ℓ_1 Optimal Control - A Polynomial Approach

by

Ing. Zdeněk Hurák

supervised by Doc. Ing. Michael Šebek, DrSc.

Dissertation

Presented to the Faculty of Electrical Engineering of

Czech Technical University in Prague

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor

in the branch of study

Control Engineering and Robotics

Czech Technical University in Prague

February 2004

To my mother

Preface

In this research work I propose a new approach to the standard ℓ_1-optimal control problem. I formulated and solved the ℓ_1-optimal control problem using linear equations with polynomials and polynomial matrices, hence the adjective *polynomial.* The core mathematical results of this thesis rely on the theory of Toeplitz operators and their truncations and hopefully this helped create a communication channel between the optimal control theory and a Toeplitz operator theory.

I started this PhD work in the fall 2001, a couple of months after I became a member of a research team lead by Professor Michael Šebek from Czech Technical University in Prague, whose major research domain is development of polynomial methods for control and filter design. This domain was successfully pioneered in the late 1970s by a world-renowned colleague or ours, Professor Vladimír Kučera. The rationale behind choosing this research topic of ℓ_1-optimal control via polynomials for my PhD thesis was quite direct - polynomial methods now present a mature bunch of convenient control design methods featuring reliable solutions to LQG, $\mathcal{H}_2$ and $\mathcal{H}_\infty$-optimal control design problems, but a solution to the ℓ_1-optimal control problem was completely missing in a designer's toolset and we knew nothing about it. But with the great polynomial expertise of my colleagues I felt confident to devote three years of my research career to this topic, even though success was not guaranteed at beginning...

Hopefully, this thesis is an evidence that now, after three years, we know a bit more; we can formulate and solve the ℓ_1-optimal control problem in the polynomial framework, using the same theoretical and computational tools that we are used to from the previous work on LQG, $\mathcal{H}_2$ and $\mathcal{H}_\infty$-optimal control problems. A complete theoretical solution to a square SISO and one-block MIMO cases were proposed, including reliable numerical algorithms. Towards the very end of this thesis we made good progress even in general multiblock MIMO design and derived the missing lower bound on the optimum value of the norm. Nontheless, bringing this procedure to computational maturity requires some more work.

First and foremost, I would like to express my gratitude to my supervisor Michael Šebek for giving me an opportunity to join his research team, for creating perfect conditions for my research, for encouragement and motivation and for introducing me to the world of research.

In the last phase of my PhD research I benefited much from collaboration with Al-

brecht Böttcher from Fakultät für Mathematik, Technische Universität Chemnitz, Germany. In fact, the key ideas of this thesis presented in the two chapters on control synthesis have been described in our two joint papers submitted for publication. My thanks go to Albrecht for helping me with this research and for showing me around in the wonderful world of Toeplitz operators. This short but intensive collaboration was very inspiring and enjoyable. I am happy to see that we are going on steadily these days.

My thanks also come to Didier Henrion, a young enthusiastic researcher from LAAS-CNRS Toulouse, France, for numerous fruitful discussions and encouragement, which helped a lot. I am lucky I can call Didier a friend and a colleague.

I also wish to thank to my colleagues at the Department of Control Engineering at Czech Technical University in Prague, especially to Martin Hromčík, Zdeněk Vlček and Radek Šindelář for making my stay in Prague more enjoyable. And I urge Martin to develop a matrix version of his nice algorithm for a stable-unstable factorization of a polynomial matrix as my further work hinges on it.

My deepest obligation comes to my parents for bringing me up in... well, a normal way...:-) under assistance of my lovely sister Blanka. Last nine years of my life were truly fantastic, happy and peacefull thanks to my beloved fiancée Kateřina, who always encouraged me to go my way.

Finally, financially...:-) This work was supported by the Ministry of Education of the Czech Republic under Project LN00B096.

Zdeněk Hurák

Czech Technical University in Prague
February 2004

ℓ_1 Optimal Control - A Polynomial Approach

Ing. Zdeněk Hurák
Czech Technical University in Prague, 2004

Supervisor: Doc. Ing. Michael Šebek, DrSc.

The thesis brings a new approach to a design of an ℓ_1-optimal feedback controller. A new theoretical and computational framework for solving this problem was developed. It relies on solving linear equations with polynomials and polynomial matrices. Advanced results from the theory of Toeplitz operators were heavily exploited. Alternative proofs for fundamental well-known results like existence of an optimal solution and finiteness of an optimal impulse response of a closed loop in SISO and one-block MIMO cases are given. Fast and reliable numerical algorithms were designed, implemented and tested. They avoid computing zeros and zero directions of polynomial matrices completely. An optimal Youla-Kučera parameter is returned as an outcome from the optimisation and can be used directly to obtain an optimal controller. Hence no need for a numerically tricky extraction of a controller from and optimal closed-loop transfer function. For general multiblock MIMO case, an iterative procedure for finding an approximate solution is proposed. It is based on solving a sequence of linear equations with polynomial matrices and at each iteration provides converging upper and lower bounds on the optimal value of the norm.

Contents

Notations

$a(\lambda)$	Laurent (two-sided) power series $a(\lambda) = \sum_{j=-\infty}^{\infty} a_j \lambda^j \quad (\lambda = e^{i\theta} \in \mathbf{T})$, 10
$a_+(\lambda)$	power series with zero coefficients associated with negative powers, 11
$\mathcal{B}(X)$	Banach algebra of all bounded linear operators mapping X into itself, 10
c_0	a subspace of ℓ_∞ with decaying sequences, 10
$\mathbf{C}$	a set of complex numbers, 10
$\det A$	determinant of matrix A
ℓ_1	a normed linear space of infinite absolutely summable real sequences, 10
ℓ_∞	a normed linear space of infinite sequences of bounded real numbers
$\mathcal{N}(T)$	nullspace of an operator T, 10
P_k	a truncation operator, 13
$\mathcal{R}(T)$	range of an operator T, 10
$T(a)$	Toeplitz operator (infinite matrix) with a symbol $a(\lambda)$, 10
$T_k(a)$	truncated Toeplitz operator (finite matrix) with a symbol $a(\lambda)$, 13
$\mathbf{T}$	a unit circle in the complex place, 10
W	Wiener algebra of functions on the unit circle with coefficients in ℓ_1, 10
W_+	functions in W with zero coefficients with the negative powers of λ, 11
$\langle x, x^* \rangle$	value of a functional x at $x*$ and vice versa, 10

Chapter 1

Introduction

1.1 ℓ_1-optimal control problem: motivation

The objective of ℓ_1-optimal control is, loosely speaking, to minimize worst-case peaks in the amplitudes of regulated variables that are induced by exogenous variables. The only assumption about exogenous variables is that they are persistent and bounded in magnitude. Notice the difference with the popular $\mathcal{H}_\infty$-optimal control, where the objective is to minimize worst-case energy carried by regulated variables, under the assumption of bounded energy of exogenous variables. This boundedness of energy implies that exogenous signals are vanishing, which is not a realistic assumption in many engineering applications.

ℓ_1-optimal control can be regarded just as another control strategy extending control engineers' toolset. It has already proven useful in applications like irrigation channel control [45], where the peaks in the regulated variable - water level - are of uppermost interest.

The objectives of this thesis were set as:

> *Develop a theoretical background and computational tools for the standard ℓ_1-optimal control problem in the polynomial framework.*

The focus of this work is on discrete-time systems. One of the reasons for this choice is the current dominance of computerized control in industrial applications even for continuous-time plants.

1.2 ℓ_1-optimal control problem: history, state of the art

The relationships between ℓ_1, $\mathcal{H}_\infty$ and Hankel norms of linear time-invariant systems were analyzed by Boyd [9]. The proposed bound on ℓ_1 system norm was used a few years later Balakrishnan [1] to compute iteratively the ℓ_1 norm of a discrete-time linear system with arbitrary precision.

The ℓ_1-optimal control problem was formulated and contrasted with the already established $\mathcal{H}_\infty$ control design methodology by Vidyasagar in late 1980s [56]. It was also

Figure 1.1: An irrigation channel control is strongly oriented on attenuation of peaks in the regulated variable.

shown that continuous-time and discrete-time cases need to be treated separately as there is no norm preserving map between the two regions of convergence (right half-plane and unit circle). Solutions to some special instances of the problem were given.

A complete solution to SISO and square MIMO problems (also called one-block problems) was proposed by Dahleh and Pearson in [17]. The primal problem is cast as a linear program with an infinite number of variables subject to a finite number of constraints. The role of these conditions is to prevent unstable zero/pole cancellation. Using standard results on duality between c_0 space of bounded and decaying sequences and ℓ_1 space of absolutely summable sequences [43], the dual problem is formulated with a finite number of variables and infinite number of constraints. It is shown however, that only finitely many constraints are active. Finite-dimensional linear program is thus obtained. Using the alignment property of the optimal solution to the dual problem, it is shown that the optimal closed loop impulse response is finite!

Vidyasagar focuses in [57] on the case where the plant has either poles or zeros on the stability boundary, i.e., the unit circle in the discrete-time case. This is a very important problem in practical control as all position control applications feature a pole at 1. Moreover, discrete-time models of strictly proper plants can have a zero at -1, depending on the discretization method. The duality-based approach proposed by Dahleh breaks down in this case, because the dual solution then lives in a ℓ_∞ space of bounded sequences instead of c_0. Vidyasagar shows that the consequences of having a pole or zeros on the imaginary

Figure 1.2: Literature survey

axis have no parallel in the case of $\mathcal{H}_\infty$ norm minimization, where this also causes troubles. Namely, in both ℓ_1 and $\mathcal{H}_\infty$ minimum distance problems, an optimal solution solution need not exist. But it is always possible to construct a sequence of $\mathcal{H}_\infty$-suboptimal controllers whose performance approaches the unattainable infimum (see e.g. [55], sec. 6.4). However, this is generally not possible in ℓ_1 case. It can be shown that the limit of a sequence of ℓ_1-suboptimal controllers can be strictly larger then the infimum. The well-known trick with weighting filters vanishing at the poles or zeros of the plant on the unit circle cannot be applied either. Vast majority of papers on ℓ_1-optimal control ignores this hot problem, which limits usability of the whole control strategy.

Meyer gives two examples [47] demonstrating that an optimal solution need not be unique and that the length of the optimal closed loop impulse response is not bounded by any function of the order n of the plant (this is in contrast with the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ problems). Indeed, changing the coefficients of a plant transfer function of a given order, the order of and optimal controller can be set arbitrarily high.

In order to relieve the high-order controller curse of ℓ_1-optimal control, Halpern gives a heuristic technique in [25]. He shows that if the order of the controller is to be less than that required for the optimal finite-impulse-response solution, a lower value of $\|.\|_1$ can be achieved if one places one of the closed loop poles somewhere in the interval $(1, \infty)$. An analytical expression is then given for the unique optimal position of this pole. Although this expression is usually nonconvex, Halpern's observation that one should give up dead-beat closed loop response when designing a ℓ_1-suboptimal control is inspiring.

A partial solution to the inverse problem of the ℓ_1-optimal control is given in [19] for a class of problems with interpolation points located in some specific region. It is also shown that every controller minimizing the $\mathcal{H}_\infty$-norm or the weighted sensitivity function is always ℓ_1-optimal for a possibly different stable weight. There are however ℓ_1-optimal controllers that are not $\mathcal{H}_\infty$-optimal.

The original solution proposed by Dahleh for SISO and one-block MIMO case was immediately extended to the general four-block case [18]. Dahleh considers a simple case with two exogenous inputs and one measured output, two regulated variables and one control signal to show that both primal and dual problems have an infinite number of constraint and variables. The reason is that besides the *zero interpolation* constraints a new type of constraints called *rank interpolation* conditions appear. Therefore it is not possible to compute the solution precisely. An optimal impulse response is no longer finite. It is then necessary to approximate an optimal impulse response by a solution to a truncated problem with only finitely many nonzero entries (in the primal domain). The sequence of optimal solutions to truncated problems gives a converging upper bound on the optimal cost to the original minimization problem. This approach is later named Finitely Many Variables (FMV) Method.

McDonald and Pearson [46] made the proofs from [18] mathematically more direct using coprime factorization and removed some minor technical assumptions. As a special case, they also considered a design of a controller under constraints on the norm with respect to some of the outputs outputs.

A missing lower bound on the optimal cost was obtained independently by Staffans [52] and Dahleh [14] by solving a truncated dual problem (with finite number of dual variables, thus finite number of primal constraints). This was named Finitely Many Equations (FME) Method. An iterative scheme FMV/FME in which two finite linear programs are solved at each truncation is proposed. In the same paper, using a particular numerical example of a scalar mixed sensitivity, Staffans explored in great detail the issue of redundancy in the corresponding linear programs. He pointed out that although the iterative truncation-based solution suggested an infinite-dimensional optimal solution, the actual solution was rational and of low order. This articulated the bad property of truncation-based approaches. Further generalizations of his conclusions to a MIMO case can be found in [53] and [54].

Conceptually different solution to a general multiblock problem was proposed by Diaz-Bobillo [20]. It was named Delay Augmentation (DA) Method because it converts the multiblock problem into a one-block problem by introducing delays (right-shifts). It attempts at reducing order inflation caused by straightforward truncation as in FMV/FME and reduces some computational burdes because only one linear program is solved at each iteration. On the other hand, additional effort is required for reordering the inputs and outputs of the system because this has significant impacts on convergence. Both lower and upper bounds converging to the optimum are provided too.

Robust stability with norm-bounded uncertainty was studied by Dahleh in [16] for unstructured uncertainty (multiplicative and additive perturbations) and in [14] for coprime stable factor perturbations. These results have been generalized by Khammash [37] for structured uncertainties.

Khammash extended these results for robust stability with unstructured and structured uncertainty and provided necessary and sufficient conditions on robust performance, including numerical synthesis procedures, in [36] and [38], respectively. In [38] he even relaxes the requirement of time-invariance of the linear plant and provides an iteration procedure of D-K type known from μ-synthesis. However, as in the case of Structured Singular Values (SSV) such procedure does not guarantee that a global optimum has been found as the problem is inherently nonconvex. In [39] a new procedure for finding a globally optimal solution was proposed that is based on a linear relaxation of a nonconvex infinite dimensional problem.

Shamma [51] addresses the question whether any improvement in disturbance rejection or robust stabilization of a plant with norm-bounded uncertainty can be achieved by using time-varying controller. The answer is, perhaps surprisingly, no, no improvement can be expected when the uncertainty is unstructured by using time-varying controller.

The major message from the work on robust stability and robust performance against norm-bounded uncertainty described above is that a lot of theoretical results from the $\mathcal{H}_\infty$-optimal control framework can be accepted in ℓ_1-optimal control framework because of validity of the celebrated small gain theorem.

Most of the above mentioned results have been obtained in the late 1980s and early 1990s and are perfectly documented in a comprehensive textbook by Dahleh and Diaz-Bobillo [15]. A major feature of these methods is that both the derivation of theoretical properties and the actual numerical computation are based on the numerically tricky interpolation, that features solution to ill-conditioned Vandermonde linear system. There are other two sources of numerical troubles. First, the computation of zeros and the so-called zero directions of polynomial matrices and, perhaps even more importantly, the extraction of an optimal controller from the closed loop-impulse response. The latter being magnified especially if one achieves some suboptimal solution only.

Late 1990s in research in ℓ_1-optimal control can be characterized by attempts to reformulate the standard problem of ℓ_1-optimal control and obviate the interpolation part. A distinguished new approach that avoids zero interpolation was proposed by Khammash in [34] and [35] and is called Q-scaled Method. In fact, this method takes some inspiration from the convex Q-parameterization relying on Ritz approximations presented in [10]. The method directly approximates the optimal Youla-Kučera parameter $Q(\lambda)$ that determines uniquely the closed loop transfer function. Khammash solves an auxiliary (regularized) problem that includes a scaled norm of $Q(\lambda)$ in the objective function, provides lower and upper bounds, and then relates the result to the solution to the original problem. Optimizing directly over $Q(\lambda)$ brings an advantage that the numerically tricky extraction of a controller from an optimal closed-loop transfer function was avoided. Moreover, although the order of the finite impulse response $Q(\lambda)$ can be high, a rational approximation of low order can be conveniently found using system identification tools. The error introduced by this approximation can be easily computed.

Another method that can do without interpolation was authored by Elia and Dahleh [21]. Their method is based on mixed-objective minimization and provides a converging lower and upper bounds on the optimal ℓ_1 norm and therefore can provide a suboptimal controller. Instead of using linear program, a semidefinite quadratic program is solved at each step. An essence of this method is that at a given step, instead of minimizing the ℓ_1 norm of the closed loop impulse response, the square of the ℓ_1 norm of the first N samples plus the square of the ℓ_2 norm of the tail is minimized. This second term amounts to solving the $\mathcal{H}_2$-optimal design first and then solving a finite-dimensional convex optimization problem. Increasing N the closed-loop behaves more like ℓ_1-optimal then $\mathcal{H}_2$-optimal. A nice feature is that an optimal controller is directly computed and need not be extracted from the optimal closed-loop transfer function.

A state-space solution to the ℓ_1-optimal control problem when all the states are available and the peak magnitudes of disturbances are known is proposed by Elia [22]. The procedure relies heavily on dynamic programming. The paper also includes important references for other state-spaces flavoured results.

A polynomial approach to the ℓ_1-optimal control has been recently proposed by Cassavola [13]. Some preliminary results on scalar mixed-sensitivity problem and scalar multiblock problem were published earlier in [11] and [12], respectively. In this method, both the closed-loop transfer function and Youla-Kučera parameter $Q(\lambda)$ are parameterized by some free term, resulting in a sequence of unconstrained and redundancy-free linear programs. This approach and Khammash's Q-scaled method share the same angle of attack with this thesis as they both consider the Youla-Kučera parameter a component of the optimization variable.

1.3 Contributions of the thesis

In accordance with the stated objectives, this thesis brings several concrete contributions

1. A major contribution of the thesis is a rigorous mathematical formulation and numerical solution to an ℓ_1-optimal control problem within the so-called polynomial framework. Advanced results from the theory of Toeplitz operators were heavily exploited. Alternative proofs for fundamental well-known results like existence of an optimal solution and finiteness of an optimal impulse response of a closed loop in SISO an one-block MIMO cases are given. Fast and reliable numerical algorithms were designed, implemented and tested. They avoid computing zeros and zero directions of polynomial matrices completely. An optimal Youla-Kučera parameter is returned as an outcome from the optimisation and can be used directly to obtain an optimal controller. Hence no need for a numerically tricky extraction of a controller from and optimal closed-loop transfer function. For general multiblock MIMO case, an iterative procedure for finding an approximate solution is proposed. It is based on solving a sequence of linear equations with polynomial matrices and at each iteration provides converging upper and lower bounds on the optimal value of the norm.

2. The core mathematical results developed in this thesis for ℓ_1-optimal controller design are of independent results and contribute to the general theory of linear equations in

the ring of (matrix) Wiener functions (power series with absolutely summable coefficients) and theory of banded (block) Toeplitz operators.

3. A numerical algorithm for computing the ℓ_∞-induced system norm of a polynomial matrix fraction was devised.

4. As a side-product of the work on algorithm for computing the norm, fast and realiable algorithms were developed for modular shift and modular multiplication of polynomial matrices.

5. All the numerical algorithms proposed in this thesis were implemented and tested in Matlab and are considered for the next release of a commercial *Polynomial Toolbox for Matlab* [42].

1.4 Author's publications related to the thesis

The major result of this thesis - a solution to the SISO version of ℓ_1-optimal control design via Diophantine equations - was submitted for publication in *SIAM Journal on Control and Optimization* [29] at December 2003. A preliminary, very simple version was presented at the *4th IFAC Symposium on Robust Control (ROCOND'03)* [31] in Milan, June 2003. The very recent results on design for MIMO systems were submitted to the *16th International Symposium on Mathematical Theory of Networks and Systems (MTNS'04)* [28] at Katholieke Universiteit Leuven, Leuven, Belgium, July 2004. The work on algorithm for the norm computation was presented at the *IEEE CCA/CACSD conference* [30] in Glasgow, September 2002. The result on modular arithmetics for polynomial matrices was submitted to *Systems and Control Letters* in June 2003 [32]. An invited talk on truncated Toeplitz operators and their use in optimal control was given to the members of the *Fakultät für Mathematik* at *Technische Universität Chemnitz* in December 2003.

1.5 Outline of the thesis

The immediately following, second chapter, contains the major contribution of the thesis: a rigorous solution to an ℓ_1-optimal control problem for a SISO plant, including a numerical example. The third chapter contains an extension to MIMO case. The fourth chapter suggests an algorithm for computing ℓ_∞-induced norm of a polynomial matrix fraction. The fifth chapter presents a new algorithm for modular shift of a polynomial matrix. Even though this result is not directly tied to ℓ_1-optimal control, it can provide an computational efficient step for the norm computation. The sixth, final chapter summarizes the achievements of this thesis and outlines immediate opportunities for improvement and further research.

