Robust System Analysis and Nonlinear System Model Reduction

Thesis by

Sonja Glavaški

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

California Institute of Technology
Pasadena, California
1998
(Submitted May 22, 1998)

To My Parents

Acknowledgements

While at Caltech I have had the pleasure of meeting and working with many interesting and talented people, who made my stay here both productive and enjoyable. First of all, I would like to thank my advisor, Professor John Doyle for agreeing to supervise my work, and helping me understand and appreciate the "big picture." I could not have asked for a more precious gift.

I would also like to thank Professor Richard Murray, for introducing me to the nonlinear system model reduction theory and making me stay in touch with the engineering point of view. I am very grateful to Professor Jerrold Marsden for lending me some of his expertise on dynamical systems, and for showing me that everything is simple, once we understand it. I wish to thank the other members of my committee: Dr. Jorge Tierno and Professor Robert McEliece.

Collaborating with various colleagues helped me carry out the work in this thesis. In that regard I want to express my gratitude to fellow graduate students at Caltech: Mathew Newlin and Jorge Tierno for their help in many ways, for numerous creative discussions, and for their constant support. I also wish to thank Professor Igor Mezić who helped in adding fluid dynamics flavor to my work. Finally I would particularly like to thank Nikola Gluhajić, who exposed me to the robust control theory in the first place.

Many colleagues have helped to put this thesis into its final form. I would especially like to thank James Primbs, Sven Khatri, and Sean Humbert, for reading early versions, and providing many useful comments. I wish to thank the other members of the Control and Dynamical Systems at Caltech for sharing their good and bad times with me. They made these years unforgettable.

Last but not least, I would like to offer my honest gratitude to my family. For the last thirty years they had always been there when I needed them. Finally, all my thanks, and all my love to Isidora and Branislav who endured with me the ups and downs of a grad student's life. It would not have been the same without them.

From the bottom of my Slavic heart, I thank you all.

Abstract

The aim of the first part of this thesis is to broaden the classes of linear systems and performance measures that numerical tools for robustness analysis can be used for. First, we consider robustness problems involving uncertain real parameters and present several new approaches to computing an improved structured singular value μ lower bound. We combine these algorithms to yield a substantially improved power algorithm.

Then, we show that both the worst case $\mathcal{H}_\infty$ performance and the worst case $\mathcal{H}_2$ performance of uncertain systems subject to norm bounded structured LTI perturbations can be written exactly in terms of the skewed μ. The algorithm for the structured singular value lower bound computation, can be extended to computing skewed μ lower bound without significant loss of performance or accuracy.

We also demonstrate how a power algorithm can be used to compute a necessary condition for disturbance rejection of both discrete and continuous time nonlinear systems. For the general case of a system with a non-optimal controller this algorithm can provide us with knowledge of the worst case disturbance.

In the second part of this thesis we explore different approaches to the model reduction of systems. First, we show that the balancing transformation and Galerkin projection commute. We also demonstrate that if the balancing transformation matrix is orthogonal, balanced truncation and Galerkin projection commute.

Next, we pursue model reduction of nonlinear systems with rotational symmetry. We separate the movement of the wave from the evolution of the wave shape using the "centering procedure," and accurately approximate the shape of the wave with just few modes. The method may be viewed as a way of implementing the Karhunen-Loeve expansion on the space of solutions of the given PDE modulo a given symmetry group. The methodology is quite general and therefore should be useful in a variety of problems.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Formulating mathematical models for physical systems is always a starting, and maybe the most crucial, point of every engineer or scientist's work. Models derived from first principles are usually a system of ordinary differential equations or partial differential equations and their goal is to maximize qualitative correctness in representing the dynamics of a physical system. However, models also have to be useful for their intended application. For example, in control system design simple models of systems and controllers are preferred since they are much easier to do analysis and synthesis with.

Given a model of a dynamical system correctly representing the system's dynamics, we would like to be able to formulate a model of reduced complexity that predicts the most of system's behavior, and is easy to work with for a particular intended application.

Systems with the simplest possible description are linear systems, whose properties can be expressed as functions on finite dimensional space. However, from the early days of the development of control theory a need to take into account uncertainty in modeling has been recognized. The small gain theorem introduced by Zames [51] in early sixties provided a first exact robust stability test with respect to unstructured uncertainty. That was a beginning of robust control theory, that treats linear system with uncertainty in a systematic way.

For linear time invariant (LTI) systems with structured uncertainty, analysis of robust performance can be reduced to search for the solution of a set of algebraic equations which give bounds on the achievable performance. One is thus able to find computationally efficient solutions, such as the power algorithm for the μ lower bound, without doing an explicit parameter search involving repeated simulation.

On the other hand performance analysis for nonlinear systems is difficult due to the wide variety of behavior and structures which can occur. Most existing tools are at a theoretical level, and software for analyzing robust performance is not widely available. The first serious attempt to extend linear systems analysis methodology, given by the structured singular value framework, to the nonlinear case has been successfully pursued in [41].

The aim of the first part of this thesis is to develop and refine theoretical and computational tools for the analysis of different classes of robust linear and nonlinear systems.

Since linear theory is quite well developed and powerful, a natural approach to the study of nonlinear systems would be to use linear spaces and a nonlinear extensions of linear analysis. One may consider a nonlinear differential equation on a linear state space, and Fourier representations are an example of such a treatment. Of course there is nothing wrong in representing solution of a nonlinear partial differential equation as a linear combination of basis functions, but linear superposition still fails, meaning that the sum of two solution of a nonlinear partial differential equation (PDE) is not a solution of that same PDE, but we can still represent individual solutions via series expansions. Karhunen-Loeve expansion (KLE), one of the major existing tools for developing a reduced order models of nonlinear systems, is also based on linear theory and represents functions in linear spaces. Since it has been introduced in the context of turbulence by Lumley [28] in the late sixties to analyze experimental data aiming to extract dominant features and trends, which are typically patterns in space and time, it has been widely used, especially in the theory of turbulence.

To our knowledge, so far no thorough attempt to better understand the nature of nonlinear system model reduction accomplished by KLE has been made. In the second part of this thesis we review the connection between KLE model development method and balanced truncation of a linear system state space model, and search for a possible extension of linear systems balanced truncation technique to nonlinear systems. As a first step, we apply Galerkin model reduction method and balanced truncation to a system driven by a linear PDE.

Physical systems may exhibit various types of both continuous and discrete

symmetries. We are interested in rotational symmetry, called homogeneity in the turbulence literature. In this case KLE will just give ordered Fourier modes as optimal basis functions. Thus, homogeneity completely determines the form of the modes used for the series expansion, and the models obtained are not really of small order. In the second part of this thesis we show how if we incorporate symmetry information in the model development, low order models indeed can be obtained.

The work we developed in this thesis covers two important areas. We thoroughly analyzed some aspects of nonlinear systems model development and we also developed some computational tools for system analysis.

1.1 Previous work

The usefulness of the structured singular value μ, introduced in [8], lies in the fact that many robustness problems can be recast as problems of computing μ with respect to some block structure.

In recent years a great deal of interest has arisen with regard to robustness problems involving uncertain parameters that are not only norm bounded, but also constrained to be real. This type of problem falls within the μ framework by extending the original definition of μ to allow both real and complex uncertainties in the block structure ([14]). This mixed μ problem has fundamentally different properties from the complex μ problem. It is well known that the general mixed μ problem has the fundamental property of being NP hard (see [3], for example), which has important implications for computation. Recent results in [6] show that computing guaranteed bounds may be NP hard as well. This strongly suggests that any scheme to compute the exact solution for the general mixed μ problem will be computationally intractable, and the best we can hope for is to get good bounds for most problems.

A standard power algorithm (SPA) for computing a mixed μ lower bound was introduced in [48]. Each iteration of the scheme is very inexpensive, involving only matrix-vector multiplications and vector inner products. Unfortunately, the lower bound power iteration does not converge on a significant number of problems. Although one can still obtain a lower bound from the scheme in such cases, the bound may be poor. The first effort to enhance the performance of the SPA is presented in [44], with encouraging results.

A single μ robustness performance analysis provides a bound β on the uncertainty under which stability as well as $\mathcal{H}_\infty$ performance level $\frac{1}{\beta}$ are guaranteed. While this approach does provide a stability and performance margin, a good estimate of the actual uncertainty bound may be available. In that case, assuming that the uncertainty bound has been normalized, a question of interest is whether the system is stable, whenever the uncertainty has size less than 1, and if that is the case, what is the worst case performance for this same uncertainty size.

A measure of the worst case $\mathcal{H}_\infty$ performance ν of an uncertain system subject to norm bounded structured LTI perturbations was introduced in [12] for the complex uncertainty case. It was suggested that the worst case $\mathcal{H}_\infty$ gain of the uncertain system could be found via an infinite sequence of μ analyses. Thus, since μ can not be computed exactly, corresponding upper and lower bounds have to be used. An upper bound for ν can be computed by solving a quasi convex optimization problem, for which efficient algorithms exist, but no algorithm for a ν lower bound computation has been suggested.

In many cases, considering a slightly different version of previously described system allows us to set additional robust performance questions in the μ framework. For example, the problem of computing the worst case $\mathcal{H}_2$ norm of an uncertain system has always been considered an important one, since many useful performance requirements are captured by it. Many recent publications have presented different approaches at solving this problem. (See [15, 37] and the references therein.) However, in all cases, the results developed provide only upper bounds on the given norm when the uncertainty is linear time invariant.

A convex condition for Robust $\mathcal{H}_2$ performance analysis under structured uncertainty, of a very similar nature to the corresponding condition for robust $\mathcal{H}_\infty$ performance, was introduced in [37]. This upper bound is shown to be necessary and sufficient for slowly linear time varying uncertainty. However no indication is given on its conservativeness when the uncertainty is LTI. Recent results show that in the MIMO case the gap can be as large as the square root of the number of inputs.

Extending ideas from linear theory onto uncertain nonlinear systems would be extremely useful. Performance analysis for nonlinear systems is difficult due

to the wide variety of behavior and structures which can occur.

The problem of disturbance rejection can be solved by general purpose nonlinear programming algorithms. However, our experience with performance analysis for linear systems suggests that specific algorithms can be designed that significantly outperform the off-the-shelf ones in the sense that they give better answers with less computational effort [32]. Recent work has shown that this approach can be successfully extended to the study of different robustness problems for nonlinear systems (see [41, 43].)

The most preferred approach to control design is via low order models. The balanced truncation model reduction method was first introduced by Moore in [30]. Applying the signal analysis to controllability and observability led to a coordinate system in which the internally balanced model has special properties. Kalman minimal realization theory was recast by responses to injected signals, and working approximations of controllable and unobservable spaces were obtained. Moore proposed using these working spaces to obtain minimal realization of a system. This was an early attempt to use principal component analysis, introduced in statistics by Hotelling [23], for coping with dynamical systems. The stability properties of the reduced order model were shown by Parnebo and Silverman in [40]. The error bound for the balanced model reduction was shown independently by Glover in [20] and by Enns in [10]. An excellent overview of balanced truncation method can be found in [52].

In the case when the system is infinite dimensional, the model approximation becomes essential. The classical approach to model reduction of nonlinear systems is using the Galerkin method and the Karhunen-Loeve expansion that attempts to find an approximate solution of a PDE in the form of a truncated series expansion. The mode functions generated by the KLE method are based on empirical data. Karhunen-Loeve expansion was introduced in [24] and the standard KLE method has been reviewed in detail in [45], [34] and [35].

Physical systems may exhibit various types of both continuous and discrete symmetries. Sirovich in [26] has advocated an approach that assumes a system is ergodic and uses its known symmetries to increase the size of the ensemble, generating symmetric data set. This insures that reduced order models will share underlying symmetry of a system. Applications of the classical approach

to model reduction of various turbulence phenomena with known symmetries can be found in [22].

One of the subjects of major interest to control engineers in recent years has became the jet engine compressor system. To avoid the development of rotating stall—a compressor instability causing a sudden drop in performance—feedback control is necessary. The simplest existing model that adequately describes the basic dynamics of rotating stall is a three state nonlinear model of Moore and Greitzer (MG3) which is a Galerkin truncation onto a first Fourier mode of the full Moore-Greitzer model developed in [11] . The equation modeling the unsteady axial flow in the compression system introduced in [29] is first model of a compressor system developed directly from the Navier-Stokes equation of an inviscid flow.

1.2 Contributions and Outline of This Work

The work presented in the first part of this thesis significantly contributes to the development of computational methods for robustness analysis of both linear and nonlinear uncertain systems, closing some existing gaps between theory and practice. The work we developed in the second part of this thesis is a first step in the extension of the linear system model reduction methodology to model reduction of nonlinear systems. First, the widely accepted method for linear system model reduction by balanced truncation and the Galerkin method for model reduction of nonlinear systems are compared. For the special class of nonlinear systems with a rotational symmetry we propose a new computationally efficient modeling method. As opposed to standard approaches, it captures the existing symmetry in a system and generates low order models.

The results are organized as follows, in Chapter 2 we give a brief overview of the main ideas in robustness analysis with an emphasis on μ theory. We introduce the notation and lexicon that is used in the first part of this thesis.

We consider the general mixed μ problem, known to be an NP hard problem, in Chapter 3 and address the problem of computing a lower bound for mixed μ. We present new approaches to computing an improved μ lower bound, based on the standard power algorithm. A comparison between the new Combined Power Algorithm (CPA) described here and previous work is

shown. These results have been reported in [32].

In Chapter 4 we consider uncertain systems with mixed structured uncertainty entering as a linear fractional transformation. Given a bound on the $\mathcal{H}_\infty$ norm of the uncertainty we try to find the worst case $\mathcal{H}_\infty$ gain of the uncertain system. Although the worst case gain cannot be computed exactly, both upper and lower bounds can be computed efficiently. At each frequency point a skewed μ problem, a mixed version of ν [12], is solved. An upper bound can be computed by solving a quasi convex optimization problem, and efficient power algorithm for the skewed-μ lower bound based on the CPA algorithm from [32] for μ is developed. These results have been introduced in [19].

In Chapter 5 we consider solving the problem of computing the worst case $\mathcal{H}_2$ norm of a system with structured linear time invariant uncertainty with bounded $\mathcal{H}_\infty$ norm. For a given MISO or SIMO uncertain system, and given a bound on the $\mathcal{H}_\infty$ norm of the uncertainty, using an extension of the structured singular value μ, both upper and lower bound for the worst case $\mathcal{H}_2$ gain are developed. The computational effort to compute these bounds is similar to the effort required to compute upper and lower bounds of the structured singular value μ in the frequency domain. The upper and lower bounds developed are based on integration over frequency. At each frequency point a skewed-μ, problem is solved. These results are introduced in [42].

A numerical algorithm for the analysis of disturbance rejection for nonlinear systems is presented in Chapter 6. This algorithm seeks solutions to the Euler-Lagrange equations and is similar to the power algorithms for a μ analysis lower bound. General purpose nonlinear programming algorithms can be used to solve this problem. However, our experience with performance analysis for linear systems suggests that specific algorithms can be designed that significantly outperform off-the-shelf algorithms. The newly developed algorithm reduces to a well studied algorithm for the lower bound of μ, when the system is linear and the algorithm is guaranteed to converge to the global optimum. The proposed analysis method is useful in generating worst case disturbances for analyzing non optimal synthesis techniques. These results were introduced in [18].

Development of a low order model that qualitatively captures the observations from a physical system governed by a partial differential equation model

is the primary goal of the research conducted in the second part of this thesis. In Chapter 7 we briefly review a classical approach to nonlinear system model reduction using the Galerkin method and Karhunen-Loeve expansion. These methods attempt to find an approximate solution of the PDE in the form of a truncated series expansion, where mode functions are based on empirical data and generated by KLE. In the case of an homogeneous system, mode functions are actually Fourier modes and the order of the reduced model determined by the truncation point is not always small.

Before presenting a remedy for model reduction of certain classes of homogeneous systems, in Chapter 8 we compare the balanced truncation method for linear systems and Galerkin method for model reduction applied to a system driven by a linear PDE, to see if there is common ground for extending the balanced truncation methodology to nonlinear systems.

The investigation conducted in Chapter 7 led to the conclusion that for any PDE having a traveling wave as a solution, the classical approach to model reduction will not give satisfactory results. In Chapter 9 we resolve this problem by incorporating symmetry information in the methodology. To generate optimal basis functions, prior to performing KLE, we process the available data set using the "centering" procedure. This involves giving an appropriate definition of the center of a wave and moving centers of all the data snapshots to a standard position. The eigenvalues of the covariance matrix of "centered" data decay rapidly and we obtain a low order model. This method may be viewed as a way of implementing the KLE on the space of solutions of the given PDE modulo a given symmetry group. The methodology is quite general and should be useful in variety of problems. We applied it to several examples, and results show that centering introduces a significant improvement when compared to the classical technique. These results were introduced in [17].

