitary vector.*

Step 2. If $\beta_j \neq 0$ *goto* 3. *Else stop.*

Step 3. $t = v$; $w = v/\beta_j$.

Step 4. $v = Ww$.

Step 5. $j = j + 1$; $\alpha_j = w^t v$; $v = v - \alpha_j w - \beta_{j-1} t$; $\beta_j = \|v\|_2$.

Step 6. Goto 2.

The tridiagonal matrix W_r is formed by taking the α_i's in the diagonal and the β_i's in the upper and lower subdiagonals. If convergence has occurred, the eigenvalues of W_r give some of the extreme eigenvalues of W. Several interesting things may happen to this iteration, and the reader is referred to [55,14] for a detailed discussion. The vector w in Step 1 may be taken from the previous eigenvalue computation, and this may help to accelerate convergence if the matrices have not varied too much. The final implementation was a variation of the above procedure called "Recursive Lanczos Iteration" [14]. The condition $\beta_j \to 0$ is associated with the convergence of the whole reduced tridiagonalization. However, it is possible to stop the algorithm after a few iterations and check for convergence of the eigenvalues, which results in a huge saving of computing time. In the case of interest, the matrix W is not sparse and hence the flop count for each iteration of Algorithm 4.2 is $(2n + 8)n$. Computer experience suggests that $o(\log n)$ iterations are needed to converge to an eigenvalue, and hence the total count adds to $o(n^2 \log n)$ flops, which is a substantial improvement over the $o(n^3)$ of the QR method.

An additional advantage of Lanczos iteration over a standard QR algorithm is that the storage requirement is much smaller. This is a general property of a Lanczos algorithm, since it only requires the storage of the β_i's and α_i's, and probably the v_i's if one wishes to compute the eigenvector of the whole matrix from the estimate of the reduced one. It is also specially true in the present case, since although the matrix $W_1(\cdot)$ is not sparse, its structure can be exploited to compute the product $W_1(\cdot) \cdot w$ without needing to store the whole matrix in memory. This saving in memory storage becomes important for large problems, and in particular when solving the general instance of the time domain constrained $\mathcal{H}_\infty$ control problem.

4.6 Starting Point for the EA

The starting point for the ellipsoid algorithm needs not be feasible, because if at least one feasible point is contained in the initial ellipsoid E_0, then the algorithm will perform feasibility iterations until no constraint is violated. The algorithm may then be started from any point, as long as the initial ellipsoid is taken big enough to contain a solution. On the other hand, from the upper bound 4.5-2 on the number of iterations required by the algorithm, one would like the initial ellipsoid to be as small as possible. It then makes sense to start the algorithm from a feasible point that is cheap to compute and which approximates the optimal solution, so that E_0 may be taken smaller. Of course, in the absence of any analytical result relating the optimal solution to the approximate one, the initial ellipsoid needs to be defined by some ad-hoc rule, and hence it cannot be guaranteed to contain a solution to the optimization problem. Therefore, upon completion of the procedure, it is necessary to check whether the final point lies strictly inside the original ellipsoid. If not, then the lower bound computed over the optimal value may not be valid and the algorithm should be restarted.

Consider the matrix function $W_s(\mathbf{q_n})$. The natural approach would be to replace the minimization of $\|W_s(\cdot)\|_2$ by $\|W_s(\cdot)\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm. The solution of the resulting problem provides an upper bound for the function of interest, and is usually much cheaper to compute. Note, however, that in the unconstrained case the solution to the F-norm minimization gives $q_0 = q_2 = \cdots = q_{n-1} = 0$, which is in general different from the actual solution. Consider instead the function

$$\phi(\mathrm{q_n}) \doteq \|W_s(\mathrm{q_n}) - W_s(\mathrm{q_n^u})\|_F^2, \qquad (4.6-1)$$

where $\mathrm{q_n^u}$ denotes the optimal unconstrained solution. The minimization of ϕ yields the desired solution in the unconstrained case, and by a well known result in perturbation theory,

$$\sum_{i=1}^{n+m} (\sigma_i(W_s(\mathrm{q_n})) - \sigma_i(W_s(\mathrm{q_n^u})))^2 \le \phi(\mathrm{q_n}). \qquad (4.6-2)$$

Recall that the matrix is symmetric, and hence $\sigma_i(W_s(\mathrm{q_n})) = |\lambda_i(W_s(\mathrm{q_n}))|$. In particular, 4.6-2 implies that $\overline{\sigma}(W_s(\mathrm{q_n})) \le \overline{\sigma}(W_s(\mathrm{q_n^u})) + \sqrt{\phi(\mathrm{q_n})}$. The bound will not be tight in general, but it becomes tighter if the singular values have similar values. From its definition,

$$\phi(\mathrm{q_n}) = \sum_{i=0}^{n-1} (q_i - q_i^*)^2 (n-i)^2 \qquad (4.6-3)$$

and therefore the initial point may be computed by solving a QP problem. Preliminary numerical experience indicates that this initial point provides a good estimate over the optimal solution. Typically, φ evaluated at the starting point is between 20 and 40 % larger than its optimal value. The estimate is tight for a small number of variables, but becomes progressively worse as the number of variables increases.

4.7 A Textbook Example

In this section, the tradeoffs involved in the design of a controller for an unstable plant are studied. The purpose is to design a controller for the uncertain single-input single-output discrete time system $\tilde{g}$ illustrated in Figure 4.1. The nominal transfer function is $g(z) = \frac{z+.2}{z^2-.6z-1.12}$, and the multiplicative uncertainty description is given by:

$$\tilde{g} = (1 + \Delta w)g \qquad (4.7-1)$$

with $\|\Delta\| \leq 1$ and $w(z) = .3705\frac{z+.986}{z+.4682}$. Let $t_f \doteq kgw/(1+kg)$. Robust stability is equivalent to $\|t_f\|_\infty < 1$. The $\mathcal{H}_\infty$-optimal controller gives $\|t_f\|_\infty = .66$ but the response to an impulse on w has an undesirable settling time and relatively large peak values for both the control action and the output (see Fig. 4.2). The objective is to study the tradeoff between control action and robustness. The problem is not one of the simple ones considered in the previous chapter, since the "time" and the "frequency" transfer functions are different, but it may still be formulated in the same framework. Let $t_t \doteq k/(1+gk)$ so that $u = t_t w$. Then the objective is to design an internally stabilizing controller that minimizes $\|t_f\|_\infty$ in order to achieve robust stability, and rejects an impulsive disturbance on w while satisfying some time domain constraints on the control action. Introducing the parametrization of all stabilizing controllers, the problem may be written as $\min_{h\in\mathcal{H}_\infty} \|r-h\|_\infty$ $s.t.$ $\mathbf{h_n} \in \Omega$, with Ω defined so that the control action lies in the envelope illustrated in Fig. 4.4. A feasible control action can attain a peak value of pk during the first 10 samples, and then has to decay to a tolerable excursion of .03 after the 20-th sample. Fig. 4.3 shows the optimal value of $\|t_f\|_\infty$ as a function of pk, with $pk = 2.35$ corresponding to the unconstrained response. This analysis clearly demonstrates what is the minimum possible control effort required to achieve robust stability. For the preliminary design, a value of $pk = .7$ was specified, the horizon length was set to 30 and the EA was stopped when $ub - lb/ub < .05$, where ub and lb denote upper and lower bounds respectively over the optimal value. Fig. 4.4 shows the output and the control action for the optimal controller; note that the time response becomes unacceptable once the constraints are cleared. As argued in Section 3.6, doubling the horizon by itself would not help in flattening up the overall response, but replacing the optimal controller by one 5 % suboptimal (i.e., one that guarantees that the norm of the robustness transfer

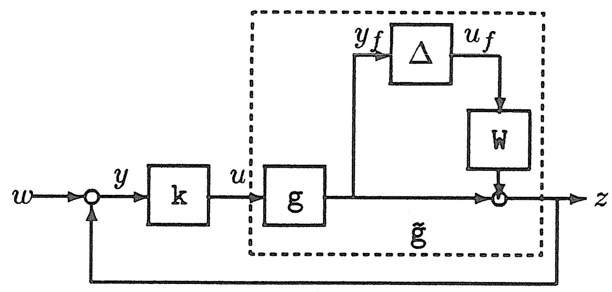

Figure 4.1: Closed-loop Configuration for the Example

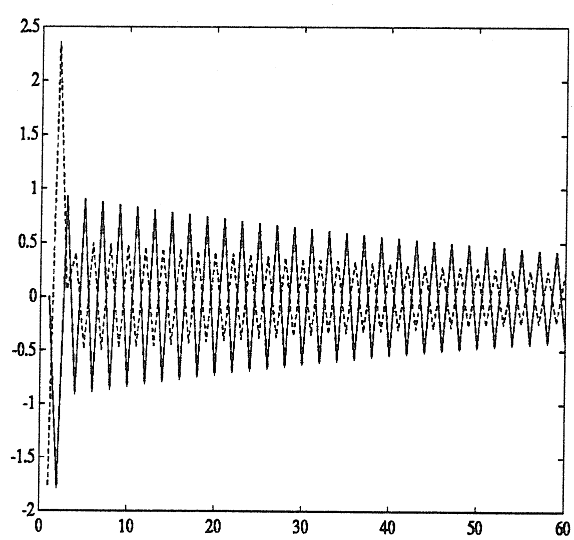

Figure 4.2: Unconstrained Behavior: Control Action (dashed), Plant Output (solid)

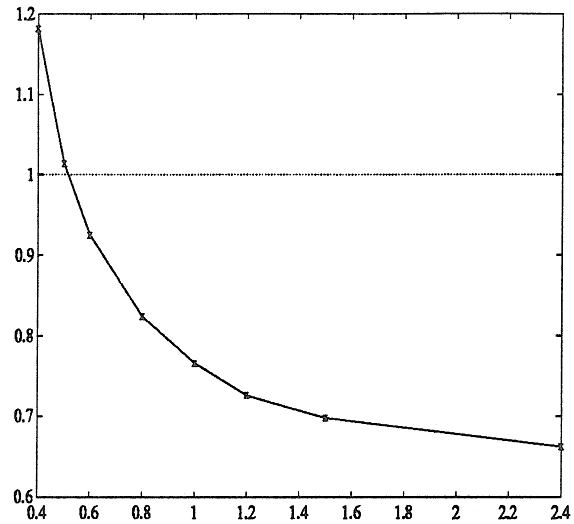

Figure 4.3: Peak of Control Effort vs. Optimal Norm

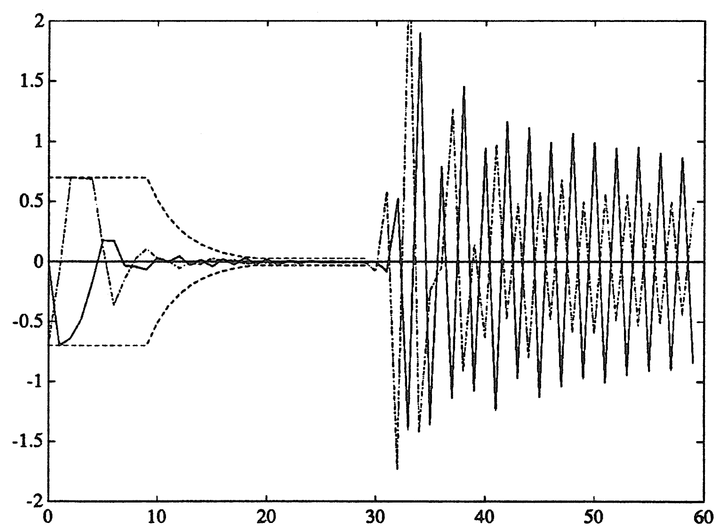

Figure 4.4: Optimal Output (full) and Control Action (dash-dotted)

function will lie within 5 % of the optimal value) improves dramatically the response after the horizon, as illustrated in Fig. 4.5. Meanwhile, the $\mathcal{H}_\infty$ norm increases from .94 to only .95. The time-domain simulations for this design are shown in Fig. 4.6. In Fig. 4.7, the absolute value of the eigenvalues of $W_s(\mathbf{q}_\mathbf{n}^{\mathbf{fin}})$ with $\mathbf{q}_\mathbf{n}^{\mathbf{fin}}$ denoting the output of the EA is plotted, showing a cluster of eigenvalues close to the maximum one. The controller has 32 states, which is undoubtedly too many for a third order plant (including the weighting), since it is well known that an unconstrained third order controller can be easily computed. However, using a simple balanced truncation scheme, it is possible to reduce the order of the controller to six. Fig. 4.8 shows the bode plot for the full and reduced order controllers; both controllers produce almost identical time domain responses and $\mathcal{H}_\infty$ norm. The poles of the reduced order controller are located at $-.0192 \pm j.8419$, $-.426 \pm j.1945$, $-.9309 \pm j.2388$, the zeros at $-.0053 \pm j1.0253$, $-.9159 \pm j.341$, $-.8677 \pm j.259$, and the gain was $-.7004$.

In order to give some idea of the computational effort involved in solving the optimization program, suppose that the first 20 samples are constrained as illustrated on Fig. 4.4, the remaining variables are chosen to verify a 5% steady-state error, a 5% tolerance stopping tolerance is selected for the EA, and the horizon length is varied from 30 to 100. The performance of the algorithm as a function of the number of variables is summarized in Fig. 4.9, where the number of iterations (in thousands), the objective function average computing time (in seconds), the initial point computing time (in minutes), and the total computing time (in minutes) are shown clockwise from top to bottom. Only function iterations are counted, since feasibility iterations amount to an average 3% of the total computing time. Note that, in particular, the computation of the initial condition is cheap compared to the total computing time. The experiment is not completely fair in the sense that after the first few, the variables are constrained to lie close to zero and hence the value they achieve on the starting point is close to their final one. Although in general

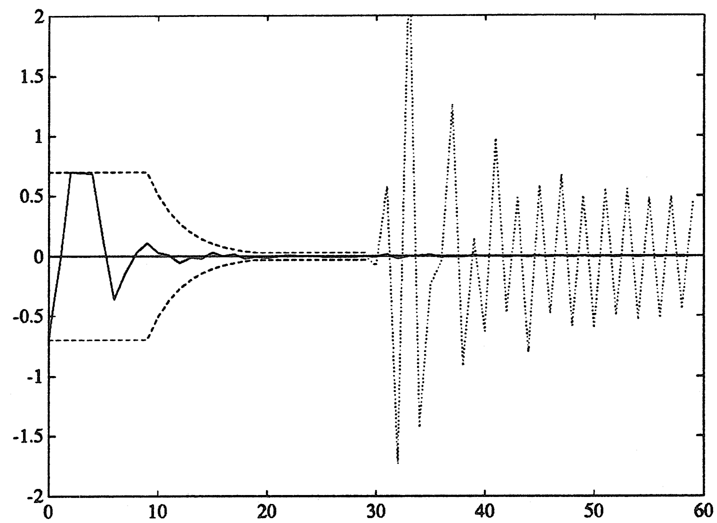

Figure 4.5: Optimal (dotted) vs Suboptimal (full) Control Action

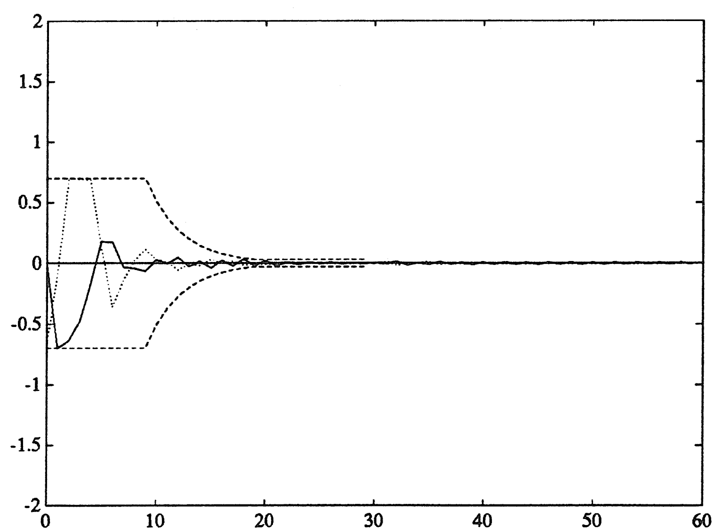

Figure 4.6: Output and Control Action for Final Design

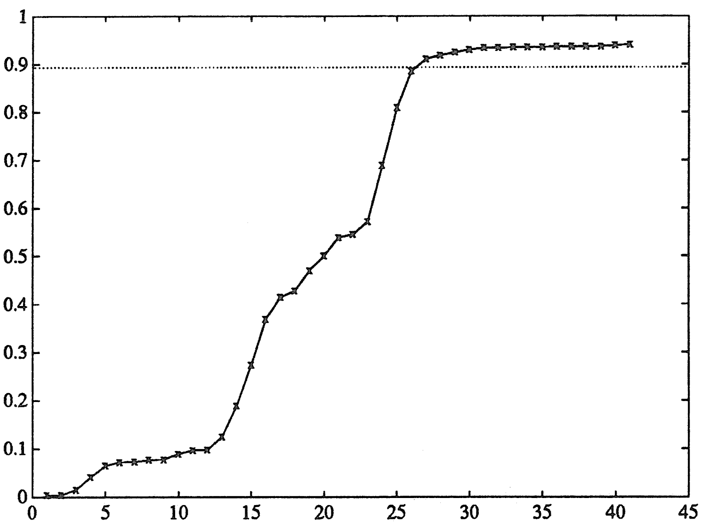

Figure 4.7: Eigenvalues of W_s

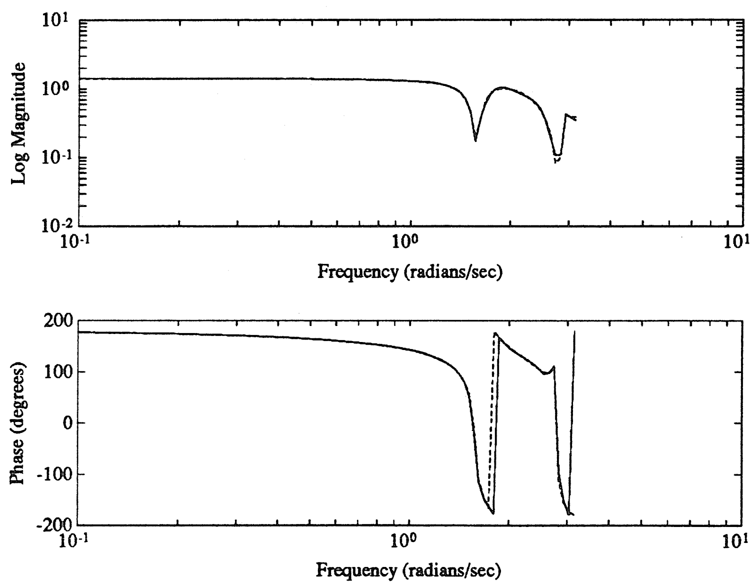

Figure 4.8: Frequency Response for Full and Reduced Order Controller

a faster growth in computing time is expected, preliminary experience indicates a satisfactory performance of the algorithm as compared to the results obtained using the approaches in [7] or [59].