Chapter 2

SISO feedback ℓ_1-optimal control

2.1 Introduction

This chapter concerns the problem of finding a discrete-time controller that minimizes the ℓ_1 norm of the impulse response of a closed-loop system that has one exogenous input and one regulated variable. A plant is described by a transfer function and the proposed design method relies on manipulation with coefficients of the numerator and denominator polynomials. It avoids computing their roots completely. This is in contrast with the existing interpolation-based approach ([17], or [15]). The presented work goes much in the *polynomial* spirit of Kučera's pioneering work [40], hence the name polynomial approach.

Basically, the proposed method relies on solving a linear equation in the ring of power series with finite sum of absolute values of the coefficients $a(\lambda)x(\lambda)+b(\lambda)y(\lambda) = c(\lambda)$, where $a(\lambda)$ is a given polynomial, $b(\lambda) = 1$ and $c(\lambda)$ is an arbitrary power series with bounded sum of absolute values of its coefficients. An optimal solution is sought that minimizes ℓ_1 norm of the coefficients sequence of $y(\lambda)$. It is shown in this chapter that such solution is guranteed to exist if and only if the polynomial $a(\lambda)$ has no zeros on the unit circle and that this solution has only finite number of nonzero terms, i.e., an optimal $y(\lambda)$ a polynomial!

Avoiding interpolation was the major incentive for this work. The presented approach follows the line of reasoning pursued by Dahleh and Pearson [17], but the problem is posed in a different setting, which leads to a different numerical algorithm. The problem is formulated mathematically as finding the distance between a given sequence in ℓ_1 and the range of a given infinite lower-triangular Toeplitz band matrix on ℓ_1. We establish necessary and sufficient conditions for the solvability of the problem and design a numerical algorithm for finding a solution. The convergence of this algorithm is rigorously proved. In contrast to previous work, which used interpolation techniques relying on the solution to numerically ill-conditioned Vandermonde systems, our algorithm is based on finding optimal solutions of overdetermined but numerically much better behaved Toeplitz systems.

2.2 The ℓ_1-optimal control problem

The objective of ℓ_1-optimal control is to design a feedback controller that guarantees minimum worst-case peaks in the regulated (error) variable in response to a bounded and persistent disturbance. Consider a standard feedback configuration with the exogenous variable corrupting the output of the plant as in Figure 2.1. The plant is described by a transfer function, say $G(\lambda) = p(\lambda)/q(\lambda)$.

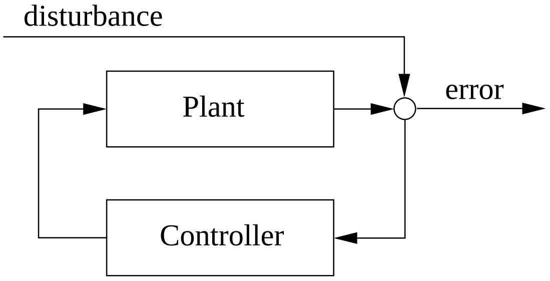

Figure 2.1: SISO disturbance rejection

It is well known [40] that the achievable internally stable closed-loop transfer functions $y(\lambda)$ of a standard feedback connection are parameterized by $y(\lambda) = b(\lambda) - a(\lambda)x(\lambda)$, where all the terms in the equation are power series with coefficient sequences residing in ℓ_1; $a(\lambda)$ and $b(\lambda)$ are derived from the description of the plant, and $x(\lambda)$, which is also called the Youla-Kučera parameter, is unknown and is to be determined. For finite-dimensional systems, $a(\lambda)$ is actually a polynomial.

Given a linear time-invariant (LTI) model of a plant, a principal task of ℓ_1-optimal control is to design a stabilizing feedback LTI controller that minimizes the ℓ_1 norm of the coefficient sequence of $y(\lambda)$. This controller will guarantee optimal attenuation of peaks in the amplitude of the error signal because the Wiener norm of $y(\lambda)$, that is ℓ_1 norm of its coefficient sequence, is equal to the ℓ_∞-induced operator norm[1] of the closed-loop system [56].

Any stabilizing controller is determined by its Youla-Kučera parameter $x(\lambda)$. As the coefficient sequence of $a(\lambda)x(\lambda)$ results from that of $x(\lambda)$ by the action of an infinite lower-triangular Toeplitz band matrix, the problem of designing an optimal controller leads to the minimum distance problem between a given sequence $b \in \ell_1$ and the range $\mathcal{R}(T(a))$ of the infinite lower-triangular Toeplitz matrix $T(a)$ in ℓ_1. A concrete design procedure will be exemplified in the next section.

A noteworthy feature of the approach proposed here is that the Youla-Kučera parameter is explicitly computed and that, consequently, there is no need to perform the numerically tricky extraction of a controller from the optimal closed-loop transfer function.

[1] This standard fact and especially its extension to operators on vector sequences is restated for a reader's convenience in the introductory section in Chapter 4, in particular in Lemma 2 on page 39.

2.3 Operator-theoretic statement of the problem

The Wiener algebra W is the Banach algebra of all complex-valued functions on the complex unit circle $\mathbf{T}$ with absolutely convergent Fourier series. Thus, a function $a : \mathbf{T} \to \mathbf{C}$ belongs to W if and only if

$$a(\lambda) = \sum_{j=-\infty}^{\infty} a_j \lambda^j \quad (\lambda = e^{i\theta} \in \mathbf{T}), \quad \|a\|_W := \sum_{j=-\infty}^{\infty} |a_j| < \infty.$$

Wiener's theorem (e.g. [7], page 6) states that if $a \in W$ has no zeros on $\mathbf{T}$, then $1/a$ is also in W. For $a \in W$, the infinite Toeplitz matrix $T(a)$ and the finite Toeplitz matrices $T_k(a)$ are defined by $T(a) = (a_{i-j})_{i,j=1}^{\infty}$ and $T_k(a) = (a_{i-j})_{i,j=1}^{k}$, respectively.

We denote by c_0 the set of all real-valued sequences $x = \{x_j\}_{j=1}^{\infty}$ with $|x_j| \to 0$ as $j \to \infty$ and by ℓ_1 the set of all real-valued sequences $x = \{x_j\}_{j=1}^{\infty}$ satisfying $\sum_{j=1}^{\infty} |x_j| < \infty$. The sets c_0 and ℓ_1 are real Banach spaces under the norms $\|x\|_\infty = \sup_{j \geq 1} |x_j|$ and $\|x\|_1 = \sum_{j=1}^{\infty} |x_j|$, respectively. Moreover, ℓ_1 is the dual space of c_0, $\ell_1 = c_0^*$, under the pairing $\langle z, b \rangle = \sum_{j=1}^{\infty} z_j b_j$, $\{z_j\} \in c_0$, $\{b_j\} \in \ell_1$.

Given a Banach space X, we let $\mathcal{B}(X)$ stand for the Banach algebra of all bounded linear operators on X. For $A \in \mathcal{B}(X)$, the norm $\|A\|_{\mathcal{B}(X)}$ is $\sup \|Ax\|$, the supremum over all $x \in X$ with $\|x\| \leq 1$, and the range and null space of A are defined by $\mathcal{R}(A) = A(X)$ and $N(A) = \{x \in X : Ax = 0\}$.

Let $a \in W$ and suppose the Fourier coefficients of a are all real. Then the infinite Toeplitz matrix $T(a)$ induces a bounded linear operator on c_0 and ℓ_1 via

$$(T(a)x)_i = \sum_{j=1}^{\infty} a_{i-j} x_j \quad (j \geq 1).$$

For $j \in \mathbf{Z}$, define the function χ_j by $\chi_j(\lambda) = \lambda^j$ $(\lambda \in \mathbf{T})$. The operators $T(\chi_1)$ and $T(\chi_{-1})$ are the forward and backward shifts acting by the rules

$$\begin{aligned} T(\chi_1) &: \{x_1, x_2, x_3, \ldots\} \mapsto \{0, x_1, x_2, \ldots\}, \\ T(\chi_{-1}) &: \{x_1, x_2, x_3, \ldots\} \mapsto \{x_2, x_3, x_4, \ldots\}. \end{aligned}$$

Clearly, $T(a) = \sum_{j=-\infty}^{\infty} a_j T(\chi_j)$. This implies that $\|T(a)\|_{\mathcal{B}(c_0)} = \|T(a)\|_{\mathcal{B}(\ell_1)} = \|a\|_W$. The adjoint operator of $T(a) : c_0 \to c_0$ is the operator $T(\overline{a}) : \ell_1 \to \ell_1$ where $\overline{a}(\lambda) = \sum_{j=-\infty}^{\infty} a_j \lambda^{-j}$ $(\lambda = e^{i\theta} \in \mathbf{T})$.

Throughout this work we suppose that $a_+(\lambda) = a_0 + a_1\lambda + \ldots + a_n\lambda^n$ with real numbers $a_0, a_1, \ldots, a_n$ and with $a_n \neq 0$. Clearly, $T(a_+)$ is a banded lower-triangular Toeplitz matrix, while $T(\overline{a}_+)$ is a banded upper-triangular Toeplitz matrix. We always think of $T(a_+)$ as acting on ℓ_1 and always consider $T(\overline{a}_+)$ as an operator on c_0. Thus, $T(a_+)$ is the adjoint of $T(\overline{a}_+)$.

This work concerns the following problem. Given $b \in \ell_1$, determine the distance

$$d := \operatorname{dist}_{\ell_1}(b, \mathcal{R}(T(a_+))) := \inf_{m \in \mathcal{R}(T(a_+))} \|b - m\|_1,$$

find out whether there is an $m_0 \in \mathcal{R}(T(a_+))$ with $\|b - m_0\|_1 = d$, and if yes, compute such an m_0. Note that once m_0 is available, we can easily solve the lower-triangular system $T(a_+)x_0 = m_0$ to get x_0.

2.4 Two results from functional analysis

We will employ the following two theorems (whose proofs can be found on pages 121 and 156 of [43]). Recall that the annihilator $M^\perp$ of a set $M \subset X$ is defined as $M^\perp = \{b \in X^* : \langle z, b\rangle = 0 \text{ for all } z \text{ in } M\}$. Furthermore, two elements $z \in X$ and $b \in X^*$ are said to be aligned if the equality $\|z\|\,\|b\| = \langle z, b\rangle$ holds.

Theorem 1. *Let M be a linear subset of a real normed space X and let $b \in X^*$. Then*

$$\inf_{m \in M^\perp} \|b - m\| = \sup_{z \in M, \|z\| \le 1} \langle z, b\rangle. \tag{2.1}$$

The infimum in (2.1) is always attained at some $m_0 \in M^\perp$. If the supremum in (2.1) is achieved for some $z_0 \in M$ with $\|z_0\| \le 1$, then z_0 and $b - m_0$ are aligned.

Theorem 2. *Let X be a Banach space and $A \in \mathcal{B}(X)$. Then $\mathcal{R}(A)$ is closed if and only if $\mathcal{R}(A^*)$ is closed, in which case $\mathcal{R}(A^*) = [N(A)]^\perp$.*

2.5 Toeplitz operators

In general, the product of two Toeplitz operators is not a Toeplitz operator. However, this happens in certain special cases. Let W_+ and W_- denote the functions in W whose Fourier coefficients with negative and positive indices vanish, respectively. Thus, if $c_\pm \in W_\pm$, then $T(c_-)$ and $T(c_+)$ are upper and lower triangular, respectively. It is easily seen by direct inspection that if $c_- \in W_-$, $f \in W$, $c_+ \in W_+$, then

$$T(c_-)T(f)T(c_+) = T(c_- f c_+). \tag{2.2}$$

The following results are known to specialists (see, e.g., [8] and [24]). We include the proofs for the reader's convenience.

Proposition 1. *The range $\mathcal{R}(T(a_+))$ is a closed subset of ℓ_1 if and only if a_+ has no zeros on $\mathbf{T}$.*

Proof. If a_+ has no zeros on $\mathbf{T}$, then a_+^{-1} belongs to W and has real Fourier coefficients. From (2.2) we obtain that $T(a_+^{-1})T(a_+) = I$. Thus, $T(a_+)$ has a bounded left inverse, which implies that the range of $T(a_+)$ is closed (see, e.g., [24, Section I.1.2]).

Now suppose $a_+(\tau) = 0$ for some $\tau \in \mathbf{T}$. Contrary to what we want, we assume that $\mathcal{R}(T(a_+))$ is a closed subset of ℓ_1. We denote by $\ell_1(\mathbf{C})$ the complex Banach space of all complex-valued sequences $x = \{x_j\}_{j=1}^\infty$ for which $\|x\|_1 = \sum_{j=1}^\infty |x_j| < \infty$. The range of $T(a_+)$ on $\ell_1(\mathbf{C})$ is $\mathcal{R}(T(a_+)) + i\mathcal{R}(T(a_+))$, which is closed whenever $\mathcal{R}(T(a_+))$ is closed.

From Theorem 2 we now infer that $T(\overline{a}_+) : c_0(\mathbf{C}) \to c_0(\mathbf{C})$ has closed range, where $c_0(\mathbf{C})$ is defined in analogy to $\ell_1(\mathbf{C})$. The operator $T(\overline{a}_+)$ is upper-triangular, and it is easily seen that the range of every nonzero upper-triangular Toeplitz operator contains all finitely supported sequences. Consequently, $T(\overline{a}_+)$ must be surjective. We may write

$$\begin{aligned}\overline{a}_+(\lambda) &= a(1/\lambda) = a_0 + a_1\frac{1}{\lambda} + \ldots + a_n\frac{1}{\lambda^n}\\ &= a_n\left(\frac{1}{\lambda} - \tau\right)\left(\frac{1}{\lambda} - z_1\right)\ldots\left(\frac{1}{\lambda} - z_{n-1}\right) = a_n(\chi_{-1}(\lambda) - \tau)d(\lambda).\end{aligned}$$

Since $T(\overline{a}_+) = T(\chi_{-1} - \tau)T(d)$ by (2.2), the operator $T(\chi_{-1} - \tau)$ is surjective together with $T(\overline{a}_+)$. The equation $T(\chi_{-1} - \tau)z = 0$ is satisfied if and only if $z_j = \tau^{j-1}z_1$ $(j \geq 1)$, and this is a sequence in $c_0(\mathbf{C})$ only for $z_1 = 0$. Thus, $T(\overline{a}_+)$ is injective on $c_0(\mathbf{C})$. In summary, $T(\chi_{-1} - \tau)$ is invertible on $c_0(\mathbf{C})$. It follows that $T(\chi_1 - 1/\tau)$ is invertible on $\ell_1(\mathbf{C})$. But the solution to $T(\chi_1 - 1/\tau)x = \{1, 0, 0, \ldots\}$ is $x_j = -\tau^j$ $(j \geq 1)$, which is not in $\ell_1(\mathbf{C})$. This contradiction proves that $\mathcal{R}(T(a_+))$ cannot be closed. □

The function $a_+(\lambda) = a_0 + a_1\lambda + \ldots + a_n\lambda^n$ is defined for all $\lambda \in \mathbf{C}$.

Proposition 2. *If a_+ has no zeros on $\mathbf{T}$, then the dimension of $N(T(\overline{a}_+))$ in c_0 is equal to the number of zeros of a_+ in the open unit disk $\mathbf{D} := \{\lambda \in \mathbf{C} : |\lambda| < 1\}$.*

Proof. Let a_+ have $\varkappa$ zeros $\delta_1, \ldots, \delta_\varkappa$ in $\mathbf{D}$ and $n - \varkappa$ zeros $\mu_1, \ldots, \mu_{n-\varkappa}$ in $\mathbf{C} \setminus (\mathbf{D} \cup \mathbf{T})$. We then have

$$\begin{aligned}\overline{a}_+(\lambda) &= a_+(1/\lambda) = a_n \prod_{k=1}^{n-\varkappa}\left(\frac{1}{\lambda} - \mu_k\right)\prod_{j=1}^{\varkappa}\left(\frac{1}{\lambda} - \delta_j\right)\\ &= \gamma\lambda^{-\varkappa}\prod_{k=1}^{n-\varkappa}\left(1 - \frac{1}{\mu_k\lambda}\right)\prod_{j=1}^{\varkappa}(1 - \delta_j\lambda)\end{aligned}$$

with $\gamma = a_n(-\mu_1)\ldots(-\mu_{n-\varkappa})$. We consider $T(\overline{a}_+)$ on $c_0(\mathbf{C})$. Let N be the null space of $T(\overline{a}_+)$ on c_0. Then $N + iN$ is the null space of $T(\overline{a}_+)$ on $c_0(\mathbf{C})$. From (2.2) we obtain that

$$T(\overline{a}_+) = \gamma T(\chi_{-\varkappa})\prod_{k=1}^{n-\varkappa}\left(I - \frac{1}{\mu_k}T(\chi_{-1})\right)\prod_{j=1}^{\varkappa}\left(I - \delta_j T(\chi_1)\right).$$

Since $\|(1/\mu_k)T(\chi_{-1})\| = 1/|\mu_k| < 1$ and $\|\delta_j T(\chi_1)\| = |\delta_j| < 1$, we conclude that the operators $I - (1/\mu_k)T(\chi_{-1})$ and $I - \delta_j T(\chi_1)$ are all invertible. Consequently, the dimension of $N + iN$ is the dimension of the null space of $T(\chi_{-\varkappa})$ on $c_0(\mathbf{C})$. It follows that the dimension of $N + iN$ over $\mathbf{C}$ is $\varkappa$, which implies that the dimension of N over $\mathbf{R}$ is also $\varkappa$. □

2.6 Existence of the solution

Here is our result on the solvability of the problem posed in Section 2.3.

Theorem 3. *The problem*

$$\|b - m\|_1 = \operatorname{dist}_{\ell_1}(b, \mathcal{R}(T(a_+))) =: d \tag{2.3}$$

has a solution $m_0 \in \mathcal{R}(T(a_+))$ for every $b \in \ell_1$ if and only if a_+ has no zeros on $\mathbf{T}$. If $a_+(\lambda) \neq 0$ for $\lambda \in \mathbf{T}$, then for every $b \in \ell_1$ there exists a $z_0 \in N(T(\overline{a}_+))$ such that

$$\|z_0\|_\infty \leq 1 \quad \textit{and} \quad d = \langle z_0, b\rangle = \sup_{z \in N(T(\overline{a}_+)), \|z\|_\infty \leq 1} \langle z, b\rangle, \tag{2.4}$$

and if $m_0 \in \mathcal{R}(T(a_+))$ is any sequence satisfying (2.3), then the sequence $b - m_0$ has only finitely many nonzero terms.

Proof. If a_+ has a zero on $\mathbf{T}$, then $\mathcal{R}(T(a_+))$ is not closed due to Proposition 1 and hence (2.3) has no solution $m_0 \in \mathcal{R}(T(a_+))$ if b is in the closure of $\mathcal{R}(T(a_+))$ but not in $\mathcal{R}(T(a_+))$.

Now suppose that a_+ has no zeros on $\mathbf{T}$. Then $\mathcal{R}(T(a_+))$ is closed by Proposition 1. From Theorem 2 we deduce that $\mathcal{R}(T(a_+)) = [N(T(\overline{a}_+))]^\perp$. The existence of an $m_0 \in \mathcal{R}(T(a_+))$ satisfying (2.3) then follows from Theorem 1. This theorem also yields the equality

$$d = \sup_{z \in N(T(\overline{a}_+)), \|z\|_\infty \leq 1} \langle z, b\rangle,$$

and since $\{z \in N(T(\overline{a}_+)) : \|z\|_\infty \leq 1\}$ is compact by virtue of Proposition 2 and the map $z \mapsto \langle z, b\rangle$ is continuous, we conclude that the supremum is attained at some $z_0 \in N(T(\overline{a}_+))$ with $\|z_0\|_\infty \leq 1$.

The last assertion of the theorem is trivial for $d = 0$. So let $d > 0$, which implies that $\|z_0\|_\infty > 0$. The sequences $b - m_0$ and z_0 are aligned by Theorem 1. Consequently, with $b - m_0 = \{e_j\}_{j=1}^\infty$ and $z_0 = \{z_j\}_{j=1}^\infty$,

$$\sum_{j=1}^\infty z_j e_j = \|z_0\|_\infty \sum_{j=1}^\infty |e_j|. \tag{2.5}$$

As $\{z_j\} \in c_0$, there is a j_0 such that $|z_j| < \|z_0\|_\infty$ for all $j \geq j_0$. From (2.5) we infer that $e_j = 0$ for $j \geq j_0$. □

2.7 Finite sections of Toeplitz operators

In this section, we quote two known theorems that will be needed when proving the convergence of and giving an error estimate for our numerical algorithm. For $k \geq 1$, we denote by P_k the projection on ℓ_1 and c_0 that acts by the rule

$$P_k : \{x_1, x_2, x_3, \ldots\} \mapsto \{x_1, \ldots, x_k, 0, 0, \ldots\}.$$

We identify $\mathcal{R}(P_k)$ and $\mathbf{R}^k$, and hence we may think of vectors in $\mathbf{R}^k$ as elements of ℓ_1 or c_0. The following theorem was established by Reich [49] and Baxter [2]. Full proofs are also in [5, Section 3.3] and [24, Section II.2].