We conclude this thesis with a summary of the work and suggestions for the future research directions.

Part I

Robust System Analysis

Chapter 2

Overview of Robustness Analysis

Robust control analysis and synthesis methods were developed to deal with different limitations of available models. They treat a system as a set of models, making up for the incompleteness and inaccuracy of a nominal model. The different models in the set will account for the errors and limitations of the measurements and limitations imposed by the model structure. Depending on noise, signal models chosen, and the nature of the class of systems studied different branches in robust control have been developed. In this chapter we will review the main ideas behind the structured singular value μ framework for robust control. The μ based methods have been proven to be useful for analyzing the performance and robustness properties both of linear and nonlinear feedback systems.

The work that we will develop in the first part of this thesis shares fundamental ideas of this framework and because of that its notation and terminology. Most of the material presented here is standard, and is based on [36] and [47].

2.1 Linear Fractional Transformations

Every robust control paradigm is based on a class of plants having simple mathematical structure, but rich enough to capture the fundamental behavior of the real system. In the first part of this thesis we will use classes of systems described by feedback interconnections, known as Linear Fractional Transformations (LFTs) for linear systems. Consider the system M shown in figure 2.1 with inputs u and w and outputs y and z defined by the equations

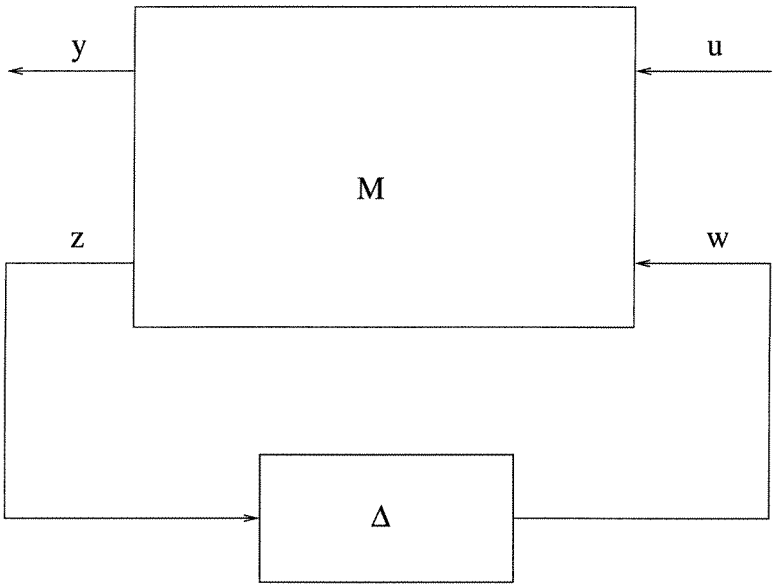

Figure 2.1: LFT interconnection of system and uncertainty.

$$\begin{aligned} y &= M_{11}u + M_{12}w \\ z &= M_{21}u + M_{22}w \\ M &= \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}. \end{aligned} \tag{2.1}$$

Let $\mathbf{\Delta}$ be a set of systems Δ having dimensions compatible with z and w. Then for each $\Delta \in \mathbf{\Delta}$ we can define the following system

$$\begin{aligned} y &= M_{11}u + M_{12}w \\ z &= M_{21}u + M_{22}w \\ w &= \Delta z. \end{aligned} \tag{2.2}$$

The system that maps u into y we will denote $\Delta * M$, which is standard notation for the Redheffer star product. The system M can be either a constant matrix or a linear operator between signal spaces.

It is instructive to consider a "feedback" interpretation of the system shown in figure 2.1 when M is a constant matrix. As long as $I - M\Delta$ is not singular, the only solutions u and w to the loop equation are $u = w = 0$. However if $I - M\Delta$ is singular, then there are infinitely many solutions to the equations, and

norms of solutions may be arbitrarily large. In connection with the stability of systems, we call this constant matrix feedback problem "unstable" when $I - M\Delta$ is singular, and systems having only zero solutions we will call "stable."

In the more general case when M is a linear operator we say that the system $\Delta * M$ is well posed if equations 2.2 have a unique solution for every u and w. Finally we define the set of systems

$$\mathbf{\Delta} * M = \{\Delta * M, \Delta \in \mathbf{\Delta}\}. \tag{2.3}$$

The sets of plants considered for robustness analysis will be of the form $\mathbf{B\Delta} * M$, where the set $\mathbf{B\Delta}$ is of the form

$$\mathbf{B\Delta} = \{blockdiag(\Delta_1, \Delta_2, \ldots, \Delta_m), \|\Delta_i\| \leq 1\}. \tag{2.4}$$

Each block Δ_i will have a very simple structure, either diagonal or a full operator, and we will refer to the structure of the elements of the set $\mathbf{\Delta}$ as the uncertainty block structure.

2.2 Structured Singular Value μ (Notation and Preliminaries)

The structured singular value μ introduced in [8] is a powerful tool for robustness analysis. Many robustness problems can be recast as one of computing μ with respect to some block structure. This structure may be defined differently for each problem depending on the uncertainty and performance objectives of the problem. Defining the structure involves specifying: the total number of blocks, the type of each block, and their dimensions. In general we consider two types of blocks: repeated scalar (both real and complex) and full blocks.

The notation used here is fairly standard and is taken from [13]. Let $\mathbf{R}$ and $\mathbf{C}$ denote the real and complex numbers respectively. For any square complex matrix M we denote the complex conjugate transpose by M^*. The largest singular value and the structured singular value are denoted by $\overline{\sigma}(M)$ and $\mu_{\mathcal{K}}(M)$ respectively. The spectral radius is denoted $\rho(M)$, and $\rho_R(M) \doteq \max\{0 \bigcup |\lambda| : \lambda \textit{ is a real eigenvalue of } M\}$. For any complex vector x, x^* denotes the complex conjugate transpose, and $|x|$ the Euclidean norm.

The definition of μ is dependent upon the underlying block structure of the uncertainties, which is defined as follows. Given a matrix $M \in \mathbf{C}^{n\times n}$ and three non-negative integers m_r, m_c, and m_C with $m \doteq m_r + m_c + m_C \leq n$, the block structure $\mathcal{K}(m_r, m_c, m_C)$ is an m-tuple of positive integers

$$\mathcal{K} = (k_1, \ldots, k_{m_r}, k_{m_r+1}, \ldots, k_{m_r+m_c}, k_{m_r+m_c+1}, \ldots, k_m) \tag{2.5}$$

where we require $\sum_{i=1}^{m} k_i = n$ so that the dimensions are compatible. Define the set of allowable perturbations as follows.

$$\begin{aligned}\mathbf{\Delta}_{\mathcal{K}} \doteq \{\Delta = \text{blockdiag}(&\delta_1^r I_{k_1}, \ldots, \delta_{m_r}^r I_{k_{m_r}}, \delta_1^c I_{k_{m_r+1}}, \ldots, \\ &\delta_{m_c}^c I_{k_{m_r+m_c}}, \Delta_1^C, \ldots, \Delta_{m_C}^C) : \\ &\delta_i^r \in \mathbf{R}, \delta_i^c \in \mathbf{C}, \Delta_i^C \in \mathbf{C}^{k_{m_r+m_c+i}\times k_{m_r+m_c+i}}\}.\end{aligned} \tag{2.6}$$

Note that $\mathbf{\Delta}_{\mathcal{K}} \subset \mathbf{C}^{n\times n}$ and that this block structure is sufficiently general to allow for repeated real scalars, repeated complex scalars, and full complex blocks. The purely complex case corresponds to $m_r = 0$.

Definition 1 ([8]) *The structured singular value, $\mu_{\mathcal{K}}(M)$, of a matrix $M \in \mathbf{C}^{n\times n}$ with respect to a block structure $\mathcal{K}(m_r, m_c, m_C)$ is defined as*

$$\mu_{\mathcal{K}}(M) \doteq \left(\min_{\Delta\in\mathbf{\Delta}_{\mathcal{K}}} \{\overline{\sigma}(\Delta) : \det(I - \Delta M) = 0\}\right)^{-1} \tag{2.7}$$

with $\mu_{\mathcal{K}}(M) \doteq 0$ if no $\Delta \in \mathbf{\Delta}_{\mathcal{K}}$ solves $\det(I - \Delta M) = 0$.

In the case of a constant matrix feedback loop, the structured singular value $\mu_{\mathcal{K}}(M)$ provides a measure of the smallest structured Δ that causes "instability" as defined in section 2.1. The norm of this "destabilizing" Δ is exactly $\frac{1}{\mu_{\mathcal{K}}(M)}$. We can also relate $\mu_{\mathcal{K}}(M)$ to familiar linear algebra quantities when $\mathbf{\Delta}$ is one of two extreme sets as shown in the following lemma.

Lemma 1 *The spectral radius $\rho(M)$ of the matrix M satisfies*

$$\rho(M) = \mu_{\mathbf{\Delta}_s},$$

where

$$\mathbf{\Delta}_s = \{\delta I, \delta \in \mathbf{C}\},$$

and the largest singular value

$$\overline{\sigma}(M) = \mu_{\mathbf{\Delta}_f},$$

where

$$\mathbf{\Delta}_f = \{\Delta_f, \Delta_f \in \mathbf{C}^{n\times n}\}.$$

We have found this lemma extremely useful in treating special cases of robustness analysis.

2.3 The Main Loop Theorem

To form the basis for most uses of μ in linear systems robustness analysis assume that two uncertainty structures $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$ are given, and define a third structure

$$\mathbf{\Delta} = \left\{ \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix} \ : \ \Delta_1 \in \mathbf{\Delta}_1, \Delta_2 \in \mathbf{\Delta}_2 \right\}. \tag{2.8}$$

There are three structures with respect to which we can compute μ. We denote $\mu_{\mathbf{\Delta}_1}$ the structured singular value computed with respect to $\mathbf{\Delta}_1$, $\mu_{\mathbf{\Delta}_2}$ the structured singular value computed with respect to $\mathbf{\Delta}_2$, and $\mu_{\mathbf{\Delta}}$ structured singular value computed with respect to $\mathbf{\Delta}$. The following theorem will state the connection between these three different structured singular values.

Theorem 1 *[36] (Main Loop Theorem)*

$$\mu_{\mathbf{\Delta}}(M) < 1 \Leftrightarrow \begin{cases} \mu_{\mathbf{\Delta}_2}(M_{22}) < 1 \\ \max_{\Delta_2 \in \mathbf{B}\mathbf{\Delta}_2} \mu_{\mathbf{\Delta}_1}(M * \Delta_2) < 1 \end{cases}$$

To better explain the importance of this theorem we will restate it as follows. Suppose that for a system M_{22} there is a property of the system and a performance level is achieved if and only if $\mu_{\mathbf{\Delta}_f}(M_{22}) < 1$. Then all the plants in the set $\mathbf{B}\mathbf{\Delta}_\mathbf{u} * M$ are well posed and satisfy the property if and only if $\mu_{\mathbf{\Delta}}(M) < 1$ where $\mathbf{\Delta}$ is the set

$$\mathbf{\Delta} = \left\{ \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix} \ : \ \Delta_1 \in \mathbf{\Delta}_u, \Delta_2 \in \mathbf{\Delta}_f \right\}. \tag{2.9}$$

As an example, consider the discrete time linear system given by the state space realization

$$\begin{aligned} x_{k+1} &= Ax_k + B_1 u_k + B_2 w_k \\ y_k &= C_1 x_k + D_{11} u_k + D_{12} w_k \\ z_k &= C_2 x_k + D_{21} u_k + D_{22} w_k \end{aligned} \tag{2.10}$$

and the associated matrices

$$M_{11} = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \end{bmatrix},$$

and

$$M = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{bmatrix}.$$

As a performance measure consider the L_2 into L_2 induced gain from u to y

$$G_{uy} = \max_{\delta_1 \in \mathbf{C},\ |\delta_1| \leq 1} \overline{\sigma}(D + C\delta_1(I - \delta_1 A)^{-1}B).$$

According to Theorem 1, the system is stable if

$$\mu_{\mathbf{\Delta}_s}(A) < 1$$

and the gain G_{uy} is less than 1

$$\max_{\Delta_f \in \mathbf{B}\mathbf{\Delta}_\mathbf{f}} \mu_{\mathbf{\Delta}_f}(M_{11} * \Delta_s) < 1$$

if and only if

$$\mu_{\mathbf{\Delta}_1}(M_{11}) < 1$$

where

$$\mathbf{\Delta}_1 = \left\{ \begin{bmatrix} \Delta_s & 0 \\ 0 & \Delta_f \end{bmatrix} \ : \ \Delta_s \in \mathbf{\Delta}_s, \Delta_f \in \mathbf{\Delta}_f \right\}.$$

Now, assuming that the system has an uncertain component described by the following equation

$$w = \delta_u z \tag{2.11}$$

with $\delta_u \in \mathbf{R}$ and $|\delta_u| \leq 1$. The system is well posed for all δ_u if and only if

$$\rho(D_{22}) < 1 \Leftrightarrow \mu_{\mathbf{\Delta}_u}(D_{22}) < 1,$$

where

$$\mathbf{\Delta}_u = \{\delta_u I : \delta_u \in \mathbf{C}\},$$

and the performance condition is met for all δ_u if and only if

$$\max_{\delta_u \in [-1,1]} \mu_{\mathbf{\Delta}_1}(\delta_u I * M) < 1.$$

Applying Theorem 1 again we see that the system given by equations 2.10 and 2.11 is stable and well posed for all $\delta_u \in [-1, 1]$ and has induced gain from u to y less than 1 if and only if

$$\mu_{\mathbf{\Delta}_2}(M) < 1$$

with

$$\mathbf{\Delta}_2 = \left\{ \begin{bmatrix} \Delta_s & 0 & 0 \\ 0 & \Delta_f & 0 \\ 0 & 0 & \delta_u I \end{bmatrix} \; : \; \Delta_s \in \mathbf{\Delta}_s, \Delta_p \in \mathbf{\Delta}_p, \delta_u \in \mathbf{R} \right\}.$$

This example is better known as state space robustness performance.

2.4 μ Upper and Lower Bound

In general, it is difficult to compute μ exactly (the general mixed μ problem is NP hard) so computation has focused on upper and lower bounds. An upper bound gives a (possibly conservative) limit on the size of allowable perturbations, and a lower bound yields a problem perturbation. Important issues are then the efficient computation of the bounds, the degree to which they approximate μ, and techniques for refining the bounds for a better approximation. Bounds are refined with transformations on M that do not affect $\mu_{\mathcal{K}}(M)$, but do affect ρ and $\overline{\sigma}$.

A significant part of the research effort in this area has been devoted to the development of fast and efficient algorithms for the computation of these bounds. The work in Part I of this thesis concentrates on improving the algorithm for the mixed μ lower bound computation, its extension to special classes of H_∞ and H_2 robustness analysis, and to H_∞ robustness analysis of a certain class of nonlinear systems.

Lower bound

In order to obtain and refine a lower bound for μ we define the following sets of block diagonal matrices, which depend on the underlying block structure.

$$\mathcal{Q}_{\mathcal{K}} \doteq \{\Delta \in \mathbf{\Delta}_{\mathcal{K}} : \delta_i^T \in [-1,\ 1], \delta_i^{c*}\delta_i^c = 1, \Delta_i^{C*}\Delta_i^C = I_{k_{m_r+m_c+i}}\}. \tag{2.12}$$

$$\begin{aligned}\mathcal{D}_{\mathcal{K}} \ \doteq\ & \{\text{blockdiag}(e^{j\theta_1}D_1, \ldots, e^{j\theta_{m_r}}D_{m_r}, D_{m_r+1}, \\ & \ldots, D_{m_r+m_c}, d_1 I_{k_{m_r+m_c+1}}, \ldots, d_{m_C} I_{k_m}) \\ & : \theta_i \in [-\pi/2,\ \pi/2], 0 < D_i = D_i^* \in \mathbb{C}^{k_i \times k_i}, 0 < d_i \in \mathbf{R}\}. \end{aligned} \tag{2.13}$$

$$\mathbf{B}_{\mathbf{\Delta}_{\mathcal{K}}} \doteq \{\Delta \in \mathbf{\Delta}_{\mathcal{K}} : \overline{\sigma}(\Delta) \leq 1\}. \tag{2.14}$$

The theoretical basis for the mixed μ lower bound is the fact that the mixed μ problem may be reformulated as a real eigenvalue maximization. The following theorem is from [48].

Theorem 2 *For any matrix $M \in \mathbf{C}^{n\times n}$, and any compatible block structure $\mathcal{K}$,*

$$\mu_{\mathcal{K}}(M) = \max_{Q\in\mathcal{Q}_{\mathcal{K}}} \rho_R(QM). \tag{2.15}$$

Note that $\rho_R(QM) = \rho_R(MQ)$. Since this maximization problem is not convex we will in general only be able to find local maxima. For any $Q \in \mathcal{Q}_{\mathcal{K}}$, $\rho_R(QM) \leq \mu_{\mathcal{K}}(M)$, so any $Q \in \mathcal{Q}_{\mathcal{K}}$ immediately gives us a lower bound for $\mu_{\mathcal{K}}(M)$. We would like this lower bound to be close to μ, therefore our goal is to find an efficient way to compute a local maximum of the function $\rho_R(QM)$ over $Q \in \mathcal{Q}_{\mathcal{K}}$.