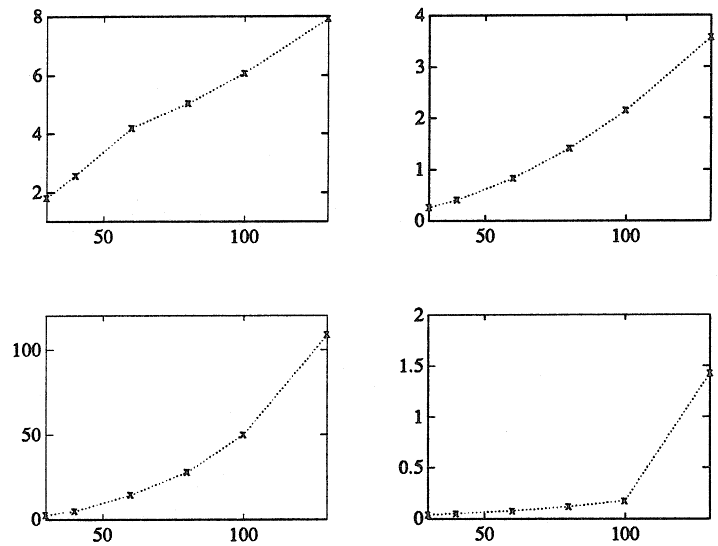

Figure 4.9: Performance of the Ellipsoid Algorithm

Chapter 5

The General Problem

In this chapter, the general Time Domain Constrained $\mathcal{H}_\infty$ Control problem is formulated and solved, by using the ideas introduced in Chapter 3, and the results on unconstrained $\mathcal{H}_\infty$-control derived in Chapter 2. After presenting a suitable framework for treating simultaneously time and frequency domain specifications, the time domain constrained one-block problem is considered. This problem is closely related to the one considered in Chapter 3, and is necessary for solving the general instance of the problem. Then, the four-block case is addressed. A result established in Chapter 2 is used to reduce the four-block to a special one-block problem, and then the preceding theory is applied. It is shown that if the unconstrained problem is solvable, then a controller solving the *constrained* $\mathcal{H}_\infty$ control problem exists if and only if the solution to an optimization problem is less than the specified norm bound. The computation of a solution and numerical aspects are also discussed.

5.1 Problem Formulation

In this section, the problem of minimizing the infinity norm of a closed-loop transfer function subject to time domain constraints over a finite horizon is formulated. Consider the basic block diagram illustrated in Fig. 5.1 Here P is obtained from the

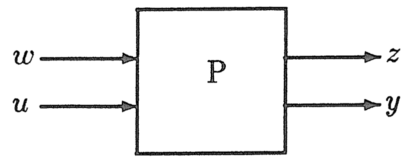

Figure 5.1: System Interconnection

nominal plant and describes the effect of the vectors of exogenous signals w and control actions u on the vector of outputs z and measurements y. In this model, inputs w and outputs z do not only include "real" signals (e.g., disturbances, commands or actual outputs), but also fictitious ones, which can be thought of as connecting unmodeled system components to the nominal model, and which can be used for robust control system design. See [7,74] for a comprehensive discussion on the role of the system interconnection matrix in modern optimal control.

Much of the work in $\mathcal{H}_\infty$ control has been concentrated on finding a controller K between y and u, which assures internal stability of the interconnection and minimizes the ∞-norm of the (weighted) transfer function between w and z. The motivation for this is that such a controller solves exactly (or approximately depending on the model uncertainty structure) the problem of robust stabilization and performance formulated in the frequency domain. In order to extend this framework to include time domain constraints, it is convenient to split the vector w of exogenous inputs into two parts: i) an exogenous input vector, denoted w_t, consisting of all input signals related to time domain constraints, and ii) an exogenous input vector, denoted w_f, consisting of all input signals related to the $\mathcal{H}_\infty$ optimization. Correspondingly, the outputs vector z is split into i) the output signal vector z_t on which one wants to impose some time domain constraints, and ii) the output signal vector z_f related to the $\mathcal{H}_\infty$ optimization. Note that it is permitted for some signals to be included both in the "time" and "frequency" signals. The purpose is then to minimize the ∞-norm of the transfer function between w_f and z_f while some time domain constraints reflected on the impulse response matrix between inputs w_t and outputs z_t are satisfied. Note that in the special case in which the length of w_t and z_t is zero, the formulation reduces to the pure $\mathcal{H}_\infty$ case.

The configuration can be represented by a transfer function P with input vector $(w_t^t \, w_f^t \, u^t)^t$ and output vector $(z_t^t \, z_f^t \, y^t)^t$, which can be partitioned as

$$P = \begin{pmatrix} P_{11} & 0 & P_{13} \\ 0 & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix}, \qquad (5.1-1)$$

so that:

$$\begin{aligned} z_t &= P_{11} w_t + P_{13} u \\ z_f &= P_{22} w_f + P_{23} u \\ y &= P_{31} w_t + P_{32} w_f + P_{33} u. \end{aligned}$$

Now suppose that the loop is closed with a controller K so that $u = Ky$. Then, solving for z_t and z_f, the closed-loop transfer functions are obtained:

$$\begin{aligned} T_{z_t w_t} &= P_{11} + P_{13} K (I - P_{33} K)^{-1} P_{31} \\ T_{z_f w_f} &= P_{22} + P_{23} K (I - P_{33} K)^{-1} P_{32}. \end{aligned}$$

Using the stabilization results in Chapter 2, the sets of internally stable closed-loop transfer functions are given by:

$$\begin{aligned} T_{z_t w_t} &= T_{11} - T_{13} Q T_{31} \\ T_{z_f w_f} &= T_{22} - T_{23} Q T_{32}, \end{aligned}$$

with $Q \in \mathcal{RH}_\infty$.

There are several reasons which make the separation between "time" and "frequency" signal convenient. First, the interconnection between frequency signals usually includes weights in order to shape the closed-loop transfer functions, which is not appropriate when specifying the time domain constraints. Remember that it is permissible for the same signal to be included both in the "time" and "frequency" signals. Secondly, it turns out that the mathematical formulation of the constrained problem becomes more clear and easier to implement when the signals are separated like this. And thirdly, as discussed in Chapter 2, the parametrization above can be chosen so that T_{23} and T_{32} are inner and co-inner transfer functions respectively. This simplifies the formulation of the $\mathcal{H}_\infty$ optimization problem and is assumed throughout.

Time Domain Specifications

Consider first the time domain specifications. Suppose that for some fixed input vector of signals w_t, one wants to constrain the j-th output of the system so that the first N samples satisfy some specified bounds. For instance, w_t can be the vector $[0\ 0\ \cdots\ w_t^i\ \cdots 0]^t$ where w_t^i is a step and one wants the j-th output to remain between some given bounds. By linearity, it is possible to assume, without loss of generality, that w_t is a signal of the form $[0\ 0\ \cdot\ w_t^i\ \cdots\ 0\ 0]^t$. This observation permits to simplify the notation, and the general case may be recovered using superposition. Then:

$$T^{ji}_{z_t w_t} \;=\; t^{ji}_{11} - t^{j\cdot}_{13} Q t^{\cdot i}_{31},$$

where lower case letters are used to stress that the functions are just scalars and $t^{j\cdot}_{13}$, $t^{\cdot i}_{31}$ denotes row and column vectors with entries t^{ji}_{13} and t^{ji}_{31} respectively. To simplify the notation, call

$$a \;\doteq\; t^{ji}_{11}$$

$$\begin{aligned} b &\doteq t_{13}^{j\cdot} \\ c &\doteq t_{31}^{\cdot j}, \end{aligned}$$

and let

$$\begin{aligned} a &= \sum_{s=0}^{\infty} a_s z^{-s} \\ b &= \sum_{s=0}^{\infty} b_s z^{-s} \\ c &= \sum_{s=0}^{\infty} c_s z^{-s} \\ Q &= \sum_{s=0}^{\infty} Q_s z^{-s}. \end{aligned}$$

Then:

$$a - bQc = \sum_{k=0}^{\infty}(a_k - \sum_{r=0}^{k}\sum_{s=0}^{r} b_{r-s}Q_s c_{k-r})z^{-k}.$$

In particular $(a - bQc)_k = a_k - \sum_{r=0}^{k}\sum_{s=0}^{r} b_{r-s}Q_s c_{k-r}$ implying that convex (linear) constraints on the input-output Markov parameters translate into convex (linear) constraints on the $Q_s's$.

Let $\mathbf{Q_n} \doteq [Q_0 \cdots Q_{n-1}]$ denote the matrix formed by the first n terms of the expansion of Q. Then, 5.1-2 implies that the time domain constraints considered in this section translate easily into constraints on the matrix $\mathbf{Q_n}$. The set where $\mathbf{Q_n}$ must lie in order to satisfy the time domain constrains is called Ω^n. See Chapter 3 for an example when all transfer functions involved are scalar.

Frequency Domain Specification

Consider now the transfer function between w_f and z_f. Then, Equation 2.4-15 gives

$$\|T_{w_f z_f}\|_\infty = \left\| \begin{matrix} G_{11} & G_{12} \\ G_{21} & G_{22} - Q^\sim \end{matrix} \right\|_\infty, \qquad (5.1-2)$$

where $G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$ is a stable transfer matrix depending on P. Recall that 5.1-2 is called a *four-block problem* [19], but if $G_{11} = G_{12} = G_{21} = 0$, then 5.1-2 is called a one-block problem. The former problem is substantially more complicated, and hence the latter is considered first. Note that the case studied in Chapter 3 corresponds to a scalar one-block problem. The frequency domain specification considered is of the form

$$\left\| \begin{matrix} G_{11} & G_{12} \\ G_{21} & G_{22} - Q^{\sim} \end{matrix} \right\|_{\infty} \leq \gamma, \qquad (5.1-3)$$

where γ is some performance level. It is well known that a number of control design specifications may be translated into a constrain like this one. In the one-block case, it is straightforward to compute the optimal value for γ; in the four-block case, this is no longer true, but necessary and sufficient conditions for the existence of a $Q \in \mathcal{RH}_\infty$ satisfying both 5.1-3 and some time domain constrains may be considered instead. Given a set of constrains Ω^n and a $\gamma > 0$, there may or may not exist a stable transfer function Q that simultaneously satisfies the frequency domain constrain and $\mathbf{Q_n} \in \Omega^n$. For instance, the problem would certainly be infeasible if γ is taken smaller than the unconstrained optimal minimum. Therefore, the time-domain constrained $\mathcal{H}_\infty$-control problem may be formulated as:

$$\textit{Find } Q(z) \in \mathcal{H}_\infty \textit{ so that 5.1-3 is satisfied and } \mathbf{Q_n} \in \Omega^n. \qquad (TDC\mathcal{H}_\infty)$$

If one is interested in the optimal case, i.e., determining the minimum γ for which Problem TDC$\mathcal{H}_\infty$ has a solution, then one can follow the usual unconstrained procedure, by iteratively adjusting γ to get as close to the optimal value as desired.

5.2 The One-Block Constrained $\mathcal{H}_\infty$-Control Problem

In this section, the solution to the time domain constrained $\mathcal{H}_\infty$ one-block control problem is studied. For notational simplicity, it is assumed that G is square with l

inputs and l outputs. The results here extend those derived for a SISO problem in 3, and are required for solving the four-block problem.

Let $G = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \in \mathcal{RH}_\infty$, and suppose that some $\mathbf{Q_n^*} = [Q_0\ Q_1\ \cdots\ Q_{n-1}]$ of appropriate dimensions is given. Consider the system defined as:

$$G(z, \mathbf{Q}_n) \doteq z^{-n}[G(z) - \sum_{i=0}^{n-1} Q_i^t z^i], \qquad (5.2-1)$$

and let $\varphi(\mathbf{Q_n}) \doteq \|G(z;\mathbf{Q_n})\|_H$. Then:

Theorem 5.1 *With the previous notation,*

a) There exists $Q(z) \in \mathcal{RH}_\infty$ such that $\mathbf{Q}_n = \mathbf{Q}_n^*$ and $\|G(z) - Q(z)^\sim\|_\infty \leq 1$ if and only if $\varphi(\mathbf{Q_n}) = \|G(z, \mathbf{Q_n^*})\|_H \leq 1$.

b) There exists an internally stabilizing controller solving Problem $TDC\mathcal{H}_\infty$, with $G = G_{22}$ and $\gamma = 1$ in 5.1-3, if and only if $\min_{\mathbf{Q_n} \in \Omega^n} \varphi(\mathbf{Q_n}) \leq 1$

Proof: It follows from the proof of Theorem 3.1.

□

Consider the realization for 5.2-1

$$G(\cdot;\mathbf{Q_n}) = \left(\begin{array}{cc|c} A & BE_1^t & 0 \\ 0 & A_f & E_n \\ \hline C & -\hat{\mathbf{Q}}_\mathbf{n} & 0 \end{array}\right) \doteq \left(\begin{array}{c|c} A_e & B_e \\ \hline C_e & D_e \end{array}\right), \qquad (5.2-2)$$

where $\hat{\mathbf{Q}}_\mathbf{n} = [Q_0^t\ Q_1^t\ \cdots\ Q_{n-1}^t]$, and let X_e and Y_e denote the controllability and observability grammians of 5.2-2. Let

$$\begin{aligned} Y_{12}^0 &= [A^tYB\ A^{2t}YB\ \cdots\ A^{nt}YB] \\ Y_{12}^l &= [C^t\ A^tC^t\ \cdots\ A^{n-1}C^t] \end{aligned}$$

$$Y_{22}^0 = \begin{bmatrix} B^tYB & B^tA^tYB & \cdots & B^tA^{(n-1)t}YB \\ B^tYAB & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ B^tYA^{n-1}B & \cdots & \cdots & B^tYB \end{bmatrix}$$

$$Y_{22}^l = \begin{bmatrix} 0 & B^tC^t & \cdots & B^tA^{(n-2)t}C^t \\ \vdots & & \ddots & \vdots \\ \vdots & & \ddots & B^tC^t \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$

$$\mathcal{Q}_n = \begin{bmatrix} Q_0^t & Q_1^t & \cdots & Q_{n-1}^t \\ 0 & Q_0^t & \cdots & Q_{n-2}^t \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & Q_0^t \end{bmatrix}.$$

Then X_e, Y_e and $\varphi(\mathbf{Q_n})$ can be computed as in the following theorem.

Theorem 5.2 *With the previous notation, the following results hold.*

a) The grammians for 5.2-2 have the form:

$$X_e = \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix}$$

$$Y_e = \begin{bmatrix} Y & (Y_{12}^0 - Y_{12}^l\mathcal{Q}_n) \\ (Y_{12}^0 - Y_{12}^l\mathcal{Q}_n)^t & Y_{22}^0 - Y_{22}^l\mathcal{Q}_n - \mathcal{Q}_n^tY_{22}^{lt} + \mathcal{Q}_n^t\mathcal{Q}_n \end{bmatrix};$$

therefore, if

$$W(\mathbf{Q_n}) = \begin{bmatrix} xYx & xY_{12}^0 \\ Y_{12}^{0t}x & Y_{22}^0 \end{bmatrix} + \begin{bmatrix} 0 & -xY_{12}^l\mathcal{Q}_n \\ -(Y_{12}^l\mathcal{Q}_n)^tx & -Y_{22}^l\mathcal{Q}_n - \mathcal{Q}_n^tY_{22}^t + \mathcal{Q}_n^t\mathcal{Q}_n \end{bmatrix} \tag{5.2-3}$$

then $\varphi(\mathbf{Q_n})^2 = \rho[W(\mathbf{Q_n})]$.

b) Let

$$W_1(\mathbf{Q_n}) = \begin{bmatrix} yA^n x & CA^{n-1}x & CA^{n-2}x & \cdots & \cdots & CAx & Cx \\ yA^{n-1}B & CA^{n-2}B & CA^{n-3}B & \cdots & \cdots & CB & -Q_0^t \\ yA^{n-2}B & CA^{n-3}B & CA^{n-4}B & \cdots & \cdots & -Q_0^t & -Q_1^t \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ yAB & CB & -Q_0^t & \cdots & \cdots & -Q_{n-3}^t & -Q_{n-2}^t \\ yB & -Q_0^t & -Q_1^t & \cdots & \cdots & -Q_{n-2}^t & -Q_{n-1}^t \end{bmatrix}. \tag{5.2-4}$$

Then $\varphi(\mathbf{Q_n}) = \overline{\sigma}(W_1(\mathbf{Q_n}))$.