Theorem 4. *If $f \in W$ and $T(f)$ is invertible on ℓ_1, then the matrices $T_k(f)$ are invertible for all sufficiently large k and $T_k^{-1}(f)P_k y$ converges in ℓ_1 to $T^{-1}(f)y$ for every $y \in \ell_1$.*

The next theorem can be proved using the asymptotic inverses presented in [5, Section 3.5] or [7, Section 2.3].

Theorem 5. *Let f be a Laurent polynomial, that is, suppose f has only finitely many nonzero Fourier coefficients, and let $T(f)$ be invertible on ℓ_1. Fix a natural number $\varkappa$. Then there exist a natural number k_0 and constants $\alpha > 0$ and $C < \infty$ such that*

$$\|P_\varkappa T_k^{-1}(f) - P_\varkappa T^{-1}(f)\|_{\mathcal{B}(\ell_1)} = \|T_k^{-1}(\overline{f})P_\varkappa - T^{-1}(\overline{f})P_\varkappa\|_{\mathcal{B}(c_0)} \leq C\, e^{-\alpha k}$$

for all $k \geq k_0$.

2.8 Numerical algorithm for the minimum distance problem

Fix $b \in \ell_1$ and a_+ as above. Suppose a_+ has exactly $\varkappa$ zeros in $\mathbf{D}$ and no zeros on $\mathbf{T}$. If $k \geq \varkappa + 1$, the operator $P_k T(a_+) P_{k-\varkappa}$ may be identified with a $k \times (k - \varkappa)$ matrix. The system $P_k T(a_+) P_{k-\varkappa} x^{(k)} = P_k b$ is overdetermined for $\varkappa \geq 1$. However, we can find an $x_0^{(k)} \in \mathbf{R}^{k-\varkappa}$ such that the residue

$$\|P_k T(a_+) P_{k-\varkappa} x^{(k)} - P_k b\|_1 \tag{2.6}$$

assumes its minimum at $x^{(k)} = x_0^{(k)}$. Let $d = \mathrm{dist}_{\ell_1}(b, \mathcal{R}(T(a_+)))$ and let d_k be the minimal value of (2.6). The following theorem reveals that d_k converges to d exponentially fast.

Theorem 6. *There are constants $E < \infty$ and $\beta > 0$ such that $|d_k - d| \leq E\, e^{-\beta k}$ for all $k \geq 1$.*

Proof. Put $f(\lambda) = \lambda^{-\varkappa} a_+(\lambda) = \lambda^{-\varkappa}(a_0 + a_1\lambda + \ldots + a_n\lambda^n)$. We claim that $T(f)$ is invertible on ℓ_1. Indeed, the proof of Proposition 2 shows that

$$T(\overline{f}) = \gamma \prod_{k=1}^{n-\varkappa} \left(I - \frac{1}{\mu_k} T(\chi_{-1}) \right) \prod_{j=1}^{\varkappa} (I - \delta_j T(\chi_1))$$

with all operators on the right being invertible on $c_0(\mathbf{C})$. It follows that $T(\overline{f})$ is invertible on c_0 and hence that $T(f)$ is invertible on ℓ_1.

From (2.2) we deduce that $T(a_+) = T(f)T(\chi_\varkappa)$. Let $x^{(k)} = \{x_1^{(k)}, \ldots, x_{k-\varkappa}^{(k)}, 0, \ldots\}$ and define $w^{(k)} \in \mathcal{R}(P_k)$ by

$$w^{(k)} = \{\underbrace{0, \ldots, 0}_{\varkappa}, x_1^{(k)}, \ldots, x_{k-\varkappa}^{(k)}, 0, \ldots\}.$$

Let Q_k be given on ℓ_1 and c_0 by $Q_k = I - P_k$, that is,

$$Q_k : \{x_1, x_2, x_3, \ldots\} \mapsto \{0, \ldots, 0, x_{k+1}, x_{k+2}, \ldots\}.$$

We have $P_{k-\varkappa}x^{(k)} = T(\chi_{-\varkappa})P_k w^{(k)}$, and since $T(\chi_{-\varkappa})T(\chi_{\varkappa}) = Q_k$ and $Q_\varkappa P_k = P_k Q_\varkappa$, we get

$$\begin{aligned}&\|P_k b - P_k T(a_+)P_{k-\varkappa}x^{(k)}\|_1 = \|P_k b - P_k T(f)T(\chi_\varkappa)T(\chi_{-\varkappa})P_k w^{(k)}\|_1 \\ &= \|P_k b - P_k T(f)P_k Q_\varkappa w^{(k)}\|_1 = \|P_k b - T_k(f)Q_\varkappa w^{(k)}\|_1.\end{aligned} \tag{2.7}$$

The minimum of (2.7) as $w^{(k)}$ ranges over $\mathbf{R}^k$ is d_k, and the minimum is attained at the $w_0^{(k)}$ corresponding to any $x_0^{(k)}$ that minimizes (2.6). Hence, by Theorems 1 and 2,

$$d_k = \sup_{z\in N(Q_\varkappa T_k(\overline{f})),\|z\|_\infty\le 1} \langle z, P_k b\rangle. \tag{2.8}$$

Theorem 4 implies that there is a k_0 such that the matrices $T_k(f)$ are invertible for all $k \ge k_0$. Let $k \ge k_0$. We have $Q_\varkappa T_k(\overline{f})z = 0$ if and only if there is a $y \in \mathbf{R}^\varkappa$ such that $T_k(\overline{f})z = P_\varkappa y$ or, equivalently, $z = T_k^{-1}(\overline{f})P_\varkappa y$ (note that $T_k(\overline{f})$ is simply the transpose of $T_k(f)$). From (2.8) we therefore obtain

$$\begin{aligned} d_k &= \sup_{z=T_k^{-1}(\overline{f})P_\varkappa y,\|z\|_\infty\le 1} \langle z, P_k b\rangle \\ &= \sup_{\|T_k^{-1}(\overline{f})P_\varkappa y\|_\infty\le 1} \langle T_k^{-1}(\overline{f})P_\varkappa y, P_k b\rangle \\ &= \sup_{\|T_k^{-1}(\overline{f})P_\varkappa y\|_\infty\le 1} \langle P_\varkappa y, P_\varkappa T_k^{-1}(f)P_k b\rangle. \end{aligned}$$

Put $\mathcal{M}_k = \{y \in \mathbf{R}^\varkappa : \|T_k^{-1}(\overline{f})P_\varkappa y\|_\infty \le 1\}$ and define $\varphi_k : \mathcal{M}_k \to \mathbf{R}$ by $\varphi_k(y) = \langle y, P_\varkappa T_k^{-1}(f)P_k b\rangle$. Then

$$d_k = \sup_{y\in\mathcal{M}_k} \varphi_k(y).$$

From Theorem 3 we know that

$$d = \sup_{z\in N(T(\overline{a}_+)),\|z\|_\infty\le 1} \langle z, b\rangle.$$

As $T(\overline{a}_+) = T(\chi_{-\varkappa})T(\overline{f})$, the equation $T(\overline{a}_+)z = 0$ is equivalent to the equation $T(\chi_{-\varkappa})T(\overline{f})f = 0$, that is, to the existence of a $y \in \mathbf{R}^\varkappa$ such that $z = T^{-1}(\overline{f})P_\varkappa y$. It follows that

$$\begin{aligned} d &= \sup_{z=T^{-1}(\overline{f})P_\varkappa y,\|z\|_\infty\le 1} \langle z, b\rangle \\ &= \sup_{\|T^{-1}(\overline{f})P_\varkappa y\|_\infty\le 1} \langle T^{-1}(\overline{f})P_\varkappa y, b\rangle \\ &= \sup_{\|T^{-1}(\overline{f})P_\varkappa y\|_\infty\le 1} \langle P_\varkappa y, P_\varkappa T^{-1}(f)b\rangle = \sup_{y\in\mathcal{M}} \varphi(y), \end{aligned}$$

where $\mathcal{M} = \{y \in \mathbf{R}^\varkappa : \|T^{-1}(\overline{f})P_\varkappa y\|_\infty \le 1\}$ and $\varphi : \mathcal{M} \to \mathbf{R}$ is given by $\varphi(y) = \langle y, P_\varkappa T^{-1}(f)b\rangle$. By Theorem 5,

$$\varphi_k(y) = \sum_{j=1}^{\varkappa} \gamma_j(k)y_j, \quad \varphi(y) = \sum_{j=1}^{\varkappa} \gamma_j y_j,$$

where $\gamma_j(k)$ converges to γ_j exponentially fast as $k \to \infty$. We remark that if $y \in \mathcal{M}_k$, then

$$\begin{array}{rcl} \|P_\varkappa y\|_\infty & \leq & \|P_k T(\overline{f}) P_k\|_{\mathcal{B}(c_0)} \|T_k^{-1}(\overline{f}) P_\varkappa y\|_\infty \\ & \leq & \|f\|_W \|T_k^{-1}(\overline{f}) P_\varkappa y\|_\infty \leq \|f\|_W. \end{array}$$

Analogously, $\|P_\varkappa y\|_\infty \leq \|f\|_W$ for $y \in \mathcal{M}$.

Now take $y_0 = (y_1^{(0)}, \ldots, y_\varkappa^{(0)}) \in \mathcal{M}$ so that $\varphi(y_0) = d$. Theorem 5 yields

$$\begin{array}{rcl} \|T_k^{-1}(\overline{f}) P_\varkappa y_0\|_\infty & \leq & \|T^{-1}(\overline{f}) P_\varkappa y_0\|_\infty + \|T_k^{-1}(\overline{f}) P_\varkappa y_0 - T^{-1}(\overline{f}) P_\varkappa y_0\|_\infty \\ & \leq & 1 + C\, e^{-\alpha k} \|P_\varkappa y_0\|_\infty \leq 1 + C\, e^{-\alpha k} \|f\|_W =: 1 + \sigma_k. \end{array}$$

Thus, $(1+\sigma_k)^{-1} y_0 \in \mathcal{M}_k$. This implies that

$$d_k \geq \varphi_k[(1+\sigma_k)^{-1} y_0] = (1+\sigma_k)^{-1} \sum_{j=1}^{\varkappa} \gamma_j(k) y_j^{(0)}.$$

Since $\{\gamma_j(k) - \gamma_j\}_{k=1}^\infty$ is exponentially decaying for each j, we have

$$\sum_{j=1}^{\varkappa} \gamma_j(k) y_j^{(0)} \geq \sum_{j=1}^{\varkappa} \gamma_j y_j^{(0)} - \tau_k = d - \tau_k$$

with some exponentially decaying sequence $\{\tau_k\}$. In summary, we have shown that $(1 + \sigma_k) d_k \geq d - \tau_k$, which gives

$$d - d_k \leq \sigma_k d_k + \tau_k \leq \sigma_k \|b\|_1 + \tau_k. \tag{2.9}$$

Again taking into account that $\{\gamma_j(k) - \gamma_j\}_{k=1}^\infty$ is exponentially decaying for each j and using that $\|P_\varkappa y\|_\infty \leq \|f\|_W$ for all $y \in \mathcal{M}_k$, we obtain

$$d_k = \sup_{y \in \mathcal{M}_k} \sum_{j=1}^{\varkappa} \gamma_j(k) y_j \leq \sup_{y \in \mathcal{M}_k} \sum_{j=1}^{\varkappa} \gamma_j y_j + \varrho_k$$

with an exponentially decaying sequence $\{\varrho_k\}$. For $y \in \mathcal{M}_k$, Theorem 5 gives

$$\begin{array}{rcl} \|T^{-1}(\overline{f}) P_\varkappa y\|_\infty & \leq & \|T_k^{-1}(\overline{f}) P_\varkappa y\|_\infty + \|T^{-1}(\overline{f}) P_\varkappa y - T_k^{-1}(\overline{f}) P_\varkappa y\|_\infty \\ & \leq & 1 + C\, e^{-\alpha k} \|P_\varkappa y\|_\infty \leq 1 + C\, e^{-\alpha k} \|f\|_W =: 1 + \sigma_k, \end{array}$$

and therefore $(1+\sigma_k)^{-1} y \in \mathcal{M}$. It follows that

$$\begin{array}{rcl} d_k & \leq & \displaystyle\sup_{(1+\sigma_k)^{-1} y \in \mathcal{M}} \sum_{j=1}^{\varkappa} \gamma_j y_j + \varrho_k = \sup_{v \in \mathcal{M}} \sum_{j=1}^{\varkappa} \gamma_j \cdot (1+\sigma_k) v_j + \varrho_k \\ & = & \displaystyle (1+\sigma_k) \sup_{v \in \mathcal{M}} \varphi(v) + \varrho_k \leq (1+\sigma_k) d + \varrho_k, \end{array}$$

whence

$$d_k - d \leq \sigma_k d + \varrho_k. \tag{2.10}$$

Combining (2.9) and (2.10) we arrive at the assertion. □

Corollary 1. *For each $k \geq 1$, let $x_0^{(k)} \in \mathcal{R}(P_{k-\varkappa})$ be an element at which (2.6) attains its minimum d_k. If $k_i \to \infty$ and $\{x_0^{(k_i)}\}_{i=1}^{\infty}$ is any sequence that converges in ℓ_1 to some $x_0 \in \ell_1$, then $\|b - T(a_+)x_0\|_1 = d$.*

Proof. If $\|P_{k_i}b - P_{k_i}T(a_+)P_{k_i-\varkappa}x_0^{(k_i)}\|_1 = d_{k_i}$ and $x_0^{(k_i)} \to x$ as $i \to \infty$, then $\|b - T(a_+)x_0\|_1 = d$ because $d_{k_i} \to d$ by Theorem 6. □

2.9 Error estimate

In practice, we have an $x_0^{(k)}$ with

$$\|P_kb - P_kT(a_+)P_{k-\varkappa}x_0^{(k)}\|_1 = d_k,$$

and $\widetilde{m}_0 = T(a_+)P_{k-\varkappa}x_0^{(k)}$ is taken as an approximate solution. The question is: How far is $\|b - \widetilde{m}_0\|_1$ away from the optimal value d? We have

$$\begin{aligned}\|b - \widetilde{m}_0\|_1 \quad &\leq \quad \|b - P_kb\|_1 + \|P_kb - P_kT(a_+)P_{k-\varkappa}x_0^{(k)}\|_1 \\ &\quad + \|P_kT(a_+)P_{k-\varkappa}x_0^{(k)} - T(a_+)P_{k-\varkappa}x_0^{(k)}\|_1. \qquad (2.11)\end{aligned}$$

Clearly, $\|b - P_kb\|_1 = \|Q_kb\|_1 = o(1)$ is given a priorily. The second term on the right of (2.11) is just d_k. Let $x_i^{(k)}$ ($i = 1, \ldots, k-\varkappa$) denote the components of $x_0^{(k)}$. The vector in the third term on the right of (2.11) is

$$-Q_kT(a_+)P_{k-\varkappa}x_0^{(k)} = -\begin{pmatrix} a_k & \ldots & a_{\varkappa+1} \\ a_{k+1} & \ldots & a_{\varkappa+2} \\ \vdots & & \vdots \end{pmatrix}\begin{pmatrix} x_1^{(k)} \\ \vdots \\ x_{k-\varkappa}^{(k)} \end{pmatrix}$$

$$= -\begin{pmatrix} 0 & \ldots & 0 & a_n & a_{n-1} & \ldots & a_{\varkappa+1} \\ 0 & \ldots & 0 & 0 & a_n & \ldots & a_{\varkappa+2} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \ldots & 0 & 0 & 0 & \ldots & a_n \end{pmatrix}\begin{pmatrix} x_1^{(k)} \\ \vdots \\ x_{k-\varkappa}^{(k)} \end{pmatrix},$$

which implies that

$$\begin{aligned}\|Q_kT(a_+)P_{k-\varkappa}x_0^{(k)}\|_1 \quad &\leq \quad (|a_{\varkappa+1}| + \ldots + |a_n|)\left(|x_{k-n+1}^{(k)}| + \ldots + |x_{k-\varkappa}^{(k)}|\right) \\ &\leq \quad \|a_+\|_W\,\|Q_{k-n}x_0^{(k)}\|_1.\end{aligned}$$

The vector $x_0^{(k)}$ is available and $\|Q_{k-n}x_0^{(k)}\|_1$ is the ℓ_1 norm of the last $n-\varkappa$ components of $x_0^{(k)}$. If k is large, then the last $n-\varkappa$ components of $x_0^{(k)}$ are expected to be small. In summary, (2.11) and Theorem 6 yield

$$\|b - \widetilde{m}_0\|_1 \leq \|Q_kb\|_1 + \|a_+\|_W\,\|Q_{k-n}x_0^{(k)}\|_1 + d + \text{ exponentially small term.}$$

If $\varkappa = n$, then $-Q_kT(a_+)P_{k-\varkappa}x_0^{(k)} = 0$, and hence we even have

$$\|b - \widetilde{m}_0\|_1 \leq \|Q_kb\|_1 + d + \text{ exponentially small term.}$$

Finally, if $\varkappa = n$ and b is finitely supported, then

$$\|b - \widetilde{m}_0\|_1 \le d + \text{ exponentially small term.}$$

2.10 Numerical example for the minimum distance problem

The following example illustrates the algorithm described above. We consider the banded lower triangular Toeplitz matrix $T(a_+)$ with the symbol

$$a_+(\lambda) = -0.1224 - 0.2906\lambda + 0.7122\lambda^2 + 2.7983\lambda^3 + 2.9168\lambda^4 + \lambda^5$$

and are looking for a sequence $x \in \ell_1$ minimizing $\|b - T(a_+)x\|_1$ for the right-hand side

$$b = \{1.8645, -0.3398, -1.1398, -0.2111, 1.1902, -1.1162, 0, 0, \ldots\}.$$

The zeros of the polynomial $a_+(\lambda)$ are all inside the open unit disk. Thus, we can proceed as in Section 7 with $\varkappa = 5$. (Notice that the algorithm of Section 7 would be applicable to $\varkappa < 5$ as well.) Accordingly, we approximate $T(a_+)$ by the finite matrices $A_k = P_{5+k}T(a_+)P_k$ $(k = 1, 2, \ldots)$. For example,

$$A_3 = \begin{pmatrix} -0.1224 & 0 & 0 \\ -0.2906 & -0.1224 & 0 \\ 0.7122 & -0.2906 & -0.1224 \\ 2.7983 & 0.7122 & -0.2906 \\ 2.9168 & 2.7983 & 0.7122 \\ 1 & 2.9168 & 2.7983 \\ 0 & 1 & 2.9168 \\ 0 & 0 & 1 \end{pmatrix}.$$

Solving the corresponding overdetermined linear system for a solution minimizing the ℓ_1-norm of the residue (using a general linear programming solver) and repeating this for increasing index k, we obtain Figures 2.2 and 2.3. In Figure 2.2 we nicely see the exponentially fast stabilization of the objective function (ℓ_1-norm of the residue) predicted by Theorem 6. Figure 2.3 reveals that for $k \ge 36$ the set length k is not decreasing any more. Thus, although we offer more and more space to y, the actual length of the optimal solution settles at 36.

Figures 2.4 and 2.5 show an optimal solution x and the residue sequence $y = b - T(a_+)x$. The very small number of nonzero terms in Figure 2.5 is a mystery we cannot yet explain.

2.11 Improvement of conditioning

Both our Toeplitz approach and the Vandermonde interpolation approach of [17] lead to linear systems. These can be tackled by invoking a linear programming solver. Numerical

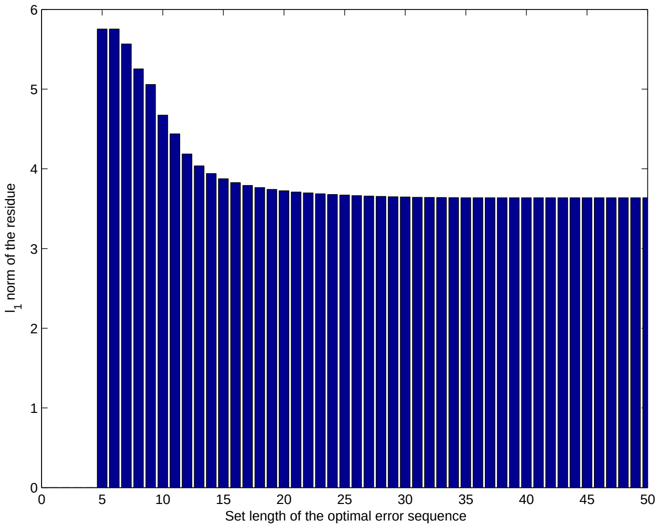

Figure 2.2: Evolvement of the ℓ_1-norm of the residue with increasing set length of the optimal error sequence

experiments show that the condition numbers of the matrices emerging in our algorithm are much smaller than those of the matrices that result from interpolation. To give a concrete example, consider the task of finding the distance between a given $x \in \ell_1$ and the range of the Toeplitz operator $T(a_+)$ with $a_+(\lambda)$ having its 10 roots equally distributed in the interval $[0.5, 0.9]$, that is, $a_+(\lambda) = 0.0238 - 0.3520\lambda + 2.3334\lambda^2 - 9.1302\lambda^3 + 23.3525\lambda^4 - 40.7975\lambda^5 + 49.3052\lambda^6 - 40.7037\lambda^7 + 21.9685\lambda^8 - 7\lambda^9 + \lambda^{10}$. Let us set the length of the approximate optimal error sequence to 13. The 2-norm condition number, that is, the ratio of the largest and the smallest singular value, of the 10×13 Vandermonde matrix V_{13} built from the roots of $a_+(\lambda)$ equals $\kappa(V_{13}) = 9.5458 \cdot 10^9$. In contrast to this, the 2-norm condition number of the matrix $A_{13} = P_{13}T(a_+)P_3$ is $\kappa(A_{13}) = 14.948$.