The maximization in (2.15) can be reformulated as

$$\mu_{\mathcal{K}}(M) = \max_{Q\in\mathcal{Q}_{\mathcal{K}}} \{\beta : det(\beta I_n - QM) = 0\}. \tag{2.16}$$

and it turns out that this maximization can often be achieved by means of a simple power iteration which is fully described in [48]. Although this iteration usually converges to a satisfactory equilibrium point, the convergence is not guaranteed. However, one can still obtain a perturbation from which a lower bound can be computed, though this need not be a local maximum. In this way the algorithm always returns a valid lower bound, regardless of convergence. The performance results presented in following chapters include problems of this kind. In all tests performed no data were excluded.

Upper bound

The whole concept of computing the upper bound for μ is based on the following theorem.

Theorem 3

$$\mu_{\mathcal{K}}(M) = \max_{Q \in \mathcal{Q}_{\mathcal{K}}} \rho_R(QM) \leq \min_{D_l, D_r} \overline{\sigma}(D_r M D_l^{-1}) \qquad D_l, D_r \in \mathcal{D}_{\mathcal{K}},$$

where D_l and D_r are invertible and satisfy

$$D_l Q = Q D_r \quad \forall Q \in \mathcal{Q}_{\mathcal{K}}.$$

Proof

For any $Q \in \mathcal{Q}_{\mathcal{K}}$, and for any pair of invertible matrices D_l and D_r, such that

$$D_l Q = Q D_r,$$

the following series of inequalities hold

$$\begin{aligned} \rho(QM) &= \rho(D_l^{-1} Q D_r M) \\ &= \rho(Q D_r M D_l^{-1}) \\ &\leq \overline{\sigma}(D_r M D_l^{-1}). \end{aligned}$$

Theorem 3 provides us with an upper bound to the structured singular value in the form of a minimization problem, which has been shown in [33] to be convex. The inequality given in the theorem can be rewritten in the form of a Linear Matrix Inequality (LMI) as shown by the following series of equivalent inequalities

$$\begin{aligned} \overline{\sigma}(D_r M D_l^{-1}) &< 1 \\ \overline{\sigma}(D_r M D_l^{-1}) - 1 &< 0 \\ u^*(M^* D_r^* D_r M - D_l^* D_l)u &< 0 \quad \forall u \\ M^* D_r^* D_r M - D_l^* D_l &< 0. \end{aligned}$$

For a complete development of this issue, especially further refinements when parts of the uncertainty structure are real, we refer the reader to [47]. Software packages for solving LMIs are now commercially available and perform quite well.

2.5 A Measure of the Worst Case $\mathcal{H}_\infty$ Performance ν

A single μ robustness performance analysis provides only a bound β on the uncertainty under which stability as well as $\mathcal{H}_\infty$ performance level $\frac{1}{\beta}$ are guaranteed. While this approach does provide a stability and performance margin, a good estimate of the actual uncertainty bound may be available. In that case, assuming that the uncertainty bound has been normalized, a question of interest is whether the system is stable, whenever the uncertainty has size less than 1, and if that is the case, what is the worst case performance for this same uncertainty size.

In [12] a quantity ν related to μ which provides answer to the following question has been introduced.

Question 1 *Determine the smallest α with the property that, for the uncertainty bounded by* 1 *, the worst case $\mathcal{H}_\infty$ performance level is better than α.*

A natural approach to answering question 1 will be via an infinite sequence of μ analyses. For any $\alpha > 0$, stability of the system 2.2 is equivalent to stability of the system M^α given by

$$\begin{aligned} y &= M_{11}u + \alpha M_{12}w \\ z &= M_{21}u + \alpha M_{22}w \\ w &= \frac{1}{\alpha}\Delta z. \end{aligned} \tag{2.17}$$

and the transfer function from u to y is equivalent for both systems. From this it follows that the system 2.2 is stable whenever $\|\mathbf{\Delta}\| \leq 1$ (i.e., whenever $\|\frac{1}{\alpha}\mathbf{\Delta}\| \leq \frac{1}{\alpha}$), with worst case performance better than α, if and only if

$$\sup_\omega \mu_{\mathbf{\Delta}_\alpha}(M^\alpha(jw)) < \alpha, \tag{2.18}$$

where

$$\mathbf{\Delta}_\alpha = \left\{ \begin{bmatrix} \Delta & 0 \\ 0 & \frac{1}{\alpha}\delta I \end{bmatrix} \; : \; \|\Delta\| \leq 1 \right\}. \tag{2.19}$$

Thus the answer to Question 1 is given by the infimum of those α satisfying 2.18, which can be found by the fixed point iteration

$$\alpha_{i+1} = \sup_{\omega} \mu_{\mathbf{\Delta}_\alpha}(M^{\alpha_i}(jw)), \quad \alpha_0 > 0. \tag{2.20}$$

In [12] a quantity providing an answer to question 1 in a single analysis has been introduced.

The definition of ν is dependent upon the underlying block structure of the uncertainties. For simplicity, and without a loss of generality, we will consider the case when block structure consists of just one performance block and one uncertainty block.

Given a matrix $M \in \mathbf{C}^{n\times n}$ and two non-negative integers k_1, and k_2 consider the multi index

$$\mathcal{K} = (k_1, k_2) \tag{2.21}$$

with $k_1 + k_2 = n$, so that the dimensions are compatible. Define the set of allowable perturbations as follows.

$$\mathbf{\Delta}^s_{\mathcal{K}} \doteq \{\Delta = \text{blockdiag}(\Delta_1, \Delta_2) : \Delta_i \in \mathbb{C}^{k_i \times k_i}, \ \overline{\sigma}(\Delta_1) \leq 1\}. \tag{2.22}$$

Definition 2 ([12]) *The quantity $\nu_{\mathcal{K}}(M)$, associated with a a complex matrix $M \in \mathbf{C}^{n\times n}$ and with respect to a block structure $\mathcal{K}$ is defined as*

$$\nu_{\mathcal{K}}(M) \doteq \left(\min_{\Delta \in \mathbf{\Delta}_{\mathcal{K}}} \{\overline{\sigma}(\Delta_2) : \det(I - \Delta M) = 0\} \right)^{-1} \tag{2.23}$$

with $\nu_{\mathcal{K}}(M) \doteq 0$ if no $\Delta \in \mathbf{\Delta}_{\mathcal{K}}$ solves $\det(I - \Delta M) = 0$.

It is important to note that in the formula for $\nu_{\mathcal{K}}(M)$, the size of Δ_1 is not minimized but just kept below 1, reflecting the fact that the required uncertainty bound is fixed.

ν is related to μ via a recursion given by the following proposition.

Proposition 1 *Suppose $\nu(M) < \infty$. Then*

$$\nu_{\mathcal{K}}(M) = \mu_{\mathcal{K}} \left(\begin{bmatrix} \nu_{\mathcal{K}}(M) I_{k_1} & 0 \\ 0 & I_{k_2} \end{bmatrix} M \right). \tag{2.24}$$

This property of ν will allow us to develop efficient algorithms for computing both lower and upper bound of skewed μ, which is a special case of ν that includes real parameters in the uncertainty, based on the computation of lower and upper bound of μ.

Chapter 3

Advances in the Computation of the μ Lower Bound

The structured singular value μ, introduced in [8], is useful because many robustness problems can be recast as problems of computing μ with respect to some block structure. In recent years a great deal of interest has arisen with regard to robustness problems involving uncertain parameters that are not only norm bounded, but also constrained to be real. This type of problem falls within the μ framework by extending the original definition of μ to allow both real and complex uncertainties in the block structure [14]. This mixed μ problem has fundamentally different properties from the complex μ problem with important implications for computation.

Recent results in [3], for example, prove that the general mixed μ problem is NP hard, while [6] indicates that computing guaranteed bounds may be NP hard as well. This strongly suggests that any scheme to compute the exact solution for the general mixed μ problem will be computationally intractable, and the best we can hope for is to get good bounds for most problems.

This chapter addresses the problem of computing a lower bound for mixed μ. A power iteration algorithm to compute a lower bound was generalized to the mixed case in [48]. Even though each iteration of the scheme is very inexpensive, involving only matrix-vector multiplications and vector inner products, its convergence properties are typically good. Unfortunately, the lower bound power iteration does not converge on a significant number of problems. Although one can still obtain a lower bound from the scheme in such cases, the bound may be poor. The first effort to enhance the performance of the standard power algorithm (SPA) is presented in [44], with encouraging results.

In this chapter we present new approaches to computing an improved μ lower bound. The SPA from [48] will be our starting point. We develop some new algorithms after examining the convergence properties of the SPA. These algorithms are then combined to yield a substantially improved power algorithm.

A comparison between the new algorithms described here and previous work including [48] is shown in Section 3.8. The results are promising and are the principal motivation for this exposition. Furthermore, the algorithms described and compared here are by no means optimized, so it is reasonable to expect that this work will lead to even better results than those in Section 3.8.

3.1 Mixed μ is NP Hard

It is well known that the general mixed μ problem has the fundamental property of being NP hard (see [3], for example), which has important implications for computation. Recent results in [6] show that computing guaranteed bounds may be NP hard as well.

It is still a fundamental open question in the theory of computational complexity to determine the exact properties of NP hard problems and we refer the reader to [16] for an in depth treatment of this subject. However, it is generally accepted that a problem being NP hard means that it cannot be computed in polynomial time in the worst case. Being NP hard is a property of the problem itself, not any particular algorithm.

The fact that the mixed μ problem is NP hard strongly suggests that it is futile to pursue methods for exactly calculating μ of general systems with mixed uncertainty for other than small problems. However, these results do not mean that one cannot develop practical algorithms to compute bounds for problems which arise in engineering applications. Typically such algorithms involve approximation, heuristics, branch and bound, or local search.

With this in mind, we see that proofs of convergence to the global maximum are *irrelevant*. Any such proof is meaningless for any computation other than those that are trivially small (and consequently easy). Any practical effort to compute an NP hard problem must be measured by its performance on a large number of typical problems, as in Section 3.8.

3.2 Characterization of Local Maxima

Efforts to compute a lower bound for μ have focused on finding a local maximum of (2.15) rather than the NP hard problem of finding a global maximum. In this section we present conditions that must be satisfied at every local maximum. Recall that by (2.15), $\mu_{\mathcal{K}}(M) = \max_{Q \in \mathcal{Q}_{\mathcal{K}}} \rho_R(QM)$. The characterization of a maximum point of $\rho_R(QM)$ at $Q = I$ will be in terms of an alignment of the right and left eigenvectors of M.

First, a bit more notation. Suppose $M \in \mathbf{C}^{n \times n}$ has an eigenvalue λ with right and left eigenvectors x and y respectively. We partition x and y compatibly with the block structure, so that $x_{r_i}, y_{r_i} \in \mathbf{C}^{k_i}$ for $i \in \{1, \ldots, m_r\}$, $x_{c_i}, y_{c_i} \in \mathbf{C}^{k_{m_r+i}}$ for $i \in \{1, \ldots, m_c\}$, and $x_{C_i}, y_{C_i} \in \mathbf{C}^{k_{m_r+m_c+i}}$ for $i \in \{1, \ldots, m_C\}$. Furthermore, for the rest of this chapter we make a non-degeneracy assumption that for every i where these subscripts are defined, $y_{r_i}^* x_{r_i} \neq 0, y_{c_i}^* x_{c_i} \neq 0, y_{C_i}^* x_{C_i} \neq 0$.

For any matrix $Q \in \mathcal{Q}_{\mathcal{K}}$ define the index sets

$$\mathcal{J}(Q) \doteq \{i \leq m_r : |\Delta_i^r| = 1\} \tag{3.1}$$

$$\hat{\mathcal{J}}(Q) \doteq \{i \leq m_r : |\Delta_i^r| < 1\} \tag{3.2}$$

and define the allowable perturbation set

$$\begin{aligned} \hat{\mathbf{B}}\Delta_\epsilon(\mathcal{J}, \hat{\mathcal{J}}) \doteq \{\Delta \in \mathbf{\Delta}_{\mathcal{K}} : |\Delta_i^r| \leq 1, i \in \mathcal{J}, |\Delta_i^r| < 1 + \epsilon, i \in \hat{\mathcal{J}}, \\ |\Delta_i^c| \leq 1, i = 1, \ldots, m_c, \overline{\sigma}(\Delta_i^C) \leq 1, i = 1, \ldots, m_C\}. \end{aligned} \tag{3.3}$$

We see that for sufficiently small $\epsilon > 0$, if $Q \in \mathcal{Q}_{\mathcal{K}}$ and $\Delta \in \hat{\mathbf{B}}\Delta_\epsilon(\mathcal{J}(Q), \hat{\mathcal{J}}(Q))$, then $Q\Delta \in \mathbf{B}_{\mathbf{\Delta}_{\mathcal{K}}}$ and $\Delta Q \in \mathbf{B}_{\mathbf{\Delta}_{\mathcal{K}}}$.

The following theorem is proved in [48].

Theorem 4 *Suppose the matrix $M \in \mathbf{C}^{n \times n}$ has a distinct real eigenvalue $\lambda_0 > 0$ with right and left eigenvectors, x and y respectively, satisfying the non-degeneracy assumption. Further suppose that $\rho_R(M) = \lambda_0$. If, for some $\epsilon > 0$, the function $\rho_R(QM)$ attains a local maximum over the set $Q \in \hat{\mathbf{B}}\Delta_\epsilon(\mathcal{J}, \hat{\mathcal{J}})$ at $Q = I$, then there exists a matrix $D \in \mathcal{D}_{\mathcal{K}}$, with $\theta_i = \pm\frac{\pi}{2}$ for every $i \in \hat{\mathcal{J}}$, and a real scalar $\psi \in (-\frac{\pi}{2}, \frac{\pi}{2})$, such that $y = e^{j\psi} D x$.*

Remark: It was shown in [48] that we have a partial converse to Theorem 4, namely that if $y = e^{j\psi}Dx$ under the above assumptions, then no directional derivative of the eigenvalue achieving $\rho_R(QM)$ over the set $Q \in \hat{\mathbf{B}}\Delta_c(\mathcal{J}, \hat{\mathcal{J}})$ is real and positive at $Q = I$.

3.3 A Lower Bound Power Algorithm

In the first part of this section, we review the SPA, which is the starting point for the research presented here. In the second part we show the results of a comparison between the SPA and standard optimization code for the same problems. These results support the assertion that the SPA is a good starting point for the development of better algorithms.

The SPA

It was shown in [48] that the problem of computing a lower bound β for $\mu_{\mathcal{K}}(M)$ can be reduced to finding matrices $Q \in \mathcal{Q}_{\mathcal{K}}$ and $D \in \mathcal{D}_{\mathcal{K}}$ with $\theta_i = \pm\frac{\pi}{2}$ for $i \in \hat{\mathcal{J}}(Q)$ and non-zero vectors b, a, z, and w such that the following set of equations holds.

$$\begin{aligned} Mb &= \beta a & M^*z &= \beta w \\ b &= Qa & b &= D^{-1}w \\ z &= Q^*QDa & z &= Q^*w. \end{aligned} \tag{3.4}$$

Finding such solutions may be attempted via the power iteration in [48]. We will not go into any of the details of the theoretical development here, but merely present the final result.