Proof: Part *a)* follows by exploiting the structure in the state-space realization 5.2-2 for solving the Lyapunov equations satisfied by X_e and Y_e. Part *b)* can be obtained by factorizing $W(\mathbf{Q_n})$. The calculations are similar to the ones issued in the proof of Theorem 3.3.

□

By the previous discussion, one has first to find $\mathbf{Q_n^*}$ that satisfies $\varphi(\mathbf{Q_n^*}) \leq 1$ —this is done by solving a convex minimization problem in $\mathbf{Q_n}$— and then to find a solution to the "standard" one-block $\mathcal{H}_\infty$ minimization problem $\|G(\cdot, \mathbf{Q_n^*}) - Q_t^\sim\|_\infty \leq 1$ — this can be done, for example, by using Theorem 2.9. Then a solution to problem TDC$\mathcal{H}_\infty$ can be computed as $Q(z) = \sum_{i=0}^{n-1} Q_i z^{-i} + z^{-n} Q_t(z)$. This procedure has some disadvantages. First, if G has degree r, then $G(z; \mathbf{Q_n})$ has degree $ln + r$ and therefore one needs to manipulate large matrices in order to compute $Q_t(z)$. This makes the computation time-consuming and potentially sensitive to numerical errors. Moreover, one would expect following Theorem 3.4, that the ln poles at 0, which are fictitiously introduced to formulate the problem, should be non-minimal. Unfortunately, the proof of Theorem 3.4 does not extend to the multivariable case. Moreover, the proof does not provide a constructive way for cancelling the non-minimal modes. The following theorem gives a realization for $Q(z) = \sum_{i=0}^{n-1} Q_i z^{-i} + z^{-n} Q_t(z)$ of order $r + nl$ that circumvents these difficulties.

Theorem 5.3 *Assume that $\mathbf{Q_n} = [Q_0 \ \cdots \ Q_{n-1}]$ is such that $\rho = \|G(z; \mathbf{Q_n})\|_H \le 1$. Let* $x_e = \begin{bmatrix} x & 0 \\ 0 & I \end{bmatrix}$, $y_e = Y_e^{1/2}$, *and*

$$A_Q = A_e - B_e C_Q$$

$$B_Q = \begin{bmatrix} AXC^t - BQ_0 \\ -Q_1 \\ \vdots \\ -Q_{n-1} \\ D_Q \end{bmatrix},$$

where C_Q satisfies

$$C_Q(I - X_e Y_e) = -D_Q C_e + B_e^t Y_e A_e$$

and $D_Q \doteq -E_D^t$, where E_D is such that the matrix

$$D_a = \begin{bmatrix} yA_e x & yB_e \\ C_e x & E_D \end{bmatrix} \qquad (5.2-5)$$

is a contraction. Then

$$Q \doteq \left(\begin{array}{c|c} A_Q & B_Q \\ \hline [0 \ -E_1^t] & Q_0 \end{array} \right) \qquad (5.2-6)$$

is such that $Q \in \mathcal{RH}_\infty$, $\|G - Q^\sim\|_\infty \le 1$ and the first n terms in the expansion of Q are $Q_0, Q_1, \cdots, Q_{n-1}$.

<u>Proof:</u> The first step is to find explicitly an optimal approximation to $G(z; \mathbf{Q_n})$. Applying Theorem 2.9 to realization 5.2-2, compute

$$A_Q = A_e - B_e C_Q = \begin{bmatrix} A & B & 0 \\ 0 & 0 & I \\ & -C_Q & \end{bmatrix} \qquad (5.2-7)$$

and

$$B_Q = \begin{bmatrix} A & BE_1^t \\ 0 & A_f \end{bmatrix} \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} C^t \\ -\mathbf{Q_n}^t \end{bmatrix} - \begin{bmatrix} 0 \\ E_n \end{bmatrix} D_Q = \begin{bmatrix} AXC^t - BQ_0 \\ -Q_1 \\ \vdots \\ -Q_{n-1} \\ D_Q \end{bmatrix}. \qquad (5.2-8)$$

Let $Q_t \doteq \left(\begin{array}{c|c} A_Q & B_Q \\ \hline C_Q & D_Q \end{array} \right)$; then Theorem 2.9 implies that $\|G(z;\mathbf{Q_n}) - Q_t(z)\|_\infty \leq 1$. As stated above, a solution to the constrained one-block problem can be computed as $Q(z) = \sum_{i=0}^{n-1} Q_i z^{-i} + z^{-n} Q_t(z)$, and therefore $Q(z)$ has the state-space realization (see 1.2-6):

$$Q = \left(\begin{array}{cc|c} A_f & E_n C_Q & \begin{bmatrix} Q_1 \\ \vdots \\ Q_{n-1} \\ D_Q \end{bmatrix} \\ 0 & A_Q & B_Q \\ \hline E_1^t & 0 & Q_0 \end{array} \right). \qquad (5.2-9)$$

Applying the similarity transformation $\begin{bmatrix} I_{nl} & \begin{bmatrix} 0 & I_{nl} \end{bmatrix} \\ 0 & I_{nl+r} \end{bmatrix}$ (recall that r denotes the dimension of A) to 5.2-9

$$Q = \left(\begin{array}{cc|c} A_f & \begin{bmatrix} 0 & I_{nl} \end{bmatrix} A_Q - A_f \begin{bmatrix} 0 & I_{nl} \end{bmatrix} + E_n C_Q & 0 \\ 0 & A_Q & \begin{bmatrix} AXC^t - BQ_0 \\ -Q_1 \\ \vdots \\ -Q_{n-1} \\ D_Q \end{bmatrix} \\ \hline E_1^t & -[0 \ E_1^t] & Q_0 \end{array} \right).$$

Using 5.2-7 it is easy to see that:

$$\begin{bmatrix} 0 & I_{nl} \end{bmatrix} A_Q - A_f \begin{bmatrix} 0 & I_{nl} \end{bmatrix} + E_n C_Q$$
$$= \begin{bmatrix} 0 \ 0 \ I \\ -C_Q \end{bmatrix} - \begin{bmatrix} 0 & 0 & I \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \ 0 \ 0 \\ C_Q \end{bmatrix} = 0,$$

and thus

$$Q = \left(\begin{array}{c|c} A_Q & \begin{bmatrix} AXC^t - BQ_0 \\ -Q_1 \\ \vdots \\ -Q_{n-1} \\ D_Q \end{bmatrix} \\ \hline -[0 E_1^t] & Q_0 \end{array} \right).$$

This concludes the proof.

□

In standard $\mathcal{H}_\infty$ optimization, the multiplicity of the largest singular value associated with the Hankel norm is generically one. As discussed in Section 3.5, in the constrained case the multiplicity can be much larger. For the limiting case in which the constraints are not binding, it was shown in 3.5 that the multiplicity increases as much as the length of the horizon. A similar property can be established for the general one-block case.

Behavior of the Overall Response

The main idea behind the approach for solving Problem TDC$\mathcal{H}_\infty$, is to constrain certain closed-loop responses to specific input over a finite horizon, and hope that the "tail" of the time response, computed as a solution to an unconstrained $\mathcal{H}_\infty$ problem, will behave nicely. From the discussion in Section 3.6, it is possible to select a sub-optimal solution to enforce a good behavior after the horizon, by using the following procedure which has the effect of placing the closed-loop poles inside a disk of radius strictly smaller than one. First introduce the change of variables $z \to z/\rho$, $\rho > 1$ in the original transfer matrix. Then compute a constrained suboptimal solution for the transformed system (for instance, the central or minimal entropy solution). Finally, transform the controller back by reversing the previous change of variables. Note that the time domain constraints must be satisfied by the final controller and not the intermediate one, and hence the constraints for the optimization problem should be scaled accordingly. It is easy to see that this scheme places the closed-loop poles inside a disk of radius $1/\rho$ and that the largest singular value of the corresponding W_1 matrix provides an upper bound over the closed-loop norm. The fact that all poles have an absolute value less than $1/\rho$ for some $\rho > 1$, introduces a damping on the time domain responses that can be used to prove that if the horizon is extended long enough, then the satisfaction of the constraints over a finite horizon implies the satisfaction of the constraints for all times (see Theorem 3.7 for a formal proof and further details).

5.3 The Four-block Constrained $\mathcal{H}_\infty$-Control Problem

In this section the four-block constrained $\mathcal{H}_\infty$-control problem is considered. The main idea is first to reduce the four-block problem to the one-block case by using the procedure in Section 2.8, and then apply the results for the constrained one-block case derived in Section 5.2.

5.3.1 Problem Transformation

Consider first the frequency domain specification 5.1-3 (recall that $\gamma = 1$). Let G have a state-space realization

$$G = \begin{array}{c} \\ \\ \end{array} \begin{array}{cc} & \begin{array}{cc} m_1 & m_2 \end{array} \\ & \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & 0 \end{array} \right) \begin{array}{c} \\ p_1 \\ p_2 \end{array} \end{array} . \qquad (5.3-1)$$

Assume, as in Section 2.8, that $\max\left\{ \|G_{11}\ G_{12}\|_\infty, \left\| \begin{matrix} G_{11} \\ G_{21} \end{matrix} \right\|_\infty \right\} < 1$. Note that this assumption holds if the unconstrained optimal norm is strictly less than one; since one expects that the constrained problem will usually achieve a minimum norm strictly larger than the unconstrained one, the condition will usually hold (unless an unconstrained solution satisfies the time domain specifications). In order to reduce the problem to the one-block case, consider the discrete-time algebraic Riccati equations associated with the pencils 2.8-14 and 2.8-15 with positive definite solutions X and Y respectively. From Theorem 2.11, there exist a $Q \in \mathcal{RH}_\infty$ such that $\left\| \begin{matrix} G_{11} & G_{12} \\ G_{21} & G_{22} - Q^\sim \end{matrix} \right\|_\infty \leq 1$ if and only if $\rho(XY) \leq 1$. Let B_3, C_3, D_{13} and D_{31} be defined by 2.8-9, 2.8-11, 2.8-8 and 2.8-10:

$$\begin{aligned} D_{13} &\doteq (I - D_{11}D_{11}^t - D_{12}D_{12}^t - C_1XC_1^t)^{1/2} \\ B_3 &\doteq -(AXC_1^t + B_1D_{11}^t + B_2D_{12}^t)D_{13}^{-t} \\ D_{31} &\doteq (I - D_{11}^tD_{11} - D_{21}^tD_{21} - B_1^tYB_1)^{1/2} \\ C_3 &\doteq -D_{31}^{-t}(B_1^tYA + D_{11}^tC_1 + D_{21}^tC_2). \end{aligned}$$

By the previous assumption, these matrices are well defined. Let G_a be

$$G_a = \left(\begin{array}{c|ccc} A & B_1 & B_2 & B_3 \\ \hline C_1 & D_{11} & D_{12} & D_{13} \\ C_2 & D_{21} & 0 & 0 \\ C_3 & D_{31} & 0 & 0 \end{array}\right) \qquad (5.3-2)$$

and consider the one-block problem associated with G_a. By Theorem 2.10 and Corollary 2.8, there exists a solution to the four-block problem if and only if there exists a solution to this one-block problem of the form $\begin{bmatrix} 0 & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{bmatrix}$. Let then $Q,\ Q_{23},\ Q_{32},\ Q_{33} \in \mathcal{RH}_\infty$ be such that

$$\left\| \begin{matrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} - Q^\sim & G_{23} - Q_{32}^\sim \\ G_{31} & G_{32} - Q_{23}^\sim & G_{33} - Q_{33}^\sim \end{matrix} \right\|_\infty \leq 1. \qquad (5.3-3)$$

Note that, for z such that $|z| = 1$,

$$\begin{aligned} &\overline{\sigma}\left(\begin{bmatrix} G_{11}(z) & G_{12}(z) & G_{13}(z) \\ G_{21}(z) & G_{22}(z) - Q(z)^\sim & G_{23}(z) - Q_{32}(z)^\sim \\ G_{31}(z) & G_{32}(z) - Q_{23}(z)^\sim & G_{33}(z) - Q_{33}(z)^\sim \end{bmatrix}\right) \\ &= \overline{\sigma}\left(\begin{bmatrix} G_{11}(z) & G_{12}(z) & G_{13}(z) \\ G_{21}(z) & G_{22}(z) - \sum_{i=0}^{n-1} Q_i^t z^i - z^n \sum_{i=0}^{\infty} Q_{i+n}^t z^i & G_{23}(z) - Q_{32}(z)^\sim \\ G_{31}(z) & G_{32}(z) - \sum_{i=0}^{n-1} Q_{23,i}^t z^i - z^n \sum_{i=0}^{\infty} Q_{23,i+n}^t z^i & G_{33} - Q_{33}(z)^\sim \end{bmatrix}\right) \\ &= \overline{\sigma}\left(\begin{bmatrix} G_{11}(z) & G_{12}(z) z^{-n} & G_{13}(z) \\ G_{21}(z) & (G_{22}(z) - \sum_{i=0}^{n-1} Q_i^t z^i) z^{-n} - Q_t(z)^\sim & G_{23} - Q_{32}^\sim \\ G_{31}(z) & (G_{32}(z) - \sum_{i=0}^{n-1} Q_{23,i}^t z^i) z^{-n} - Q_{23t}(z)^\sim & G_{33}(z) - Q_{33}(z)^\sim \end{bmatrix}\right). \end{aligned}$$

Here $Q_t(z) \doteq \sum_{i=0}^{\infty} Q_{i+n} z^{-i}$, $Q_{23t}(z) \doteq \sum_{i=0}^{\infty} Q_{23,i+n} z^{-i}$, and the last expression is obtained from the previous one by multiplying on the left by the unitary matrix

$\begin{bmatrix} I & 0 & 0 \\ 0 & z^{-n}I & 0 \\ 0 & 0 & I \end{bmatrix}$. Since equality holds for every z such that $|z| = 1$,

$$\left\| \begin{matrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} - Q^{\sim} & G_{23} - Q_{32}^{\sim} \\ G_{31} & G_{32} - Q_{23}^{\sim} & G_{33} - Q_{33}^{\sim} \end{matrix} \right\|_{\infty} = \left\| \begin{matrix} G_{11} & G_{12}^{n} & G_{13}(z) \\ G_{21} & G_{22}(\cdot; \mathbf{Q_n}) - Q_t^{\sim} & G_{23} - Q_{32}^{\sim} \\ G_{31} & G_{32}(\cdot; \mathbf{Q_{23n}}) - Q_{23t}^{\sim} & G_{33} - Q_{33}^{\sim} \end{matrix} \right\|_{\infty},$$

where $G_{12}^{n} \doteq G_{12}(z)z^{-n}$, $G_{22}(\cdot; \mathbf{Q_n}) \doteq [G_{22}(z) - \sum_{i=0}^{n-1} Q_i^t z^i]z^{-n}$ and $G_{32}(\cdot; \mathbf{Q_{23n}}) \doteq [G_{32}(z) - \sum_{i=0}^{n-1} Q_{23,i}^t z^i]z^{-n}$. These calculations and Theorem 5.1, establish the following result.

Theorem 5.4 *Suppose that Q_i, $i = 1, 2, \cdots, n-1$ are given. Then there exist Q_i, $i = n+1, n+2, \cdots$ such that $Q(z) \doteq \sum_{i=0}^{\infty} Q_i z^{-i}$ satisfies*

$$\left\| \begin{matrix} G_{11} & G_{12} \\ G_{21} & G_{22} - Q(z)^{\sim} \end{matrix} \right\|_{\infty} \leq 1,$$

if and only if there exist $\mathbf{Q_{23n}} = [Q_{23,0}, \cdots Q_{23,n-1}]$ such that $\varphi(\mathbf{Q_n}; \mathbf{Q_{23n}}) \leq 1$, where

$$\varphi(\mathbf{Q_n}; \mathbf{Q_{23n}}) = \|G_a(\mathbf{Q_n}; \mathbf{Q_{23n}})\|_H \tag{5.3-4}$$

$$= \left\| \begin{matrix} G_{11} & G_{12}^{n} & G_{13} \\ G_{21} & G_{22}(\cdot; \mathbf{Q_n}) & G_{23} \\ G_{31} & G_{32}(\cdot; \mathbf{Q_{23n}}) & G_{33} \end{matrix} \right\|_{H}. \tag{5.3-5}$$

□

5.3.2 State-Space Computation of the Objective Function

In this section, an expression for $\varphi(\cdot)$ in terms of the state-space realization of G_a and the matrices Q_i, $Q_{23,i}$ is given, following the formulas derived for the one-block case. Although the number of variables in the problem has been enlarged by the inclusion of the $Q_{23,i}$'s, it is possible to show that condition 5.3-3 implies that these variables are uniquely determined by the Q_i's.