2.12 Numerical algorithm: minimization of ℓ_1 norm of sensitivity function

We consider the standard feedback configuration of Figure 2.1 with a discrete-time plant $G(\lambda)$ and a negative sign in the feedback loop. Our aim is to construct a stabilizing discrete-time controller $C(\lambda)$ that minimizes the Wiener norm of the sensitivity function of the closed-loop system, that is, of the transfer function $1/(1 + C(\lambda)G(\lambda))$ between the disturbance and the error or, equivalently, the ℓ_1 norm of the impulse response.

We suppose that the plant is given as the quotient of two polynomials $p(\lambda)$ and $q(\lambda)$ without common zeros and with no zeros on the unit circle, $G(\lambda) = p(\lambda)/q(\lambda)$. The

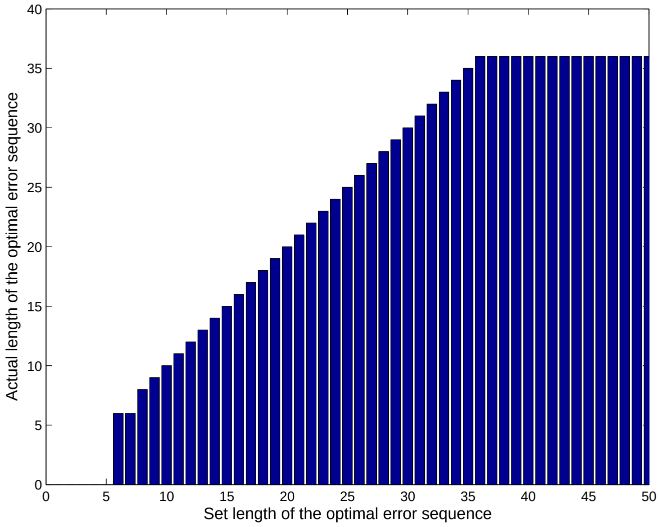

Figure 2.3: Evolvement of the actual length of the optimal error sequence with increasing set length of the optimal error sequence

Youla-Kučera parametrization of all stabilizing controllers is

$$C(\lambda) = \frac{v(\lambda) + q(\lambda)x(\lambda)}{w(\lambda) - p(\lambda)x(\lambda)}, \tag{2.12}$$

where $v(\lambda), w(\lambda)$ are polynomials determined by $G(\lambda)$ and $x(\lambda)$ is a function we can freely choose in the Wiener algebra. The entire procedure can be done in four steps.
Step 1. Find stable-unstable factorizations $p(\lambda) = p_s(\lambda)p_u(\lambda)$ and $q(\lambda) = q_s(\lambda)q_u(\lambda)$. Here the indices s and u label polynomials with all zeros inside and outside the unit circle, respectively. Efficient algorithms for stable-unstable factorization are known (see, e.g., [4] and the references cited there). In particular, reliable FFT-based algorithms are available from [3], [27].
Step 2. Find polynomials $x_0(\lambda)$ and $y_0(\lambda)$ satisfying the Diophantine equation $q(\lambda)x_0(\lambda) + p(\lambda)y_0(\lambda) = 1$. This problem can be conveniently solved using the Polynomial Toolbox [42].
Step 3. The polynomials $v(\lambda), w(\lambda)$ in (2.12) are

$$v(\lambda) = q_u(\lambda)p_u(\lambda)y_0(\lambda), \quad w(\lambda) = q_u(\lambda)p_u(\lambda)x_0(\lambda).$$

Step 4. Inserting the result of Step 3 in (2.12) we obtain

$$\frac{1}{1+CG} = \frac{1}{1 + \dfrac{q_u p_u y_0 + qx}{q_u p_u y x_0 - px}\,\dfrac{p}{q}} = \frac{q q_u p_u x_0 - qpx}{q_u p_u (q x_0 + p y_0)},$$

which equals $qx_0 - q_s p_s x$ by virtue of Step 2. Thus, the final task is to minimize $\|y(\lambda)\|_W = \|q(\lambda)x_0(\lambda) - q_s(\lambda)p_s(\lambda)x(\lambda)\|_W$ or, in terms of the coefficient sequences, to minimize $\|y\|_1 =$

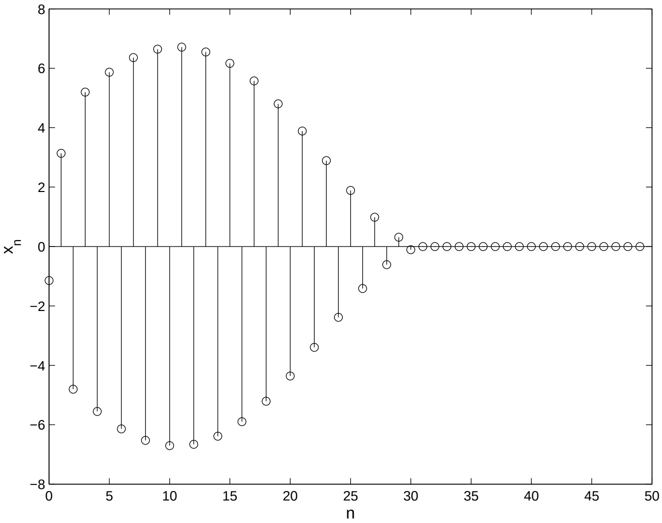

Figure 2.4: A sequence $x \in \ell_1$ minimizing $\|b - T(a_+)x\|_1$

$\|b - T(a_+)x\|_1$, where $b \in \ell_1$ is the coefficient sequence of $q(\lambda)x_0(\lambda)$ and $a_+(\lambda) = q_s(\lambda)p_s(\lambda)$. This problem can be solved using the algorithm of Section 2.8. The desired optimal controller is given by (2.12) with $v(\lambda), w(\lambda)$ from Step 3 and $x(\lambda)$ from Step 4.

To have a numerical example, let

$$G(\lambda) = \frac{p(\lambda)}{q(\lambda)} = \frac{-45\lambda - 132\lambda^2 + 9\lambda^3}{-20 - 48\lambda + 5\lambda^2}.$$

The above procedure yields the Youla-Kučera parameter $x(\lambda) = 0.1321 - 0.0052\lambda$, the sensitivity function $y(\lambda) = 1.0000 - 12.5000\lambda - 37.5000\lambda^2$, and the optimal controller

$$C(\lambda) = \frac{-41.6667 + 4.1667\lambda}{-7.5000 + 113.0000\lambda - 7.5000\lambda^2}.$$

A simulation result is shown in Figure 2.6. The horizontal axis represents the discrete time k.

The disturbance is only known to be bounded in magnitude. The response of the closed-loop system to a disturbance bounded in magnitude by 1 is compared for the ℓ_1-optimal controller computed above and some random stabilizing controller. Similar results will be obtained even with other more sophisticated controllers like LQG, $\mathcal{H}_2$- and $\mathcal{H}_\infty$-optimal.

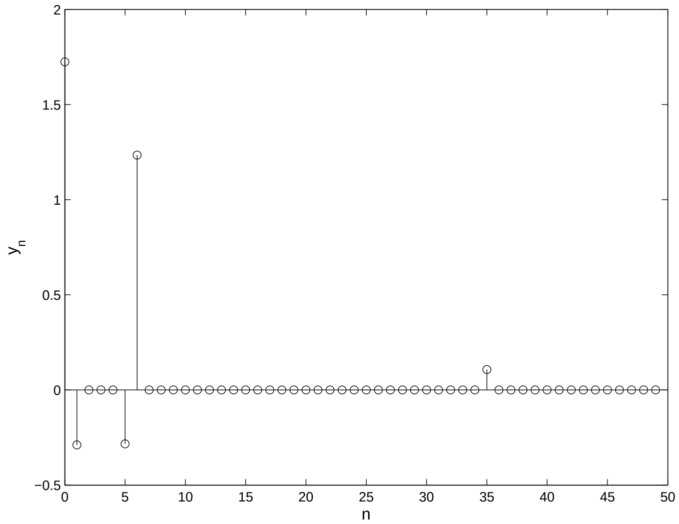

Figure 2.5: The optimal error sequence (residue) $y = b - T(a_+)x$

2.13 Summary

In this chapter we described a new theoretical framework for solving the standard ℓ_1-optimal control problem. The objective of ℓ_1-optimal control is to design a discrete-time feedback controller that will attenuate optimally the worst-case peaks in the amplitude of the regulated variable induced by a persistent disturbce that is only known to be bounded in magnitude. Motivation for this kind of control can be found in applications, where large peaks in the amplitude of a regulated variable are not acceptable.

The proposed method accepts a description of an LTI model of a plant in the form of a rational transfer function, i.e., a ratio of two univariate polynomials. In the theoretical derivation, advanced results from the theory of banded Toeplitz operators were invoked. Formulating the optimal control problem as searching for a minimum distance between a given sequence in ℓ_1 and a range of a given Toeplitz operator, an alternative proof has been given for the fact that an optimal controller exists if and only if the plant has neither pole nor zero on the unit circle. Moreover, an optimal controller need not be unique. Additionally, the optimal closed-loop impulse response is finite.

A numerical procedure for solving this design problem relies on solving linear equations with polynomials. Since, the corresponding linear program is build directly from the coefficients of polynomials, computation of roots is avoided and the problem is much better conditioned. Moreover, the optimal Youla-Kučera parameter is returned as an outcome of the optimization and there is no need for a numerically tricky extraction of an optimal controller from the optimal closed-loop transfer function.

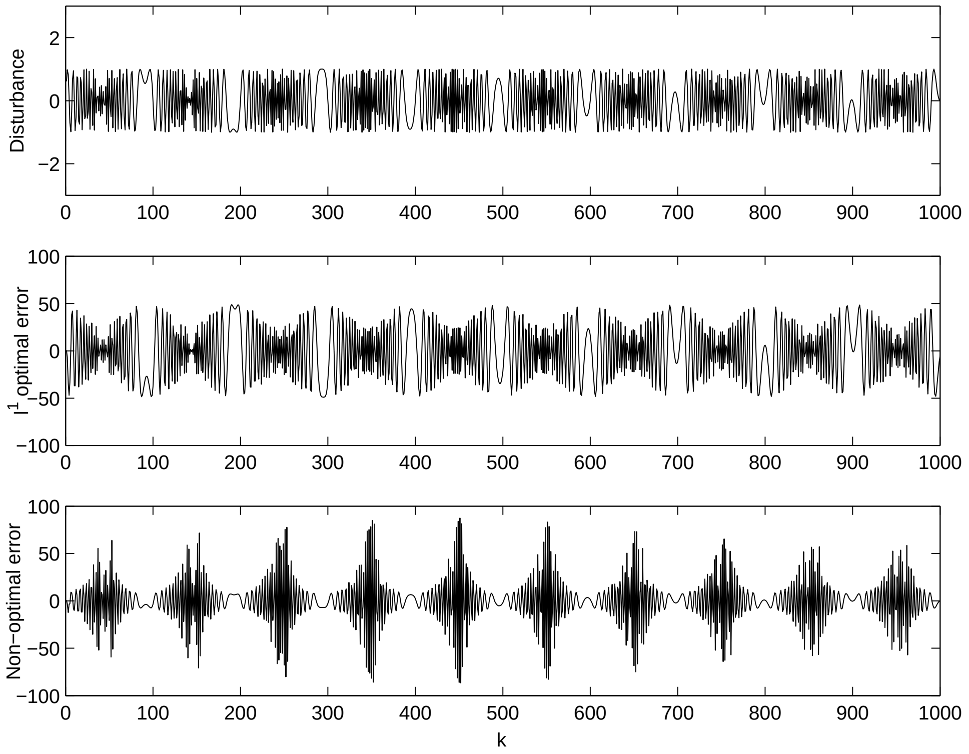

Figure 2.6: Simulation of a disturbance rejection with ℓ_1-optimal and non-optimal controllers

Chapter 3

MIMO feedback ℓ_1-optimal control

3.1 Introduction

In this chapter we present an extension of the ℓ_1-optimal control design procedure to a MIMO case. This is a useful direction of extension not only for the sake of real systems with many inputs and many outputs but also for SISO systems with more refined control requirements. For example, among the exogenous inputs to an artificial *generalised plant* one may include a reference angular position for an elevation axis of a telescope as well as a disturbing torque induced by wind buffeting. Among the regulated variables it is possible to include an error between the required position and a true (measured) position and a control voltage applied to the armature of the motor (possibly frequency-weighted). The objective of ℓ_1-optimal control in this particular case is then to minimise the peaks in the positionning error and the control voltage, induced by the disturbing torque and changes in reference position.

The major theoretical achievement presented in this chapter is that for systems with no poles or zeros on the unit circle an optimal controller is always guaranteed to exist, morever in the square MIMO case an optimal closed loop transfer function is a polynomial matrix and an optimal controller is rational. In fact, these results have been known since Dahleh's seminal paper [17] but we provide alternative mathematical framework for theoretical derivation and numerical computation. The proposed numerical procedure avoids computing zeros and zero directions of polynomial matrices, hence enjoys better numerical properties. Algorithmically speaking, the key problem solved in this section is: given a triple of stable rational matrices $A(\lambda)$, $B(\lambda)$ and $C(\lambda)$, find a solution to the linear equation $A(\lambda)X(\lambda)B(\lambda) + Y(\lambda) = C(\lambda)$ that minimizes row-sum norm of $Y(\lambda)$.

In a general multiblock case an optimal solution $Y(\lambda)$ need not be a polynomial matrix. Solving a sequence of the above equations for a polynomial matrix $Y(\lambda)$ of an increasing degree gives an upper bound on the optimal norm, whose converge is rigorously proven. A relaxation procedure for computation of a converging lower bound is proposed.

In this work we often invoked both standard and advanced results for block Toeplitz operators. Therefore, let's start with some basic definitions.

3.2 Block Toeplitz operators

By a proper transfer function we mean a function $g(\lambda) = \sum_{k=0}^{\infty} g_k \lambda^k$ with real coefficients g_k that is analytic in a neighborhood of the origin. The Toeplitz matrix $T(g)$ associated with a proper transfer function g is the infinite lower triangular matrix

$$T(g) = \begin{pmatrix} g_0 & & & \\ g_1 & g_0 & & \\ g_2 & g_1 & g_0 & \\ \dots & \dots & \dots & \dots \end{pmatrix}.$$

For a real sequence $s = \{s_j\}_{j=0}^{\infty}$, we define the sequence $T(g)s$ as the sequence $\sigma = \{\sigma_j\}_{j=0}^{\infty}$ given by

$$\begin{pmatrix} \sigma_0 \\ \sigma_1 \\ \sigma_2 \\ \dots \end{pmatrix} = \begin{pmatrix} g_0 & & & \\ g_1 & g_0 & & \\ g_2 & g_1 & g_0 & \\ \dots & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} s_0 \\ s_1 \\ s_2 \\ \dots \end{pmatrix}.$$

We denote by ℓ_∞ the real Banach space of all real sequences $s = \{s_j\}_{j=0}^{\infty}$ for which

$$\|s\|_\infty := \sup_{j \geq 0} |s_j| < \infty.$$

It is well known that $T(g)$ induces a bounded operator on ℓ_∞ if and only if

$$\|g\|_W := \sum_{k=0}^{\infty} |g_k| < \infty. \tag{3.1}$$

The set of all proper transfer functions g satisfying (3.1) is called the (real and analytic) Wiener algebra and is denoted by W_+. Clearly, functions in W_+ are analytic for $|\lambda| < 1$ and continuous for $|\lambda| \leq 1$. If $g \in W_+$, then the norm of $T(g)$ on ℓ_∞ is known to be just $\|g\|_W$.

Now let $G(\lambda) = \sum_{k=0}^{\infty} G_k \lambda^k$ be a matrix function with coefficients in $\mathbf{R}^{m \times n}$ that is analytic in a neighborhood of the origin. We refer to such matrix functions as proper transfer functions as well. We write $G = (G_{ij})_{i=1,j=1}^{m,n}$ and define the block Toeplitz matrix $T(G)$ by

$$T(G) = \begin{pmatrix} T(G_{11}) & \dots & T(G_{1n}) \\ \vdots & & \vdots \\ T(G_{m1}) & \dots & T(G_{mn}) \end{pmatrix}.$$

This matrix transforms n-tuples $(s^1, \dots, s^n)$ of real sequences into m-tuples $(\sigma^1, \dots, \sigma^m)$ of real sequences in the natural fashion. We denote by ℓ_∞^k the real Banach space of all k-tuples

$(s^1, \dots, s^k)$ of sequences $s^1, \dots, s^k \in \ell_\infty$ with the norm

$$\|s\|_\infty := \max(\|s^1\|_\infty, \dots, \|s^k\|_\infty).$$

From the preceding paragraph we know that $T(G)$ generates a bounded operator of ℓ_∞^n to ℓ_∞^m if and only if $G \in W_+^{m\times n}$. In that case the norm of $T(G)$ as an operator of ℓ_∞^n to ℓ_∞^m,

$$\|T(G)\|_{\mathcal{B}(\ell_\infty^n, \ell_\infty^m)} := \sup_{\|s\|_\infty \le 1} \|T(G)s\|_\infty,$$

equals the row-sum norm

$$\|G\|_W := \max_{1\le i\le m} \sum_{j=1}^{n} \|G_{ij}\|_W. \tag{3.2}$$

3.3 The ℓ_1-optimal control problem

We consider the configuration of Figure 3.1 with discrete-time linear time-invariant systems. The inputs and outputs are real sequences $\{s_j\}_{j=0}^\infty$. We suppose that we have n_w exogenous inputs, n_u control inputs, n_z regulated outputs, and n_y measured variables. The entire feedback system can be written in the form

$$\begin{aligned} z &= T(P_{wz})w + T(P_{uz})u, \\ y &= T(P_{wy})w + T(P_{uy})u, \\ u &= T(C)y. \end{aligned}$$

Here $P_{wz}, P_{uz}, P_{wy}, P_{uy}$ are given proper transfer functions of appropriate sizes and C is a proper transfer function of appropriate size that has to be designed.

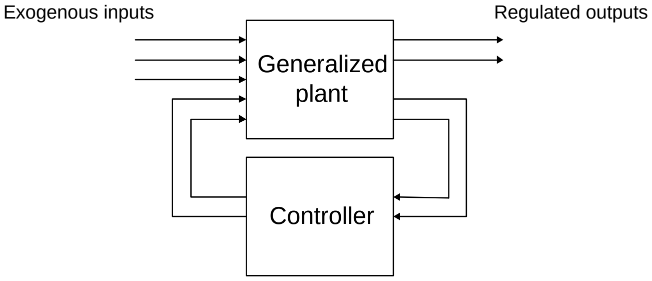

Figure 3.1: Standard feedback control configuration

We have $z = T(G)w$ with the proper transfer function

$$G = P_{wz} + P_{uz}C(I - P_{uy}C)^{-1}P_{wy}. \tag{3.3}$$

The objective of ℓ_1 control is to find the C's such that $T(G)$ is a bounded linear operator of $\ell_\infty^{n_w}$ to $\ell_\infty^{n_z}$, and in ℓ_1 optimal control we look for the C's for which

$$\|T(G)\|_{\mathcal{B}(\ell_\infty^{n_z},\ell_\infty^{n_w})} = \|G\|_W$$

is minimal or close to the infimum.