In this section we explicitly write the formulae only for the simple block structure with $m_r = m_c = m_C = 1$. This is for notational simplicity only. The formulae for an arbitrary block structure are obtained simply by repeating the formulae for each block type appropriately. Except for equations (3.7) and (3.9), the blocks are updated independently. Given $\mathcal{K} = (k_1, k_2, k_3)$ the appropriate scaling set becomes

$$\begin{aligned} \mathcal{Q}_{sub} = \{\text{blockdiag}(q^r I_{k_1}, q^c I_{k_2}, Q^C) : q^r \in [-1\ 1], q^{c*}q^c = 1, \\ Q^{C*}Q^C = I_{k_3}\}. \end{aligned} \tag{3.5}$$

Partition the four vectors b, a, z, and $w \in \mathbf{C}^n$ compatibly with this block structure as

$$b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}, \quad w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \tag{3.6}$$

where b_i, a_i, z_i, and $w_i \in \mathbf{C}^{k_i}$. Allow these vectors to evolve via the following power iteration:

$$\tilde{\beta}_{k+1} a_{k+1} = M b_k \ : \ \tilde{\beta}_{k+1} \in \mathbf{R}^+, \ |a_{k+1}| = 1 \tag{3.7}$$

$$\begin{aligned} z_{1_{k+1}} &= \tilde{q}_{k+1} w_{1_k} \\ z_{2_{k+1}} &= \frac{w_{2_k}^* a_{2_{k+1}}}{|w_{2_k}^* a_{2_{k+1}}|} w_{2_k} \\ z_{3_{k+1}} &= \frac{|w_{3_k}|}{|a_{3_{k+1}}|} a_{3_{k+1}} \end{aligned} \tag{3.8}$$

$$\hat{\beta}_{k+1} w_{k+1} = M^* z_{k+1} \ : \ \hat{\beta}_{k+1} \in \mathbf{R}^+, \ |w_{k+1}| = 1 \tag{3.9}$$

$$\begin{aligned} b_{1_{k+1}} &= \hat{q}_{k+1} a_{1_{k+1}} \\ b_{2_{k+1}} &= \frac{a_{2_{k+1}}^* w_{2_{k+1}}}{|a_{2_{k+1}}^* w_{2_{k+1}}|} a_{2_{k+1}} \\ b_{3_{k+1}} &= \frac{|a_{3_{k+1}}|}{|w_{3_{k+1}}|} w_{3_{k+1}} \end{aligned} \tag{3.10}$$

where $\tilde{q}_{k+1}$ and $\hat{q}_{k+1}$ evolve as

$$\begin{aligned} \tilde{\alpha}_{k+1} &= \operatorname{sgn}(\hat{q}_k) \frac{|b_{1_k}|}{|a_{1_{k+1}}|} + \operatorname{Re}(a_{1_{k+1}}^* w_{1_k}) \\ \tilde{q}_{k+1} &= \min(\max(-1, \tilde{\alpha}_{k+1}), 1) \\ \hat{\alpha}_{k+1} &= \operatorname{sgn}(\tilde{q}_{k+1}) \frac{|b_{1_k}|}{|a_{1_{k+1}}|} + \operatorname{Re}(a_{1_{k+1}}^* w_{1_{k+1}}) \\ \hat{q}_{k+1} &= \min(\max(-1, \hat{\alpha}_{k+1}), 1). \end{aligned} \tag{3.11}$$

Note that all used relationships are written in a manner that does not involve the matrices Q and D.

It was shown in [48] that if the above iteration converges to an equilibrium point then we have a matrix $Q \in \mathcal{Q}_{sub}$ such that $QMb = \tilde{\beta} b$ and $w^* QM = \hat{\beta} w^*$, so that $\max(\tilde{\beta}, \hat{\beta})$ gives us a lower bound for $\mu_{\mathcal{K}}(M)$. Furthermore if $\tilde{\beta} = \hat{\beta}$

then this bound corresponds to a local maximum of (2.15). In a significant number of cases the iteration does not converge, and the resulting guess for $Q \in \mathcal{Q}_{\mathcal{K}}$ yields a poor lower bound for μ. These are precisely the cases which need improvement.

The SPA vs. Standard Optimization

In order to examine the average computational requirements both for the SPA implemented in Matlab[1] (with no .mex files) and for directly solving (2.15) via the standard optimization techniques of NPSOL[2], also implemented in Matlab (with .mex files), we ran both algorithms 100 times on random complex matrices with independent normally distributed elements, and collected statistical data. This was repeated for problems of various sizes.

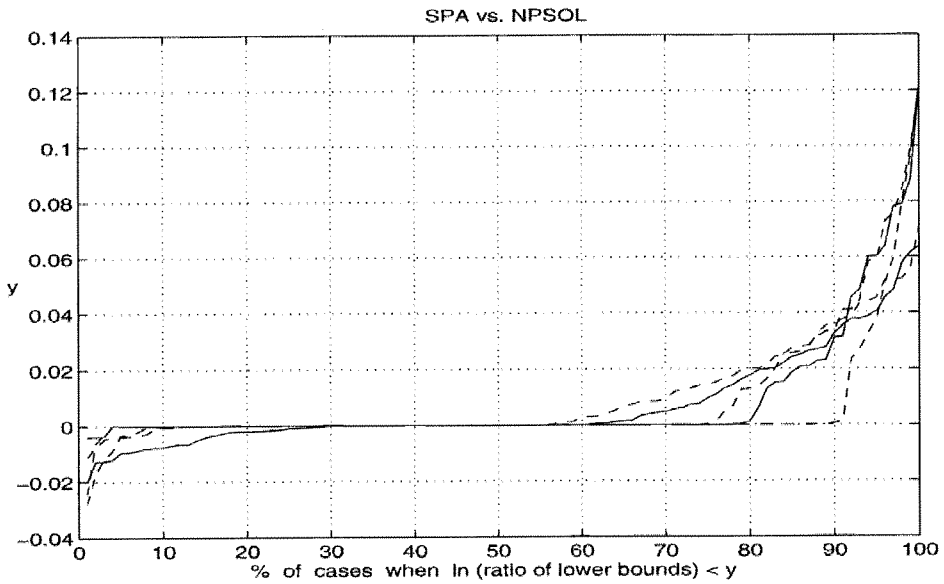

Figure 3.1: ln((SPA lower bound)/(NPSOL lower bound)) for problems of size 4 through size 30. The SPA computes better bounds: its bounds are often significantly better, and seldom significantly worse than NPSOL's bounds.

The ratio of computed lower bounds and computational requirement versus matrix size are shown in Figure 3.1 and Figure 3.2 respectively for a block

[1]Matlab is available from The MathWorks, Inc., Cochituate Place, 24 Prime Park Way, Natick, MA 01760

[2]Information about NPSOL is available from the Stanford Office of Technology Licensing, 350 Cambridge Ave, Suite 250, Palo Alto, CA 94306

structure consisting of complex scalar uncertainties. The flop count for the SPA includes an upper bound computation, while the flop count for the standard optimization doesn't include the .mex file flops, so the difference between the algorithms is even more pronounced than that shown in Figure 3.2.

It can be seen that the SPA is much faster, even though the lower bound obtained is better. This advantage becomes more significant as the size of problem grows. Note that NPSOL is finding local maximums for these problems; the SPA computes better bounds because it tends to find larger local maximums.

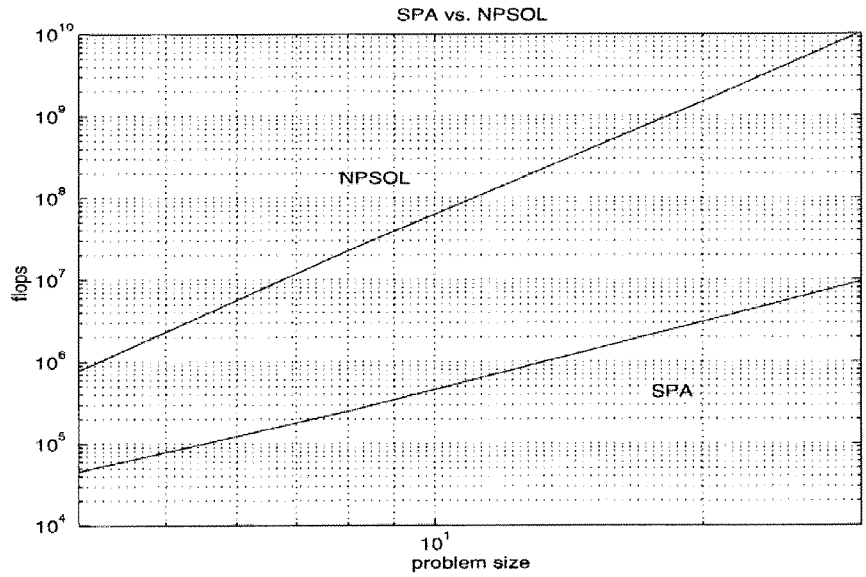

Figure 3.2: SPA flop count and a partial NPSOL flop count for problems of size 4 through size 30. Clearly, the SPA can solve much larger problems than NPSOL can.

3.4 The SPA Using the Rank One Solution

In this section we explain a connection between local maximums of $\rho_R(QM)$ and the solution of a rank one μ problem. The power iteration attempts to force the right and left eigenvectors of QM to satisfy the alignment condition of Theorem 4 associated with a local maximum of $\rho_R(QM)$. It turns out that this alignment condition is also associated with the solution to a certain rank one μ problem formed from these vectors.

The resulting rank one μ problem leads to a new way of updating Q in the SPA. The new algorithm is nearly as fast as the SPA and provides better bounds. This is the simplest improvement over the SPA in this chapter.

Rank One $\mu_{\mathcal{K}}(M)$

The rank one mixed μ problem is the special case of computing $\mu_{\mathcal{K}}(M)$ where M is a dyad. This is a much easier problem as any non-zero local maximum of $\rho_R(QM)$ is global ([5]). The following theorem from [46] characterizes the solution to the rank one problem in terms of the alignment conditions of Theorem 4.

Theorem 5 *Suppose we have $M = uv^*$, with $u, v \in \mathbf{C}^n$ satisfying the non-degeneracy assumptions, and a compatible block structure $\mathcal{K}$. Further suppose we have $Q \in \mathcal{Q}_{\mathcal{K}}$ with $q_i^r \neq 0$ for $i = 1, \ldots, m_r$ such that $\rho_R(QM) = \beta > 0$. Then we have $\beta = \mu_{\mathcal{K}}(M)$ iff there exists $D \in \mathcal{D}_{\mathcal{K}}$ with $\theta_i = \pm\frac{\pi}{2}$ for $|q_i^r| < 1$ and $\psi \in (-\frac{\pi}{2}, \frac{\pi}{2})$ such that $v = e^{j\psi}DQu$.*

The simplicity of the rank one mixed μ problem is apparent in the following graphical interpretation in the complex plane. Let $M = uv^*$, $\Delta \in \mathbf{B}_{\mathbf{\Delta}_{\mathcal{K}}}$, and $0 < \alpha \in \mathbf{R}$. Then

$$\begin{aligned} 0 &= \det(I - \frac{\Delta}{\alpha}M) = \det(1 - v^*\frac{\Delta}{\alpha}u) \\ &\Leftrightarrow \; v^*\Delta u = \alpha \\ &\Leftrightarrow \; \sum_{i=1}^{m_r} \delta_i^r v_{r_i}^* u_{r_i} + \sum_{i=1}^{m_c} \delta_i^c v_{c_i}^* u_{c_i} + \sum_{i=1}^{m_C} v_{C_i}^* \Delta_i^C u_{C_i} = \alpha. \end{aligned}$$

This equation is the basis of the graphical interpretation of the rank one mixed μ problem. The problem is solved simply by choosing δ_i^r, δ_i^c, and Δ_i^C so that vectors $\delta^r v_r^* u_r$, $\delta^c v_c^* u_c$, and $v_C^* \Delta^C u_C$ add up to the largest possible positive real number. The contributions of the complex blocks cannot be optimal unless they are collinear; they must all have the same phase. So, we have

$$\sum_{i=1}^{m_c} \delta_i^c v_{c_i}^* u_{c_i} + \sum_{i=1}^{m_C} v_{C_i}^* \Delta_i^C u_{C_i} = e^{j\Psi} L_C, \tag{3.12}$$

and the rank one mixed μ problem reduces to choosing real numbers $\delta_i^r \in [-1, 1]$ and $\Psi \in [-\pi, \pi]$ to maximize α where

$$\sum_{i=1}^{m_r} \delta_i^r v_{r_i}^* u_{r_i} + e^{j\Psi} L_c = \alpha. \tag{3.13}$$

By thinking of (3.13) as a vector sum in the complex plane it is apparent that the solution to this problem can easily be obtained geometrically.

The following algorithm is guaranteed to compute μ exactly for a rank one mixed μ problem. It also gives us an optimal perturbation $Q \in \mathcal{Q}_\mathcal{K}$.

- Choose starting values for the real perturbations as $\delta_i^r = \text{sgn}(\text{Re}(v_{r_i}^* u_{r_i}))$
- Compute $S = \text{sgn}(\Im(\sum_{i=1}^{m_r} \delta_i^r v_{r_i}^* u_{r_i}))$
- Rank all the components $\delta_i^r v_{r_i}^* u_{r_i}$ by argument $\angle(S\delta_i^r v_{r_i}^* u_{r_i})$ in descending order
- For the highest rank component not yet considered, assign the optimal value of this δ_i^r unrestricted in sign or magnitude to Δ_{opt} (the optimal value depends on L_C and $\angle(S\delta_i^r v_{r_i}^* u_{r_i})$)
- If $\text{sgn}(\Delta_{opt}) = -\text{sgn}(\delta_i^r)$ and $|\Delta_{opt}| > 1$ and this is not the last possible component, then go back to the previous step, else proceed. In either case, reassign $\delta_i^r = \max(-1, \min(\Delta_{opt}, 1))$
- We now have an optimal δ_i^r and can easily compute the optimal value of Ψ, and then the perturbation Q

The algorithm requires at most a search over real parameters, the cost of which grows linearly with m_r. Thus we have a closed form solution with trivial computational requirements for both $\mu_\mathcal{K}(M)$ and the associated $Q \in \mathcal{Q}_\mathcal{K}$.

The Rank One Algorithm (ROA)

The relationship between each μ problem and a corresponding rank one μ problem is shown by the following Theorem, which connects the alignment conditions of Theorem 4 with a particular rank one μ problem.

Theorem 6 *Suppose we have the matrix $M \in \mathbf{C}^{n\times n}$ and $Q \in \mathcal{Q}_{\mathcal{K}}$ such that QM has a real positive eigenvalue. Further suppose that $q_i^r \neq 0$ for $i = 1, \ldots, m_r$, and that the corresponding right and left eigenvectors of QM, denoted by x and y respectively, satisfy the non-degeneracy assumption. Then there exists $D \in \mathcal{D}_{\mathcal{K}}$ with $\theta_i = \pm\frac{\pi}{2}$ for $|q_i^r| < 1$ and $\psi \in (-\frac{\pi}{2}, \frac{\pi}{2})$ such that $y = e^{j\psi}Dx$ iff the matrix $Q \in \mathcal{Q}_{\mathcal{K}}$ solves the rank one μ problem $\max_{\hat{Q}\in\mathcal{Q}_{\mathcal{K}}} \rho_R(\hat{Q}M_{r1})$ where $M_{r1} = \hat{x}y^*$ and $\hat{x} = Q^{-1}x$.*

Proof By assumption we have $y^*x > 0$ so that $y^*Q\hat{x} > 0$ and hence $\rho_R(QM_{r1}) > 0$. The result then follows from Theorem 5. ■

The resulting rank one problem is easily solved, and allows us to choose $Q \in \mathcal{Q}_{\mathcal{K}}$ which is consistent with the alignment condition. This is used to modify the power iteration as follows

- Start with initial guesses for $b, w \in \mathbf{C}^n$
- ▶ Update a with the power step $\tilde{\beta}a = Mb$
- Compute the $Q \in \mathcal{Q}_{\mathcal{K}}$ that maximizes $\rho_R(Qaw^*)$
- Update z with $z = Q^*w$
- Update w with the power step $\hat{\beta}w = M^*z$
- Compute the $Q \in \mathcal{Q}_{\mathcal{K}}$ that maximizes $\rho_R(Qaw^*)$
- Update b with $b = Qa$
- If converged, then stop, else go to ▶

Sometimes the rank one problem constructed from the eigenvectors of a local maximum will have multiple solutions. Some of these solutions of the rank one mixed μ problem might not correspond to equilibrium points of the original problem. This can happen when two or more products $w_{r_i}^* a_{r_i}$ have the same phase or opposite phase. A procedure that chooses the correct solution is currently being investigated.

.5 The Wrap in Reals Algorithm (WRA)

This section presents a more substantial modification to the SPA. The principal difficulty with the SPA is that if $\rho_R(QM) \ll \rho(QM)$ then the algorithm often doesn't converge. One way of thinking of this is that the "power" steps (3.7) and (3.9) are increasing the component corresponding to $\rho(QM)$ faster than the rest of the steps can move towards the solution corresponding to $\rho_R(QM)$. The idea in this section is to utilize the good convergence properties of the SPA on complex blocks, while proceeding more cautiously on the real blocks.

For the remainder of the chapter we introduce the following notation. We separate the perturbation Q into its real and complex components, Q_r and Q_c, as follows.

$$Q = \begin{pmatrix} Q_r & 0 \\ 0 & Q_c \end{pmatrix}.$$

We partition the vectors a, b, z, and w and the matrix M compatibly:

$$a = \begin{pmatrix} a_r \\ a_c \end{pmatrix}, \quad b = \begin{pmatrix} b_r \\ b_c \end{pmatrix}, \quad z = \begin{pmatrix} z_r \\ z_c \end{pmatrix}, \quad w = \begin{pmatrix} w_r \\ w_c \end{pmatrix},$$

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

The following theorem, proved in [44], gives alignment conditions in terms of the real and complex block components in a way that allows for an algorithm that updates the real blocks independently of the complex blocks.