Consider the realization for $G_a(\cdot;\mathbf{Q_n},\mathbf{Q_{23n}})$

$$G_a(\cdot;\mathbf{Q_n},\mathbf{Q_{23n}}) = \left(\begin{array}{cc|ccc} A & B_2E_1^t & B_1 & 0 & B_3 \\ 0 & A_f & 0 & E_n & 0 \\ \hline C_1 & D_{12}E_1^t & D_{11} & 0 & D_{13} \\ C_2 & -\hat{\mathbf{Q}}_{\mathbf{n}} & D_{21} & 0 & 0 \\ C_3 & -\hat{\mathbf{Q}}_{\mathbf{23n}} & D_{31} & 0 & 0 \end{array}\right) = \left(\begin{array}{c|c} A_e & B_e \\ \hline C_e & D_e \end{array}\right), \quad (5.3-6)$$

where

$$\begin{aligned} \hat{\mathbf{Q}}_{\mathbf{n}} &\doteq [Q_0^t\ Q_1^t\ \cdots\ Q_{n-1}^t] \\ \hat{\mathbf{Q}}_{\mathbf{23n}} &\doteq [Q_{23,0}^t\ Q_{23,1}^t\ \cdots\ Q_{23,n-1}^t]. \end{aligned}$$

The realization for G_a in 5.3-6 is similar to the one in 5.2-2; in fact, by following the same procedure as in the proof of Theorem 5.2, explicit formulas for the grammians of 5.3-6 can be found. Let X_e and Y_e denote the controllability and observability grammians of G_a, and let

$$\begin{aligned} Y_{12}^0 &= [(A^tYB_2 + C_1^tD_{12})\ A^t(A^tYB_2 + C_1^tD_{12})\ \cdots\ A^{(n-1)t}(A^tYB_2 + C_1^tD_{12})] \\ Y_{12}^l &= [C_2^t\ A^tC_2^t\ \cdots A^{(n-1)t}C_2^t\ C_3^t\ A^tC_3^t\ \cdots A^{(n-1)t}C_3^t] \\ Y_{22}^l &= \begin{bmatrix} 0 & B_2^tC_2^t & B_2^tA^tC_2^t & \cdots & B_2^tA^{(n-2)t}C_2^t & B_2^tC_3^t & B_2^tA^tC_3^t & \cdots & B_2^tA^{(n-2)t}C_3^t \\ 0 & 0 & B_2^tC_2^t & \cdots & B_2^tA^{(n-3)t}C_2^t & 0 & B_2^tC_3^t & \cdots & B_2^tA^{(n-3)t}C_3^t \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & B_2^tC_2^t & 0 & 0 & \cdots & B_2^tC_3^t \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \end{aligned}$$

$$Y_{22}^0 =$$

$$\begin{bmatrix} B_2^tYB_2 + D_{12}^tD_{12} & B_2^t(A^tYB_2 + C_1^tD_{12}) & \cdots B_2^tA^{(n-2)t}(A^tYB_2 + C_1^tD_{12}) \\ (B_2^tYA + D_{12}^tC_1)B_2 & B_2^tYB_2 + D_{12}^tD_{12} & \cdots B_2^tA^{(n-3)t}(A^tYB_2 + C_1^tD_{12}) \\ \vdots & \vdots & \ddots \quad \vdots \\ (B_2^tYA + D_{12}^tC_1)A^{(n-3)}B_2 & (B_2^tYA + D_{12}^tC_1)A^{(n-4)}B_2 \cdots & B_2^t(A^tYB_2 + C_1^tD_{12}) \\ (B_2^tYA + D_{12}^tC_1)A^{(n-2)}B_2 & (B_2^tYA + D_{12}^tC_1)A^{(n-3)}B_2 \cdots & B_2^tYB_2 + D_{12}^tD_{12} \end{bmatrix} \quad (5.3-7)$$

and

$$\mathcal{Q}_n = \begin{bmatrix} Q_0^t & Q_1^t & \cdots & Q_{n-1}^t \\ 0 & Q_0^t & \cdots & Q_{n-2}^t \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & Q_0^t \\ Q_{23,0}^t & Q_{23,1}^t & \cdots & Q_{23,n-1}^t \\ 0 & Q_{23,0}^t & \cdots & Q_{23,n-2}^t \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Q_{23,0}^t \end{bmatrix}. \quad (5.3\text{-}8)$$

Lemma 5.1 *For G_a in 5.3-6,*

$$X_e = \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix}$$

$$Y_e = \begin{bmatrix} Y & (Y_{12}^0 - Y_{12}^l\mathcal{Q}_n) \\ (Y_{12}^0 - Y_{12}^l\mathcal{Q}_n)^t & Y_{22}^0 - Y_{22}^l\mathcal{Q}_n - \mathcal{Q}_n^tY_{22}^{lt} + \mathcal{Q}_n^t\mathcal{Q}_n \end{bmatrix},$$

where X and Y are the stabilizing solutions for the pencils 2.8-14 and 2.8-15 respectively. Let $W_0(\mathbf{Q_n}, \mathbf{Q_{23n}})$ be defined by

$$W_0(\mathbf{Q_n}, \mathbf{Q_{23n}}) = \begin{bmatrix} xYx & x(Y_{12}^0 - Y_{12}^l\mathcal{Q}_n) \\ (Y_{12}^0 - Y_{12}^l\mathcal{Q}_n)^tx & Y_{22}^0 - Y_{22}^l\mathcal{Q}_n - \mathcal{Q}_n^tY_{22}^{lt} + \mathcal{Q}_n^t\mathcal{Q}_n \end{bmatrix}; \quad (5.3-9)$$

then $\varphi(\cdot,\cdot) = \rho[W_0(\cdot,\cdot)]$.

Proof: A procedure for finding the formulas for the grammians was discussed above. The expression for φ holds because $\varphi(\cdot,\cdot)$ denotes the square of the Hankel norm of $G_a(\cdot;\cdot,\cdot)$.

□

Lemma 5.2 *Let W_1 be defined as:*

$$W_1(\mathbf{Q_n},\mathbf{Q_{23n}}) = \begin{bmatrix} yA^n x & yA^{n-1}B_2 & yA^{n-2}B_2 & \cdots & yAB_2 & yB_2 \\ C_1A^{n-1}x & C_1A^{n-2}B_2 & C_1A^{n-3}B_2 & \cdots & C_1B_2 & D_{12} \\ C_1A^{n-2}x & C_1A^{n-3}B_2 & C_1A^{n-4}B_2 & \cdots & D_{12} & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ C_1Ax & C_1B_2 & D_{12} & 0 & \cdots & 0 \\ C_1x & D_{12} & 0 & \cdots & \cdots & 0 \\ C_2A^{n-1}x & C_2A^{n-2}B_2 & C_2A^{n-3}B_2 & \cdots & C_2B_2 & -Q_0^t \\ C_2A^{n-2}x & C_2A^{n-3}B_2 & C_2A^{n-4}B_2 & \cdots & -Q_0^t & -Q_1^t \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ C_2Ax & C_2B_2 & -Q_0^t & \cdots & \cdots & -Q_{n-2}^t \\ C_2x & -Q_0^t & -Q_1^t & \cdots & \cdots & -Q_{n-1}^t \\ C_3A^{n-1}x & C_3A^{n-2}B_2 & C_3A^{n-3}B_2 & \cdots & C_3B_2 & -Q_{23,0}^t \\ C_3A^{n-2}x & C_3A^{n-3}B_2 & C_3A^{n-4}B_2 & \cdots & -Q_{23,0}^t & -Q_{23,1}^t \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ C_3Ax & C_3B_2 & -Q_{23,0}^t & \cdots & \cdots & -Q_{23,n-2}^t \\ C_3x & -Q_{23,0}^t & -Q_{23,1}^t & \cdots & \cdots & -Q_{23,n-1}^t \end{bmatrix}.$$

Then

$$\varphi(\mathbf{Q_n},\mathbf{Q_{23n}}) = \overline{\sigma}[W_1(\mathbf{Q_n},\mathbf{Q_{23n}})]^2 \qquad (5.3-10)$$

Proof: Using the equations and remarks in Theorem 2.10 and after some algebra, one gets $W = W_1^t W_1$ and therefore the proof follows.

□

The expressions for the objective function obtained so far, depend on the original variables of the problem and on the $Q_{23,i}$'s that were introduced in the transformation of the problem discussed in the previous section. However, the structure of the problem imposes constraints on the value that the $Q_{23,i}$'s can take, once the Q_i's are fixed. To see this, recall from the remark preceding Theorem 2.10 that $\begin{bmatrix} G_{11}^{\sim} & G_{21}^{\sim} & G_{31}^{\sim} \end{bmatrix} \begin{bmatrix} G_{11} \\ G_{21} \\ G_{31} \end{bmatrix} = I$. Therefore, condition 5.3-3 implies that:

$$\begin{bmatrix} G_{11}^{\sim} & G_{21}^{\sim} & G_{31}^{\sim} \end{bmatrix} \begin{bmatrix} G_{12} \\ G_{22} - Q^{\sim} \\ G_{32} - Q_{23}^{\sim} \end{bmatrix} = 0 \qquad (5.3-11)$$

and hence $Q_{23}^{\sim} = (G_{31}^{\sim})^{-1}[G_{11}^{\sim}G_{12} + G_{21}^{\sim}(G_{22} - Q^{\sim})] + G_{32}$. More specifically, the following result holds.

Lemma 5.3 *The auxiliary variables Q_{23}^{i}, $i = 0, \cdots, n-1$ satisfy the recursion*

$$\begin{aligned} Q_{23,0}^{t} &= D_{31}^{-t}(B_1^t Y B_2 + D_{11}^t D_{12} - D_{21}^t Q_0^t) \\ Q_{23,i}^{t} &= D_{31}^{-t}(B_1^t A^{it} Y B_2 + B_1^t A^{(i-1)t} C_1^t D_{12} - D_{21}^t Q_i^t \\ &\quad - \sum_{j=0}^{i-1} B_1^t A^{(i-1-j)t}(C_2^t Q_j^t + C_3^t Q_{23,j}^t)) \end{aligned}$$

Explicitly,

$$Q_{23,i} = [B_2^t Y A_{13}^i B_1 + D_{12}^t C_{13} A_{13}^i B_1 - Q_0 D_{21} - \sum_{j=0}^{i-1} Q_j C_{23} A_{13}^{i-1-j} B_1] D_{31}^{-1}$$

where $A_{13} \doteq (A - B_1 D_{31}^{-1} C_3)$, $C_{13} \doteq (C_1 - D_{11} D_{31}^{-1} C_3)$ and $C_{23} = (C_2 - D_{21} D_{31}^{-1} C_3)$.

Proof: See Appendix.

□

Remark: By the remark preceding Theorem 2.10, A_{13} is stable and therefore the recursion is stable.

Using the result in the Lemma, it is straightforward to write $\mathbf{Q_{23n}}$ as a linear combination of the Q_i's, and therefore to rewrite matrices W and W_1 as functions of $\mathbf{Q_n}$ only. The following theorem gives an alternative objective function, that depends only on the original variables of the problem.

Theorem 5.5 *Let*

$$W_2(\mathbf{Q_n}) = \left[\begin{array}{c|cccc|cccccc}
yA^n x & yA^{n-1}B_1 & \cdots & yAB_1 & yB_1 & yA^{n-1}B_2 & yA^{n-2}B_2 & \cdots & yAB_2 & yB_2 \\
\hline
C_1A^{n-1}x & C_1A^{n-2}B_1 & \cdots & C_1B_1 & D_{11} & C_1A^{n-2}B_2 & C_1A^{n-3}B_2 & \cdots & C_1B_2 & D_{12} \\
C_1A^{n-2}x & C_1A^{n-3}B_1 & \cdots & D_{11} & 0 & C_1A^{n-3}B_2 & C_1A^{n-4}B_2 & \cdots & D_{12} & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
C_1x & D_{11} & 0 & \cdots & 0 & D_{12} & 0 & 0 \quad \cdots & & 0 \\
\hline
C_2A^{n-1}x & C_2A^{n-2}B_1 & \cdots & C_2B_1 & D_{21} & C_2A^{n-2}B_2 & C_2A^{n-3}B_2 & \cdots & C_2B_2 & -Q_0^t \\
C_2A^{n-2}x & C_2A^{n-3}B_1 & \cdots & D_{21} & 0 & C_2A^{n-3}B_2 & C_2A^{n-4}B_2 & \cdots & -Q_0^t & -Q_1^t \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
C_2x & D_{21} & 0 & \cdots & 0 & -Q_0^t & -Q_1^t & -Q_2^t & \cdots & -Q_{(n-1)}^t
\end{array}\right]. \tag{5.3 – 12}$$

Then $\rho(W_0) \leq 1$ *if and only if* $\overline{\sigma}(W_2) \leq 1$.

Proof: See Appendix.

□

With a slight abuse of notation, let $W_0(\mathbf{Q_n})$ and $W_1(\mathbf{Q_n})$ denote $W_0(\mathbf{Q_n}, \mathbf{Q_{23n}}(\mathbf{Q_n}))$ and $W_1(\mathbf{Q_n}, \mathbf{Q_{23n}}(\mathbf{Q_n}))$ respectively. Theorem 5.6 summarizes the results as follows.

Theorem 5.6 *Consider the problems*

$$\begin{aligned}
\mu_0 &= \min_{\mathbf{Q_n} \in \Omega^n} \rho[W_0(\mathbf{Q_n})] \\
\mu_1 &= \min_{\mathbf{Q_n} \in \Omega^n} \overline{\sigma}[W_1(\mathbf{Q_n})]
\end{aligned}$$

$$\mu_2 = \min_{\mathbf{Q_n} \in \Omega^n} \overline{\sigma}[W_2(\mathbf{Q_n})],$$

and assume that Ω^n is a convex set. Then:

a) There exist a solution to Problem TDC$\mathcal{H}_\infty$ if and only if either one of the μ_i's is less than or equal to one.

b) The minimization problems defined by the μ_i's and Ω^N are convex.

5.3.3 State-Space Formula for a Solution

Theorem 5.6 shows how to check whether the four-block constrained $\mathcal{H}_\infty$-control problem has a solution (i.e., μ_i has to be less than or equal to one), and how to calculate the first n terms $Q_0,\ Q_1, \cdots, Q_{n-1}$ of the expansion of such a solution. With this Q_i's, and the corresponding $Q_{23,i}$'s computed using Lemma 5.3, Theorem 2.10 can be applied to compute a solution Q_{at} that solves the one-block problem $\|G_a(\cdot;\mathbf{Q_n},\mathbf{Q_{23n}}) - Q_{at}^{\sim}\|_\infty \leq 1$. Taking the compression $Q_t = \begin{bmatrix} 0 & I & 0 \end{bmatrix} Q_{at} \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix}$, gives $Q(z) = \sum_{i=0}^{n-1} Q_i z^{-i} + z^{-n} Q_t(z)$ solves the four-block TDC$\mathcal{H}_\infty$ problem. Theorem 5.3 can be extended to obtain a realization for $Q(z)$ without the fictitious poles at zero introduced in the formulation of the problem. To see this, consider again the realization 5.3-6 for G_a, and suppose that a solution to the one-block problem is computed by using Theorem 2.9. From Theorem 2.10, and in order to get a solution to the four-block problem, take

$$E_D = \begin{bmatrix} D_{11} & 0 & D_{13} \\ D_{21} & E_{22} & E_{23} \\ D_{31} & E_{32} & E_{33} \end{bmatrix}$$

for some $E_{ij},\ i,j =\ 2,\ 3$ that make

$$D_a = \begin{bmatrix} y_e A_e x_e & y_e \begin{bmatrix} B_1 \\ 0 \end{bmatrix} & y_e \begin{bmatrix} 0 \\ E_n \end{bmatrix} & y_e \begin{bmatrix} B_3 \\ 0 \end{bmatrix} \\ \begin{bmatrix} C_1 X & D_{12} E_1^t \end{bmatrix} & D_{11} & 0 & D_{13} \\ \begin{bmatrix} C_2 X & -\hat{\mathbf{Q}}_{\mathbf{n}} \end{bmatrix} & D_{21} & E_{22} & E_{23} \\ \begin{bmatrix} C_3 X & -\hat{\mathbf{Q}}_{\mathbf{23n}} \end{bmatrix} & D_{31} & E_{32} & E_{33} \end{bmatrix}$$

a contraction. Here $x_e = \begin{bmatrix} x & 0 \\ 0 & I \end{bmatrix}$ and $y_e = Y_e^{1/2}$. Let Q_{at} be such that $\|G_a - Q_{at}^{\sim}\|_\infty \leq 1$ and let

$$Q_{at} = \left(\begin{array}{c|c} A_Q & B_Q \\ \hline C_Q & D_Q \end{array} \right) = \left(\begin{array}{c|ccc} A_Q & B_{Q_1} & B_{Q_2} & B_{Q_3} \\ \hline C_{Q_1} & 0 & 0 & 0 \\ C_{Q_2} & 0 & -E_{22}^t & -E_{32}^t \\ C_{Q_3} & 0 & -E_{23}^t & -E_{33}^t \end{array} \right), \tag{5.3 - 13}$$

with A_Q, B_Q and C_Q constructed as in Theorem 2.10. Using the definitions of C_3, B_3, D_{13} and D_{31}, and Equation 5.6-1 in the Appendix, it is straightforward to verify that $C_{Q_1} = 0$ and $B_{Q_1} = 0$. Moreover, from 2.6-27

$$\begin{bmatrix} C_{Q_2} \\ C_{Q_3} \end{bmatrix} (I - X_e Y_e) = \begin{bmatrix} C_{Q_{21}} & C_{Q_{22}} \\ C_{Q_{31}} & C_{Q_{32}} \end{bmatrix} (I - X_e Y_e) = \tag{5.3-14}$$

$$- \begin{bmatrix} 0 & E_{22}^t & E_{32}^t \\ D_{13}^t & E_{23}^t & E_{33}^t \end{bmatrix} \begin{bmatrix} C_1 & D_{12} E_1^t \\ C_2 & -\hat{\mathbf{Q}}_{\mathbf{n}} \\ C_3 & -\hat{\mathbf{Q}}_{\mathbf{23n}} \end{bmatrix} + \begin{bmatrix} 0 & E_n^t \\ B_3^t & 0 \end{bmatrix} Y_e \begin{bmatrix} A & B_2 E_1^t \\ 0 & A_f \end{bmatrix},$$

and from 2.6-23

$$A_Q = \begin{bmatrix} A - B_3 C_{Q_{31}} & B_2 E_1^t - B_3 C_{Q_{32}} \\ 0 & [0 \ \ I] \\ -C_{Q_{21}} & -C_{Q_{22}} \end{bmatrix} \tag{5.3 - 15}$$

and

$$B_{Q_2} = \begin{bmatrix} AX & B_2E_1^t \\ o & A_f \end{bmatrix} \begin{bmatrix} C_2^t \\ -\hat{\mathbf{Q}}_{\mathbf{n}}^t \end{bmatrix} + \begin{bmatrix} B_1 & 0 & B_3 \\ 0 & E_n & 0 \end{bmatrix} \begin{bmatrix} D_{21}^t \\ E_{22}^t \\ E_{23}^t \end{bmatrix}$$

$$= \begin{bmatrix} AXC_2^t - B_2^tQ_0 + B_1D_{21}^t + B_3E_{23}^t \\ -Q_1 \\ -Q_2 \\ \vdots \\ -Q_{n-1} \\ E_{22}^t \end{bmatrix}. \quad (5.3\text{-}16)$$

Taking $Q_t = \begin{bmatrix} 0 & I & 0 \end{bmatrix} Q_{at} \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix}$ gives

$$Q_t = \left[\begin{array}{c|c} A_Q & B_{Q_2} \\ \hline [C_{Q_{21}} \quad C_{Q_{22}}] & -E_{22}^t \end{array}\right], \quad (5.3-17)$$

and therefore if $Q(z) = \sum_{i=0}^{n-1} Q_i z^{-i} + z^{-n}Q_t(z)$, $Q(z)$ has a state-space representation:

$$Q = \left(\begin{array}{c|c} \begin{bmatrix} A_f & E_n[C_{Q_{21}} \quad C_{Q_{22}}] \\ 0 & A_Q \end{bmatrix} & \begin{array}{c} \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q^{n-1} \\ -E_{22}^t \end{bmatrix} \\ B_{Q_2} \end{array} \\ \hline \begin{bmatrix} E_1^t & 0 \end{bmatrix} & Q_0 \end{array}\right). \quad (5.3-18)$$

Taking the similarity transformation $\begin{bmatrix} I & \begin{bmatrix} 0 & I \end{bmatrix} \\ 0 & I \end{bmatrix}$ as in the proof of Theorem 5.3, establishes the following result.