We make the usual assumptions. Thus, we assume that P_{wz}, P_{uz}, P_{wy} are rational matrix functions with entries in W_+ and that P_{uy} is a rational matrix function that may have poles in the closed unit disk. Then there exist (real) matrix polynomials $B_R, A_R, B_L, A_L, Y_R^0, X_R^0, Y_L^0, X_L^0$ such that

$$P_{uy} = B_R A_R^{-1} = A_L^{-1} B_L \tag{3.4}$$

and

$$\begin{pmatrix} X_L^0 & -Y_L^0 \\ -B_L & A_L \end{pmatrix} \begin{pmatrix} A_R & Y_R^0 \\ B_R & X_R^0 \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}. \tag{3.5}$$

Under these assumptions, the set of all controllers C for which $\|G\|_W$ is finite is given by the Youla-Kučera parametrization:

$$C = Y_R X_R^{-1} = X_L^{-1} Y_L \tag{3.6}$$

where

$$X_R = X_R^0 + B_R\widetilde{Q}, \quad Y_R = Y_R^0 + A_R\widetilde{Q}, \tag{3.7}$$

$$X_L = X_L^0 + \widetilde{Q}B_L, \quad Y_L = Y_L^0 + \widetilde{Q}A_L, \tag{3.8}$$

and $\widetilde{Q}$ is an arbitrary matrix function in $W_+^{n_u \times n_y}$. Inserting (3.7) in (3.6) and taking into account (3.4) and (3.5) we obtain

$$C(I - P_{uy}C)^{-1} = (Y_R^0 + A_R\widetilde{Q})A_L,$$

and inserting this in (3.3) we arrive at the representation[1]

$$G = H + \widetilde{U}\widetilde{Q}\widetilde{V}$$

with

$$H = P_{wz} + P_{uz}Y_R^0 A_L P_{wy},$$
$$\widetilde{U} = P_{uz}A_R, \quad \widetilde{V} = A_L P_{wy}.$$

The rational matrix functions $\widetilde{U}$ and $\widetilde{V}$ can be written as

$$\widetilde{U} = -UU_R^{-1}, \quad \widetilde{V} = V_L^{-1}V$$

[1] We really have the plus sign in the formula for G. The minus sign in (6.116) of [44] is actually false, because already (6.60) has the wrong sign.

with (real) matrix polynomials U, U_R, V_L, V such that $\det U_R(\lambda) \neq 0$ and $\det V_L(\lambda) \neq 0$ for $|\lambda| \leq 1$. We put $Q = U_R^{-1}\widetilde{Q}V_L^{-1}$ and have

$$G = H - UQV.$$

Thus, our problem is as follows: we are given a rational matrix function $H \in W_+^{n_z \times n_w}$ and two matrix polynomials $U \in W_+^{n_z \times n_u}$ and $V \in W_+^{n_y \times n_w}$, and we look for a matrix function $Q \in W_+^{n_u \times n_y}$ such that $\|H - UQV\|_W$ is minimal or close to the infimum. Furthermore, it is desirable to find a rational matrix function Q with this property.

3.4 Existence of the solution

In practically relevant control problems we always have $n_z \geq n_u$ and $n_w \geq n_y$. Throughout what follows we assume that these two inequalities are satisfied. Our main assumption is that the two matrix polynomials U and V have full rank on the unit circle $\mathbf{T}$, that is,

$$\operatorname{rank} U(\lambda) = n_u \quad \text{and} \quad \operatorname{rank} V(\lambda) = n_y \quad \text{for all } \lambda \in \mathbf{T}. \tag{3.9}$$

The following lemma is well known known and can be easily proved using the Smith normal form. For the reader's convenience, we cite it with an absolutely elementary proof.

Lemma 1. *Under assumption (3.9), there exist rational (real) matrix functions L and K without poles on $\mathbf{T}$ such that $LU = I$ and $VK = I$.*

Proof. To avoid heavy notation, let us consider the case where

$$V = \begin{pmatrix} v_1 & v_2 & v_3 & v_4 \\ v_5 & v_6 & v_7 & v_8 \end{pmatrix}.$$

We denote by $V_1, \ldots, V_6$ the 2×2 submatrices of V. From (3.9) we infer that

$$\sum_{k=1}^{6} |\det V_k(\lambda)|^2 > 0 \quad \text{for all } \lambda \in \mathbf{T}.$$

Put

$$h_j(\lambda) = \frac{\overline{\det V_j(\lambda)}}{\sum_{k=1}^{6} |\det V_k(\lambda)|^2},$$

the bar denoting complex conjugation. Then h_j is a rational function without poles on $\mathbf{T}$ and

$$\sum_{j=1}^{6} (\det V_j) h_j = 1.$$

The following trick is from [6, Proposition 13.9] and [50, Lemma 3.1]. Let $\operatorname{adj} V_j$ denote the adjugate matrix of V_j. Then

$$(\det V_j)I = V_j(\operatorname{adj} V_j)$$

and letting $K_j = (\operatorname{adj} V_j)h_j$ we get

$$\sum_{j=1}^{6} V_j K_j = \sum_{j=1}^{6} V_j(\operatorname{adj} V_j)h_j = \sum_{j=1}^{6}(\det V_j)h_j I = I.$$

Write

$$K_j = \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}.$$

We have

$$\begin{aligned} &\begin{pmatrix} v_1 & v_2 \\ v_5 & v_6 \end{pmatrix}\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} + \begin{pmatrix} v_1 & v_3 \\ v_5 & v_7 \end{pmatrix}\begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} \\ &+ \begin{pmatrix} v_1 & v_4 \\ v_5 & v_8 \end{pmatrix}\begin{pmatrix} a_3 & b_3 \\ c_3 & d_3 \end{pmatrix} + \begin{pmatrix} v_2 & v_3 \\ v_6 & v_7 \end{pmatrix}\begin{pmatrix} a_4 & b_4 \\ c_4 & d_4 \end{pmatrix} \\ &+ \begin{pmatrix} v_2 & v_4 \\ v_6 & v_8 \end{pmatrix}\begin{pmatrix} a_5 & b_5 \\ c_5 & d_5 \end{pmatrix} + \begin{pmatrix} v_3 & v_4 \\ v_7 & v_8 \end{pmatrix}\begin{pmatrix} a_6 & b_6 \\ c_6 & d_6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

and hence

$$\begin{pmatrix} v_1 & v_2 & v_3 & v_4 \\ v_5 & v_6 & v_7 & v_8 \end{pmatrix}\begin{pmatrix} a_1 + a_2 + a_3 & b_1 + b_2 + b_3 \\ c_1 + a_4 + a_5 & d_1 + b_4 + b_5 \\ c_2 + c_4 + a_6 & d_2 + d_4 + b_6 \\ c_3 + c_5 + c_6 & d_3 + d_5 + d_6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which completes the proof. □

It will be convenient to write the matrix $H - UQV$ as a column. This can be done in a standard fashion, or better in two standards manners, namely by column stacking on the one hand and by row stacking on the other. As we are concerned with the row-sum norm (3.2), we stack matrices by rows. For example, the equality

$$\begin{pmatrix} H_1 & H_2 \\ H_3 & H_4 \end{pmatrix} - \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}\begin{pmatrix} Q_1 & Q_2 \end{pmatrix}\begin{pmatrix} v_1 & v_2 \\ v_3 & v_4 \end{pmatrix} = \begin{pmatrix} E_1 & E_2 \\ E_3 & E_4 \end{pmatrix}$$

is equivalent to the equality

$$\begin{pmatrix} H_1 \\ H_2 \\ H_3 \\ H_4 \end{pmatrix} - \begin{pmatrix} u_1v_1 & u_1v_3 \\ u_1v_2 & u_1v_4 \\ u_2v_1 & u_2v_3 \\ u_2v_2 & u_2v_4 \end{pmatrix}\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{pmatrix}. \tag{3.10}$$

Let ℓ_1 be the usual real Banach space of real sequences $s = \{s_j\}_{j=0}^{\infty}$ with

$$\|s\|_1 := \sum_{j=0}^{\infty} |s_j| < \infty.$$

We denote by $h_j \in \ell_1$ the sequence of the Taylor coefficients of H_j and define q_j and e_j analogously. (By Taylor coefficients we always mean the Taylor coefficients at the origin.) Then (3.10) can also be written in the form

$$\begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{pmatrix} - T \begin{pmatrix} u_1v_1 & u_1v_3 \\ u_1v_2 & u_1v_4 \\ u_2v_1 & u_2v_3 \\ u_2v_2 & u_2v_4 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}. \tag{3.11}$$

We have

$$\begin{aligned} \left\| \begin{pmatrix} E_1 & E_2 \\ E_3 & E_4 \end{pmatrix} \right\|_W &= \max(\|E_1\|_W + \|E_2\|_W, \|E_3\|_W + \|E_4\|_W) \\ &= \max(\|e_1\|_1 + \|e_2\|_1, \|e_3\|_1 + \|e_4\|_1) \end{aligned} \tag{3.12}$$

Thus, when viewing (3.11) as an equality in ℓ_1^4, we must define the norm in ℓ_1^4 by (3.12).

In the general case we write $H - UQV = E$ as

$$\operatorname{vec} H - (U \otimes V^\top) \operatorname{vec} Q = \operatorname{vec} E,$$

where $\otimes$ is the Kronecker product of matrices and vec is defined in the obvious way, and then we pass to $\ell_1^{n_z n_w}$ by writing

$$h - T(F)q = e$$

with $F = U \otimes V^\top$. The norm in $\ell_1^{n_z n_w}$ is defined in analogy to (3.12), that is, if

$$e = \begin{pmatrix} e_{11} & \dots & e_{1n_z} & \dots & e_{n_w 1} & \dots & e_{n_w n_z} \end{pmatrix}^\top \in \ell_1^{n_z n_w}$$

then

$$\|e\|_1 := \max_{1 \le i \le n_w} (\|e_{i1}\|_1 + \ldots + \|e_{in_z}\|_1). \tag{3.13}$$

Clearly,

$$\|H - UQV\|_W = \|h - T(F)q\|_1. \tag{3.14}$$

The block Toeplitz operator $T(F)$ acts from $\ell_1^{n_u n_y}$ to $\ell_1^{n_z n_w}$. While the norm in $\ell_1^{n_z n_w}$ is given by (3.13) we need not and do not specify a concrete norm in $\ell_1^{n_u n_y}$ - we equip this space with any vector norm of the ℓ_1 norms of the components.

Theorem 7. *Under assumption (3.9), the operator $T(F) : \ell_1^{n_u n_y} \to \ell_1^{n_z n_w}$ has closed range.*

Proof. Suppose $\|z - T(F)x_n\|_1 \to 0$. There are $Z \in W_+^{n_z \times n_w}$ and $X_n \in W_+^{n_u \times n_y}$ such that z and x_n are the Taylor coefficients of $\operatorname{vec} Z$ and $\operatorname{vec} X_n$, respectively. By (3.14), $\|Z - UX_nV\|_W \to 0$. We denote by W the (real and full) Wiener algebra of all function g on $\mathbf{T}$ with real Fourier coefficients and absolutely convergent Fourier series, Thus, $g \in W$ if and only if

$$g(\lambda) = \sum_{k=-\infty}^{\infty} g_k \lambda^k \ (|\lambda| = 1), \quad g_k \in \mathbf{R}, \quad \|g\|_W := \sum_{k=-\infty}^{\infty} |g_k| < \infty.$$

Let L and K be the rational matrix functions of Lemma 1. Obviously, $L \in W^{n_u \times n_z}$ and $K \in W^{n_w \times n_y}$. Therefore,

$$\|LZK - X_n\|_W = \|L(Z - UX_nV)K\|_W \leq \|L\|_W \|Z - UX_nV\|_W \|K\|_W \to 0,$$

which implies that $X_n \to LZK =: X$ in $W^{n_u \times n_y}$. As $X_n \in W_+^{n_u \times n_y}$, it follows that $X \in W_+^{n_u \times n_y}$. Let $x \in \ell_1^{n_u n_y}$ be the sequence of the Taylor coefficients of $\operatorname{vec} X$. Since $x_n \to x$ in $\ell_1^{n_u n_y}$, we obtain that $z = T(F)x$ is in the range of $T(F)$. □

Corollary 2. *Under assumption (3.9), there exists a* $Q \in W_+^{n_u \times n_y}$ *at which* $\|H - UQV\|_W$ *attains its minimum.*

Proof. Let $d = \inf\{\|H - UQV\|_W : Q \in W_+^{n_u \times n_y}\}$. From (3.14) we infer that

$$d = \inf\{\|h - m^*\|_1 : m^* \in \mathcal{R}(T(F))\}, \tag{3.15}$$

where $\mathcal{R}(T(F))$ is the range of $T(F)$. Let c_0 be the real Banach space of all real sequences that converge to zero. The norm in c_0 is the ℓ_∞ norm and c_0 is known to be a closed subspace of ℓ_∞. Since $\ell_1^{n_z n_w}$ is the dual space of $c_0^{n_z n_w}$, the norm in $c_0^{n_z n_w}$ being,

$$\begin{aligned} &\left\| \begin{pmatrix} e_{11} & \dots & e_{1n_z} & \dots & e_{n_w 1} & \dots & e_{n_w n_z} \end{pmatrix}^\top \right\|_{c_0} \\ &\quad := \sum_{i=1}^{n_w} \max(\|e_{i1}\|_\infty, \dots, \|e_{in_z}\|_\infty), \end{aligned} \tag{3.16}$$

and, by Theorem 7, the range of $T(F)$ is closed, the infimum in (3.15) is attained by virtue of a well known duality result (see, e.g., Theorem 2 on page 121 of [43]). □

3.5 The square case

We now consider the case where U and V are square. Thus, suppose $n_z = n_u$ and $n_y = n_w$. Condition (3.9) is then equivalent to the requirement

$$\det U(\lambda) \neq 0 \quad \text{and} \quad \det V(\lambda) \neq 0 \quad \text{for all } \lambda \in \mathbf{T}. \tag{3.17}$$

Theorem 8. *Let (3.17) be satisfied. If* $Q \in W_+^{n_u \times n_y}$ *is any matrix function at which* $\|H - UQV\|$ *attains its minimum, then* Q *is a rational matrix function and the residue* $H - UQV$ *is a matrix polynomial.*

Proof. Given a scalar transfer function $g(\lambda) = \sum_{k=0}^{\infty} g_k \lambda^k$, we define the Toeplitz matrix $T(g^*)$ as the upper triangular matrix

$$T(g^*) = \begin{pmatrix} g_0 & g_1 & g_2 & \cdots \\ & g_0 & g_1 & \cdots \\ & & g_0 & \cdots \\ & & & \cdots \end{pmatrix}.$$

Let $F = U \otimes V^\top$ and write $F = (F_{ij})_{i=1,j=1}^{M,N}$ with scalar functions F_{ij}. Notice that $M = n_z n_w$ and $N = n_u n_y$. We define the block Toeplitz matrix $T(F^*)$ by

$$T(F^*) = \begin{pmatrix} T(F_{11}^*) & \dots & T(F_{M1}^*) \\ \vdots & & \vdots \\ T(F_{1N}^*) & \dots & T(F_{MN}^*) \end{pmatrix}.$$

Obviously, $T(F) : \ell_1^N \to \ell_1^M$ is the adjoint of $T(F^*) : c_0^M \to c_0^N$. Let $\mathcal{N}(T(F^*))$ be the null space of $T(F^*)$ on c_0^M. Due to Theorem 7, $\mathcal{R}(T(F)) = \mathcal{N}(T(F^*))^\perp$. Consequently, by Theorem 2 on page 121 of [43], the number (3.15) equals

$$d = \sup\{\langle z, h\rangle : z \in \mathcal{N}(T(F^*)), \|z\|_{c_0} \leq 1\}. \tag{3.18}$$

We have

$$\det F(\lambda) = \det\left(U(\lambda) \otimes V^\top(\lambda)\right) = (\det U(\lambda))^{n_y} (\det V(\lambda))^{n_u}$$

and hence $\det F(\lambda) \neq 0$ for $\lambda \in \mathbf{T}$. This implies that $\mathcal{N}(T(F^*))$ is finite-dimensional (see, e.g.,[29, Proposition 13.3] or [24, Section VIII.4]). It follows that the supremum in (3.18) is a maximum. Assume this maximum is attained at z_0. Again by Theorem 2 on page 121 of [43], z_0 and $e := h - T(F)q$ are aligned, that is, $\langle z_0, e\rangle = \|z_0\|_{c_0}\|e\|_1$. Taking into account definitions (3.13) and (3.16), this easily gives that e is finitely supported. Hence $E = H - UQV$ is a matrix polynomial, which shows that Q is rational. □

3.6 A numerical algorithm

By (3.14), the problem $\|H - UQV\|_W \to \min$ is equivalent to the problem

$$\|h - T(F)q\|_1 \to \min \tag{3.19}$$

with a matrix polynomial

$$F(\lambda) = F_0 + F_1\lambda + \ldots + F_r\lambda^r, \quad F_j \in \mathbf{R}^{n_z n_w \times n_u n_y} =: \mathbf{R}^{M\times N}.$$

We replace (3.19) by the finite problem

$$\|P_n h - P_n T(F) P_{n-r} q\|_1 \to \min. \tag{3.20}$$

Here $P_n : \ell_1 \to \ell_1$ is projection onto the first coordinates, that is,

$$P_n : \{s_0, s_1, s_2, \ldots\} \mapsto \{s_0, s_1, \ldots, s_{n-1}, 0, \ldots\}.$$

For a k-tuple $s = (s^1, \ldots, s^k) \in \ell_1^k$, we define $P_n s = (P_n s^1, \ldots, P_n s^k)$. Let $Q_n = I - P_n$. In the special case (3.11), for example, problem (3.20) amounts to minimizing

$$\left\| \begin{pmatrix} P_n h_1 \\ P_n h_2 \\ P_n h_3 \\ P_n h_4 \end{pmatrix} - \begin{pmatrix} P_n T(u_1 v_1) P_{n-r} & P_n T(u_1 v_3) P_{n-r} \\ P_n T(u_1 v_2) P_{n-r} & P_n T(u_1 v_4) P_{n-r} \\ P_n T(u_2 v_1) P_{n-r} & P_n T(u_2 v_3) P_{n-r} \\ P_n T(u_2 v_2) P_{n-r} & P_n T(u_2 v_4) P_{n-r} \end{pmatrix} \begin{pmatrix} P_{n-r} q_1 \\ P_{n-r} q_2 \end{pmatrix} \right\|_1.$$

Notice that $P_nT(u_iv_j)P_{n-r}$ may be identified with an $n \times (n-r)$ matrix, and hence we may regard the problem as the problem of finding an ℓ_1^4 optimal solution of an overdetermind system with $4n$ equations and $2(n-r)$ variables. In the general case, (3.20) involves Mn linear expressions (= "equations") and $N(n-r)$ variables.

We denote the minima in (3.19) and (3.20) by d and d_n, respectively.

Theorem 9. *We have*

$$d \le d_n + \|Q_nh\|_1$$

for all $n \ge 1$. If (3.9) is satisfied, then $d_n \to d$ as $n \to \infty$.

Proof. Let

$$d_n = \|P_nh - P_nT(F)P_{n-r}q_n^*\|_1.$$

We may think of $P_{n-r}q_n^*$ as the Taylor coefficients of a matrix polynomial D of degree at most $n-r-1$ and we know that F is a matrix polynomial of degree r. This implies that DF has at most the degree $n-1$ and hence $Q_nT(F)P_{n-r}q_n^* = 0$. It follows that

$$\begin{aligned} d_n &= \|Ph - P_nT(F)P_{n-r}q_n^*\|_1 \\ &= \|P_nh - T(F)P_{n-r}q_n^*\|_1 \\ &\ge \|h - T(F)P_{n-r}q_n^*\|_1 - \|Q_nh\|_1 \\ &\ge d - \|Q_nh\|_1, \end{aligned} \tag{3.21}$$

as claimed.

If (3.9) holds, we deduce from Corollary 2 that there is a q_0 such that

$$d = \|h - T(F)q_0\|_1. \tag{3.22}$$

Since $\|P_ny\|_1 \le \|y\|_1$ for every y, we get

$$\begin{aligned} d_n &= \|P_n(h - T(F)P_{n-r}q_0)\|_1 \\ &\le \|h - T(F)P_{n-r}q_0\|_1 \\ &\le \|h - T(F)q_0\|_1 + \|T(F)Q_{n-r}q_0\|_1 \\ &\le d + \|T(F)\|\,\|Q_{n-r}q_0\|_1. \end{aligned} \tag{3.23}$$

Combining (3.21) and (3.23) we arrive at the conclusion that $d_n \to d$. □

We remark that since H is a rational matrix function with entries in W_+, the term $\|Q_nh\|_1$ goes to zero exponentially fast. In the square case, we can say even more.

Corollary 3. *If U and V are square matrix polynomials satisfying (3.17), then there are constants $C < \infty$ and $\delta > 0$ such that*

$$d_n - Ce^{-\delta n} \le d \le d_n + Ce^{-\delta n}$$

for all $n \ge 1$.

Proof. Let $\|H - UQV\|_W = d$ and, accordingly, $\|h - T(F)q_0\|_1 = d$. We already noted that $h \in \ell_1^M$ is exponentially decaying, and hence the estimate $d \le d_n + Ce^{-\delta n}$ follows from Theorem 9. By Theorem 8, the Taylor coefficients of Q are exponentially decaying, which implies that $q_0 \in \ell_1^M$ is also exponentially decaying. This in conjunction with (3.23) gives the estimate $d_n - Ce^{-\delta n} \le d$. □

It is clear that (3.20) is the simpler to handle the smaller the defect r between n and $n - r$ is. In [29], we considered the SISO case and showed that then Corollary 3 is true with (3.20) replaced by

$$\|P_n h - P_n T(F) P_{n-\kappa} q\|_1 \to \min,$$

where κ is the number of zeros (counted with multiplicities) of the scalar polynomial F in the open unit disk. Clearly, $\kappa \le r$. However, the proof of this result is much more involved than the proofs of Theorems 9 and Corollary 3.