Theorem 7 *For a given matrix $M \in \mathbf{C}^{n\times n}$ and a given uncertainty structure $\mathcal{K}$, suppose we have $Q \in \mathcal{Q}_\mathcal{K}$, non-zero vectors a, b, z, w, and a positive scalar β such that $\beta I - M_{11}Q_r$ is nonsingular. Further assume that $b_c \neq 0$, $z_c \neq 0$ and all the diagonal elements of Q_r are nonzero.*

Then the following conditions hold:

$$Mb = \beta a \qquad M^*z = \beta w \tag{3.14}$$

$$b = Qa \qquad b = D^{-1}w \tag{3.15}$$

$$z = Q^*QDa \qquad z = Q^*w$$

iff there exist non-zero vectors a_c, b_c, z_c, w_c *satisfying the following conditions:*

$$M_C b_c = \beta a_c \qquad M_C^* z_c = \beta w_c \tag{3.16}$$

$$b_c = Q_c a_c \qquad b_c = D_c^{-1} w_c \tag{3.17}$$

$$z_c = Q_c^* Q_c D_c a_c \qquad z_c = Q_c^* w_c$$

$$Q_r(\beta I - M_{11}Q_r)^{-1}M_{12}b_c = D_r^{-1}(\beta I - M_{11}^* Q_r)^{-1}M_{21}^* z_c \tag{3.18}$$

where $M_C = (M_{22} + M_{21}Q_r(\beta I - M_{11}Q_r)^{-1}M_{12})$.

An iterative algorithm which separates the real and complex blocks in this way is as follows:

- Start with some given values for b, w, β, Q_r, and M
- ▶ Update a_c with the power step $\tilde{\beta} a_c = M_C b_c$
- Update z_c as in the SPA
- Update w_c with the power step $\hat{\beta} w_c = M_C^* z_c$
- Update b_c as in the SPA
- If converged, then go to the next step, else go to ▶
- Compute $a_r = (\beta I - M_{11}Q_r)^{-1}M_{12}b_c$ and $w_r = (\beta I - M_{11}^* Q_r)^{-1}M_{21}^* z_c$
- Update Q_r
- Update β
- Compute $M_C = (M_{22} + M_{21}Q_r(\beta I - M_{11}Q_r)^{-1}M_{12})$
- If converged, then stop, else go to ▶

This specifies a class of algorithms, which may update Q_r in a variety of ways. Additionally, we run one rank one type iteration first, in order to get the starting values of b, w, β, and Q_r needed for the WRA.

Compared to the ROA, the WRA is computationally more expensive, but the convergence properties are much better, even with a simple Q_r update. It is thus reasonable to mix the two procedures, and only use the slower and more reliable one in the cases where the faster scheme fails to converge.

In the SPA at a local maximum of $\rho_R(MQ)$, the vector a is not necessarily in the direction of the largest eigenvalue of MQ, and the algorithm might not converge if $\rho_R(MQ) \ll \rho(MQ)$. In the WRA, however, at a local maximum of $\rho_R(MQ)$ the vector a_c always corresponds to the largest eigenvalue of $M_C Q_c$. The WRA may be thought of as first implementing the SPA on the complex part of the problem with the real part fixed (the SPA has very good convergence properties on complex problems), then improving the real part of the problem with the complex part fixed. While [44] uses this idea of separating the real and complex parts of the problem, they do not exploit the convergence properties of the SPA on complex problems. Thus it is not surprising that the WRA has significantly better convergence properties than both the SPA and the algorithm in [44].

3.6 The Shift and Inverse Algorithm (SIA)

Another way to find an eigensolution that does not correspond to $\rho(MQ)$ is based on the observation that if $\lambda x = MQx$ and β^s is not an eigenvalue, then

$$(\lambda - \beta^s)^{-1} x = (MQ - \beta^s I)^{-1} x.$$

When λ and β^s are close to each other, $(\lambda - \beta^s)^{-1}$ is a very large eigenvalue of $(MQ - \beta^s I)^{-1}$, and a power iteration based on this equation will converge to the eigensolution $\lambda x = MQx$ of MQ. Such an iteration is called a shift and inverse iteration.

We use this idea to modify the SPA. The Q updates are as in the SPA, except that now Q must be formed explicitly. This algorithm, called the SIA, proceeds as follows

- Start with a Q, a β^s close to an eigenvalue of MQ, an a close to a right eigenvector of MQ, and a w close to a left eigenvector of QM.
- ▶ Update a_k and $\tilde{\beta}$ with the inverse power step $(\tilde{\beta} - \beta^s)^{-1} a_k = (MQ - \beta^s I)^{-1} a_{k-1}$ with $\tilde{\beta} \in \mathbf{R}^+$ and $|a_k| = 1$.

- Update $Q \in \mathcal{Q}_{\mathcal{K}}$ and β^s.
- Update w_k and $\hat{\beta}$ with the inverse power step $(\hat{\beta} - \beta^s)^{-1} w_k = (M^* Q^* - \beta^s I)^{-1} w_{k-1}$ with $\hat{\beta} \in \mathbf{R}^+$ and $|w_k| = 1$.
- Update $Q \in \mathcal{Q}_{\mathcal{K}}$ and β^s.
- If converged, then stop, else go to ▶ .

This algorithm converges very well when we start close to an eigensolution. When we do not start close to an eigensolution, however, it performs very poorly. Thus it is appropriate to use it only in conjunction with some other algorithm that gets us near a solution first.

3.7 Combining the Algorithms

In previous sections we introduced three new algorithms for mixed μ lower bound computation. Here, we aim to combine them in a way that utilizes their respective good qualities and avoids their bad qualities. All the algorithms have enhanced convergence properties compared to the SPA, but are also more computationally costly. Since the ROA is computationally the least expensive—nearly as inexpensive as the SPA—we want to preserve the speed of the ROA for those problems where the ROA converges, while improving convergence (and therefore accuracy) for those problems where the ROA fails to converge.

In the cases where the ROA fails to converge, we need to continue looking for a local maximum with one of our other algorithms. Note that all our algorithms start with some guess for the perturbation or the vectors as input. If these are close to a local maximum then the algorithms perform particularly well. Consequently, when we decide to switch to another algorithm in the middle of a problem, we can take advantage of the computation already performed. Knowing that the SIA has particularly good properties only when we are quite close to the equilibrium point, we choose to continue with the WRA, which works better than the SIA when we aren't so close to an equilibrium point. In the case that the WRA also fails to converge—typically it just slows down, often near the optimum (this might be remedied with a better Q_r update)—we will continue with the SIA.

The following combination of the algorithms defines a new algorithm which we denote by CPA, for combined power algorithm. This is the CPA algorithm used in the next section, (i.e. this is the scheduling used).

- Run the ROA for up to 50 iterations
- If not yet converged, then run the WRA for up to 50 iterations
- If not yet converged, then run the SIA for up to 50 iterations
- If not converged, then construct a lower bound from the current perturbation

This scheduling reflects experience from testing not presented here. Certainly, there is room for improvement. In particular, the scheduling should depend on a priori knowledge of the problem.

3.8 Numerical Experience

The nature of the mixed μ problem is such that the only meaningful way to evaluate an algorithm is by testing it on a large number of representative problems. In this section we present a comparison of several algorithms, each run on the same type of problems. We also show how the performance of the best algorithm depends on problem size.

For the purpose of testing algorithms it is desirable to be able to generate nontrivial problems for which we know μ. A procedure given in [50] allows us to construct matrices where μ is equal to some specified value, and the Q achieving the maximum of $\rho_R(QM)$ typically satisfies $\rho_R(QM) \ll \rho(QM)$. We denote these matrices as the set $\mathcal{R}_{\mathcal{B}}$. We emphasize that these are particularly difficult problems for mixed μ lower bound computation. This is desirable because existing algorithms work well on most problems; the point of this research is to do better in the cases where existing code is inadequate.

Figure 3.3 shows a comparison between the algorithms described here and also the algorithm in [44] which is called the APA. The SIA is not included as it performs poorly unless it is given a good starting point; it is only meaningful in conjunction with another algorithm. We tested the algorithms on 500 matrices in the set $\mathcal{R}_{\mathcal{B}}$ each with 4 real parameters, 2 scalar complex blocks and one

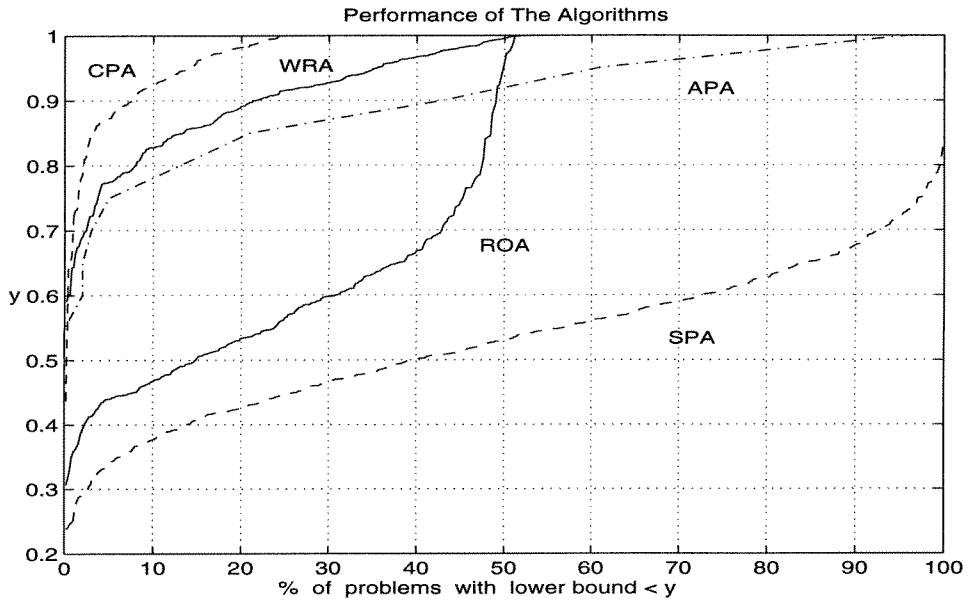

Figure 3.3: Algorithm performance on 500 hard problems.

2×2 complex block in the block structure, and with $\mu = 1$. The results are dramatic.

Figure 3.4 shows how the bound computation using the best algorithm, CPA, becomes more difficult as the problem size increases. The best performance shown is with 2 real parameters, and the worst is with 12. All problems were hard problems from the set $\mathcal{R}_{\mathcal{B}}$ with $\mu = 1$, and with 2 scalar complex blocks and one 2×2 complex block in the block structure. These results further motivate research in better bounds computation.

3.9 Conclusion

Even though the general mixed μ problem is NP hard, which strongly suggests that the exact computation for the general mixed μ problem will be computationally intractable, reasonable computation of upper and lower bounds is possible.

Several new approaches to computing an improved μ lower bound have been presented here. These algorithms have been combined to yield a substantially improved power algorithm. The improvement on a large set of particularly difficult problems is shown in Figure 3.3 and is the justification for this work.

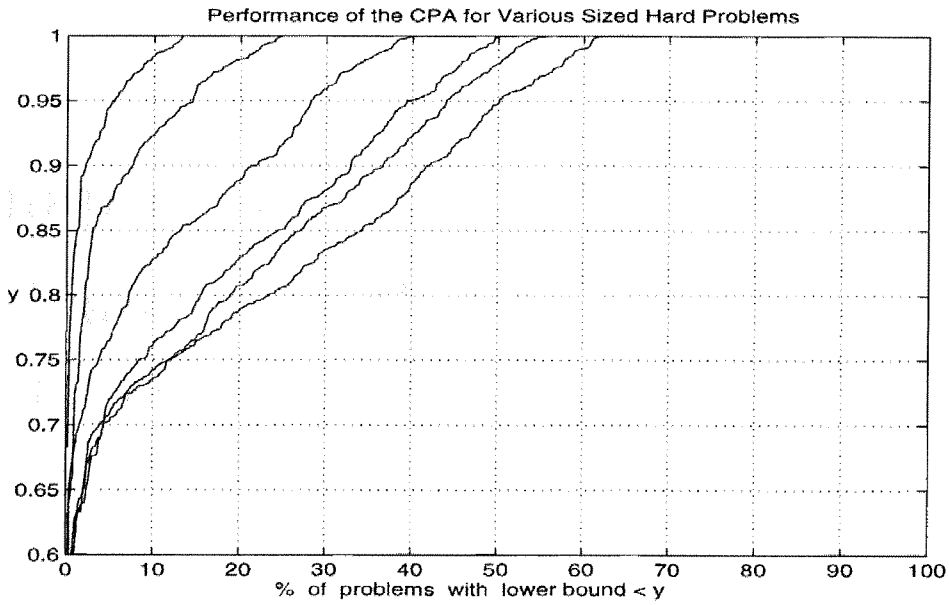

Figure 3.4: Computation is more difficult as problem size increases from 2 to 12 real parameters. CPA on $\mathcal{R}_B$ matrices.

Furthermore, the algorithms described and compared here are by no means optimized, so it is reasonable to expect that this work will lead to even better results than those in Section 3.8.

Chapter 4

Advances in Worst-case $\mathcal{H}_\infty$ Performance Computation

In this chapter we consider uncertain systems with mixed structured uncertainty entering as a linear fractional transformation and ask the following question: given an uncertain system, and given a bound on the norm of the uncertainty, find the worst case $\mathcal{H}_\infty$ gain. This problem was introduced in [12] for the complex uncertainty case.

Although this worst case gain can not be computed exactly, we will show that both upper and lower bounds on this worst case gain can be computed efficiently. The computational effort required by these bounds is similar to the one required to compute upper and lower bounds of the structured singular value μ in the frequency domain.

At each frequency point a skewed-μ problem, a mixed version (i.e. the block structure contains both real and complex uncertainties) of ν [12], is solved. An upper bound can be computed by solving a quasi convex optimization problem, for which efficient algorithms exist. The main contribution of this chapter is to show that an efficient power algorithm can be developed for the skewed-μ lower bound, based on the corresponding algorithm for the mixed-μ lower bound. We test the algorithm extensively, both on systems generated at random, and on systems derived from real life engineering applications.

The rest of this chapter is organized as follows: In the next section we define the skewed structured singular value. In Section 4.2 we show how a ν lower bound can be posed as a maximization problem and generalize this approach to the skewed μ lower bound.

In Section 4.3 we explain the connection between ν and μ and propose the

algorithm used for computation of the skewed μ lower bound. In Section 4.4 we present the corresponding skewed μ upper bound. Finally in Section 4.5 we study the behavior of the algorithms when tested on the model of an airplane control system during the flare phase of an automatic landing.

4.1 Skewed-μ

In this section we will define the skewed structured singular value as a mixed version (i.e. the block structure contains both real and complex uncertainties) of ν introduced in [12] and presented in Section 2.5. Given a matrix $M \in \mathbf{C}^{(n+n_s)\times(n+n_s)}$ and three non-negative integers m_r, m_c, and m_C with $m \doteq m_r + m_c + m_C \leq n$, the block structure $\mathcal{K}_s(m_r, m_c, m_C + 1)$ is an $(m+1)$-tuple of positive integers

$$\mathcal{K}_s(m_r, m_c, m_C + 1) = (\mathcal{K}(m_r, m_c, m_C), n_s)$$

where we require $\sum_{i=1}^{m} k_i = n$ so that the dimensions are compatible. Define the set of allowable perturbations as follows.

$$\mathbf{\Delta}^s_{\mathcal{K}_s} \doteq \{\Delta_s = \text{blockdiag}(\Delta_f, \Delta) : \Delta_f \in \mathbf{\Delta}_{\mathcal{K}}, \bar{\sigma}(\Delta_f) \leq 1, \Delta \in \mathbf{C}^{n_s \times n_s}\}. \quad (4.1)$$

We partition a matrix M in accordance with Δ_s, as shown in figure 4.1.

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

Definition 3 *The skewed structured singular value, $\mu^s_{\mathcal{K}_s}(M)$, of a matrix $M \in \mathbf{C}^{(n+n_s)\times(n+n_s)}$, such that $\mu_{\mathcal{K}}(M_{11}) < 1$, with respect to a block structure $\mathcal{K}_s(m_r, m_c, m_C + 1)$ is defined as*

$$\mu^s_{\mathcal{K}_s}(M) \doteq \left(\min_{\Delta_s \in \mathbf{\Delta}^s_{\mathcal{K}_s}} \{\bar{\sigma}(\Delta) : \det(I - \Delta_s M) = 0, \Delta_s = \text{blockdiag}(\Delta_f, \Delta)\} \right)^{-1}$$

with

$\mu^s_{\mathcal{K}_s}(M) \doteq 0$ *if no* $\Delta_s \in \mathbf{\Delta}^s_{\mathcal{K}_s}$ *solves* $\det(I - \Delta_s M) = 0$.