Theorem 5.7 *Suppose Q_i, $i = 0,\ 1,\ \cdots, n-1$ are such that $\overline{\sigma}[W_2(\mathbf{Q_n})] \leq 1$. Let*

$$A_Q = \begin{bmatrix} A - B_3 C_{Q_{31}} & B_2 E_1^t - B_3 C_{Q_{32}} \\ -E_n C_{Q_{21}} & A_f - E_n C_{Q_{22}} \end{bmatrix} \qquad (5.3-19)$$

and

$$B_{Q_2} = \begin{bmatrix} (A - B_3 D_{13}^{-1} C_1) X C_2^t - (B_2 + B_3 D_{13}^{-1} D_{12}) Q_0 + (B_1 + B_3 D_{13}^{-1} D_{11} D_{21}^t) \\ -Q_1 \\ -Q_2 \\ \vdots \\ -Q_{n-1} \\ E_{22}^t \end{bmatrix}. \qquad (5.3-20)$$

Then,

$$Q = \left[\begin{array}{c|c} A_Q & B_{Q_2} \\ \hline -\begin{bmatrix} 0 & E_1^t \end{bmatrix} & Q_0 \end{array}\right] \qquad (5.3-21)$$

is such that

$$\left\| \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} - Q^{\sim} \end{bmatrix} \right\|_\infty \leq 1.$$

Moreover, $Q \in \mathcal{RH}_\infty$ and $Q(z) = \sum_{i=0}^{\infty} Q_i z^{-i} + z^{-n} Q_t(z)$, where $Q_t \in \mathcal{RH}_\infty$.

A bound on the order of a controller that solves the problem is not hard to derive. For simplicity, assume that the frequency domain condition 5.1-3 holds with strict inequality.

Theorem 5.8 *Let P be an m-th order plant, and suppose that there exist an internally stabilizing controller solving Problem TDC$\mathcal{H}_\infty$ over a horizon of length n, under the additional conditions cited above. Then there exist an internally stabilizing controller of order at most $ln + m$ that solves the problem, with $l = \min(m_2, p_2)$.*

Proof: Let K_{tdc} be a controller solving Problem TDC$\mathcal{H}_\infty$. From Theorem 4.1 in [39], all controllers that satisfy the $\mathcal{H}_\infty$ norm bound (but not necessarily the time domain constraints) can be parametrized as $K = \mathcal{F}_l(K_o, Q)$, where $Q \in \mathcal{RH}_\infty$, $\|Q\|_\infty < 1$, and K_o has a state-space representation:

$$K_o = \left(\begin{array}{c|cc} A_K & B_{K1} & B_{K2} \\ \hline C_{K1} & D_{K11} & D_{K12} \\ C_{K2} & D_{K21} & 0 \end{array}\right), \qquad (5.3-22)$$

where $A_K \in R^{m \times m}$. In particular, there exist $Q_{tdc} \in \mathcal{RH}_\infty$, $\|Q_{tdc}\|_\infty < 1$ such that $K_{tdc} = \mathcal{F}_l(K_o, Q_{tdc})$. Assume that $l = m_2 \leq p_2$. Let $Q^{\sim}_{tdc}(z) = \sum_{i=0}^{\infty} Q_i^t z^i$, and consider the transfer matrix $Q^n_{tdc} = z^{-n} \sum_{i=0}^{n-1} Q_i^t z^i$. Then, Q^n_{tdc} is stable and has a realization:

$$Q^n_{tdc} = \left(\begin{array}{c|c} A_f & E_n \\ \hline \hat{\mathbf{Q}}_{\mathbf{n}} & 0 \end{array}\right)$$

with $\hat{\mathbf{Q}}_{\mathbf{n}} = [Q_0 \ Q_1 \ \cdots Q_{n-1}]$. By construction, there exist $Q_t \in \mathcal{RH}_\infty$ such that $\|Q^n_{tdc} - Q^{\sim}_t\|_\infty < 1$. Moreover, from Theorem 5.3 there exist a $\hat{Q}_{tdc}$, $\|\hat{Q}_{tdc}\|_\infty < 1$, such that $\hat{Q}_{tdc} = \sum_{i=0}^{n-1} Q_i z^{-i} + z^{-n} Q_t(z)$ and $\hat{Q}_{tdc}$ has a realization with ln states. Taking $\hat{K}_{tdc} = \mathcal{F}_l(K_o, \hat{Q}_{tdc})$ then establishes the result.

□

5.4 Numerical Computation

From the approach to the time domain constrained $\mathcal{H}_\infty$-control problem presented in the previous section, a solution can be computed solving essentially the same convex minimization problem as in the SISO case, except that the number of variables and the cost of computing the objective function are [62]. This was extensively discussed in Chapter 4. Theorem 5.6 shows that in the four-block case, it is possible to choose between minimizing the largest eigenvalue of W_0 or the largest singular value of W_1 or W_2. In order to compare these alternatives, the following estimate of the computational cost is useful.

Lemma 5.4 *If $A \in R^{n \times n}$, the cost associated with an eigenvalue Lanczos iteration is $(2k+8)n$ flops, were k denotes the average number of non-zero elements per row. If $A \in R^{m \times n}$, the cost associated with a singular value Lanczos iteration is $4n_a^2 + 5(m+n)$, where n_a denotes the square root of the number of non-zero elements in A.*

Proof: See [31].

□

A comparison of the efforts involved in the computation of the largest singular value for W_1 and W_2 shows that, if the sparsity pattern is taken into account, then the flop count for both cases is approximately the same. Since the latter does not require the computation of the auxiliary variables $\mathbf{Q_{23n}}$, a task that requires $o(n^3)$ flops, it is preferable over the former. For W_2 in 5.3-12, and with the notation of the lemma,

$$n_a^2 = (p_1 m_1 + p_1 m_2 + p_2 m_1) n^2/2 + p_2 m_2 n^2 + rn(m_1 + m_2 + p_1 + p_2).$$

Hence the Lanczos iteration flop count for the computation of the largest singular value of W_2 is approximately equal to

$$fc_{sv} = 2n^2(p_1 m_1 + p_1 m_2 + p_2 m_1 + 2p_2 m_2) + (5+r)n(m_1 + m_2 + p_1 + p_2).$$

On the other hand, the Lanczos iteration flop count for the computation of the spectral radius of W is approximately equal to

$$fc_{ev} = 2(r + p_2 n)^2 + 8(r + p_2 n).$$

Based solely on this flop count, the largest eigenvalue of W is to be preferred over the largest singular value of W_2. However, two additional facts should be considered before discarding the second objective function. First, note that the computation of W requires $o(n^3)$ flops (note that using Lemma 5.3 it is possible to write W as a quadratic function of the Q_i's alone) and the evaluation of a gradient requires

$o(n^2)$ flops. On the contrary, W_2 is available and requires no computation, while the evaluation of the gradient is linear in n. Second, the storage requirement for W_2 is approximately $rn(m_1 + m_2 + p_1 + p_2) + n(p_1 m_1 + p_1 m_2 + p_2 m_1 + 2p_2 m_2)$, while that for W is $rnm_2 + n^2 m_2^2/2$. This is because the symmetric matrix W needs to be stored but the product of W_2 or its transpose and a vector can be computed if $yA^n x$, $yA^i B_k$, $C_j A^i x$, $C_j A^i B_k$, Q_i are stored, in the memory, with $i = 0, \cdots, n-1$, $j = 1, 2$, $k = 1, 2$. This suffices for computing the largest singular value using vector Lanczos iterations, that were found to be the convenient choice in [62]. From these observations, the most efficient objective function will also depend on the implementation and on the actual system in which the computations are performed.

Given that the aim is to reduce the value of the objective function below one, it is not necessary (or even desirable) to wait until the optimization algorithm has converged to a minimum. In principle, one can stop the iterations when the condition is met, but it is useful to perform some additional iterations and try to obtain a further reduction on the value of the objective function. This guarantees that the condition is met even in the presence of numerical errors. Recall, for instance, that the Lanczos iteration gives an approximation to the actual value of the objective function, and numerical experience has shown that the discrepancy can be relatively large (sometime close to .5 %) if the matrix has a cluster of largest singular values or eigenvalues. Moreover, from Equation 5.3-14, it is convenient to have $X_e Y_e < I$ to facilitate the computation of the optimal approximation.

5.5 Example

Consider again the example described in Section 4.7, and suppose that the bandwidth achieved there by the constrained optimal controller is unacceptable. In order to reduce it, the weighted infinity norm of the transfer function between w and z is penalized, thus adding a new objective to the robust performance problem. The weighting function selected is a high-pass filter W_{hp} with cut-off frequency 2.5 rad/sec. and frequency response magnitude illustrated in Fig. 5.2. The first approach would be to minimize the $\mathcal{H}_\infty$ norm between inputs w, u_1 and outputs y_2, y_1 in the configuration shown in Fig. 5.3. This gives a four-block problem with an unconstrained $\mathcal{H}_\infty$ optimal of 1.6, and therefore it is not clear whether the specifications can or cannot be met. Note, however, that the aim is not to minimize the above objective function, but rather to jointly minimize the $\mathcal{H}_\infty$ norm of the transfer functions between w and y_2 and between u_1 and y_1. Instead of trying to solve this hard problem for which no closed-form solution exists even for the unconstrained case, one can attempt to minimize the structured singular value of the problem, by performing some *D-K iterations* [4]. In the example, after designing one $\mathcal{H}_\infty$ controller and computing the D-scalings associated with the closed-loop, the value of the optimal $\mathcal{H}_\infty$ norm reduces to .8 and hence the scaled plant may be used to attempt a design, using the same time domain specifications as in Chapter 4. In Fig. 5.4, the frequency response of the resulting controller is illustrated, showing that the objective of reducing the bandwidth of the controller has been achieved. It is apparent that a simpler controller, i.e., one with less than 6 states, may suffice to satisfy the time domain constrains and achieve robust stability. Indeed, the controller may be reduced to third order, with poles at $.0426 \pm j.8059$, $-.0925$, zeros at $.0213 \pm j.9398$, $-.8315$ and a gain of $-.6909$. This controller yields an $\mathcal{H}_\infty$ norm from u_1 to y_1 of .97, with a time domain response shown in Fig. 5.5. During the course of the optimization, the objective function became less than one after 15 iterations. The ellipsoid algorithm was stopped one hundred iterations afterwards, although convergence was not achieved; indeed, there was almost 20% difference between the upper and lower bounds.

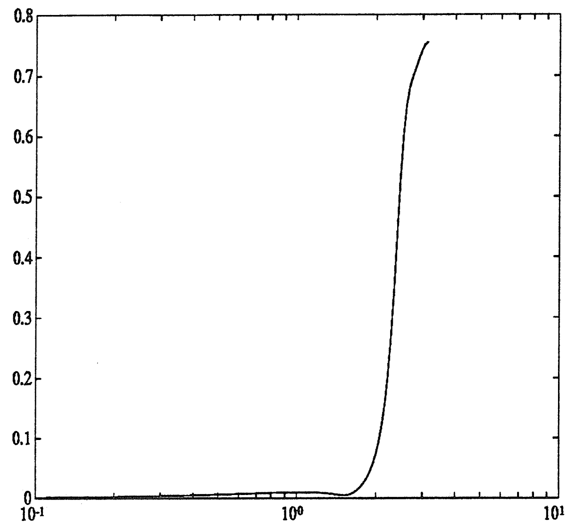

Figure 5.2: Magnitude of the Weighting Function

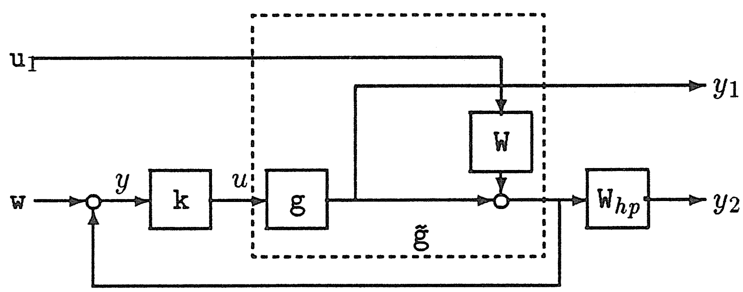

Figure 5.3: Configuration for the Four Block Example

5.6 Appendix: Additional Proofs

Proof of Lemma 5.3

Although it is possible to prove Lemma 5.3 by performing the calculations suggested in the discussion preceding the statement of the lemma, an alternate proof is given which requires fewer algebraic manipulations, and is based on the following fact.

Lemma 5.5 *Let* $g_1 \doteq \left(\begin{array}{c|c} A & B_1 \\ \hline C & D \end{array}\right) \in \mathcal{RH}_\infty$, $g_2 \doteq \left(\begin{array}{c|c} A & B_2 \\ \hline C & 0 \end{array}\right) \in \mathcal{RH}_\infty$, *and assume that* (A, B_2) *is controllable. Then* $[g_1^\sim g_2]_- = 0$ *if and only if* $D^tC + B_1^tYA$.

<u>Proof:</u> Using some standard system manipulations (see [63]), it is not hard to obtain the state-space realization $[g_1^\sim g_2]_- = \left(\begin{array}{c|c} A & B_2 \\ \hline D^tC + B_1^tYA & 0 \end{array}\right)$. By the controllability condition the result follows.

□

<u>Proof of Lemma 5.3</u>: Let

$$g_1 = \begin{bmatrix} G_{11} \\ G_{21} \\ G_{31} \end{bmatrix} \qquad \hat{g}_2 = \begin{bmatrix} G_{12}^n \\ G_{22}(\cdot; \mathbf{Q_n}) - Q_t^\sim \\ G_{32}(\cdot; \mathbf{Q_{23n}}) - Q_{23t}^\sim \end{bmatrix}.$$

From the discussion preceding the statement of Lemma 5.3, $g_1^\sim \hat{g}_2 = 0$. This implies that $[g_1^\sim \hat{g}_2]_- = 0$, and hence, by the stability of g_1, that $[g_1^\sim g_2]_- = 0$, where $g_2 \doteq [\hat{g}_2]_-$. From 5.3-6

$$g_1 = \left(\begin{array}{cc|c} A & B_2E_1^t & B_1 \\ 0 & A_f & 0 \\ \hline C_1 & D_{12}E_1^t & D_{11} \\ C_2 & -\hat{\mathbf{Q}}_\mathbf{n} & D_{21} \\ C_3 & -\hat{\mathbf{Q}}_\mathbf{23n} & D_{31} \end{array}\right) \qquad g_2 = \left(\begin{array}{cc|c} A & B_2E_1^t & 0 \\ 0 & A_f & E_n \\ \hline C_1 & D_{12}E_1^t & 0 \\ C_2 & -\hat{\mathbf{Q}}_\mathbf{n} & 0 \\ C_3 & -\hat{\mathbf{Q}}_\mathbf{23n} & 0 \end{array}\right).$$

It is easy to show that $\left(\begin{bmatrix} A & B_2E_1^t \\ 0 & A_f \end{bmatrix}, \begin{bmatrix} 0 \\ E_n \end{bmatrix}\right)$ is a controllable pair; therefore, by the previous lemma and using the formula for Y_e in Lemma 5.1, the condition $[g_1^{\sim} g_2]_- = 0$ implies

$$\begin{bmatrix} D_{11}^tC_1 + D_{21}^tC_2 + D_{31}^tC_3 & D_{11}^tD_{12}E_1^t - D_{21}^t\hat{\mathbf{Q}}_\mathbf{n} - D_{31}^t\hat{\mathbf{Q}}_\mathbf{23n} \end{bmatrix}$$
$$+\begin{bmatrix} B_1^tY & B_1^t(Y_{12}^0 - Y_{12}^l\mathcal{Q}_n) \end{bmatrix}\begin{bmatrix} A & B_2E_1^t \\ 0 & A_f \end{bmatrix} = 0.$$

In particular,

$$D_{11}^tD_{12}E_1^t - D_{21}^t\hat{\mathbf{Q}}_\mathbf{n} - D_{31}^t\hat{\mathbf{Q}}_\mathbf{23n} + B_1^tYB_2E_1^t + B_1^t(Y_{12}^0 - Y_{12}^l\mathcal{Q}_n)A_f = 0. \quad (5.6-1)$$

Multiplying this equation on the right by E_1 and using the formulas for Y_{12}^0 and Y_{12}^l,

$$D_{11}^tD_{12} - D_{21}^tQ_0^t - D_{31}^tQ_{23}^{0t} + B_1^tYB_2 = 0$$

Since D_{13} is invertible, the formula for Q_{23}^0 follows. Multiplying on the right successively by E_i, $i = 2, \cdots, n-1$, the recursion is obtained. The explicit formula can be computed from the recursion by substitution.