Theorem 9 provides us with pretty good upper bounds for d. The search for tight lower bounds motivates the following modification of problem (3.20). Note that (3.20) is equivalent to finding

$$d_n := \min_{Q_{n-r}P_n q = 0} \|P_n h - P_n T(F) P_n q\|_1,$$

because the constraint $Q_{n-r}P_n q = 0$ forces the last r components of P_n to be zero. We now fix a number $\varepsilon > 0$ and consider the problem of determining

$$\tilde{d}_n := \min_{\|Q_{n-r}P_n q\| \le \varepsilon} \|P_n h - P_n T(F) P_n q\|_1. \tag{3.24}$$

As, obviously, $\tilde{d}_n \le d_n$ and $d_n + \|Q_n h\|_1$ is only slightly larger than d, there is some hope that $\tilde{d}_n$ is a close lower bound for d. The vector-valued sequence q is living in $\ell_1^{n_u n_y}$ and as said in Section 3.4, there is no need for specifying a concrete norm in $\ell_1^{n_u n_y}$. We now may take advantage of this freedom. In fact, the requirement $\|Q_{n-r}P_n q\| \le \varepsilon$ is a constraint for Nr variables and we may choose $\|\cdot\|$ to be any vector norm on $\mathbf{R}^{Nr}$. For example, in the context of (3.11), we can take

$$\|Q_{n-r}P_n q\| = \max(\|Q_{n-r}P_n q_1\|_\infty, \|Q_{n-r}P_n q_2\|_\infty),$$

and if we write

$$P_n q_i = (q_0^{(i)}, q_1^{(i)}, \dots, q_{n-1}^{(i)}),$$

then $\|Q_{n-r}P_n q\| \le \varepsilon$ is the constraint

$$|q_{n-r}^{(1)}| \le \varepsilon, \quad \dots, \quad |q_{n-1}^{(1)}| \le \varepsilon, \quad |q_{n-r}^{(2)}| \le \varepsilon, \quad \dots, \quad |q_{n-1}^{(2)}| \le \varepsilon.$$

Since $Q_n T(F) P_{n-r} q = 0$, we have the estimate

$$\|Q_n T(F) P_n q\|_1 = \|Q_n T(F) Q_{n-r} P_n q\|_1 \le \|T(F)\| \, \|Q_{n-r}P_n q\|, \tag{3.25}$$

and $\|T(F)\|$ depends on the norm on $\mathbf{R}^{Nr}$ we have chosen but not on n.

Let q_0 satisfy $\|h - T(F)q_0\|_1 = d$ and put $e = h - T(F)q_0$.

Theorem 10. *Suppose (3.9) holds. If* $\|Q_{n-r}P_n q_0\| \le \varepsilon$ *then*

$$\tilde{d}_n - \|Q_n e\|_1 \le d,$$

and for arbitrary $n \ge 1$ *we have*

$$d \le \tilde{d}_n + \|Q_n h\|_1 + \|T(F)\|\,\varepsilon.$$

Proof. Our assumptions guarantee that

$$\begin{aligned}
\tilde{d}_n &\le \|P_n h - P_n T(F) P_n q_0\|_1 \\
&= \|P_n h - P_n T(F) q_0\|_1 \\
&= \|P_n (h - T(F) q_0)\|_1 \\
&= \|P_n e\|_1 = \|e - Q_n e\|_1 \\
&\le \|e\|_1 + \|Q_n e\|_1 = d + \|Q_n e\|_1.
\end{aligned}$$

On the other hand, let q_n^* be a solution of the minimum problem (3.24):

$$\|Q_{n-r} P_n q_n^*\| \le \varepsilon, \quad \|P_n h - P_n T(F) P_n q_n^*\|_1 = \tilde{d}_n.$$

Then

$$\begin{aligned}
d &\le \|h - T(F) P_n q_n^*\|_1 \\
&= \|P_n h + Q_n h - P_n T(F) P_n q_n^* - Q_n T(F) P_n q_n^*\|_1 \\
&\le \|P_n h - P_n T(F) P_n q_n^*\|_1 + \|Q_n h\|_1 + \|Q_n T(F) P_n q_n^*\|_1 \\
&\le \tilde{d}_n + \|Q_n h\|_1 + \|T(F)\|\varepsilon,
\end{aligned}$$

the last estimate resulting from (3.25). □

We know that $\|Q_n h\|_1 \to 0$ (exponentially fast) and $\|Q_n e\|_1 \to 0$ (in the square case even $\|Q_n e\|_1 = 0$ for all sufficiently large n). Consequently, we certainly have $\|Q_n h\|_1 \le \varepsilon$ and $\|Q_n e\|_1 \le \varepsilon$ if only n is large enough. Thus, Theorem 10 shows that $\tilde{d}_n - \varepsilon$ is a lower bound for d whenever n is large enough and, moreover, that this bound is at a distance of at most $2\varepsilon + \|T(F)\|\varepsilon$ to d. If e (or equivalently, q_0) decays exponentially, then the n's for which the lower bound $\tilde{d}_n - \varepsilon$ is applicable are certainly not of astronomic dimensions.

3.7 Numerical example

We consider the triple of polynomial matrices U, V, and H given by

$$\begin{aligned}
U(\lambda) &= \begin{pmatrix} 4.2 + 6.5\lambda + 0.0025\lambda^2 + 0.39\lambda^3 & -1.4 - 2.1\lambda + 0.11\lambda^2 - 0.17\lambda^3 \\ 2.1 - 1.3\lambda + 0.32\lambda^2 & -0.79 + 0.66\lambda - 0.13\lambda^2 \end{pmatrix} \\
V(\lambda) &= \begin{pmatrix} 0.2 + 4.5\lambda + 4.1\lambda^2 + 0.92\lambda^3 & 0.51 + 5.8\lambda + 7.7\lambda^2 + 2.4\lambda^3 \\ 0.27 + 3.5\lambda + 1.4\lambda^2 & 0.7 + 4.5\lambda + 3.6\lambda^2 \end{pmatrix} \\
H(\lambda) &= \begin{pmatrix} -3 + \lambda + 4\lambda^2 & -5 + 3\lambda + 9\lambda^2 \\ 13\lambda - 3\lambda^2 & 14 + \lambda^2 \end{pmatrix}.
\end{aligned}$$

We look for a pair of matrix functions Q and E in $W_+^{2\times 2}$ that solve the equation $UQV+E=H$ and minimize the row-sum norm of the matrix function E.

For a matrix function $S(\lambda)=\sum_{j\geq 0} S_j\lambda^j$, we define the matrix polynomial $\pi_n S$ by

$$(\pi_n S)(\lambda)=\sum_{j=0}^{n-1} S_j\lambda^j.$$

With this notation, the equation

$$P_nT(F)P_{n-r}q+P_ne=P_nh$$

is equivalent to the equation

$$(\pi_n U)(\pi_{n-r}Q)(\pi_n V)+\pi_n E=\pi_n H. \tag{3.26}$$

We solve (3.26) successively for $n=r,r+1,r+2,\ldots$. In the case at hand, the matrix polynomial $F=U\otimes V^\top$ has the degree $r=6$ and since U, V, H are of degree 4, they are not affected by π_n for $n\geq r=6$. Thus,we have to find matrix polynomials $\pi_{n-6}Q$ and $\pi_n E$ such that

$$U(\pi_{n-6}Q)V+\pi_n E=H, \quad \|\pi_n E\|_W\to\min. \tag{3.27}$$

We call n the expected degree of E and $n-6$ the expected degree of Q. If (3.27) is uniquely solvable, we refer to the degrees of the solutions $\pi_n E$ and $\pi_{n-6}Q$ as the true degrees of E and Q, respectively. Table 3.1 reports the results. The relative error is defined as

$$\|U(\pi_{n-6}Q)V+\pi_n E-H\|_W/\|(U,V,H)\|_W.$$

Thus, as seen from Table 3.1, the optimal solution pair can be found within 5 steps. This pair is

$$\begin{aligned} Q(\lambda) &= \begin{pmatrix} -6.4-15\lambda-0.5\lambda^2+0.09\lambda^3-0.6\lambda^4 & 7.6+23\lambda+11\lambda^2+0.11\lambda^3+0.8\lambda^4+0.4\lambda^5 \\ -26-59\lambda-6.9\lambda^2-0.8\lambda^3-1.5\lambda^4 & 29+92\lambda+48\lambda^2+5.3\lambda^3+2.5\lambda^4+0.97\lambda^5 \end{pmatrix} \\ E(\lambda) &= \begin{pmatrix} -2.6-6.6\lambda-9.8\lambda^2 & -4+3.3\lambda^4 \\ 0.51+7.8\lambda & 15-2.6\lambda \end{pmatrix}. \end{aligned}$$

3.8 Numerical algorithm for a solution to a square MIMO problem

The equation $U(\lambda)Q(\lambda)V(\lambda)+E(\lambda)=H(\lambda)$ analyzed in the previous sections consitutes a major computational step in solving the one-block ℓ_1-optimal control problem and the contribution of this thesis. However, the whole control design procedure is outlined here for completeness. Given a discrete-time linear time-invariant model of a generalized plant with a transfer matrix

$$P(\lambda)=\begin{pmatrix} P_{wz}(\lambda) & P_{uz}(\lambda) \\ P_{wy}(\lambda) & P_{uy}(\lambda) \end{pmatrix} \tag{3.28}$$

Table 3.1: Sequence of the optimal solutions to (3.27) for increasing $n \geq 6$.

Expected deg E	True deg E	Expected deg Q	True deg Q	$\|\pi_n E\|_W$	Relative error
6	6	0	0	25.72	1.24e-10
7	7	1	1	16.99	6.03e-10
8	8	2	2	16.27	1.65e-08
9	6	3	3	15.14	1.88e-10
10	2	4	4	14.94	1.37e-10
11	2	5	4	14.94	4.85e-10
12	2	6	4	14.94	3.12e-10
13	2	7	4	14.94	1.94e-09
14	2	8	4	14.94	2.40e-09
15	2	9	4	14.94	2.11e-09
16	2	10	4	14.94	1.86e-10
17	2	11	4	14.94	3.74e-11
18	2	12	4	14.94	5.47e-11
19	2	13	4	14.94	4.86e-11
20	2	14	4	14.94	5.76e-11

Step 1. Express the block relating the control and measured variables as a left and right ratio of polynomial matrices

$$P_{uy}(\lambda) = B_R(\lambda)A_R(\lambda)^{-1} = A_L(\lambda)^{-1}B_L(\lambda) \tag{3.29}$$

Step 2. Solve the linear equation with polynomial matrices

$$A_L(\lambda)X_R(\lambda) + B_L(\lambda)Y_R(\lambda) = I \tag{3.30}$$

Step 3. Build the three rational matrices expressed as ratios of polynomial matrix fractions

$$H(\lambda) = P_{wz}(\lambda) + P_{uz}(\lambda)Y_R(\lambda)A_L(\lambda)P_{wy}(\lambda) \tag{3.31}$$

$$U(\lambda) = P_{uz}(\lambda)A_R(\lambda) \tag{3.32}$$

$$V(\lambda) = A_L(\lambda)P_{yw}(\lambda) \tag{3.33}$$

Step 4. Keep only the numerator polynomial matrices of matrix fractions $U(\lambda)$ and $V(\lambda)$ and perform stable-unstable factorization of these polynomial matrices. A possible procedure is to convert them into Smith form. Store the result in $U(\lambda)$ and $V(\lambda)$.
Step 5. Solve the equation

$$G(\lambda) = H(\lambda) - U(\lambda)Q(\lambda)V(\lambda) \tag{3.34}$$

for $G(\lambda)$ and $Q(\lambda)$ with $Q(\lambda)$ of a given degree and minimum row-sum. Repeat this step until there is no improvement in the achieved norm of $Y(\lambda)$.

STEP 5. Substitute $Q(\lambda)$ into the expression for the controller

$$C(\lambda) = (Y_R(\lambda) + A_R(\lambda)Q(\lambda))(X_R(\lambda) - B_R(\lambda)Q(\lambda))^{-1} \tag{3.35}$$

3.9 Summary

In this chapter we extended the new theoretical and computational framework for solving the standard ℓ_1-optimal control problem introduced in the previous chapter to a MIMO case.

The problem was formulated using linear equations with polynomial matrices and therefore supplies a missing tool a designer's polynomial control design toolset, with reliable procedures for LQG, $\mathcal{H}_2$ and $\mathcal{H}_\infty$-optimal control already available.

At a mathematical level, advanced results from the theory of block Toeplitz operators were employed. The optimal control problem was formulated as a search for a minimum distance between a given vector sequence in ℓ_1 and a range of a block Toeplitz operator. Alternative derivation of existence conditions for an optimal controller was provided, which shows that an optimal controller is guaranteed to exist unless the generalized plant has poles or zeros on the unit circle. An optimal controller need not be unique. An alternative proof has also been given for the fact that an optimal impulse response of a square system is finite. These alternative theoretical derivations not only lend new insight into the problem, but also lead to new numerical algorithm, which is conceptually extremely simple, yet very reliable compared to the existing interpolation based approaches. No need to compute zeros and zero directions of polynomial matrices and no need to extract an optimal controller from a closed-loop transfer function.

For a general multiblock MIMO case, the optimal closed-loop impulse response is not finite and therefore an iterative scheme for obtaining an approximate solution was proposed. At each step, a linear equation with polynomial matrices is solved and both lower and upper bounds on the optimal norm of a closed-loop transfer function are provided. These bounds are shown to converge to the optimum value for increasing degree of a polynomial matrix approximating the optimal closed-loop transfer function.

Chapter 4

Computing the ℓ_1 norm of a polynomial matrix fraction

4.1 Introduction

This chapter presents a secondary result of this thesis, a numerical algorithm for computing the ℓ_∞-induced norm of a system described by a polynomial matrix fraction. For practical control design, knowing the exact value of this norm is not crucial, but it can be useful for comparison of an optimal design with a non-optimal one.

The modern optimal and robust control takes much advantage of viewing (the model of) a plant as an operator mapping some normed space of infinite real sequences into the same space. An operator norm is then a measure of how much the plant magnifies the exogenous (input) variables when mapping them into the regulated (output) variables.

This thesis is confided to linear time-invariant (LTI) systems for which both the exogenous and the regulated variables are known to be bounded and persistent, that is to say, they can be modeled as real sequences living in ℓ_∞. The corresponding system norm then expresses the worst-case magnification of the amplitude of disturbing exogenous variables. It is a standard result that in scalar case the ℓ_∞-induced operator norm is equal to the ℓ_1 of the impulse response of the system. Using the operator-theoretic terminology, this operator norm is equal to the Wiener norm of a generating function, called *symbol.* MIMO systems can be described as operators mapping the space ℓ_∞^n into ℓ_∞^m, where the norm of an infinite sequence $u = \{u_0, u_1, \ldots\}$ of real n-tuples in ℓ_∞^n is $\|u\|_\infty = \sup_{k\in\mathbb{Z}_+} \max_{i=1,2,\ldots,n} |u_k^i|$.

Lemma 2. *Consider a causal LTI system T with n inputs and m outputs described by a transfer matrix $G(z) = \sum_{k=0}^{\infty} G_k z^{-k}$. Assuming that input signals are persistent and bounded, the greatest possible factor, by which the system T magnifies the amplitude of the input signals, i.e., the ℓ_∞-induced norm of an operator mapping ℓ_∞^n into ℓ_∞^m, is*

$$\|T(G)\|_{\ell_{\infty-induced}} = \|G(z)\|_W = \max_{1\le i\le m} \sum_{j=1}^{n} \|G^{ij}(z)\|_W \tag{4.1}$$

where $G^{ij}(z)$ is a scalar transfer function and its norm is equal to a sum of absolute values of its impulse response.

Proof. This result is standard (see, e.g. [15]) and follows directly from systematic application of triangle inequality for norms. □

A note on notation: Note that we encounter three types of objects here, that describe the same thing. First, we use the concept of an operator/system, which can be measured using ℓ_∞-induced norm. Second, we work with functions of a complex variable, called symbols/transfer functions, that uniquely represent these operators/systems, and we measure them with Wiener norm. Third, in time domain there are infinite sequence of real matrices that can be measured using the ℓ_1 norm. To relax these notational issues, with some abuse of notation, we will use the notation ℓ_1 norm for operators, symbols and the corresponding sequences. This is generally accepted in the control community. Also, we express transfer funtions as functions of a variable $z = 1/\lambda$.

The main idea about the computation of this norm is identical to [1] but the computation is based on polynomial equations and therefore does not require conversion to state space format. Consider a stable discrete-time linear time-invariant system described by a transfer matrix $G(z) = G_0 + G_1 z^{-1} + G_2 z^{-2} + \ldots$ Its state-space realization is

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k), \;\; x(0) = 0 && (4.2) \\ y(k) &= Cx(k) + Du(k) && (4.3) \end{aligned}$$

with n inputs, m outputs and d-dimensional state space. The matric coefficients G_k in the power series G can be easily obtained by

$$G_k = \left\{ \begin{array}{r} D, \;\; k = 0 \\ CA^{k-1}B, \;\; k = 1, 2, \ldots \end{array} \right.$$

A simple method for computing the ℓ_∞-induced gain of the operator $T(G)$ is to approximate it by considering the first N_s terms only, with N_s being sufficiently large. How large? Just large enough to guarantee the required precision of the norm. A procedure to compute the norm of the *tail* system $G_{tail}(z) = G_{N_s+1} z^{-1} + G_{N_s+2} z^{-2} + \ldots$ is proposed in [1]. This norm directly gives an error introduced by truncation $\|G_{tail}\|_1 = \|G\|_1 - \|G_{FIR-approx}\|_1$. The method is based on the important Theorem 2 from [9] (with slight extension to MIMO case in [15]) that relates the ℓ_∞-induced system norm and the singular values of the associated Hankel operator. For the tail $\{CA^{N_s+1}B, CA^{N_s+2}B, \ldots\}$ having the state-space realization

$$\left[\begin{array}{c|c} A & A^{N_s}B \\ \hline C & 0 \end{array} \right] \tag{4.4}$$

the ℓ_1 norm of the coefficient sequence of the tail is bounded by

$$\sigma_{H1} \leq \|G_{tail}\|_1 \leq 2\sqrt{m}(\sigma_{H1} + \sigma_{H2} + \ldots + \sigma_{Hd}) \tag{4.5}$$

where $\sigma_{H1} \geq \sigma_{H2} \geq \ldots \geq \sigma_{Hd}$ are the singular values of the associated Hankel operator and d is the McMillan degree of the system. It is shown in [1], that with increasing N_s, the difference between the lower and upper bounds converges monotonically to zero and therefore a binary search can be used to find the least N_s guaranteeing the required precision.

4.2 Discussion of computational steps for polynomial matrices

Two versions of a complete algorithm for l_1 norm computation differing in a way they compute the polynomial matrix fraction description of the plant are given at the next two sections. Both algorithms are composed of several computational steps. These are briefly discussed below.

4.2.1 Computing the polynomial matrix fraction for the tail

Computing the ℓ_1 norm of a polynomial matrix fraction via impulse response truncation necessarily introduces some error. To evaluate how large the error is for a given number N_s of samples, the polynomial matrix fraction description of the tail must be obtained. Basically, there are two ways to accomplish this. Either exactly via division of polynomial matrices or approximately using FFT-based long division of polynomial matrices and a matrix version of the well-know relations between Markov parameters and coefficients of the transfer function.

Shift the impulse response (and consequently also the system description) forward by N_s steps. The resulting polynomial matrix fraction is obviously noncausal. Divide the polynomial matrices to extract the strictly proper part, which now represents the tail of the system. To state this more formally, let $N(z)$ and $D(z)$ are polynomial matrices of appropriate dimensions defining the transfer function via the left polynomial matrix fraction $\hat{G}(z) = D(z)^{-1}N(z)$. The polynomial matrix $D(z)$ is assumed Schur stable, nonsingular and row reduced. The tail $\hat{G}_{tail}(z) = D_{tail}(z)^{-1}N_{tail}(z)$ of $\hat{G}(z)$ obtained by truncating the first N_s terms of the impulse response is given by

$$\begin{aligned} N_{tail}(z) &= z^{N_s}N(z) - D(z)Q(z) \\ D_{tail}(z) &= D(z) \end{aligned} \tag{4.6}$$

where $Q(z)$ is a polynomial matrix determined uniquely by the condition $\deg N_{tail}(z) < \deg D_{tail}(z)$ with the inequality extending to the corresponding row degrees of $N_{tail}(z)$ and $D_{tail}(z)$, i.e., $D_{tail}(z)$ is row-reduced.

4.2.2 Computing the singular values of Hankel operator

Hankel operator describes relation between past inputs and future outputs and as such it doesn't depend on a particular state-space realization. A well-established method for computing the singular values of Hankel operator in state-space is due to Glover [23]. Young [58] presents an attempt to circumvent the realization step using directly the coefficients of the polynomial matrices. Kwakernaak [41] then gives a revised and improved version of Young's algorithm based on two-sided polynomial equations. The algorithm will not be described here, the reader is referred to the original paper. Nonetheless, the computational steps of the algorithm are listed in the algorithm summary in the next section. This is aimed to give a picture about the numerical properties of the proposed solution to ℓ_1 norm computation.