According to (2.18) The supremum over frequency of the skewed structured singular value can then be used to answer the following question:

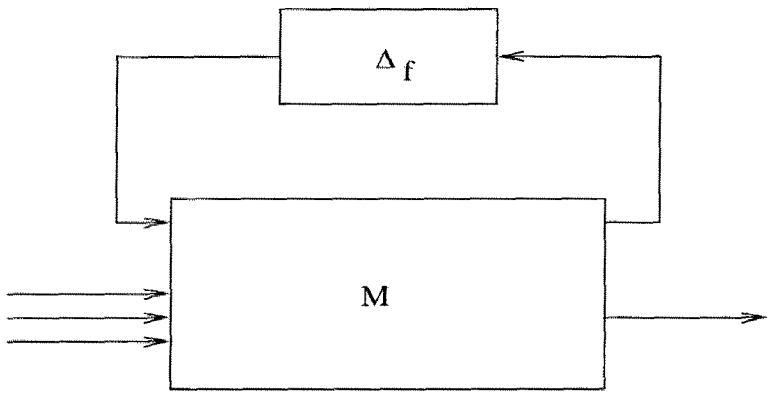

Figure 4.1: LFT interconnection of system and uncertainty.

Question 2 *What is the worst case $\mathcal{H}_\infty$ gain of the system if the uncertainty has size 1 or less?*

Definition 3 suggests that $\mu^s_{\mathcal{K}_s}$ may be related to μ . This is indeed the case as stated in Proposition 1 which is a mixed version of the proposition from [12].

Proposition 1 *Suppose $\mu^s_{\mathcal{K}_s} < \infty$. Then*

$$\mu^s_{\mathcal{K}_s}(M) = \mu\left(\begin{bmatrix} \mu^s_{\mathcal{K}_s}(M) I_n & 0 \\ 0 & I_{n_s} \end{bmatrix} M\right). \tag{4.2}$$

Proof Let $\mu^s_{\mathcal{K}_s} = \mu^s_{\mathcal{K}_s}(M)$ and let μ_s be the right hand side of (4.2). We show that $\mu_s \geq \mu^s_{\mathcal{K}_s}$ and $\mu^s_{\mathcal{K}_s} \geq \mu_s$. The former holds trivially if $\mu^s_{\mathcal{K}_s} = 0$. Thus, assume $\mu^s_{\mathcal{K}_s} > 0$.From the definition of $\mu^s_{\mathcal{K}_s}(\cdot)$, there exists $\Delta_s = diag\{\Delta_f, \Delta\} \in \mathbf{\Delta}^s_{\mathcal{K}_s}$ such that $\overline{\sigma}(\Delta) = (\mu^s_{\mathcal{K}_s})^{-1}$ and

$$det(I - \Delta_s M) = det\left(I - \begin{bmatrix} (\mu^s_{\mathcal{K}_s}(M))^{-1}\Delta_f & 0 \\ 0 & \Delta \end{bmatrix}\begin{bmatrix} \mu^s_{\mathcal{K}_s}(M) I_n & 0 \\ 0 & I_{n_s} \end{bmatrix} M\right) = 0.$$

Since

$$\overline{\sigma}\left(\begin{bmatrix} (\mu^s_{\mathcal{K}_s}(M))^{-1}\Delta_f & 0 \\ 0 & \Delta \end{bmatrix}\right) = (\mu^s_{\mathcal{K}_s})^{-1},$$

in view of the definition of $\mu(\cdot)$, we have

$$\mu_s \geq \mu^s_{\mathcal{K}_s}.$$

Again, $\mu^s_{\mathcal{K}_s} \geq \mu_s$ holds trivially if $\mu_s = 0$. Thus, assume $\mu_s > 0$. Let $\hat{\Delta} = diag\{\hat{\Delta}_f, \hat{\Delta}\}$ such that $\overline{\sigma}(\hat{\Delta}_f) = \overline{\sigma}(\hat{\Delta}) = (\mu_s)^{-1}$ and

$$\det\left(I - \hat{\Delta}\begin{bmatrix} \mu^s_{\mathcal{K}_s}(M)I_n & 0 \\ 0 & I_{n_s} \end{bmatrix} M\right) = \det\left(I - \begin{bmatrix} \mu^s_{\mathcal{K}_s}(M)\hat{\Delta}_f & 0 \\ 0 & \hat{\Delta} \end{bmatrix} M\right) = 0.$$

Since $\mu_s \geq \mu^s_{\mathcal{K}_s}$, we have $diag\{\mu^s_{\mathcal{K}_s}\hat{\Delta}_f, \hat{\Delta}\} \in \mathbf{\Delta}^s_{\mathcal{K}_s}$, and from the definition of $\mu^s_{\mathcal{K}_s}(\cdot)$,

$$\mu^s_{\mathcal{K}_s} \geq (\overline{\sigma}(\hat{\Delta}))^{-1} = \mu_s.$$

■

4.2 The Skewed μ Lower Bound

In Section 2.5 we suggested that Question 2 could be answered via an infinite sequence of μ analyses. First we will make it precise here how ν can indeed be obtained from μ, and then we will show how this approach can be generalized to obtaining skewed μ from μ.

Define the function $f : \mathbf{R} \longrightarrow \mathbf{R}$ as

$$f(\alpha) = \mu\left(\begin{bmatrix} \alpha I_{k_1} & 0 \\ 0 & I_{k_2} \end{bmatrix} M\right). \tag{4.3}$$

It has been shown in [31] that the function $f(\alpha)$ is continuous and nondecreasing.

The following proposition is taken from [12].

Proposition 2 *(a) $\beta > \nu(M)$ implies that $f(\beta) < \beta$,*
(b) $0 < \beta < \nu(M) < \infty$ implies that $f(\beta) > \beta$.

Theorem 8 *([12]) Let $\{\alpha_k\}$ be the sequence generated by the fixed point iteration*

$$\alpha_{k+1} = f(\alpha_k), \quad k = 1, 2, \ldots, \tag{4.4}$$

with α_0 any positive number. Then

$$\lim_{k\to\infty} \alpha_k = \nu(M). \tag{4.5}$$

Proof Follows directly from Proposition 2 and the continuity of f. ∎

The theoretical basis for the skewed-μ lower bound is the fact that the μ problem may be reformulated as an eigenvalue maximization (2.15) which can also be posed in the following form

$$\mu_{\mathcal{K}}(M) = \max_{Q \in \mathcal{Q}_{\mathcal{K}}} \{\beta : det(\beta I_n - QM) = 0\}.$$

In order to obtain and refine a lower bound for $\mu^s_{\mathcal{K}_s}$ we define the following set of block diagonal matrices, which depend on the underlying block structure.

$$\mathcal{Q}^s_{\mathcal{K}_s} \doteq \{Q_s = \text{blockdiag}(Q_f, Q) : Q_f \in \mathcal{Q}_{\mathcal{K}} \;\; Q^*Q = I_{n_s}\}.$$

According to the Theorem 2 the $\mu^s_{\mathcal{K}_s}$ may then be reformulated as a following maximization

$$\begin{aligned} \mu^s_{\mathcal{K}_s}(M) &= \max_{Q \in \mathcal{Q}^s_{\mathcal{K}_s}} \{\beta : det(\beta I_{n+n_s} - Q \begin{bmatrix} \beta I_n & 0 \\ 0 & I_{n_s} \end{bmatrix} M) = 0\} \\ &= \max_{Q \in \mathcal{Q}^s_{\mathcal{K}_s}} \{\beta : det(\beta I_{n+n_s} - QM_\beta) = 0\} \end{aligned} \tag{4.6}$$

where

$$M_\beta = \begin{bmatrix} \beta I_n & 0 \\ 0 & I_{n_s} \end{bmatrix} M.$$

Since this maximization problem is not convex we will in general only be able to find local maxima. For any $Q \in \mathcal{Q}^s_{\mathcal{K}_s}$ for which exist β such that $det(\beta I_{n+n_s} - QM_\beta) = 0$, $\beta \leq \mu^s_{\mathcal{K}_s}$, and any such $Q \in \mathcal{Q}^s_{\mathcal{K}}$ immediately gives us a lower bound for $\mu^s_{\mathcal{K}_s}(M)$. Efficient computation of a local maximum of the function $\rho(QM_\beta)$ over $Q \in \mathcal{Q}^s_{\mathcal{K}_s}$ can be achieved by modification of the power iteration fully described in [32].

In the following subsection we show how the power algorithm for computing a lower bound for μ can be modified to compute a lower bound for $\mu^s_{\mathcal{K}_s}$. This work is based on [49], with additional improvements made by the authors.

4.3 Skewed μ Lower Bound Power Algorithm

To solve the maximization problem (4.6) and compute a lower bound β for $\mu^s_{\mathcal{K}_s}$ we developed an algorithm starting from a combined power algorithm for

computing a lower bound for μ. The following combination of the algorithms, introduced and explained in detail in the previous chapter, defines a new algorithm which we denote by SCPA, for skewed combined power algorithm.

- Run the ROA for up to 50 iterations
- If not yet converged, then run the WRA for up to 50 iterations
- If not yet converged, then run the SIA for up to 50 iterations
- If not converged, then construct a lower bound from the current perturbation

Every stage of SCPA consisting of a different power iteration scheme has been modified in the following way.

1. Start with initial guesses for $b, w \in \mathbf{C}^n$
2. Update a with the power step $\tilde{\beta}a = M_\beta b$
3. Compute $Q \in \mathcal{Q}_\mathcal{K}$
4. Update z with $z = Q^* w$
5. Update w with the power step $\hat{\beta}w = M_\beta^* z$
6. Compute $Q \in \mathcal{Q}_\mathcal{K}$
7. Update b with $b = Qa$
8. If converged, then stop, else go to 2

With every new update for β we actually change matrix M_β used in the power iteration. This is definitely going to reduce the efficiency of the SCPA algorithm compared to the CPA algorithm where in every power iteration we use the same matrix M.

The nature of the skewed-μ problem is such that the only meaningful way to evaluate an algorithm is by testing it on a large number of representative problems. In the rest of this section we present a performance analysis for the proposed algorithm.

For the purpose of testing the power algorithm it is desirable to be able to generate nontrivial problems for which we know $\mu^s_{\mathcal{K}_s}$. A procedure given in [50] allows us to construct matrices where μ is equal to some specified value. We denote these matrices as the set $\mathcal{R}_{\mathcal{B}}$. Despite the fact that we have no convergence guarantees, the algorithm in fact converges most of the time. We tested the algorithm on 100 matrices in the set $\mathcal{R}_{\mathcal{B}}$ each with k real parameters, two complex scalar blocks and one (2×2) complex block in the block structure. We fixed the size of the first $k+2$ blocks to be less than or equal to 1, having then that $\mu^s_{\mathcal{K}_s} = 1$, and varied k from 2 to 12. Table 4.1 shows how the percentage of cases in which our algorithm converges in less than 100 iterations varies with the block size.

It is important to note that, even when the power algorithm does not converge, a lower bound on $\mu^s_{\mathcal{K}_s}$ is still obtained. Figure 4.2 shows the distribution of the answers given by the power algorithm (The true value of $\mu^s_{\mathcal{K}_s}$ is 1). The behavior of the algorithm degrades with increase of a number of real parameters, but is still more than satisfactory.

Table 4.1: Numerical evaluation of the Power Algorithm

Size of uncertainty blocks	2	4	6	8	10	12
% cases that converge	99	99	93	91	91	89

4.4 The Skewed μ Upper Bound

To obtain a tractable upper bound for $\mu^s_{\mathcal{K}_s}$ we introduce the following set of positive definite block diagonal scaling matrices, which commute with the elements in $\mathbf{\Delta}^s_{\mathcal{K}_s}$.

$$\mathcal{D}^s_{\mathcal{K}_s} \doteq \{D_s = \text{blockdiag}(D_f, D) : D_f \in \mathcal{D}_{\mathcal{K}}, D = D^* \in \mathbf{C}^{n_s \times n_s}\}$$

Once again we use the familiar complex μ upper bound technique to obtain the following upper bound for skewed-μ.

$$\mu^s_{\mathcal{K}_s} \leq \inf_{D \in \mathcal{D}^s_{\mathcal{K}_s}} \min_{0 \leq \beta \in \mathbf{R}} \{\beta : M^*_\beta D M_\beta - \beta^2 D \leq 0\}$$

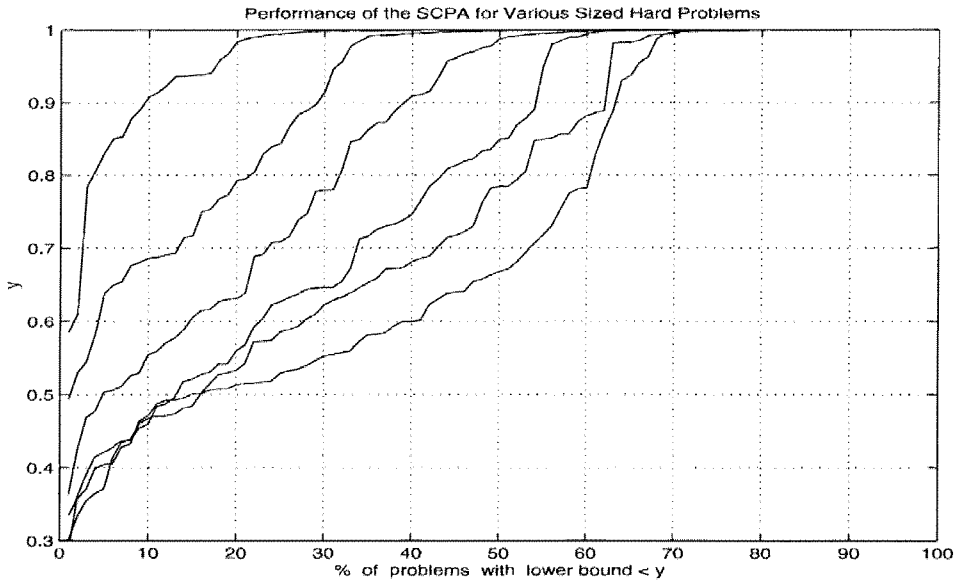

Figure 4.2: Computation is more difficult as problem size increases from 2 to 12 real parameters. SCPA on $\mathcal{R}_{\mathcal{B}}$ matrices.

$$= \inf_{D \in \mathcal{D}^s_{\mathcal{K}_s}} \min_{0 \leq \beta \in \mathbf{R}} \{\beta : M^* D M - \begin{bmatrix} I_n & 0 \\ 0 & \beta^2 I_{n_s} \end{bmatrix} D \leq 0\}$$

Computationally this reduces to a Linear Matrix Inequality. The interior point methods described in [2] can be used to efficiently compute this upper bound.

4.5 Worst-case $\mathcal{H}_\infty$ Performance Computation Results

We will study in this section the noise rejection capabilities of an airplane control system during the flare phase of an automatic landing. The disturbances considered are gusts in the longitudinal and vertical directions. The performance outputs are altitude and vertical speed. In this phase it is particularly important to minimize the effect of gusts on vertical speed, since deviation from the nominal can result in impact higher than acceptable at touchdown. The uncertainty corresponds to modeling error in certain aerodynamic coefficients. The system includes a 4 state longitudinal aerodynamic model of the airplane, plus the flare control law, and Dryden filters for the

wind gusts.

We computed $\mu^s_{\mathcal{K}_s}$'s upper and lower bounds for the airplane closed loop model over the $[0.01, 10]Hz$ frequency range. The model has 13 states, 5 inputs and 5 outputs. The fixed size part of uncertainty, Δ_f, consists of three real scalars, and we separately analyzed the $\mathcal{H}_\infty$ norm of the system for the two performance outputs.

Figures 4.3 and 4.4 show the results of computation. The gap between the upper and the lower bound is very small, meaning that the computation is very accurate.

4.6 Conclusion

We showed in this chapter that the worst case $\mathcal{H}_\infty$ gain of an uncertain system subject to norm bounded structured LTI perturbations can be written exactly in terms of the skewed structured singular value. Although, like μ, the skewed structured singular value can not be computed exactly, we discussed efficient algorithm to compute corresponding upper and lower bounds.

The algorithm was shown to perform satisfactorily both for test matrices generated artificially, and for systems derived from engineering applications. The results presented in this chapter thus show that the enhanced algorithm developed recently for the structured singular value, can be extended to the problem of computing worst case gains under fixed size uncertainty, without significant loss of performance or accuracy.