□

Proof of Theorem 5.5

The proof of Theorem 5.5 is cumbersome because it involves the use of matrices that are difficult to manipulate (let alone, write down). Instead of giving the general proof, the case $n = 2$ is treated in detail.

<u>Proof:</u> Let $n = 2$. Then

$$W_2(\mathbf{Q_n}) = \left[\begin{array}{c|cc|cc} yA^2x & yAB_1 & yB_1 & yAB_2 & yB_2 \\ \hline C_1Ax & C_1B_1 & D_{11} & C_1B_2 & D_{12} \\ C_1x & D_{11} & 0 & D_{12} & 0 \\ \hline C_2Ax & C_2B_1 & D_{21} & C_2B_2 & -Q_0^t \\ C_2x & D_{21} & 0 & -Q_0^t & -Q_1^t \end{array}\right]$$

and consider the matrix

$$W_a = \begin{bmatrix} yA^2x & yAB_1 & yB_1 & yAB_2 & yB_2 \\ C_1Ax & C_1B_1 & D_{11} & C_1B_2 & D_{12} \\ C_1x & D_{11} & 0 & D_{12} & 0 \\ C_2Ax & C_2B_1 & D_{21} & C_2B_2 & -Q_0^t \\ C_2x & D_{21} & 0 & -Q_0^t & -Q_1^t \\ \hline C_3Ax & C_3B_1 & D_{31} & Z_{11} & Z_{12} \\ C_3x & D_{31} & 0 & Z_{21} & Z_{22} \end{bmatrix},$$

where the Z_{ij} denote some free variables. Assume that $\overline{\sigma}[W_2(\mathbf{Q_2})] \leq \mathbf{1}$. Since

$$\begin{bmatrix} yA^2x & yAB_1 & yB_1 \\ C_1Ax & C_1B_1 & D_{11} \\ C_1x & D_{11} & 0 \\ C_2Ax & C_2B_1 & D_{21} \\ C_2x & D_{21} & 0 \\ \hline C_3Ax & C_3B_1 & D_{31} \\ C_3x & D_{31} & 0 \end{bmatrix}^t \begin{bmatrix} yA^2x & yAB_1 & yB_1 \\ C_1Ax & C_1B_1 & D_{11} \\ C_1x & D_{11} & 0 \\ C_2Ax & C_2B_1 & D_{21} \\ C_2x & D_{21} & 0 \\ \hline C_3Ax & C_3B_1 & D_{31} \\ C_3x & D_{31} & 0 \end{bmatrix} = \begin{bmatrix} yXy & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \qquad (5.6-2)$$

and $\rho(XY) \leq 1$, from a well-known result in dilation theory, there exist Z_{ij}'s such that $\overline{\sigma}(W_a) = 1$. Suppose that the Z_{ij}'s are chosen so that $\overline{\sigma}(W_a) = 1$. Then, it follows from 5.6-2 that $[B_1^t y \; D_{11}^t \; 0 \; D_{21}^t \; 0 \; D_{31}^t \; 0]^t$ has to be orthogonal to any other column of W_a. In particular,

$$\begin{bmatrix} B_1^t y & D_{11}^t & 0 & D_{21}^t & 0 & D_{31}^t & 0 \end{bmatrix} \begin{bmatrix} yAB_2 & yB_2 \\ C_1B_2 & D_{12} \\ D_{12} & 0 \\ C_2B_2 & -Q_0^t \\ -Q_0^t & -Q_1^t \\ Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} = 0.$$

Therefore, from 2.8-11 and Lemma 5.3, $Z_{11} = C_3 B_2$ and $Z_{12} = -Q^t_{23,0}$. Proceeding similarly with the second block column of W_a, it follows that $Z_{21} = -Q^t_{23,0}$ and $Z_{22} = -Q^t_{23,1}$. Moreover, after some simple manipulations:

$$W_a^t W_a = \begin{bmatrix} xYx & 0 & 0 & x(Y^0_{12} - Y^l_{12}Q_2) \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ (Y^0_{12} - Y^l_{12}Q_2)^t x & 0 & 0 & Y^0_{22} - Y^l_{22}Q_2 - Q^t_2 Y^{lt}_{22} + Q^t_2 Q_2 \end{bmatrix}.$$

This proves that W in 5.3-9 satisfies $\overline{\sigma}(W) \leq 1$. The converse follows immediately since W_2 is the upper part of W_a. The proof for the case of $n > 2$ can be established following an identical procedure and it only requires a more elaborate notation.

□

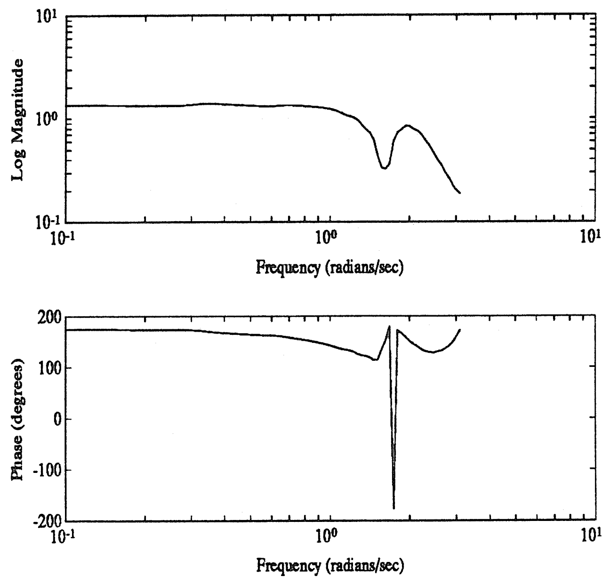

Figure 5.4: Frequency Response of the Controller

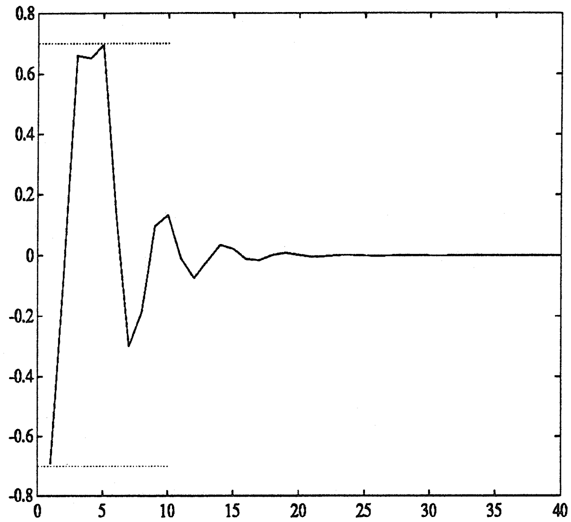

Figure 5.5: Time Domain Response: Third Order Controller. Control Action (ful). Output (dash-dot).

Chapter 6

Solution to a Benchmark Problem for Robust Control

In the 1990 American Control Conference, Wie and Bernstein formulated a benchmark problem with the purpose of highlighting the issues of different robust control design methods [73]. The system, depicted in Fig. 6.1 and consisting of two masses coupled by a spring with no damping, is claimed to be a generic model of an uncertain dynamical system with noncolocated sensor and actuator. It is noncolocated because the control force acts on body 1, while the position of body 2 is measured. The problem is interesting since it is challenging and at the same time simple to describe and with specifications broad enough to allow a variety of design strategies. It has been the focus of special sessions in the 1990, 1991 and 1992 American Control Conference, and a special forthcoming issue of the AIAA Journal on Guidance and Control, where numerous alternative solutions have been given. Assuming that the measurements are corrupted by some noise, it is not hard to derive the state-space representation [73]:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -k/m_1 & k/m_1 & 0 & 0 \\ k/m_2 & -k/m_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1/m_1 \\ 0 \end{bmatrix} u + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1/m_2 \end{bmatrix} w$$

$$y = x_2 + v,$$

where

x_1 = position of body 1.

x_2 = position of body 2.

x_3 = position of body 3.

u = control input.

w = plant disturbance.

y = sensor measurement ($= x_2 + v$).

v = sensor noise.

z = performance variable ($= x_2$).

Wie and Bernstein proposed the following design problems:

Design 1. Design a controller with the following properties:

i The closed-loop system is stable for $m_1 = m_2 = 1$ and $.5 < k < 2.0$.

ii For $w(t)$ = unit impulse at $t = 0$, the performance variable z has a settling time of about 15 seconds for the nominal system $m_1 = m_2 = k = 1$.

iii The measurement noise $v(t)$ is to be characterized by each designer to reflect realism and practical control design.

iv Achieve reasonable performance/stability robustness.

v Minimize controller effort.

vi Minimize controller complexity.

Design 2. Same as *Design 1*, except that in place of i insert:

i Maximize a stability performance measure with respect to the three uncertain parameters m_1, m_2, k whose nominal values are $m_1 = m_2 = k = 1$.

Following the framework described in Chapter 5, it is possible to draw the block diagram illustrated in Fig. 6.2, where the signals are:

w plant disturbance (input).

v sensor noise (input).

u_i (for i=0, 1 or 2) fictitious input used to assess robust stability.

y_i (for i=0, 1 or 2) fictitious output used to assess robust stability.

y sensor measurement.

z position of the second mass.

u control action.

Define the "time" and "frequency" inputs and outputs as follows:

$$u_t = w \qquad y_t = \begin{bmatrix} z \\ u \end{bmatrix}$$

$$u_f = \begin{bmatrix} w \\ v \\ u_0 \\ u_1 \\ u_2 \end{bmatrix} \qquad y_f = \begin{bmatrix} z \\ u \\ y_0 \\ y_1 \\ y_2 \end{bmatrix}.$$

Note that u is included both as an output and an input so that control actions can be constrained, and that w, z and u are repeated as time and frequency signals. The corresponding transfer matrix between these inputs and outputs is continuous-time, and therefore, in order to apply the results on this thesis, it is necessary to compute a discrete time model, by adding a sampler and a hold device at each output and input

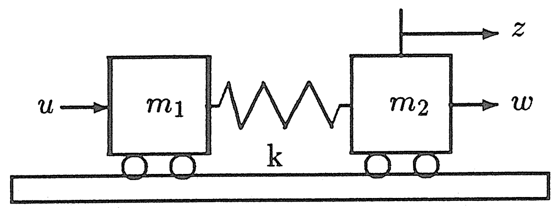

Figure 6.1: The Benchmark Problem

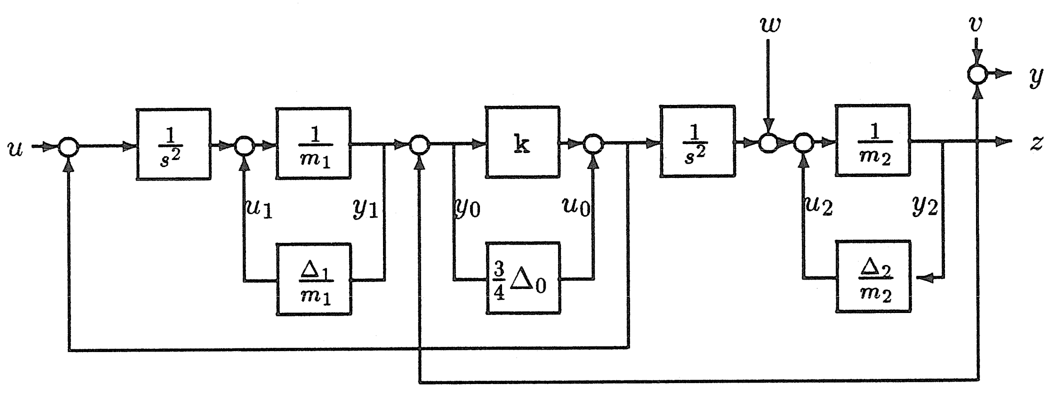

Figure 6.2: Block Diagram

respectively. A sampling time of .1 secs was chosen, corresponding to approximately fifteen times the bandwidth of the open-loop system. This selection also involved a practical issue, since a smaller sampling time implies a larger number of samples, and hence larger optimization programs for computing a constrained $\mathcal{H}_\infty$ solution. An anti-aliasing filter was included at the output of the plant, and it was checked that relatively large variations, e.g., 30%, in the sampling rate left the results essentially invariant.

Design 1

For the first design, the only uncertainty considered is in the value of the spring constant. In order to treat the problem in the $\mathcal{H}_\infty$ framework, it is necessary to assume that the uncertain parameter k lies inside a disk of radius δ_r in the complex plane, for each frequency ω. This assumption makes the approach conservative, since the resulting robustness problem is possibly much more restrictive than the actual one. The first design considered was a standard $\mathcal{H}_\infty$ one, where the objective was to minimize the $\mathcal{H}_\infty$ norm between u_0 and y_0. As the continuous time system has poles on the $j\omega$-axis, the discretized version has poles on the unit circle and hence is not well posed for the $\mathcal{H}_\infty$ design. This difficulty is overcome in [11] by introducing a bilinear transform that removes the ill-posedness and provides tunning parameters that can be adjusted to achieve the design specifications. Similarly, it is possible to use the change of variables introduced in Section 3.6. If the value ρ is set to 1.03 (recall that this corresponds to placing all closed-loop poles on a circle of radius 1/1.03), the resulting norm is slightly above one. However, as shown in Fig. 6.3, the control action is too large to be applicable.

In order to investigate the tradeoff between robustness (as measured by the $\mathcal{H}_\infty$ norm) and tolerable control actions, several constrained controllers were designed, with $\rho = 1.001$ and a horizon of 150 samples; note that this horizon corresponds to the 15 seconds specified for the settling time. Results are summarized in Fig. 6.4a; the main conclusions is that, modulo the conservatism introduced by treating the real parametric uncertainty as being complex, a control bound of at least .7 is necessary in order to achieve robust stability. Similarly, the tradeoff between robustness and the tolerable excursion of the mass m_2 was investigated, by fixing a settling time tolerace of .05 (i.e., at $t = 150$, the variable x_2 has to be within $\pm .05$), bounding the control action by one, and varying the tolerable excursion. Results are summarized in Fig. 6.4b, which shows that, under the conditions of robust stability and control action bounded by one, the excursion of the second body cannot be constrained to be less than 2.7. Unfortunately, this result is not conclusive due to the conservatism cited above and also because the constrains are largely violated after the horizon is cleared. From the discussion in Chapter 3, one can attempt to improve the tail behavior by enlarging the horizon and increasing the value of ρ. Fig. 6.5 shows the effect of increasing the value of ρ on both the result obtained from the minimization (solid line) and the actual $\mathcal{H}_\infty$ norm of the transfer function between u_0 and y_0. Note that both functions are sensitive on ρ and also the gap between them grows fast, implying that even for ρ slightly larger than one, the upper bound obtained by the minimization can be a poor estimate on the actual value of the norm. As an illustration, a controller was designed by setting the horizon length to 250 and taking $\rho = 1.03$; the upper bound found by the optimization is 2.3 but the actual $\mathcal{H}_\infty$ norm of the robustness transfer function is 1.6, giving a 40 % gap between both. If the fact that the uncertainty on k is real is taken into account, the resulting controller turns out to be robustly stabilizing. However, the controller has very large gains at high frequencies, and the resulting control action may then be sensitive to measurement noise.