4.2.3 Long division of polynomial matrices via FFT

Once the (lowest necessary) number N_s of samples that is necessary to guarantee the required precision is known, the only remaining task is to compute the ℓ_1 norm of the Finite Impulse Response (FIR) approximation of the original operator. As the number N_s of samples is usually quite high, it is reasonable to compute the impulse response of the FIR system only approximately using Fast Fourier Transform (FFT). This is also why it is advantageous to consider $N_s = 2^w - 1,\ w > 0,\ w \in \mathbb{Z}$ only. The FFT-based approach towards manipulation with polynomial matrices is now standard and widely employed in modern algorithms. The idea is to use FFT to evaluate the polynomial matrix fraction $D(z)^{-1}N(z)$ at (complex) points along the unit circle and then to apply inverse FFT on the sequence of N_s constant matrices $D(z_i)^{-1}N(z_i),\ i = 1, 2, \ldots, 2^w$. See [26] for details.

4.2.4 Identification from the truncated long division

Let $\hat{G}(z) = G_0 + G_1 z^{-1} + G_2 z^{-2} + \ldots$ is a matrix semiinfinite formal power series in indeterminate z obtained by (left) long division of polynomial matrices $D(z)$ and $N(z)$. Set $G_i = 0,\ i = 0, 1, \ldots, N_s$ to obtain the tail $G_{tail}(z) = z^{N_s}(G_{N_s+1} z^{-(N_s+1)} + G_{N_s+2} z^{-(N_s+2)} + \ldots)$. Let $N_{tail}(z)$ and $D_{tail}(z)$ be left coprime polynomial matrices such that $G_{tail}(z) = D_{tail}(z)^{-1} N_{tail}(z)$. From (4.6) it is clear that $D_{tail}(z) = D(z) = D_0 + D_1 z + \ldots + D_d z^d$ and therefore the only polynomial matrix that remains to be computed is $N_{tail}(z) = N_{0_{tail}} + N_{1_{tail}} z + \ldots + N_{d_{tail}} z^d$. can be obtained by straightforward extension of the standard relation between Markov parameters and coefficients of a scalar transfer function. Equating the coefficient matrices of equal powers in the equation $D_{tail}(z) G_{tail}(z) = N_{tail}(z)$ we get $N_d = 0$ and

$$\begin{pmatrix} N_{0_{tail}}^T \\ N_{1_{tail}}^T \\ \vdots \\ N_{d-1_{tail}}^T \end{pmatrix}^T = \begin{pmatrix} D_1^T \\ D_2^T \\ \vdots \\ D_d^T \end{pmatrix}^T \begin{pmatrix} G_{N_s+1} & 0 & \cdots & 0 \\ G_{N_s+2} & G_{N_s+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ G_{N_s+d} & G_{N_s+d-1} & \cdots & G_{N_s+1} \end{pmatrix}$$

4.3 Algorithm l1norm-I

Two versions of the algorithm for computing the ℓ_1 norm of a polynomial matrix fraction are presented differing in the way they compute the polynomial matrix fraction description of the tail of the truncated sequence of matrices. The algorithm L1NORM-I described in this section computes the tail exactly via standard Euclidean division algorithm, while the algorithm L1NORM-II given in the next section computes the polynomial matrix fraction description of the tail approximately via the FFT and identification routines.

INPUT.

1. Polynomial matrices $N(z)$ and $D(z)$ of appropriate dimensions with $D(z)$ square, roots inside the unit circle, row reduced. The fraction $D(z)^{-1}N(z)$ is an $m \times n$ stable, strictly proper transfer matrix.

2. Tolerance ε (a reasonable value is 10^{-4} or 10^{-6}).

OUTPUT. The ℓ_1 norm of a polynomial matrix fraction computed within a given tolerance ε.

STEP 1. Set $N_s = 2^w - 1$ (reasonable initial guess is $N_s = 63$).

STEP 2. Apply N_s times the forward shift operator z to G:

$$\hat{G}_{shifted}(z) = z^{N_s} D(z)^{-1} N(z) \tag{4.7}$$

STEP 3. Extract the strictly proper part from $\hat{G}_{shifted}(z)$ performing polynomial matrix division:

$$z^{N_s} D(z)^{-1} N(z) = Q(z) + D(z)^{-1} N_{tail}(z)$$

The tail is then described by the polynomial matrices $N_{tail}(z) = z^{N_s}N(z) - Q(z),\ D_{tail}(z) = D(z)$ defining the strictly proper transfer matrix

$$\hat{G}_{tail}(z) = D_{tail}(z)^{-1} N_{tail}(z) \tag{4.8}$$

with $D_{tail}(z)$ row reduced.

STEP 4. Perform left-to-right conversion

$$D(z)^{-1}N(z) = \bar{N}(z)\bar{D}(z)^{-1}$$

with $\bar{D}(z)$ column reduced.

STEP 5. Compute the singular values $\sigma_{H1}, \sigma_{H2}, \ldots, \sigma_{Hk}$ of Hankel operator associated with $\hat{G}_{tail}(z)$ (for details see [41]).

1. Perform left-to-right polynomial matrix fraction conversion

$$\bar{D}^{\sim}(z)^{-1}V(z) = \tilde{V}(z)\tilde{D}(z)^{-1}$$

where $\bar{D}^{\sim}(z)^{-1}V(z)$ denotes adjoint $\bar{D}^T(\frac{1}{z})$ multiplied by $z^{deg(D(z))}$, $V(z)$ is a polynomial matrix with columns $e_i z^v$, $v = 0, 1, \ldots, c_i - 1$, $i = 1, 2, \ldots, m$, c_i is a column degree of $\bar{D}(z)$, and e_i is the ith m-dimensional unit vector.

2. Solve the bilateral linear polynomial matrix equation

$$N_{tail}(z)\tilde{V}(z) = A(z)\tilde{D}(z) + D(z)B(z)$$

for polynomial matrices $A(z)$ and $B(z)$ such that $D(z)^{-1}A(z)$ is strictly proper.

3. Perform left-to-right polynomial matrix fraction conversion

$$D(z)^{-1}A(z) = \check{A}(z)\check{D}(z)^{-1}$$

4. Solve two bilateral symmetric linear polynomial matrix equations

$$\begin{aligned} \tilde{V}^{\sim}(z)\tilde{V}(z) &= \tilde{D}^{\sim}(z)C_c(z) + C_c^{\sim}(z)\tilde{D}(z) \\ \check{A}^{\sim}(z)\check{A}(z) &= \check{D}^{\sim}(z)C_r(z) + C_r^{\sim}(z)\check{D}(z) \end{aligned}$$

for square polynomial matrices $C_c(z)$ and $C_r(z)$ such that $C_c(z)\tilde{D}(z)^{-1}$ and $C_r(z)\check{D}(z)^{-1}$ are strictly proper.

5. Compute the Gramians Γ_c and Γ_r of the bases of the cokernel and the range of Hankel operator respectively as

$$\begin{aligned} \Gamma_c &= C_{cl}\tilde{D}_l^{-1} \\ \Gamma_r &= C_{rl}\check{D}_l^{-1} \end{aligned}$$

where $\tilde{D}_l$ and $\check{D}_l$ are leading coefficient matrices of the column reduced matrices $\tilde{D}(z)$ and $\check{D}(z)$ respectively. The matrices C_{cl} and C_{rl} are the associated leading coefficient matrices of $C_c(z)$ and $C_r(z)$ respectively.

6. Factor the Gramians as $\Gamma_c = T_c^T T_c$ and $\Gamma_r = T_r^T T_r$.

7. Compute the singular value decomposition $T_r T_c^{-1} = U\Sigma W^H$, where $\Sigma = diag(\sigma_{H1}, \sigma_{H2}, \ldots, \sigma_{Hk}, 0, \ldots, 0)$

Step 6. Compute the difference δ_{bounds} between the lower and upper bounds on the ℓ_1 norm of the operator described (in frequency domain) by $\hat{G}_{tail}(z)$:

$$\delta_{bounds} = 2\sqrt{m}(\sigma_{H1} + \sigma_{H2} + \ldots + \sigma_{Hk}) - \sigma_{H1} \tag{4.9}$$

Step 7. Check the achieved precision and if insufficient,increase the number of samples N_s: if $\delta_{bounds} \geq \epsilon$ then $w = 2w,\ N_s = 2^w - 1$, goto *Step 2.*

Step 8. Compute the first N_s terms of the long division of the original polynomial matrix fraction $D(z)^{-1}N(z)$ using FFT algorithm:

1. Compute the FFT transform $N(z_i)$ $i = 1, 2, \ldots, 2^w$ of the sequence of $n \times m$ coefficient matrices N_i padded with zero trailing matrices.

2. Compute the FFT transform $D(z_i)$ $i = 1, 2, \ldots, 2^w$ of the sequence of $n \times n$ coefficient matrices D_i padded with zero trailing matrices.

3. Compute the inverse FFT transform G_i $i = 1, 2, \ldots, 2^w$ of the sequence of $n \times m$ matrices $D(z_i)^{-1}N(z_i)$. The first N_s terms of the long division are now given by $G_0 + G_1 z^{-1} + G_2 z^{-2} + \ldots + G_{N_s} z^{-N_s}$.

Step 9. Find the system ℓ_1 norm of a FIR (truncated) approximation of the operator G:

$$\|G\|_1 \approx \|G_{FIR-approx}\|_1 = \max_{1\leq i\leq m} \sum_{j=1}^{n} \sum_{k=0}^{N_s} |G_{ij}(k)| \tag{4.10}$$

4.4 Algorithm l1norm-II

Input.

1. Polynomial matrices $N(z)$ and $D(z)$ of appropriate dimensions with $D(z)$ square, roots outside the unit circle, row reduced. The fraction $D(z)^{-1}N(z)$ is an $m \times n$ strictly proper transfer matrix.

2. Tolerance ε (reasonable value is 10^{-4} or 10^{-6}).

Output. The ℓ_1 norm of a polynomial matrix fraction computed within a given tolerance ε.
Step 1. Set $N_0 = 2^w - 1$ for some w large enough (reasonable initial guess is $N_0 = 1023$).
Step 2. Perform long division $D(z)^{-1}N(z)$ of the two polynomial matrices $N(z)$ and $D(z)$ using FFT interpolation approach to obtain a sequence of matrices $G(0), G(1), \ldots, G(N_0)$.
Step 3. Use the last $d = N_0 - N_s$ samples to compute the polynomial matrix fraction description of the tail $\hat{G}_{tail}(z) = D_{tail}(z)^{-1}N_{tail}(z)$ via solving

$$\begin{pmatrix} N^T_{0_{tail}} \\ N^T_{1_{tail}} \\ \vdots \\ N^T_{d-1_{tail}} \end{pmatrix}^T = \begin{pmatrix} D_1^T \\ D_2^T \\ \vdots \\ D_d^T \end{pmatrix}^T \begin{pmatrix} G_{N_s+1} & 0 & \cdots & 0 \\ G_{N_s+2} & G_{N_s+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ G_{N_s+d} & G_{N_s+d-1} & \cdots & G_{N_s+1} \end{pmatrix}$$

for $N_{tail}(z) = N_{0_{tail}} + N_{1_{tail}} z + \ldots + N_{d-1_{tail}} z^d$ with $D_{tail}(z) = D(z)$.
Step 4. Compute the singular values $\sigma_{H1}, \sigma_{H2}, \ldots, \sigma_{Hd}$ of Hankel operator associated with $\hat{G}_{tail}(z)$ using the same Kwakernaak's algorithm [41] as in *Step 4* in the algorithm l1norm-I.
Step 5. Compute the difference δ_{bounds} between the lower and upper bounds on the ℓ_1 norm of the operator described (in frequency domain) by $\hat{G}_{tail}(z)$:

$$\delta_{bounds} = 2\sqrt{m}(\sigma_{H1} + \sigma_{H2} + \ldots + \sigma_{Hd}) - \sigma_{H1} \tag{4.11}$$

STEP 6. Check the achieved precision and if insufficient, increase the number of samples N_0: if $\delta_{bounds} \geq \epsilon$ then $w = 2w,\ N_0 = 2^w - 1$, goto *Step 2.*
STEP 7. Compute the ℓ_1 system norm of a FIR approximation corresponding to the (already computed) first N_s samples $G(k)$ by

$$\|G\|_1 \approx \|G_{FIR-approx}\|_1 = \max_{1 \leq i \leq m} \sum_{j=1}^{n} \sum_{k=0}^{N_s} |G_{ij}(k)|$$

Remark: In practice, there is almost no computational effort associated with considering all the (already computed) $N_0 = N_s + d$ terms in the above relation. The precision can be further improved.

4.5 Example

For the ease of comparison, the data is taken from [15] because the state-space solution to the ℓ_1 norm computation is given there. Consider the following 2 × 2 MIMO transfer function with state-space realization

$$\left[\begin{array}{ccc|cc} 1 & -0.6 & 0.8 & 0 & 0 \\ 0.6 & 0.4 & -0.5 & 1 & 0 \\ -0.3 & 0.1 & -0.9 & 0 & 1 \\ \hline 1 & 0 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & -0.5 & 0 \end{array}\right] \tag{4.12}$$

The left polynomial matrix fraction $D(z)^{-1}N(z)$ corresponding to the state-space realization (4.12) is

$$N(z) = \left[\begin{array}{ll} 0.9 + 0.57z - 0.73z^2 & -0.79 - 0.25z - 0.9z^2 \\ -0.34 + 0.16z & 1 - 0.53z \end{array}\right]$$

$$D(z) = \left[\begin{array}{ll} 0.21 - 0.25z - 0.9z^2 & -0.59 - 0.26z + 0.56z^2 \\ 1 - 0.53z & 0.021 - 0.85z \end{array}\right]$$

The ℓ_1 norm computed using L1NORM-I algorithm is

$$\|D^{-1}(z)N(z)\|_1 = 9.7441 \pm 10^{-4}$$

using $N = 128$ samples of the impulse response.

4.6 Summary

A new algorithm for computing ℓ_1 norm of a polynomial matrix fraction has been described in this chapter. The algorithm uses directly the coefficients of the polynomial matrix fraction description and thus avoids the realization step. Two variants of the algorithm were developed differing in the way they compute the polynomial matrix fraction description of the so-called tail system. Numerical example has been given to show that the results agree with those obtained using the state-space algorithm.

Obviously, the most critical step is the computation of the bases of the Hankel operator. This requires solving two symmetric and one general two-sided linear equations with polynomial matrices. Currently, this cannot compete with the state space approach based on solving two Lyapunov equations to obtain grammians.

Chapter 5

Modular arithmetics for polynomial matrices

5.1 Introduction

As a side-product of the work on ℓ_1-optimal control, this chapter gives two efficient algorithms for modular arithmetics for polynomial matrices: one for modular shift of a polynomial matrix and one for modular multiplication of two polynomial matrices. Both algorithms avoid division by another polynomial matrix. The algorithms are are just extensions of well-known principles for scalar polynomials. It is shown in this work that row and column reducedness of a polynomial matrix are the right concepts for this extension.

The relevance of modular shift for the ℓ_1-optimal control is that the presented algorithm can replace division of two polynomial matrices in those steps where a polynomial matrix fraction of a tail after truncation must be obtained. Superiority of the algorithm over the naive division-based approach is demonstrated using a numerical example with random data. The whole story with modular arithmetics for polynomial matrices is of independent control-theoretic interest, though.

In scalar case, the modular shift operation describes evolution of a state vector of a linear system. This is known as *Nerode equivalence* [48] (also described for example in [33], pp. 316). In matrix case, a strictly proper part of the polynomial matrix fraction $D(z)^{-1}\hat{N}(z)$ for $\hat{N}(z) = z^k N(Z)$ describes a *tail system*, i.e. the system with the same impulse response as the error sequence that is cut off after FIR truncation, which is a useful computational step in ℓ_1-optimal control [30], [21].

The chapter is organized in several sections. The following, second section gives some definitions used in the chapter. The third section presents the numerical algorithm for a left modular shift of a polynomial matrix. The fourth section extends the idea used for modular shift to modular multiplication of two polynomial matrices and offers a fast algorithm. The final, fifth section reports on numerical experiments and compares the performance of the presented algorithms with the naive ones.

5.2 Basic definitions

Definition 1 (k-step shift)**.** *Let $N(z)$ be a polynomial matrix in complex variable z. A shifted polynomial matrix $\hat{N}(z)$ is defined as $N(z)$ with the powers of the complex variable increased by k*

$$\hat{N}(z) = z^k N(z) \tag{5.1}$$

Definition 2 (Left modular k-step shift)**.** *Let $\hat{N}(z) = z^k N(z)$ be a shifted polynomial matrix and $D(z)$ be another square nonsingular polynomial matrix such that the left polynomial matrix division $D(z)^{-1}\hat{N}(z)$ is well defined. Left $D(z)$-modular shift of a polynomial matrix $\bar{N}(z)$ is the (unique) remainder in the left division of the shifted polynomial matrix $\hat{N}(z)$ and $D(z)$*

$$\begin{aligned} \bar{N}(z) &= \hat{N}(z) \mod D(z) && (5.2) \\ &= zN(z) - D(z)Q(z) && (5.3) \end{aligned}$$

where $Q(z)$ is the quotient determined uniquely by the polynomial matrices $\hat{N}(z)$ and $D(z)$, with the rational matrix $D^{-1}(z)\bar{N}(z)$ strictly proper, i.e. $\lim_{z\to\infty} D^{-1}(z)\bar{N}(z) = 0$*.*

Definition 3 (Left modular multiplication of two polynomial matrices)**.** *Consider three polynomial matrices $A(z)$, $B(z)$ and square nonsingular $C(z)$ such that a left polynomial matrix division $C(z)^{-1}A(z)B(z)$ is well defined. The result of left modular multiplication of $A(z)$ and $B(z)$ is the unique remainder after left division of $A(z)B(z)$ by $C(z)$.*

5.3 Algorithm for left modular shift of a polynomial matrix

Consider two polynomial matrices $N(z)$ and $D(z)$ of appropriate sizes such that the left polynomial matrix division $D(z)^{-1}N(z)$ defines a matrix of strictly proper rational functions. In other words, the polynomial matrix $N(z) = N(s) \mod D(z)$. Assume that $D(z)$ is row-reduced. If it is not, it can be always made so [33] at some additional computational cost. Now, write the polynomial matrix $D(z)$ as follows

$$\begin{aligned} D(z) \;=\; & \begin{pmatrix} z^{k_1} & 0 & \dots & 0 \\ 0 & z^{k_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & z^{k_n} \end{pmatrix} D_q + \begin{pmatrix} z^{k_1-1} & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & z^{k_n-1} \end{pmatrix} D_{q-1} + \dots \\ & + \begin{pmatrix} * & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & * \end{pmatrix} D_0 \end{aligned} \tag{5.4}$$

where $k_i,\ i = 1, \dots, n$ are the row degrees of $D(z)$, $q = \max_{1 \leq i \leq n} k_i$ and D_q is the leading row coefficient matrix and the star denotes an element that is either 1 or 0. Under the assumption of row-reducedness of $D(z)$, the constant matrix D_q is nonsingular, i.e., $detD_q \neq 0$. Perform the same decomposition of the polynomial matrix $N(z)$

$$
\begin{aligned}
N(z) &= \begin{pmatrix} z^{k_1-1} & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & z^{k_n-1} \end{pmatrix} N_{q-1} + \begin{pmatrix} z^{k_1-2} & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & z^{k_n-2} \end{pmatrix} N_{q-2} + \dots \\
&+ \begin{pmatrix} * & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & * \end{pmatrix} N_0
\end{aligned}
\tag{5.5}
$$

The result that follows uses the arguments used for scalar polynomials in [33], pp.337, with the role of the leading coefficient of the scalar polynomial taken over by the leading row coefficient matrix that is guaranteed to be nonsingular.