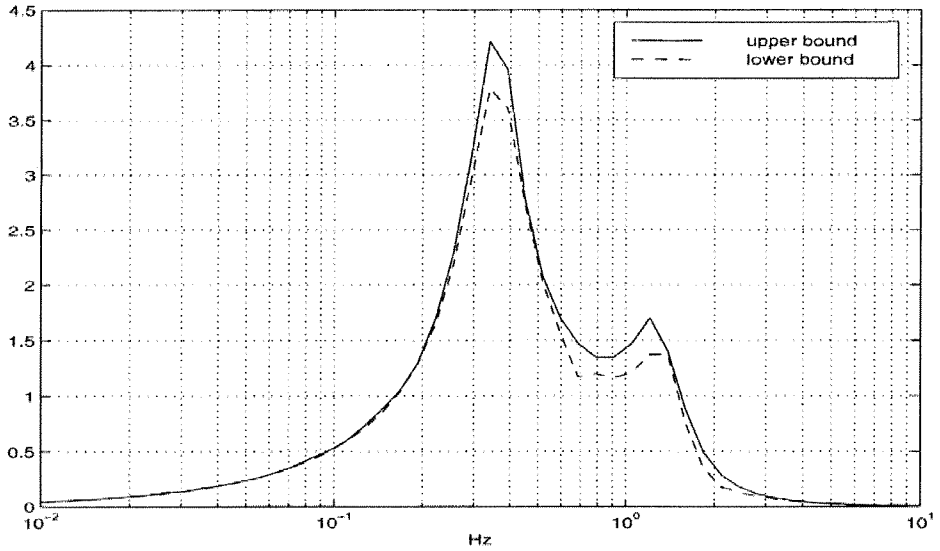

Figure 4.3: $\mu^s_{\mathcal{K}_s}$ upper and lower bound for the first output $\mathcal{H}_\infty$ gain of a landing airplane.

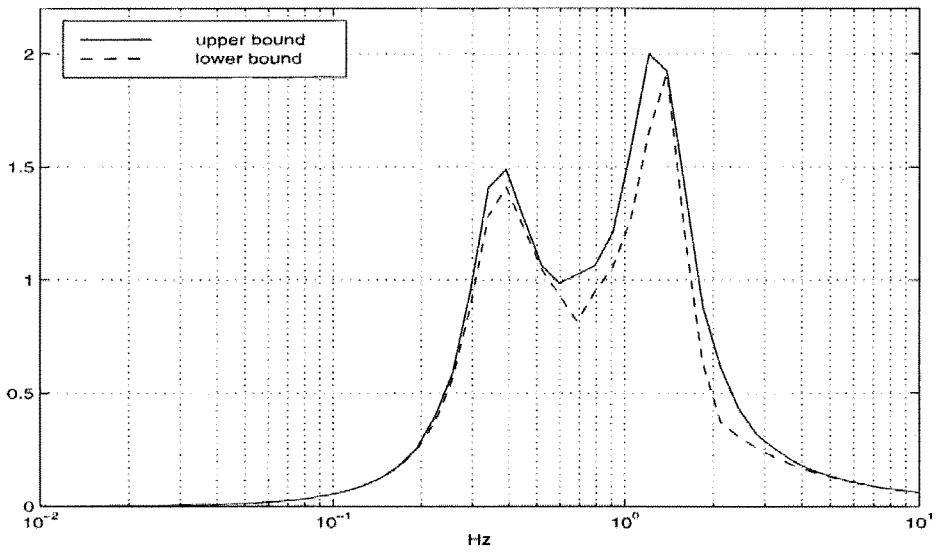

Figure 4.4: $\mu^s_{\mathcal{K}_s}$ upper and lower bound for the second output $\mathcal{H}_\infty$ gain of a landing airplane.

Chapter 5

Practical Upper and Lower Bounds for Robust $\mathcal{H}_2$ Performance under LTI Perturbations

The problem of computing the worst case $\mathcal{H}_2$ norm of an uncertain system has always been considered an important one, since many useful performance requirements are captured by it. Many recent publications have presented different approaches at solving this problem. (See [15, 37] and the references therein.) However, in all cases, the results developed provide only upper bounds on the given norm when the uncertainty is linear time invariant.

A convex condition for Robust $\mathcal{H}_2$ performance analysis under structured uncertainty, of a very similar nature to the corresponding condition for robust $\mathcal{H}_\infty$ performance, was introduced in [37]. Computationally, it reduces to solving a Linear Matrix Inequality (LMI) as a function of frequency, plus an integral over frequency for computing an upper bound on the $\mathcal{H}_2$ cost over a set of plants. This upper bound is shown to be necessary and sufficient for slowly linear time varying uncertainty. However no indication is given on its conservativeness when the uncertainty is LTI. Recent results show that in the MIMO case the gap can be as large as the square root of the number of inputs.

In this chapter we consider uncertain systems with structured complex uncertainty entering as a linear fractional transformation and ask the following question: given a MISO or SIMO uncertain system, and given a bound on the norm of the uncertainty, find the worst case $\mathcal{H}_2$ gain. We will show that both upper and lower bounds on this worst case gain can be computed. The

computation effort required by these bounds is similar to the one required to compute upper and lower bounds of the structured singular value μ in the frequency domain.

Since these robustness results are used mainly for analysis purposes, considering only MISO or SIMO systems is not a significant restriction. In a typical noise rejection application, for analysis purposes it is sufficient to consider one output at a time. Combining the outputs in a single performance index usually reduces the amount of information being looked at. (On the other hand, for synthesis purposes, it is clear that the MIMO case is the most interesting.)

The upper and lower bounds developed are based on integration over frequency as in [37]. At each frequency point a complex skewed-μ, a special case of ν [12] problem is solved. An upper bound can be computed by solving a quasi convex optimization problem, for which efficient algorithms exist. In this chapter we also show that an efficient power algorithm can be developed for the complex skewed-μ lower bound.

As an important caveat, it must be said that the conditions developed do not impose causality on the perturbations. We are currently studying ways of imposing the causality constraint. In particular we believe there are large classes of problems in which the worst case perturbation, as developed in this chapter, is causal.

The rest of this chapter is organized as follows: In the next section we define the complex skewed structured singular value, and the algorithms used for computation of the corresponding upper and lower bounds. In Section 5.2 we show how the complex skewed structured singular value can be used to compute the worst case $\mathcal{H}_2$ norm of a MISO or SIMO system. Finally in Section 5.3 we show some computational results.

5.1 Complex Skewed-μ

In this section we will define the complex skewed structured singular value as a special case of ν presented in section 2.5. Given a matrix $M \in \mathbf{C}^{(n+n_s)\times(n+n_s)}$ and two non-negative integers m_c, and m_C with $m \doteq m_c + m_C \leq n$, the block

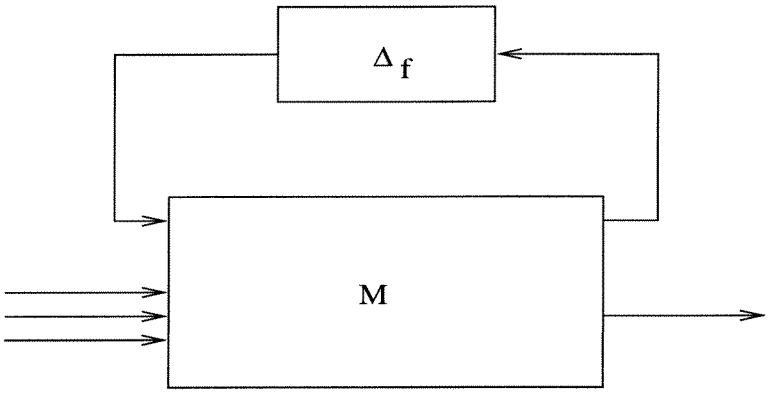

Figure 5.1: LFT interconnection of system and uncertainty.

structure $\mathcal{K}_s(m_c, m_C, n_s)$ is an $(m+1)$-tuple of positive integers

$$\begin{aligned} \mathcal{K}_s(m_c, m_C + 1) &= (\mathcal{K}(m_c, m_C), n_s) \\ &= (k_1, \ldots, k_{m_c}, k_{m_c+1}, \ldots, k_m, n_s) \end{aligned}$$

where we require $\sum_{i=1}^{m} k_i = n$ so that the dimensions are compatible. Define the set of allowable perturbations as follows.

$$\mathbf{\Delta}^s_{\mathcal{K}_s} \doteq \{\Delta_s = \text{blockdiag}(\Delta_f, \Delta) : \Delta_f \in \mathbf{\Delta}_{\mathcal{K}}, \bar{\sigma}(\Delta_f) \leq 1, \Delta \in \mathbf{C}^{n_s \times n_s}\}. \quad (5.1)$$

As shown in figure 5.1 we also partition a matrix M in accordance with Δ_s.

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

Definition 4 *[12] The complex skewed structured singular value, $\mu^s_{\mathcal{K}_s}(M)$, of a matrix $M \in \mathbf{C}^{(n+n_s)\times(n+n_s)}$, such that $\mu_{\mathcal{K}}(M_{11}) \leq 1$, with respect to a block structure $\mathcal{K}_s(m_c, m_C + 1)$ is defined as*

$$\mu^s_{\mathcal{K}_s}(M) \doteq \left(\min_{\Delta_s \in \mathbf{\Delta}^s_{\mathcal{K}_s}} \{\overline{\sigma}(\Delta) : \det(I - M\Delta_s) = 0, \Delta_s = \text{blockdiag}(\Delta_f, \Delta)\} \right)^{-1}$$

with

$\mu^s_{\mathcal{K}_s}(M) \doteq 0$ *if no* $\Delta_s \in \mathbf{\Delta}^s_{\mathcal{K}_s}$ *solves* $\det(I - \Delta_s M) = 0$.

The supremum over frequency of the complex skewed structured singular value can then be used to answer the following question:

Question 3 *What is the worst case $\mathcal{H}_2$ gain of the system if the uncertainty has size 1 or less?*

The Lower Bound

The complex skewed-μ lower bound is a special case of an eigenvalue maximization (4.6) and can also be posed as

$$\mu^s_{\mathcal{K}_s}(M) = \max_{Q \in \mathcal{Q}^s_{\mathcal{K}_s}} \{\beta : det(\beta I_{n+n_s} - QM_\beta) = 0\}. \tag{5.2}$$

where

$$M_\beta = \begin{bmatrix} \beta I_n & 0 \\ 0 & I_{n_s} \end{bmatrix} M.$$

As before, since this maximization problem is not convex we will in general only be able to find local maxima by finding matrices $Q \in \mathcal{Q}^s_{\mathcal{K}_s}$ and $D \in \mathcal{D}^s_{\mathcal{K}_s}$ non-zero vectors b, a, z, and w such that the following set of equations holds.

$$\begin{aligned} M_\beta b = \beta a \qquad & M^*_\beta z = \beta w \\ b = Qa \qquad & b = D^{-1} w \\ z = Q^* QDa \qquad & z = Q^* w. \end{aligned} \tag{5.3}$$

To compute a lower bound β for $\mu^s_{\mathcal{K}_s}$ we modified the standard power algorithm from [48] for computing a lower bound for complex μ. Without loss of generality we will explicitly write the formulae only for the simple block structure with $m_c = m_C = 1$. The formulae for an arbitrary block structure are obtained simply by repeating the formulae for each block type appropriately. Except for equations (5.6) and (5.8), the blocks are updated independently. Given $\mathcal{K} = (k_1, k_2)$ the appropriate scaling set becomes

$$\mathcal{Q}_\mathcal{K} = \{\text{blockdiag}(q^c I_{k_2}, Q^C) : q^{c*} q^c = 1, Q^{C*} Q^C = I_{k_3}\}. \tag{5.4}$$

Partition the four vectors b, a, z, and $w \in \mathbf{C}^n$ compatibly with this block structure as

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \quad w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{5.5}$$

where b_i, a_i, z_i, and $w_i \in \mathbf{C}^{k_i}$. Allow these vectors to evolve via the following power iteration:

$$\tilde{\beta}_{k+1} a_{k+1} = M_\beta b_k \ : \ \tilde{\beta}_{k+1} \in \mathbf{R}^+, \ |a_{k+1}| = 1 \tag{5.6}$$

$$z_{1_{k+1}} = \frac{w_{1_k}^* a_{1_{k+1}}}{|w_{1_k}^* a_{1_{k+1}}|} w_{1_k} \tag{5.7}$$

$$z_{2_{k+1}} = \frac{|w_{2_k}|}{|a_{2_{k+1}}|} a_{2_{k+1}}$$

$$\hat{\beta}_{k+1} w_{k+1} = M_\beta^* z_{k+1} \ : \ \hat{\beta}_{k+1} \in \mathbf{R}^+, \ |w_{k+1}| = 1 \tag{5.8}$$

$$b_{1_{k+1}} = \frac{a_{1_{k+1}}^* w_{1_{k+1}}}{|a_{1_{k+1}}^* w_{1_{k+1}}|} a_{1_{k+1}} \tag{5.9}$$

$$b_{2_{k+1}} = \frac{|a_{2_{k+1}}|}{|w_{2_{k+1}}|} w_{2_{k+1}}.$$

If the above iteration converges to an equilibrium point then we have a matrix $Q \in \mathcal{Q}_\mathcal{K}$ such that $QM_\beta b = \tilde{\beta} b$ and $w^* Q M_\beta = \hat{\beta} w^*$, so that $\max(\tilde{\beta}, \hat{\beta})$ gives us a lower bound for $\mu_{\mathcal{K}_s}^s(M)$. Furthermore if $\tilde{\beta} = \hat{\beta}$ then this bound corresponds to a local maximum of (5.2). In a significant number of cases the iteration does converge quite quickly, although it is somewhat less efficient compared to the SPA for computing the lower bound of complex μ.

Upper Bound

To obtain a tractable upper bound for $\mu_{\mathcal{K}_s}^s$ we introduce the following set of positive definite block diagonal scaling matrices, which commute with the elements in $\mathbf{\Delta}_\mathcal{K}^s$.

$$\mathcal{D}_{\mathcal{K}_s}^s \doteq \{D_s = \text{blockdiag}(D_f, D) : D_f \in \mathcal{D}_\mathcal{K}, D = D^* \in \mathbf{C}^{n_s \times n_s}\}.$$

Once again we use the familiar complex μ upper bound technique to obtain the following upper bound for the complex skewed-μ.

$$\begin{aligned} \mu_{\mathcal{K}_s}^s &\leq \inf_{D \in \mathcal{D}_{\mathcal{K}_s}^s} \min_{0 \leq \beta \in \mathbf{R}} \{\beta : M_\beta^* D M_\beta - \beta^2 D \leq 0\} \\ &= \inf_{D \in \mathcal{D}_{\mathcal{K}_s}^s} \min_{0 \leq \beta \in \mathbf{R}} \{\beta : M^* D M - \begin{bmatrix} I_n & 0 \\ 0 & \beta^2 I_{n_s} \end{bmatrix} D \leq 0\}. \end{aligned}$$

Computationally this reduces to a Linear Matrix Inequality. The interior point methods described in [2] can be used to efficiently compute this upper bound.

5.2 Robust $\mathcal{H}_2$ as a Complex Skewed μ Problem.

A standard setup for robustness analysis consists of a nominal LTI map M and a structured LTI perturbation Δ_f which enters the system in a feedback fashion. In this case the closed loop system $(\Delta_f * M)$, where $*$ denotes the Redheffer star-product, is LTI also. The $\mathcal{H}_2$ norm of an LTI system $H(j\omega)$ is defined by

$$\|H\|_2 \doteq \left(\int_{-\infty}^{\infty} \text{trace}(H(j\omega)^* H(j\omega))\frac{dw}{2\pi}\right)^{\frac{1}{2}}.$$

The system will be said to have robust $\mathcal{H}_2$ performance if it is robustly stable and

$$\sup_{\|(\Delta_f)\|_\infty \leq 1} \|(M * \Delta_f)\|_2 < 1$$

This appears to be a desirable design specification from both the performance and uncertainty points of view, especially when we are interested in rejection of white signals in the worst case.

In this chapter we will simply analyze the worst-case $\mathcal{H}_2$ cost over a set of perturbations for a special class of LTI plants for which $(M * \Delta_f)$ is SIMO or MISO system. The main result of this chapter is given in the following theorem. To simplify the notation we assume that $(M * \Delta_f)$ has n inputs and 1 output.