In order to produce a design that meets all the specifications of Design 1 (except for the complexity of the controller issue, to be treated a posteriori), the frequency

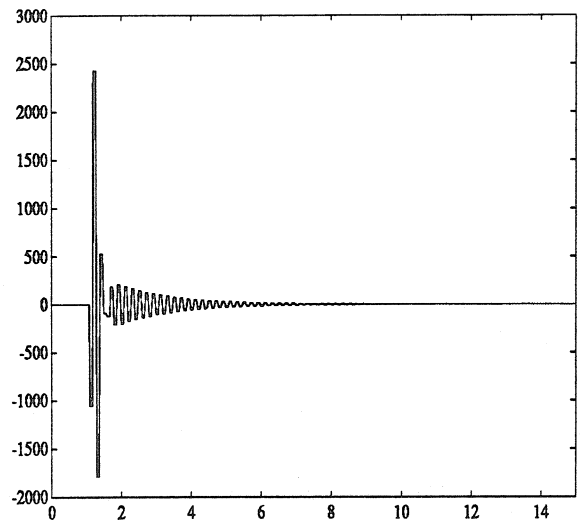

Figure 6.3: Unconstrained Control Action

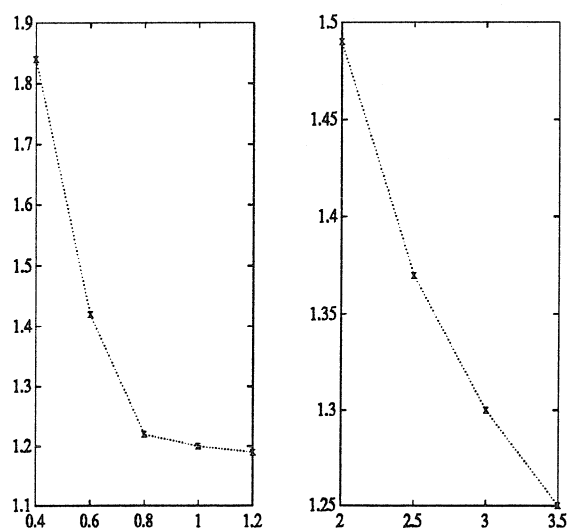

Figure 6.4: $\mathcal{H}_\infty$-norm vs. Bounds in Control Action and $\max|z_k|$.

domain objective was modified to take into account the transfer functions between w and z and between v and u, i.e., the weighted transfer matrix between $\begin{bmatrix} w & v & u_0 \end{bmatrix}^t$ and $\begin{bmatrix} z & u & y_0 \end{bmatrix}^t$ was minimized. The output z was multiplied by a constant of .1 to reduce the impact on the overall norm, while the control action was weighted by a high-pass filter with cutoff frequency of 2.5. The idea behind this scheme is to improve the behavior of the second body (since displacements are now penalized) and reduce the gain of the controller at high frequencies, thus reducing the sensitivity to measurement noise. An $\mathcal{H}_\infty$ controller with $\rho = 1.01$ was designed, followed by D-K iterations to make the result less conservative (see, for instance, [4]). The value of the objective function did not seem to improve after one of such iterations, giving $\gamma = 1.4$; the $\mathcal{H}_\infty$ norm of the resulting transfer function between u_0 and y_0 is 1.2 and hence the controller achieves robust stability. Fig. 6.6 shows that, although the time domain responses were closer to the specifications, they are far from satisfying the time domain constraints. These, and the time responses to follow, were computed by simulating the continuous-time system with a sample-data implementation of the controller; in all cases, the dash-dot, solid and dash lines, correspond to the displacement of the first and second masses and the control action respectively. The horizon length was fixed to 250, and after several iterations, it was found that $1.63 \leq \gamma_{opt} \leq 1.65$, with γ_{opt} denoting the optimal $\mathcal{H}_\infty$-norm. The norm of the transfer function between u_0 and y_0 for the resulting controller is 1.4; however, the Nyquist plot in Figure 6.7 shows that the closed-loop system is robustly stable for real uncertainty. The displacement of the two bodies and the control action are shown in Fig. 6.8, while the Bode plot of the transfer function between sensor noise and control action is given in Fig. 6.9. Note that all the specifications are achieved, but the order of the resulting controller is extremely high and therefore a procedure for model reduction was needed. The order reduction problem is not hard, becuase the controller is stable, and its frequency response seems to be fairly "smooth" (see

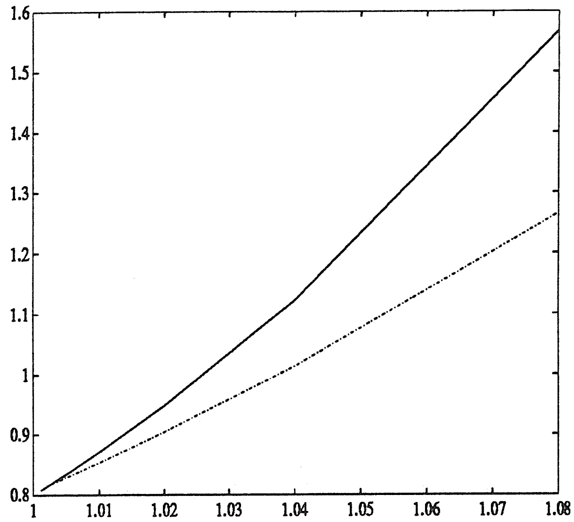

Figure 6.5: $\mathcal{H}_\infty$-norm vs. ρ

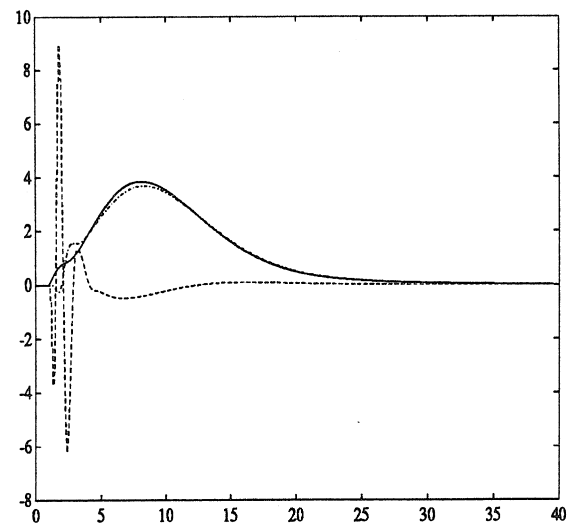

Figure 6.6: Time Responses for Optimal Unconstrained Controller

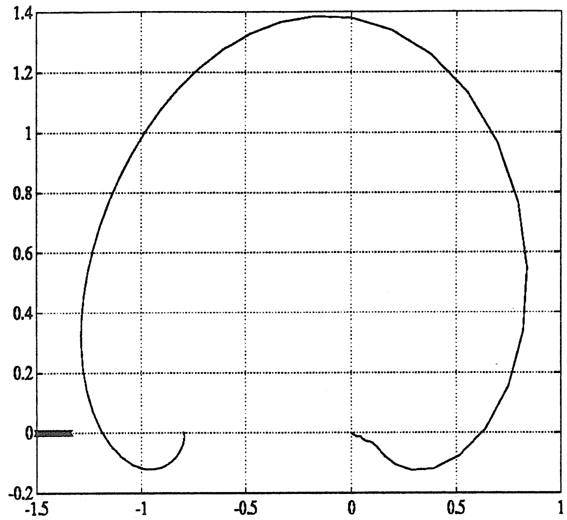

Figure 6.7: Robustness Transfer Function

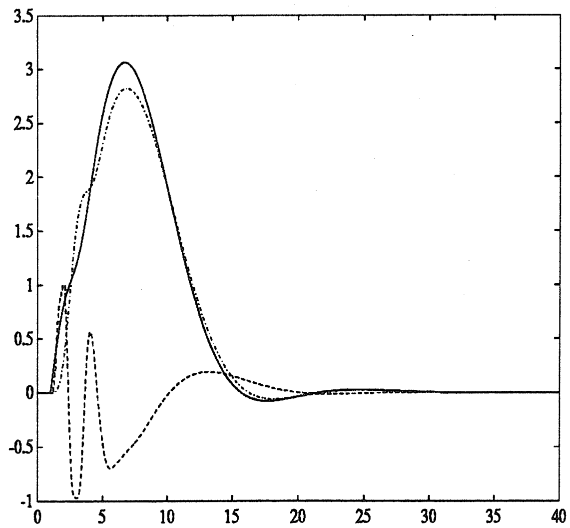

Figure 6.8: Time Responses for the Constrained Controller (design 1)

the Bode plot in Fig. 6.10, solid line). It was therefore reasonable to expect that the established methods for order reduction in the frequency domain, should produce a low-order controller without any significant effect in the closed-loop behavior. Indeed, a controller with 12 states was computed using the balanced truncation method; its frequency response has no noticeable difference with the full order one. Moreover, since robustness becomes critical only at the low-frequency range, this controller could be further model reduced, by using the weighted balanced truncation method and a low-pass filter with cutoff frequency .3 as weight. The resulting controller has 7 states, with poles at $.7106 \pm .5947i$, $.7252 \pm .2736i$, $.8372 \pm .3283i$, $.5531$, zeros at $.8368 \pm .7054i$, $.8602 \pm .4225i$, $.9896$, 1.041, 10.6446, and a gain of $-.4377$; its transfer function is shown in dash line in Fig. 6.10. The difference in the time responses between this controller and the full order one is negligible.

Design 2

The objective of this design is to maximize robustness, assuming that the three parameters k, m_1 and m_2 are uncertain and with nominal values $k = m_1 = m_2 = 1$. In the absence of any further information about the spring or the masses, if Δk, Δm_1 and Δm_2 denote the amount of uncertainty, the objective is to design a controller that remains stable for $|\Delta k| < \gamma_r$, $|\Delta m_1| < 1/\gamma_r$ and $|\Delta m_2| < 1/\gamma_r$, with γ_r as small as possible, while satisfying the time domain specifications. In order to use the $\mathcal{H}_\infty$ framework, it was assumed that each Δ lay on a disk in the complex plane, and the system with four inputs (the sensor noise v and three fictitious inputs u_0, u_1, u_2) and four outputs (the control action u and fictitious outputs y_0, y_1, y_2) was considered. The uncertainty blocks where connected around the fictitious outputs and inputs. The transfer function from v to u was weighted by a high-pass filter with cutoff frequency 1, in order to penalize the sensor noise amplification at high frequencies. After one D-K iteration, the norm of the unconstrained transfer function was found

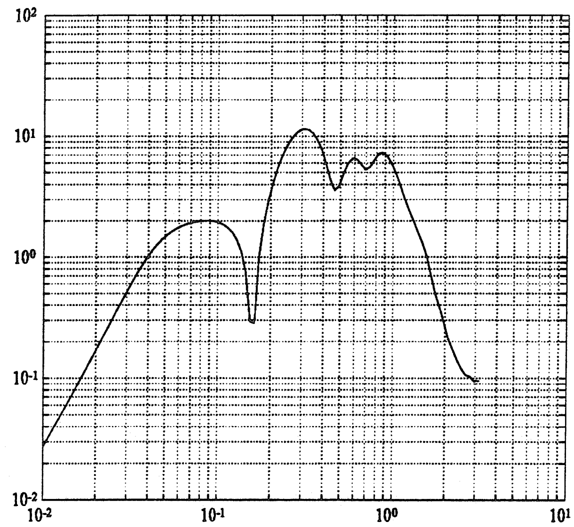

Figure 6.9: Transfer Function Between Sensor Noise and Control Action

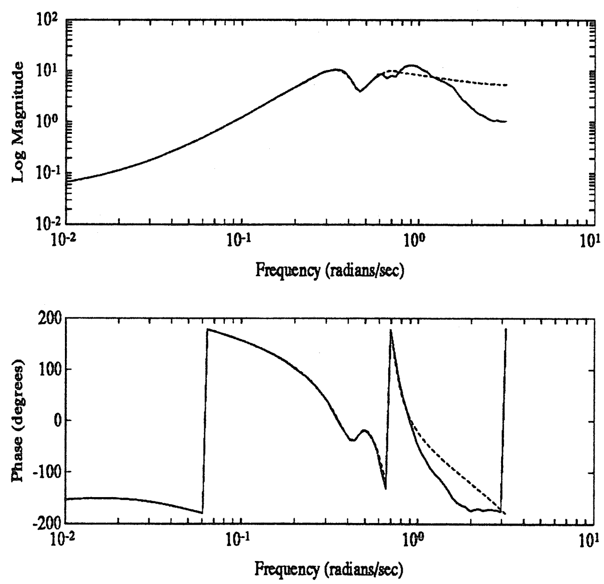

Figure 6.10: Frequency Response of Constrained Controller (design 1)

to be slightly above 4. Fixing γ to 5.5, a constrained controller was designed for a horizon of length 250. The structured singular value of the resulting closed-loop system was found to be 4.5, suggesting a 22 % tolerable uncertainty on the Δ's. If the real nature of the uncertainty is taken into account, then $1/\gamma_r$ is found to be close to .3, implying that the system tolerates 30 % uncertainty on the three parameters before going unstable. The displacement in the two masses and control action are shown in Fig. 6.11. The transfer function from sensor noise to control action is shown in Fig. 6.12 and has a bandwidth of 1.5 (or 15 Hz for the continuous time system). The order reduction problem was not much harder than in Design 1. Using balanced model truncation weighted by a low-pass filter with cutoff frequency at .8, it was possible to compute a 6-th order controller with characteristics almost identical to the full order one, which had 274 states. The poles are located at $.4472 \pm .55971i$, $.6927 \pm .1360i$, $.7940 \pm .3337i$, the zeros at 7.7097, $-.5506$, $.7151 \pm .55961$, 1.0543, .9885, and the gain is -0.2334. The frequency response of the reduced order controller is shown in Fig. 6.13.

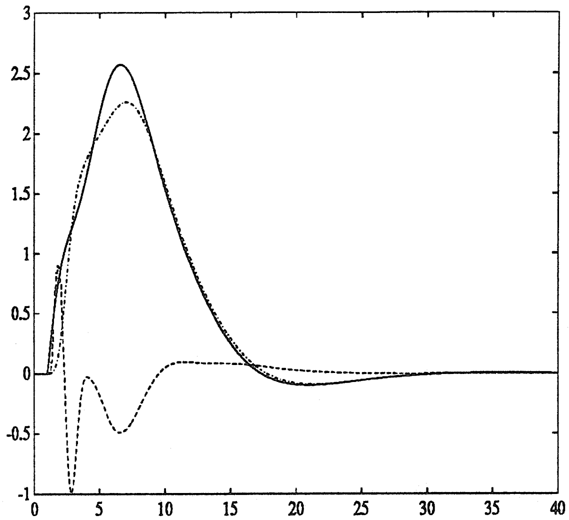

Figure 6.11: Time Responses for the Constrained Controller (design 2)

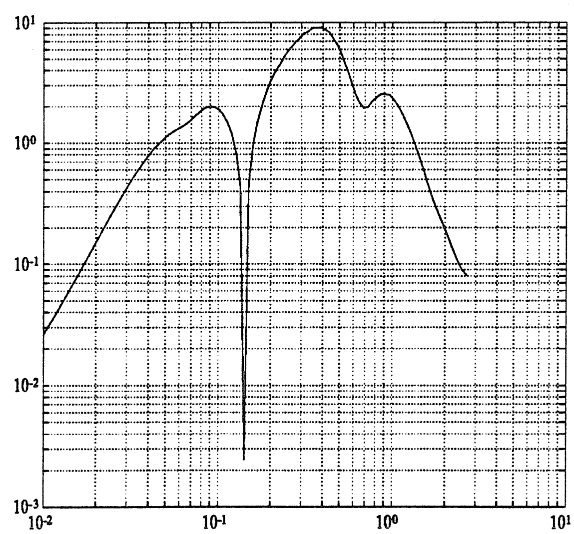

Figure 6.12: Transfer Function from Sensor Noise to Control Action

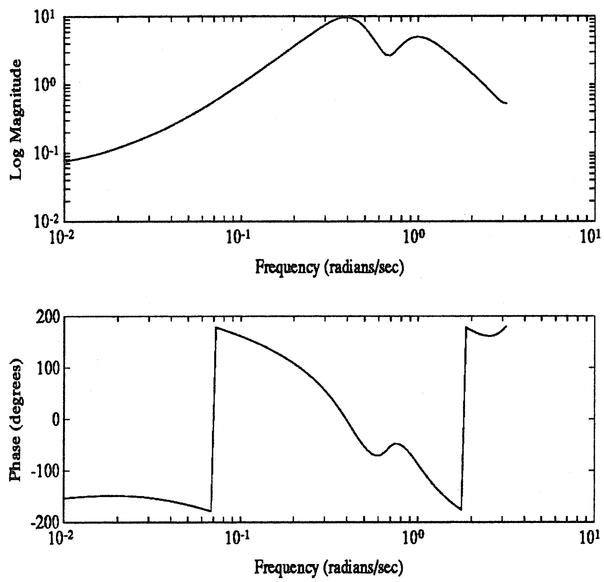

Figure 6.13: Frequency Response of Constrained Controller (design 2)

Chapter 7

Summary and Future Directions

This chapter concludes the thesis summarizing the results and suggesting some future lines of investigation.

7.1 Summary

During the last few years, $\mathcal{H}_\infty$-optimal control theory has proved to be a very useful and flexible tool for designing closed-loop controllers for a variety of applications. On a typical $\mathcal{H}_\infty$ design session, one starts by taking a number of input and output signals to a system, for which specifications are to be imposed besides the control actions and measurements. These inputs and outputs usually include actual signals and also some fictitious ones that are used to isolate model uncertainty. After building the transfer matrix that relates the inputs to the outputs, one proceeds to design weighting functions on these signals so that the minimization of the $\mathcal{H}_\infty$ norm of the system, results in the satisfaction of all specifications. Although some specifications, like robustness against unstructured uncertainty, may be exactly formulated in terms of an $\mathcal{H}_\infty$ norm, some others can be much more challenging, hence requiring a number of trial and error iterations with no guarantee of producing a satisfactory result. In that case, one must conclude that either the specifications were too tight and no feasible solution exists or that the process of translating the specification introduced a new -and undesirable- tradeoff.

The main motivation for the present thesis was to study a paradigm for controller design, that allows for the explicit specification of time domain constraints in the $\mathcal{H}_\infty$-optimal control formulations. These constraints are treated *exactly*, i.e., it is not necessary to translate them into the frequency domain, and therefore the approach is more apt for designing a controller that satisfies an $\mathcal{H}_\infty$ norm bound and some hard time-domain constraints or study the related tradeoff involved in a design. The first case considered was a single input-single output that requires a minimum background but still can be used to present the main ideas and results. Specifically, the design problem was first reduced (by using the Youla parametrization) to the minimization of the norm of a transfer function which depends affinely on the the design parameter subject to constraints on the coefficients of its Taylor expansion. The fact that time domain constraints are imposed over a horizon of length N, implies that only the first N coefficients of the expansion of the parameter are constrained; then, using an idea originally proposed in [38], the problem was divided into first computing these parameters and then calculating the "tail" so as to minimize the $\mathcal{H}_\infty$ norm. The novelty in the present treatment is that an objective function is formulated in Chapter 3, which depends only on the first part of the response. By minimizing this function subject to the constraints, and then adding the tail computed from an unconstrained $\mathcal{H}_\infty$ problem, a solution to the constrained problem is obtained. A simple formula for the objective function in terms of the state-space data of the problem was derived, which expresses it as the norm of a matrix that depends linearly on the parameters. In particular, this implies that the function is convex in the parameters, and therefore if the constraints are convex, then the whole optimization problem is so and can be effectively solved. Unfortunately, minimizing the maximum singular value of a matrix usually produces a nondifferentiable problem, and indeed it was shown that in the limiting case when constraints are non-binding, the multiplicity of the largest singular value is far larger than the upper bound discussed for generic matrices in [26]. The fact that constraints are imposed only over a finite horizon, was shown to cause no major trouble, if one is willing to sacrifice optimality; this is because the optimal solution may not behave well after the horizon, but suboptimal ones, if appropriately chosen, are guaranteed to satisfy a steady-state constraint. A numerical procedure for solving the optimization problem was discussed in Chapter 4. The two main conclusions of this chapter, were that the Ellipsoid Algorithm seems to be a good candidate for computing a solution –giving its ability to deal with large-scale problems– and that vector Lancszos iterations provide the best tool for the computation of the objective function. This is because it is well suited for dealing with large matrices and a cluster of largest singular values or eigenvalues. The general instance of the problem can be solved by combining the machinery reviewed and developed in Chapter 2 for the unconstrained problem with the ideas presented in Chapter 3. This is done in Chapter 5, which is technically more involved but otherwise parallels the treatment for of simple case.