Lemma 3 (k-step left modular shift of a polynomial matrix)**.** *Consider two polynomial matrices $N(z)$ and $D(z)$ of appropriate sizes such that the left polynomial matrix division $D(z)^{-1}N(z)$ defines a matrix of strictly proper rational functions, $D(z)$ being row-reduced. Using the* row degree *decomposition of the polynomial matrices as in (5.4) and (5.5), the relationship between the constant matrices of $N(z)$ and $\bar{N}(z) = z^k N(z) \mod D(z)$ is given*

$$
\begin{pmatrix} \bar{N}_0 \\ \bar{N}_1 \\ \vdots \\ \bar{N}_{q-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & -D_0 D_q^{-1} \\ I & 0 & 0 & \dots & 0 & -D_1 D_q^{-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & I & -D_{q-1} D_q^{-1} \end{pmatrix}^k \begin{pmatrix} N_0 \\ N_1 \\ \vdots \\ N_{q-1} \end{pmatrix}
\tag{5.6}
$$

Proof. First, prove the lemma for one-step shift. Combining the definition (5.1) of a shifted polynomial matrix $\hat{N}(s)$ and the equation (5.5)

$$
\begin{aligned}
zN(s) &= \begin{pmatrix} z^{k_1} & 0 & \dots & 0 \\ 0 & z^{k_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & z^{k_n} \end{pmatrix} N_{q-1} + \begin{pmatrix} z^{k_1-1} & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & z^{k_n-1} \end{pmatrix} N_{q-2} + \dots \\
&+ \begin{pmatrix} *z & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & *z \end{pmatrix} N_0
\end{aligned}
\tag{5.7}
$$

Substituting for the diagonal matrix $diag(z^{k_1}, z^{k_2}, \dots, z^{k_n})$ in (5.7) from (5.4) and comparing the matrix terms with equal (diagonal) powers the lemma follows for k=1. The extension to $k > 1$ can be devised easily because the resulting constant matrix $\bar{N}$ can be placed at the position of matrix N in (5.6) to obtain a two-step modular shift. By induction, the lemma follows. □

5.4 Algorithm for left modular multiplication of polynomial matrices

The idea is fairly simple yet computationally efficient: express the multiplication of two polynomial matrices as a sequence of modular shifts and apply the fast algorithm from the previous section.

Lemma 4 (Algorithm for left modular multiplication of two polynomial matrices)**.** *Given polynomial matrices $A(z)$, $B(z)$ and $C(z)$ of appropriate sizes with $C(z)$ square and nonsingular and of degree q, the left modular multiplication can be computed in the following steps*

1. *express the $C(z)$ matrix in the row degree form (5.4) and build the block companion matrix M according to (5.6)*

2. *initialize the power index $i = 1$, the actual companion matrix $\bar{M} = M$ and the resulting polynomial matrix $Y(z) = 0$*

3. *compute $A_b(z) = A(z)B_i$*

4. *express the auxiliary polynomial matrix $A_b(z)$ in the row degree form (5.5) and concatenate the matrix coefficients into a block vector*

$$A_b = \begin{pmatrix} A'_{b0} & A'_{b1} & \dots & A'_{b(q-1)} \end{pmatrix}'$$

5. *compute $A_b(z)z^i \mod C(z)$ as $\bar{A}_b = \bar{M}A_b$*

6. *convert the polynomial matrix in row degree form represented by the block column vector and row degrees to the standard matrix polynomial format and add this to the polynomial matrix $Y(z)$ storing the result.*

7. *if $i < \deg B(z)$ then $i = i + 1$, update the block companion matrix $\bar{M} = M\bar{M}$ and go to step 3 otherwise end*

Proof. Self-evident. The only issue that might not be obvious at first is that the second step cannot break the strict properness of the polynomial matrix fraction. Indeed, $A(z)$ is assumed to have row degrees strictly less than the corresponding row degrees of $C(z)$. Multiplying $A(z)$ by an arbitrary constant matrix from the right cannot increase the row degrees. The strict properness is thus preserved. □

Table 5.1: Typical computation times [s] and relative errors for MTIMESMOD and MTIMES-LDIV algorithms.

Size and degree	MTIMESMOD	MTIMES-LDIV	Relative error
m=n=5, d=5	0.15	0.04	2.8681e-014
m=n=5, d=15	0.34	0.09	3.8701e-015
m=n=10, d=10	0.31	0.19	2.2448e-014
m=n=10, d=20	0.79	1.03	7.7952e-016
m=n=20, d=10	0.58	1.10	1.7120e-014
m=n=20, d=20	2.30	6.98	2.1058e-015

Remark: It appears wise to take advantage of the special structure of the companion matrix for the update in step 7. This multiplication operation can be efficiently computed as:

$$\bar{M} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & -C_0C_q^{-1} \\ I & 0 & 0 & \dots & 0 & -C_1C_q^{-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & I & -C_{q-1}C_q^{-1} \end{pmatrix}^k \begin{pmatrix} \bar{M}_0 \\ \bar{M}_1 \\ \vdots \\ \bar{M}_{q-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{M}_0 \\ \vdots \\ \bar{M}_{q-2} \end{pmatrix} + \begin{pmatrix} -C_0C_q^{-1}\bar{M}_{q-1} \\ -C_1C_q^{-1}\bar{M}_{q-1} \\ \vdots \\ -C_{q-1}C_q^{-1}\bar{M}_{q-1} \end{pmatrix} \tag{5.8}$$

5.5 Numerical experiments

The platform on which the numerical experiments were performed was: PC, Intel Pentium 4, CPU 1300MHz, 512 MB RAM, Microsoft Windows 2000, Matlab 6.5, Polynomial Toolbox 3.0 [42]. Accuracy and computational speed were compared with random data between the proposed algorithm MTIMESMOD and the standard approach MTIMES-LDIV based on multiplicaton followed by Euclidean division. Some results are reported in Table 5.1 and visualised in in Figures 5.1 and 5.2.

5.6 Summary

Two algorithms for modular arithmetics with polynomial matrices were described: modular shift and modular multiplication. Numerical experiments confirm computational superiority over naive approaches based on division of two polynomial matrices. This is significant for larger problems (size and degree of polynomial matrices, number of steps of a modular shift). The presented algorithms avoid division altogether. Practical interpretation of the modular shift operation was given. A research is under way that aims at utilizing the modular multiplication in solving linear Diophantine equations with polynomial matrices.

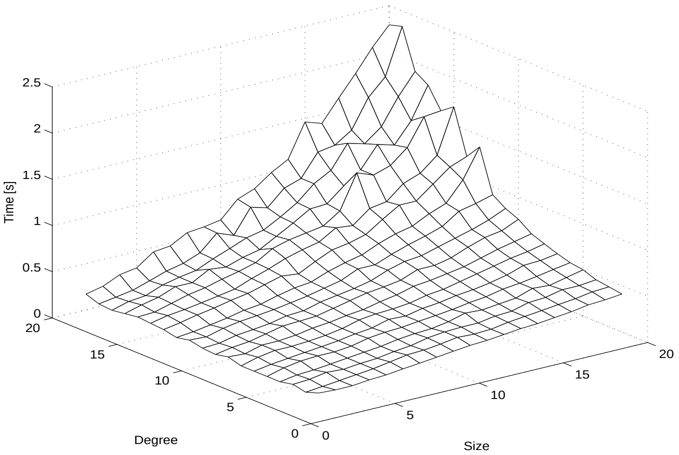

Figure 5.1: Typical computation times for the new MTIMESMOD algorithm for modular multiplication

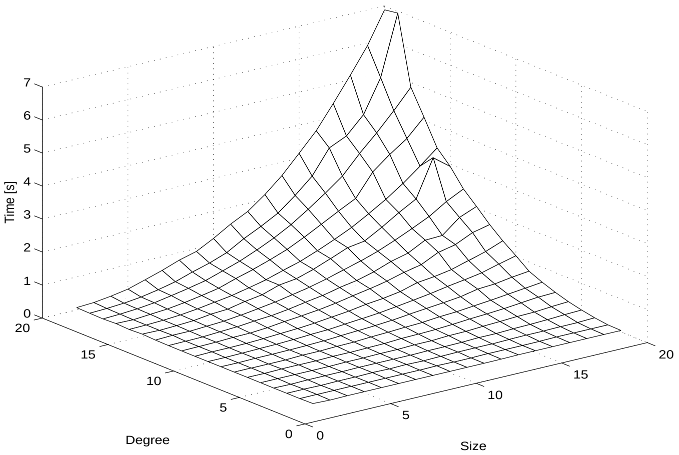

Figure 5.2: Computation times for the naive MTIMES-LDIV algorithm for modular shift

Chapter 6

Conclusions

6.1 Summary

In this thesis I formulated and solved the standard problem of a design of an ℓ_1-optimal feedback controller within the convenient framework of polynomials and polynomial matrices. I provided alternative proofs for existence of an optimal controller and proposed reliable interpolation-free numerical algorithms. In SISO and one-block MIMO cases I showed that an optimal solution can be achived by solving a linear equation with polynomial matrices. In a general multiblock MIMO case, I proposed an approximation method that solves a sequence of linear equations with polynomial matrices and gives a converging upper and lower bounds on the optimal value of the norm. Superiority of the proposed methodology over the interpolation-based algorithms was shown by means of numerical experiments for a scalar plant. Addidionaly, I devised efficient and reliable numerical algorithms for computing the norm of a polynomial matrix fraction. Finally, stating this important control problem in the language of polynomials and polynomial matrices, I believe that I opened a new angle of attack to this challenging problem and contributed to better understanding of this promising yet not mature optimal control strategy.

6.2 Future research

This thesis laid a theoretical and computational foundation of a direct polynomial approach to ℓ_1-optimal control, but it was out of my reach during this PhD period to bring this methodology to maturity and turn it into a viable control design tool. A lot of work remains to be done. A few hot topics of immediate need are outlined bellow.

- **Reliable algorithms for stable-unstable factorization of polynomial matrix.** While avoiding the computation of zeros of a polynomial matrix turned out numerically very pleasing, separation of stable and unstable part of a polynomial matrix is currently not a reliable step. Fast and reliable routines are available for scalar polynomials but when it comes to polynomial matrices, the only available procedure is a conversion to the Smith form and obtaining the scalar stable-unstable factorization for each invariant factor. This is numerically very weak.

- **Handling the case of zeros or poles at 1 and -1.** From a practical point of view, this is a very common situation. Either the original plant has an integrator that maps to a pole at 1 after discretization or it is strictly proper, which shows as a zero a -1 when using Tustin discretization method. However, a general cure to this problem seems unobtainable in ℓ_1-optimal control for the reasons outlined in the first chapter.

Bibliography

[1] V. Balakrishnan and S. Boyd. On computing the worst-case peak gaing of linear systems. *Systems and Control Letters*, 19:265–269, 1992.

[2] G. Baxter. A norm inequality for a finite-section Wiener-Hopf equation. *Illinois J. Math.*, 7:97–103, 1963.

[3] D. A. Bini. Using FFT-based techniques in polynomial and matrix computations: recent advances and applications. *Numer. Funct. Anal. Optim.*, 21:47–66, 21.

[4] D. A. Bini and A. Böttcher. Polynomial factorization through Toeplitz matrix computations. *Linear Algebra Appl.*, 366:25–37, 2003.

[5] A. Böttcher and S. M. Grudsky. *Toeplitz matrices, Asymptotic Linear Algebra, and Functional Analysis.* Birkhauser Verlag, 2000.

[6] A. Böttcher, Yu. I. Karlovich, and I. M. Spitkovsky. *Convolution Operators and Factorization of Almost Periodic Matrix Functions.* Birkhäuser Verlag, Basel, 2002.

[7] A. Böttcher and B. Silbermann. *Introduction to large truncated Toeplitz matrices.* Springer-Verlag New York, 1999.

[8] A. Böttcher and B. Silbermann. *Spectral Properties of Toeplitz Band Matrices.* a book to appear, 2004.

[9] S. Boyd and J. Doyle. Comparison of peak and rms gains for discrete-time systems. *Systems and Control Letters*, 9:1–6, 1987.

[10] S. P. Boyd and C. H. Barratt. *Linear Controller Design: Limits of Performance.* Prentice-Hall, 1991.

[11] A. Casavola. A polynomial approach to the ℓ_1-mixed sensitivity optimal control problem. *IEEE Transactions on Automatic Control*, 41:751–756, May 1996.

[12] A. Casavola and D. Famularo. Q domain sub/super-optimization linear programming methods for MIMO ℓ_1 control problems. In *Proceedings of the 4th IFAC Symposium on Robust Control Design ROCOND'00*, Prague, Czech Republic, 2000.

[13] A. Casavola and D. Famularo. MIMO ℓ^1 optimal control problems via the polynomial equation approach. *International Journal of Control*, 76(8):823–835, August 2003.

[14] M. A. Dahleh. BIBO stability robustness in the presence of coprime factor perturbations. *IEEE Transactions on Automatic Control*, 37(3):352–355, March 1992.

[15] M. A. Dahleh and I. J. Diaz-Bobillo. *Control of Uncertain Systems: A Linear Programming Approach.* Prentice-Hall, Englewood Cliffs, NJ 07632, USA, 1995.

[16] M. A. Dahleh and Y. Ohta. A necessary and sufficient condition for robust BIBO stability. *Systems and Control Letters*, 11:271–275, 1988.

[17] M. A. Dahleh and J.B. Pearson, Jr. ℓ^1-optimal feedback controllers for MIMO discrete-time systems. *IEEE Transactions on Automatic Control*, AC-32(4):314–322, April 1987.

[18] M. A. Dahleh and J.B. Pearson, Jr. Optimal rejection of persistent disturbances, robust stability, and mixed sensitivity minimization. *IEEE Transactions on Automatic Control*, 33(8):722–731, August 1988.

[19] G. Deodhare and M. Vidyasagar. ℓ_1-optimality of feedback control systems: The SISO discrete-time case. *IEEE Transactions on Automatic Control*, 35(9):1082–1085, September 1990.

[20] I. J. Diaz-Bobillo and M. A. Dahleh. Minimization of the maximum peak-to-peak gain: The general multiblock problem. *IEEE Transactions on Automatic Control*, 38(10):1459–1482, October 1993.

[21] N. Elia and M. A. Dahleh. A quadratic programming approach for solving the ℓ_1 multiblock problem. *IEEE Transactions on Automatic Control*, 43(9):1242–1252, September 1998.

[22] N. Elia and M. A. Dahleh. Minimization of the worst case peak-to-peak gain via dynamic programming: State feedback case. *IEEE Transactions on Automatic Control*, 45(4):687–701, April 2000.

[23] K. Glover. All optimal hankel-norm approximations of linear multivariable systems and their $\mathcal{L}_\infty$-error bounds. *International Journal of Control*, 39:1115–1193, 1984.

[24] I. Gohberg and I. A. Feldman. *Convolution Equations and Projection Methods for Their Solution.* American Mathematical Society, Providence, RI, 1974.

[25] M. E. Halpern, R. J. Evans, and R. D. Hill. Pole placement for ℓ_1-suboptimal design. In 34^{th} *Conference on Decision & Control*, pages 1201–1204, New Orleans, LA, December 1995.

[26] M. Hromčík and M. Šebek. New algorithm for polynomial matrix determinant based on FFT. In *European Control Conference ECC'99*, Karlsruhe, Germany, September 1999.

[27] M. Hromčík and M. Šebek. FFT based algorithm for polynomial plus-minus factorization. In *European Control Conference ECC'03*, Cambridge, UK, September 2003.

[28] Z. Hurák and A. Böttcher. MIMO ℓ_1-optimal control via block toeplitz operators. In *Proceedings of 16th International Symposium on Mathematical Theory of Networks and Systems MTNS'04*, Leuven, Belgium, July 2004. Katholieke Universiteit Leuven. Submitted for publication.

[29] Z. Hurák, A. Böttcher, and M. Šebek. Minimum distance to the range of a banded lower triangular toeplitz operator in $\boldsymbol{\ell^1}$ and application in $\boldsymbol{\ell^1}$-optimal control. *SIAM Journal on Control and Optimization*, 2003. Submitted for publication.

[30] Z. Hurák and M. Šebek. On computing the ℓ_1 norm of a polynomial matrix fraction. In *Proc. of IEEE CCA/CACSD Conference*, Glasgow, UK, September 2002.

[31] Z. Hurák and M. Šebek. Algebraic approach to the ℓ^1-optimal control. In *Proceedings of 4th IFAC Symposium on Robust Control Design ROCOND'03*, Milan, Italy, September 2003.

[32] Z. Hurák and M. Šebek. Modular shift of a polynomial matrix. *Systems and Control Letters*, 2003. Submitted for publication.

[33] T. Kailath. *Linear Systems*. Prentice-Hall, 1980.

[34] M. Khammash. Solution of the ℓ_1 MIMO control problem without zero interpolation. In 35^{th} *Conference on Decision & Control*, pages 4040–4045, Kobe, Japan, December 1996.

[35] M. Khammash. A new approach to the solution of the ℓ_1 control problem: the scaled-Q method. *IEEE Transactions on Automatic Control*, 45(2):180–187, February 2000.

[36] M. Khammash and J. B. Pearson. Robust disturbance rejection in ℓ^1 optimal control systems. *Systems and Control Letters*, 14:93–101, 1990.

[37] M. Khammash and J. B. Pearson. Performance robustness of discrete-time system with structured uncertainty. *IEEE Transactions on Automatic Control*, 36:398–412, 1991.

[38] M. Khammash and J. B. Pearson. Analysis and design for robust performance with structured uncertainty. *Systems and Control Letters*, 20:179–187, 1993.

[39] M. Khammash, M. V. Salapaka, and T. Van Voorhis. Robust synthesis in ℓ_1: A globally optimal solution. *IEEE Transactions on Automatic Control*, 46(11):1744–1754, February 2001.

[40] V. Kučera. *Discrete Linear Control: The Polynomial Approach*. John Wiley and Sons, Chichester, 1979.

[41] H. Kwakernaak. Polynomial computation of Hankel singular values. In *Proceedings of the* 31^{st} *IEEE Conference on Decision and Control*, volume 4, pages 3595–3599, 1992.

[42] H. Kwakernaak and M. Šebek. Polynomial Toolbox for Matlab, version 2.5. `http://www.polyx.com`, 2000.

[43] D. G. Luenberger. *Optimization by Vector Space Methods.* John Wiley and Sons, 1969.

[44] J. M. Maciejowski. *Multivariable Feedback Design.* Addison-Wesley Publishing Company, 1989.

[45] P.-O. Malaterre and M. Khammash. ℓ_1 controller design for a high-order 5-pool irrigation canal system. In *IEEE-CDC conference*, Sydney, Australia, December 2000.

[46] J. S. McDonald and J. B. Pearson. ℓ^1-optimal control of multivariable systems with output norm constraints. *Automatica*, 27:317–329, 1991.

[47] D. G. Meyer. Two properties of ℓ_1-optimal controllers. *IEEE Transactions on Automatic Control*, 33(9):876–878, September 1988.

[48] A. Nerode. Linear automaton transformations. *Proc. American Math. Society*, 1958.

[49] E. Reich. On non-Hermitian Toeplitz matrices. *Math. Scand.*, 10:145–152, 1962.

[50] L. Rodman and I. M. Spitkovsky. Almost periodic factorization and corona theorem. *Indiana Univ. Math. J.*, 47:1234–1256, 1998.

[51] J. S. Shamma and M. A. Dahleh. Time-varying versus time-invariant compensation for rejection of persistent bounded disturbances and robust stabilization. *IEEE Transactions on Automatic Control*, 36(7):838–847, July 1991.

[52] O. J. Staffans. Mixed sensitivity minimization problems with rational ℓ^1-optimal solutions. *Journal of Optimization Theory and Applications*, 70(1):173–189, July 1991.

[53] O. J. Staffans. Mimo ℓ^1-optimization with a scalar control. *Journal of Optimization Theory and Applications*, 74(3):545–564, September 1992.

[54] O. J. Staffans. The four-block model matching problem in ℓ^1 and infinite-dimensional linear programming. *SIAM Journal of Control and Optimization*, 31(3):747–779, May 1993.

[55] M. Vidyasagar. *Control Systems Synthesis: A Factorization Approach.* MIT Press, 1985.

[56] M. Vidyasagar. Optimal rejection of persistent bounded disturbances. *IEEE Transactions on Automatic Control*, AC-31(6):527–534, June 1986.

[57] M. Vidyasagar. Further results on the optimal rejection of persistent bounded disturbances. *IEEE Transactions on Automatic Control*, 36(6):642–652, June 1991.

[58] N. J. Young. A polynomial method for the singular value decomposition of block hankel operators. *Systems & Control Letters*, 14:103–112, 1990.

Vita

Zdeněk Hurák was born on May 1974 in Krnov, Czech Republic. He got his Ing. degree (Czech equivalent to MSc.) in in Aviation Electrical Engineering, with major in avionics and weapon systens, at Air Force Faculty, Military Academy in Brno, Czech Republic, in 1997. After the graduation he got a position of a teaching and lab assistant at the same institution. In the spring of 1999 he was a Boeing Research Fellow at the Department of Electrical and Computer Engineering, Iowa State University, Ames, USA. Since 2000 he has been a member of the Institute of Information Theory and Automation, Czech Academy of Science. Since the fall 2000 he has been employed as a full-time researcher at the Center for Applied Cybernetics, Czech Technical University in Prague, Czech Republic. Since the fall 2001 he has been Dr. Michael Šebek's PhD student at the Department of Control Engineering, Czech Technical University in Prague.

Zdeněk Hurák won a best paper award at the 6th International Student Conference on Electrical Engineering in Prague, in 2001 and was granted a prize by Czech Airlines.

Address: Center for Applied Cybernetics,
Faculty of Electrical Engineering,
Czech Technical University
Karlovo náměstí 13/E, 12135, Prague
Tel: +420 224357683, Fax: +420 224916648
E-mail: z.hurak@c-a-k.cz
Web: `http://ar.c-a-k.cz/hurak`

This dissertation was typeset with LaTeX 2_ε[1] by the author.

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society.