To summarize, in this thesis it has been shown that the time domain constrained $\mathcal{H}_\infty$ problem may be formulated as a convex optimization, that can be solved using a numerical procedure which is effective and easy to implement.

7.2 Future Directions

Among the points that deserve further study, there are the following.

- Numerical Optimization. Although in this thesis, an algorithm has been given that has been used successfully to solve actual problems, the large-scale nature of the optimization constitutes an obstacle for solving large multivariable problems. Some aspects of this problem (specifically, the ones related with storage requirements or the time involved in the evaluation of the objective function) may be addressed by using more powerful computers, but the slow converge rate of the ellipsoid algorithm cannot be improved in this way. Recently, a number of optimization algorithms for minimizing the spectral radius that are based on penalty methods and Newton-type iterations have appeared in the literature [52,40,8], that are guaranteed to have a much faster convergence rate. The application of these algorithms for solving the minimization problem seems promising, although it is yet to be established how this algorithms perform on large-scale problem [9].

- Order of the controller. The bounds obtained for the order of the controller are reminiscent to the ones usually found in l_1 optimal control [15], in the sense that they depend on the order of the plant *and* on the size of the minimization that needs to be solved in order to compute a solution to the problem. However, the objective function is an $\mathcal{H}_\infty$ norm with a direct connection with the closed-loop frequency response, and therefore all the methods available for doing model reduction in the frequency domain can be tried on the controller. For instance, weighted or unweighted balanced model truncation usually produce good results and also provide a tight estimate on the degradation of norm caused by the reduction. Note that this is not true for the l_1 case, since although some bounds can be computed, they usually give only poor estimates of the actual degradation in an l_1 sense [46]. Of course, the procedures must be applied with caution, since they do not guarantee that the time domain constraints will be satisfied by the reduced order controller. On the other hand, some methods that guarantee a good matching of the first few terms of a Taylor expansion for the controller exist in the literature, but they provide no estimate on the error between the full and the reduced order transfer function, measured in the $\mathcal{H}_\infty$-norm. The effect of the techniques for order reduction on the time domain behavior of the closed-loop system, and the formulation of methods that provide guaranteed behavior for both the time and frequency domain aspects of the design, should be studied further. In particular, the limited experience with some of the problems solved suggests that there may exist a connection between the order of the controller and the order of the plant *plus* the number of active time domain constraints.

- Robust performance. The theory presented in this thesis allows the inclusion of time domain constraints on the $\mathcal{H}_\infty$ design methodology, so that for fixed inputs, the *nominal* behavior of the outputs can be specified. Although this is a very desirable characteristic, as discussed above it falls short of being completely satisfactory, because no guarantees can be given on the time domain behavior of the uncertain system which is the object of the design. This characteristic is shared by all other methods for constrained $\mathcal{H}_\infty$ design found in the literature, and deserves more study.

Bibliography

[1] V. Adamjan, D. Arov and M. Krein (1978). "Infinite Block Hankel Matrices and their Connection with the Interpolation Problem," *Amer. Math. Soc. Transl.* vol. 111, pp. 133-156 (Russian original, 1971).

[2] W. Arnold and A. Laub (1984). "Generalized Eigenproblem Algorithms and Software for Algebraic Riccati Equations." *Proceedings of the IEEE*, vol. 72, pp. 1746-1754.

[3] B. Bamieh, J. Pearson, B. Francis and A. Tannenbaum (1991). "A Lifting Technique for Linear Periodic Systems with Applications to Sampled-data Control," *Systems & Control Letters*, vol. 17, No. 1, pp. 79-88.

[4] G. Balas, J. Doyle, K. Glover, A. Packard and R. Smith (1991). *μ-Tools. Matlab Functions for the Analysis and Design of Robust Control Systems*, Musyn Inc., P.O. Box 1337, Minneapolis, MN 55414-5377.

[5] J. Ball and A. Ran (1987). "Optimal Hankel Norm Model Reduction and Winer-Hopf Factorization I: the Canonical Case," *SIAM Journal on Control and Optimization*, vol. 25, pp. 362-382.

[6] R. Bland, D. Goldfarb and M. Todd (1981). "The Ellipsoid Method: A Survey," *Operations Research*, Vol. 29, No. 6.

[7] S. Boyd and C. Barratt (1990). *Linear Controllers Design - Limits of Performance*, Prentice-Hall, Englewood Cliffs.

[8] S. Boyd and L. El Ghaoui (1992). "Method of Centers for Minimizing Generalized Eigenvalues." Preprint.

[9] S. Boyd (1991). Personal Communication.

[10] S. Boyd, V. Balakrishnan, C. Barratt, N. Khraishi, X. Li, D. Meyer and S. Norman (1988). "A New CAD Method and Associated Architectures for Linear Controllers," *IEEE Transactions on Automatic Control*, vol. 33, No.3.

[11] R. Chiang and M. Safonov, "H_∞ Robust Control Synthesis for an Undamped, Non-colocated Spring-Mass System," *Proceedings of the ACC*, San Diego, CA, 1990.

[12] C. Chu, J. Doyle and E. Lee (1986). "The General Distance Problem in $\mathcal{H}_\infty$ Optimal Control Theory," *International Journal of Control*, vol. 44, pp. 565-596.

[13] F. Clarke (1983). *Optimization and Nonsmooth Analysis*, Wiley, New York.

[14] J. Cullum and R. Willoughby (1985). *Lanczos Algorithms for Large Symmetric Eigenvalue Computations. Vol. 1: Theory. Vol. 2: Programs*, Birkh äuser.

[15] M. Dahleh and J. Pearson (1987). "L1-Optimal Feedback Controllers for MIMO Discrete-Time-Systems," *IEEE Transactions on Automatic Control*, vol. 32, 314-322.

[16] J. Doyle (1982). "Analysis of Feedback Systems with Structured Uncertainties," *IEE Proceedings, Part D*, Vol. 133, pp. 45-56.

[17] C. Davis, W. Kahan and H. Weinberger (1982). "Norm-preserving Dilations and their Application to Optimal Error Bounds," *SIAM J. Numer. Anal.*, vol. 19, pp. 445-469.

[18] J. Doyle, K. Glover, P. Khargonekar and B. Francis (1989). "State-Space Solutions to Standard H^2 and H_∞ Control Problems," *IEEE Transactions on Automatic Control*, vol. 34, No.8.

[19] J. Doyle (1984). *Lecture Notes in Advances in Multivariable Control*, ONR/Honeywell Workshop, Minneapolis.

[20] T. Dziuban, J. Ecker and M. Kupfershmid (1985). "Using Deep Cuts in an Ellipsoid Algorithm for Nonlinear Programming," *Mathematical Programming Study*, No. 25.

[21] K. Fegley, S. Blum, J. Bergholm, A. Calise, J. Marowitz, G. Porcelli and L. Sinha (1971). "Stochastic and Deterministic Design and Control Via Linear and Quadratic Programming," *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 759-766.

[22] A. Frazho and D. Augenstein (1991). "A Remark on Two Solutions to the Rational Nehari Interpolation Problem," *Integral Equations and Operator Theory*, vol. 14, pp. 304-310.

[23] B. Francis and J. Doyle (1987). "Linear Control Theory with an $\mathcal{H}_\infty$ Optimally Criterion," *SIAM Journal on Control and Optimization*, vol. 25, pp. 815-844.

[24] C. Foias and A. Frazho (1989). *The Commutant Lifting Approach to Interpolation Problems*, Birkhäuser Verlag.

[25] B. Francis (1987). *A course in H_∞ Control Theory*, in Lecture Notes in Control and Information Science, M. Thoma and A. Wyner, Eds. New York: Springer Verlag, vol. 88.

[26] S. Friedland, J. Nocedal and M. Overton (1987). "The Formulation and Analysis of Numerical Methods for Inverse Eigenvalue Problems," *SIAM Journal on Numerical Analysis*, Vol. 24, pp. 634-667.

[27] K. Furuta and S. Phoojaruenchanachai (1990). "An Algebraic Approach to Discrete-time $\mathcal{H}_\infty$ Control Problems," *Proceedings of the 1990 ACC*, San Diego, California, vol. 3, pp. 3067-3072,

[28] K. Glover (1984). "All Optimal Hankel-norm Approximation of Linear Multivariable Systems and their $\mathcal{L}_\infty$-error Bounds," *International Journal of Control*, vol. 39, pp. 1115-1193.

[29] K. Glover and J. Doyle (1989). "A State-space Approach to $\mathcal{H}_\infty$ Optimal Control," in *Three Decades of Mathematical System Theory: A Collection of Surveys at the Occasion of the 50th Birthday of Jan C. Willems'*, H. Nijmeijer and J. M. Schumacher eds., Springer-Verlag, Lecture Notes in Control and Information Sciences, vol. 135, pp. 179-218.

[30] C."Gustafson and C. Desoer (1983). "Controller Design for Linear Multivariable Feedback Systems with Stable Plants, Using Optimization with Inequality Constraints," *International Journal of Control*, vol. 37, no. 5, pp. 881-907.

[31] G. Golub and C. Van Loan (1985). *Matrix Computations*, Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press.

[32] M.J. Grimble and D. Fragopoulos (1990). "Solution of Discrete $\mathcal{H}_\infty$ Optimal Control Problems Using a State-space Approach," *Proceedings of the 29th CDC Conference*, Honolulu, Hawaii, pp. 1775-1780.

[33] I. Ghoberg, M. Kaashoek and F. van Schagen (1988). "Rational Contractive and Unitary Interpolants in Realized Form," *Integral Equations and Operator Theory*, vol. 11, pp. 105-127.

[34] K. Glover (1989). "A Tutorial on Hankel-norm Approximation," in *From Data to Model*, Jan C. Willems ed. Springer-Verlag.

[35] K. Glover, D. Limebeer, J. Doyle, E. Kasenally and M. Safonov (1991). "A Characterization of all solutions to the 4 block general distance problem," Siam Journal on Control and Optimization, vol. 29, pp. 283-324.

[36] D. Gu, M. Tsai, S. O'Young and I. Postlethwaite (1989). "State-space Formulae for Discrete-time $\mathcal{H}_\infty$ Optimization," *International Journal of Control*, vol. 49, No.5, pp. 1683-1723.

[37] I. Horowitz (1963). *Synthesis of Feedback Systems.* Academic Press.

[38] J. Helton and A. Sideris (1989). "Frequency Response Algorithms for H_∞ Optimization with Time Domain Constraints," *IEEE Transactions on Automatic Control*, vol. 34, No.4.

[39] P. Iglesias and K. Glover (1991). "State-Space Approach to Discrete-time $\mathcal{H}_\infty$ Control," *International Journal of Control*, vol. 54, pp. 1031-1073.

[40] F. Jarre (1991). "An Interior-point Method for Minimizing the Maximum Eigenvalue of a Linear Combination of Matrices," to appear in SIAM J. on Control and Optimization.

[41] J. Kelly (1960). "The Cutting Plane Method for Solving Convex Programs," *Journal of the SIAM*, Vol. 8, pp. 703-712.

[42] K. Kiwiel (1983). "An Agreggate Subgradient Method for Nonsmooth Convex Minimization," *Mathematical Programming*, Vol. 27, pp. 320-341.

[43] K. Kiwiel (1985). *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics 1333, Springer, Berlin.

[44] K. Kiwiel (1985). "An Algorithm for Linearly Constrained Convex Nondifferentiable Minimization Problems," *Journal of Mathematical Analysis and Applications*, vol. 105, No. 3.

[45] D. Kavranoglu and A. Sideris (1989). "A simple solution to $\mathcal{H}_\infty$ optimization problems," *Proceedings of the 1989 ACC*, Pittsburgh, pp. 753-758.

[46] J. Lam and B. Anderson (1992). "L_1 Impulse-response Error Bound for Balanced Truncation," *System and Control Letters*, vol. 18, No. 2, pp. 129-137.

[47] C. Lemarechal (1982). "Numerical Experiments in Nonsmooth Optimization," in E. Nurminski, ed., *Progress in Nondifferentiable Optimization*, International Institute for Applied System Analysis, Laxenburg, Austria.

[48] D. Limebeer, M. Green and D. Walker (1989). "Discret-time $\mathcal{H}_\infty$ Control," *Proceedins of the 1989 CDC*, Tampa, Florida, pp. 392-396.

[49] D. Limebeer and Y. Hung (1987). "An Analysis of Pole Zero Cancellations in $\mathcal{H}_\infty$-optimal Control Problems of the First Kind," *SIAM Journal on Control and Optimization*, vol. 25, no. 5.

[50] K. Liu, T. Mita and H. Kimura (1990). "Complete Solution to the Standard $\mathcal{H}_\infty$ Control Problem of Discrete Time Systems," *Proceedins of the 1990 CDC*, Honolulu, Hawai, pp. 1786-1793.

[51] B. Molinari (1975). "The Stabilizing Solution of the Discrete Algebraic Riccati Equation," *IEEE Transactions on Automatic Control*, vol. 20, pp. 396-399.

[52] Y. Nesterov and A. Nemirovsky (1990). *Optimization Over Positive Semidefinite Matrices: Mathematical Background and User's Manual*, USSR Academy of Sciences Central Economic & Mathematical Institute, 32 Krasikova St., Moscow 117418 USSR.

[53] M. Overton (1988). "On minimizing the maximum eigenvalue of a symmetric matrix," *SIAM Journal on Matrix Analysis and Applications*, vol. 9, no. 2.

[54] M. Overton (1992). "Large-scale Optimization of Eigenvalues," *SIAM Journal on Optimization*, pp. 88-120.

[55] B. Parlett (1980). *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs.

[56] T. Pappas, A. Laub and N. Sandell Jr. (1980). "On the Numerical Solution of the Discrete-Time Algebraic Riccati Equation," *IEEE Transactions on Automatic Control*, vol. 20, pp. 631-641.

[57] E. Polak, "On the Mathematical Foundations of Nondifferentiable Optimization in Engineering Design," *SIAM Review*, March 1987.

[58] E. Polak, D. Mayne and Y. Wardi (1983). "On the Extension of Constrained Optimization Algorithms from Differentiable to Nondifferentiable Problems," *SIAM Journal on Control and Optimization*, Vol. 21, No. 2, pp. 179-203.

[59] E. Polak and S. Salcudean (1989). "On the Design of Linear Multivariable Feedback Systems Via Constrained Nondifferentiable Optimization in H_∞ Spaces," *IEEE Transactions on Automatic Control*, Vol.34, No.3.

[60] S.C. Power (1982). *Hankel Operators on Hilbert Space.* Pitman.

[61] H. Rosenbrock (1970). *State-space and Multivariable Systems.* John Wiley & Sons.

[62] A. Sideris and H. Rotstein (1992). "Single Input-Single Output $\mathcal{H}_\infty$-Control with Time Domain Constraints," submitted for publication.

[63] H. Rotstein and A. Sideris (1992). "Recursive Computation of the Optimal Nehari Approximation," submitted for publication.

[64] A. Sideris (1990). "$\mathcal{H}_\infty$ Optimal Control as a Weighted Wiener-Hopf Problem," *IEEE Transactions on Automatic Control*, vol. 35, No. 3, pp. 1272-1276.

[65] L. Silverman and M. Bettayeb (1980). "Optimal Approximation of Linear Systems," *Proceedings of the American Control Conference.*

[66] N. Shor and V. Gershovich (1979). "Family of Algorithms for Solving Convex Programming Problems," *Cybernetics*, Vol. 15, No. 4, pp. 502-507.

[67] M. Safonov, D.J.N. Limebeer and R. Chiang (1989). "Simplifying the $\mathcal{H}^\infty$ Theory via Loop-shifting, Matrix-pencil and Descriptor Concepts," *International Journal of Control*, vol. 50, No.6.

[68] N. Shor (1985). *Minimization Methods for Non-Differentiable Functions*, Springer Series in Computational Mathematics, Springer-Verlag, 1985.

[69] A. Stoorvogel (1992). "The Discrete-time $\mathcal{H}_\infty$ Control Problem with Measurement Feedback," *Siam Journal on Control and Optimization*, vol. 30, pp. 182-202.

[70] J. Truxal (1955). *Automatic Feedback Control System Synthesis.* MacGraw-Hill.

[71] P. Van Dooren (1981). "A Generalized Eigenvalue Approach for Solving Riccati Equations," *Siam Journal on Scientific Statistic Computations*, vol. 2, pp. 121-135.

[72] L. Wang and G. Zames (1990). "Lipschitz Continuity of $\mathcal{H}_\infty$ Interpolation," *System and Control Letters*, vol. 14, No. 5, pp. 381-387.

[73] B. Wie and D. Bernstein, "A Benchmark Problem for Robust Control Design," *Proceedings of the ACC*, San Diego, CA, 1990.

[74] K. Zhou, J. Doyle and K. Glover (1992). *Robust and Optimal Control*, preprints.

[75] J. Zowe (1985). "Nondifferentiable Optimization," in K. Schittkowski, ed., *Computational Mathematical Programming*, Springer, Berlin, pp. 323-356.