Theorem 9 *Consider the uncertain linear time invariant system of Figure 5.1, where Δ_f is LTI, for each ω, $\Delta_f(j\omega)$ is in $X_\mathcal{K}$, and $\|\Delta_f\|_\infty \leq 1$. Assume that the system $M * \Delta_f$ is either SIMO or MISO and well posed for all Δ_f. Consider also the uncertainty structure*

$$\mathbf{\Delta}^s_{\mathcal{K}_s} = \left\{\text{blockdiag}(\Delta_f, \Delta), \Delta_f \in X_\mathcal{K}, \bar{\sigma}(\Delta_f) \leq 1, \Delta \in \mathbf{C}^{n\times 1}\right\}.$$

Then for any positive ϵ there exists a finite sequence of frequency points $\{\omega_i\}_{i=1}^N$, such that

$$\sum_{i=1}^{N} \mu^s_{\mathcal{K}_s}(M(j\omega_i))^2 \Delta\omega_i - \epsilon \leq \sup_{\|(\Delta_f)\|_\infty \leq 1} \|M * \Delta_f\|_2^2 \leq \sum_{i=1}^{N} \mu^s_{\mathcal{K}_s}(M(j\omega_i))^2 \Delta\omega_i + \epsilon. \tag{5.10}$$

Proof

In this case $M * \Delta_f$ is a row vector. Since for any vector u,

$$\text{trace}(u^* u) = \text{trace}(uu^*) = \bar{\sigma}(u)^2$$

we will have,

$$\text{trace}((M * \Delta_f)(j\omega)(M * \Delta_f)^*(j\omega)) = \overline{\sigma}^2((M * \Delta_f)(j\omega)).$$

We can now write the worst case gain of the system as

$$\sup_{\|(\Delta_f)\|_\infty \leq 1} \|M * \Delta_f\|_2 = \sup_{\|(\Delta_f)\|_\infty \leq 1} \left(\int_{-\infty}^{\infty} \overline{\sigma}^2 \left((M * \Delta_f)(j\omega) \right) \frac{d\omega}{2\pi} \right)^{\frac{1}{2}}.$$

It can be shown that, if we consider Δ's both causal **and noncausal** we can interchange the integral and supremum:

$$\sup_{\|(\Delta_f)\|_\infty \leq 1} \|M * \Delta_f\|_2 = \left(\int_{-\infty}^{\infty} \sup_{\Delta_f} \overline{\sigma}^2 \left((M * \Delta_f)(j\omega) \right) \frac{d\omega}{2\pi} \right)^{\frac{1}{2}}.$$

Continuity arguments, together with the existence of a bandwidth beyond which the gain of the systems decreases rapidly, can be used to prove that for any positive ϵ a sequence of frequency points $\{\omega_i\}_{i=0}^N$ exists such that

$$\int_{-\infty}^{\infty} \sup_{\Delta_f} \overline{\sigma}^2 \left((M * \Delta_f)(j\omega) \right) \frac{d\omega}{2\pi} \leq \sum_{i=1}^{N} \sup_{\Delta_f} \overline{\sigma}^2 \left((M * \Delta_f)(j\omega_i) \right) \Delta\omega_i + \epsilon \quad (5.11)$$

$$\sum_{i=1}^{N} \sup_{\Delta_f} \overline{\sigma}^2 \left((M * \Delta_f)(j\omega_i) \right) \Delta\omega_i - \epsilon \leq \int_{-\infty}^{\infty} \sup_{\Delta_f} \overline{\sigma}^2 \left((M * \Delta_f)(j\omega) \right) \frac{d\omega}{2\pi}. \quad (5.12)$$

Since Δ is a full complex block, the definition of the complex skewed structured singular value and the small gain theorem imply

$$\sup_{\Delta_f} \overline{\sigma}^2 \left((M * \Delta_f)(j\omega_i) \right) = \mu_{\mathcal{K}_s}^s (M(j\omega_i))^2.$$

Substituting in Equations (5.11) and (5.12)

$$\int_{-\infty}^{\infty} \sup_{\Delta_f} \overline{\sigma}^2 \left((M * \Delta_f)(j\omega) \right) \frac{d\omega}{2\pi} \leq \sum_{i=1}^{N} \mu_{\mathcal{K}_s}^s (M(j\omega_i))^2 \Delta\omega_i + \epsilon \quad (5.13)$$

$$\sum_{i=1}^{N} \mu_{\mathcal{K}_s}^s(M(j\omega_i))^2 \Delta\omega_i - \epsilon \leq \int_{-\infty}^{\infty} \sup_{\Delta_f} \overline{\sigma}^2 \left((M * \Delta_f)(j\omega)\right) \frac{d\omega}{2\pi} \tag{5.14}$$

and thus

$$\sum_{i=1}^{N} \mu_{\mathcal{K}_s}^s(M(j\omega_i))^2 \Delta\omega_i - \epsilon \leq \sup_{\|(\Delta_f)\|_\infty \leq 1} \|M * \Delta_f\|_2^2 \leq \sum_{i=1}^{N} \mu_{\mathcal{K}_s}^s(M(j\omega_i))^2 \Delta\omega_i + \epsilon. \tag{5.15}$$

■

Remark: Although the proof of this theorem does not construct the sequence $\{\omega_i\}_{i=1}^N$, engineering knowledge of the system can be used to select these points.

5.3 Worst-case $\mathcal{H}_2$ Performance Computation Results

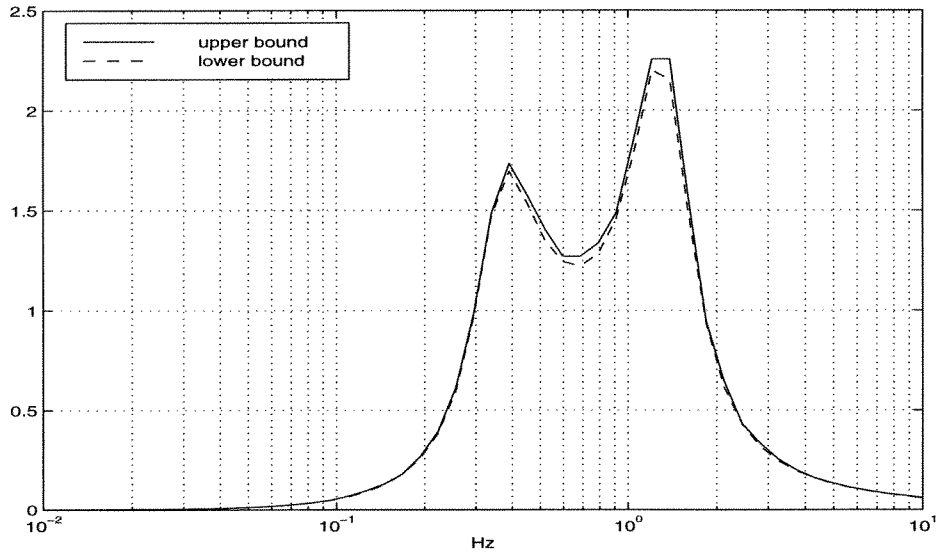

Figure 5.2: $\mu_{\mathcal{K}_s}^s$ upper and lower bound for the second output $\mathcal{H}_2$ gain of a landing airplane.

To be able to determine the worst case $\mathcal{H}_2$ gain of an airplane in the last phase of landing we computed $\mu_{\mathcal{K}_s}^s$ upper and lower bounds for its model over

the $[0.01, 10]Hz$ frequency range. The model used has been introduced in Section 4.5, and has 13 states, 5 inputs and 5 outputs. The fixed size of uncertainty Δ_f consists of three complex scalars, and we analyzed the $\mathcal{H}_2$ norm of the system for one of the outputs.

Figure 5.2 shows the results of computation. The gap between the $\mu^s_{\mathcal{K}_s}$ upper and the lower bound is very small, meaning that our computation is very accurate.

5.4 Conclusion

We showed in this chapter that the worst case $\mathcal{H}_2$ norm of an uncertain system subject to norm bounded structured LTI perturbations can be written exactly in terms of the complex skewed structured singular value. Although, like μ, the complex skewed structured singular value can not be computed exactly, we discussed efficient algorithms to compute corresponding upper and lower bounds.

Even though computation of these bounds implies numerical integration over frequency, we do not believe this to be a restriction in any practical sense. Since the computational effort required to compute the bounds on the worst case $\mathcal{H}_2$ performance is the same as in a frequency by frequency evaluation of the structured singular value (the current practice for robustness analysis in industry), this new approach can be effortlessly integrated into the current robustness analysis tools.

The solution presented here is more accurate than the one presented in [37] on two accounts. First it uses the true definition of the worst case $\mathcal{H}_2$ norm, instead of the more conservative definition used there. Second under this formulation both upper and lower bounds can be computed. Comparisons with other methods, such as the one presented in [15] need to be carried out, to establish the relative merits of the different methods. Comparisons between [37] and [15] can be found in [38], where it is shown that examples can be constructed to favor each of the methods with respect to the other.

Still to be resolved is the issue of causality of the perturbations, and how much conservativeness is introduced by not imposing it. It is also necessary to investigate under which circumstances the perturbation that achieves the worst case norm, as presented in this chapter, is causal.

Chapter 6

Nonlinear H_∞ Robustness Analysis

Robust controllers for linear systems generated with existing analysis and synthesis computational tools provide guaranteed performance in the presence of structured uncertainty. These tools can also find the worst case disturbances for a given controller.

For linear time invariant (LTI) systems with structured uncertainty, analysis of robust performance can be reduced to searching for the solution of a set of algebraic equations which give bounds on the achievable performance. One is thus able to find computationally efficient solutions, such as the power algorithm for the μ lower bound, without doing an explicit parameter search involving repeated simulation.

Performance analysis for nonlinear systems is difficult due to the wide variety of behavior and structures which can occur, and the most of existing tools are at a theoretical level. Our experience with performance analysis for linear systems suggests that specific algorithms can be designed that significantly outperform the off-the-shelf ones in the sense that they give better answers with less computational effort [32]. Recent work has shown that this approach can be successfully extended to the study of different robustness problems for nonlinear systems (see [41, 43] for a complete discussion.)

In this chapter we will consider the problem of disturbance rejection. General purpose nonlinear programming algorithms can be used to solve this problem, but we adopt the methodology of [41], and develop the same type of a power algorithm, for the performance index different from one used in [41]. This algorithm can be tested via the converse HJB method, which can gen-

erate the worst-case disturbance for the optimal controller. For the first time results obtained by this type of algorithm are compared with exact solutions given by different synthesis methods.

For nonoptimal controllers there may be disturbances that are even worse than ones obtained by the converse HJB method. The analysis method in this chapter is useful in generating worst-case disturbances for analyzing non optimal synthesis techniques. The results in this chapter are joint work with Jorge Tierno and have been reported in [18].

6.1 The Disturbance Rejection Problem

Let u be the noise signal perturbing the system, and let y be the output signal, that is the difference between the nominal and the actual trajectories. The equations describing the system will then be

$$\begin{aligned} x_{k+1} &= f(x_k, u_k, k) \\ y_k &= g(x_k, u_k, k) \qquad 0 \leq k \leq N-1 \end{aligned} \tag{6.1}$$

with given initial conditions x_0.

We wish to find

$$\begin{aligned} \max_u \quad & J(u, x_0) \\ J(u, x_0) &= (\|y\|^2 - \gamma^2 \|u\|^2), \end{aligned} \tag{6.2}$$

where for a sequence $y = (y_0, y_1, \ldots)$ such that $\sum_{i=1}^{\infty} |y_i|^2 < \infty$, the associated norm is

$$\|y\| = \left(\sum_{i=1}^{\infty} |y_i|^2 \right)^{\frac{1}{2}}. \tag{6.3}$$

The preceding problem is a nonlinear constrained optimization problem. However, it is in general non-convex and an exact solution is thus out of the question: we have to settle for upper and lower bounds. In this chapter we will show the existence of a power-type algorithm to compute a lower bound for the performance index (6.2), based on the search for locally worst case signals.

6.2 Necessary Conditions for Worst Case Signals

Any evaluation of the function $J(u, x_0)$, for given initial conditions and valid values of the disturbance signal is a lower bound on (6.2). So a simple way of getting lower bounds is through repeated simulation of the system for different values of the uncertain signals in the model. This is at present the state of the art of nonlinear analysis as applied in industry: good simulation models are developed and designs are tested through extensive simulation, usually selecting the uncertain signals at random. This approach is practical, since it requires information from the plant that is usually available, and often gives reasonable results. The algorithm we present in this chapter will improve on this approach without sacrificing in simplicity or in the generality of the information required. Instead of simulating at random points, we would look for points that are good candidates for being local maximums. We will do this search through a "power-like" algorithm. In order to develop this algorithm we first have to establish the necessary conditions signals have to meet in order to be the worst case.

Theorem 10 [4] *For a dynamical system described by the equations:*

$$x_{k+1} = f(x_k, u_k) \qquad 0 \leq k \leq N-1 \tag{6.4}$$

with x_0 given and performance index

$$J = \sum_{k=0}^{N-1} L(x_k, u_k), \tag{6.5}$$

if the signal sequence $u_0, \cdots, u_{N-1}$ achieves an extremum of J, then there exists a solution to the two point boundary value problem:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) \\ \lambda_k &= \left(\frac{\partial f}{\partial x_k}\right)^T \lambda_{k+1} + \left(\frac{\partial L}{\partial x_k}\right)^T \\ 0 &= \left(\frac{\partial f}{\partial u_k}\right)^T \lambda_{k+1} + \left(\frac{\partial L}{\partial u_k}\right)^T \\ 0 &\leq k \leq N-1 \end{aligned} \tag{6.6}$$

with boundary conditions:

$$\begin{aligned} x_0 &\quad \text{given} \\ \lambda_N &= 0. \end{aligned} \tag{6.7}$$

The disturbance rejection problem can be written in the form of Theorem 10. First we consider the performance condition. Define

$$L(x_k, u_k) = \frac{1}{2}(y_k^* y_k - \gamma^2 u_k^* u_k). \tag{6.8}$$

Then optimizing $\|y\|^2 - \gamma^2\|u\|^2$ is equivalent to optimizing the performance index

$$J = \sum_{k=0}^{N-1} L(x_k, u_k) =: \frac{1}{2}(\|y\|^2 - \gamma^2\|u\|^2), \tag{6.9}$$

for the system satisfying the difference equation

$$x_{k+1} = f(x_k, u_k) \tag{6.10}$$

where

$$y_k = g(x_k, u_k)$$

with given initial conditions x_0.

This problem is in the form of equation (10). So a sequence of signals achieves the worst case value of the performance index J only if there exists a sequence $\lambda_o, \cdots, \lambda_{N-1}$, satisfying

$$\lambda_k = \left(\frac{\partial f}{\partial x_k}\right)^T \lambda_{k+1} + \left(\frac{\partial g}{\partial x_k}\right)^T y_k \tag{6.11}$$

with final state conditions

$$\lambda_N = 0 \tag{6.12}$$

and satisfying the following alignment conditions

$$u_k = \frac{1}{\gamma^2}\left(\left(\frac{\partial f}{\partial u_k}\right)^T \lambda_{k+1} + \left(\frac{\partial g}{\partial u_k}\right)^T y_k\right). \tag{6.13}$$

Remarks: Equation (6.11) describes a linear time varying dynamical system whose inputs are the outputs of the original system. We will refer to this system as the adjoint or co-system.

Equations (6.13) can be interpreted as an alignment condition between the outputs of the adjoint system and the inputs to the original dynamical system. Thus, these equations describe two dynamical systems interconnected in a feedback loop.

If we consider both the equations for the system, co-system, and the alignment conditions together, we have a two point boundary value problem, i.e. a set of difference equations with boundary conditions at two distinct instants in time.

Several methods for solving the general two point boundary value problem have been devised and thoroughly studied. (See for example [25], [1].) However, the standard methods are based on gradient descent. In what follows we present a method to solve this particular instance of the two point boundary value problem that avoids the problems of gradient descent methods. The algorithm is a generalization of the power algorithm for the lower bound of μ.

6.3 A Power Algorithm

For a trajectory that meets the necessary conditions for a critical point, the Euler Lagrange conditions can be naturally separated into (i) a dynamical system with initial conditions only; (ii) a dynamical system with final conditions only; (iii) aligning conditions between the inputs and outputs of the two systems.

So, if the perturbation signals achieve the necessary conditions, the following composition of mappings yields the identity map:

- Integrate the system equations with initial conditions x_o.
- Compute the co-system along the current trajectory. Integrate these equations backwards in time with final state condition 0.
- From the alignment conditions in (6.13) compute updated values for u.

Denote this composition by

$$u^1 = \Phi(u^o). \tag{6.14}$$

The following iterative algorithm searches for fixed points of Φ, by evaluating it repeatedly.

1. Simulate the system with some initial u^0.

2. **Repeat**

$$u^{i+1} := \Phi(u^i)$$

3. **until**

$$u^{i+1} = u^i$$

Remarks: If the algorithm converges, it converges to a fixed point of Φ and thus to a signal that meets the necessary conditions for a critical point.

In order to prove convergence we would have to prove that Φ is a contraction around fixed points. For structured robustness analysis problems, that has not been proved even for the simpler case when the system is linear (otherwise P=NP!). However, for the disturbance rejection problem considered in this chapter, we can easily prove global convergence in the linear case. For the nonlinear problem, it is unlikely that we will be able to go beyond local convergence.

6.4 The Linear Case

Since the aim of this work is to extend the analysis methodology for linear systems given by the structured singular value framework to nonlinear systems, it is useful to understand the above power algorithm when specialized to linear systems. In this case it turns out that the proposed algorithm reduces to the standard power algorithm for μ as described by Young and Doyle [48]. Because of the special nature of the disturbance rejection problem, we can prove global convergence for the algorithm.

By Theorem 10, for a discrete linear time invariant system, given by the equations:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \\ 0 &\leq k \leq N-1 \end{aligned} \tag{6.15}$$

with given initial conditions x_0 and performance index of the form

$$\begin{aligned} J &= \sum_{k=0}^{N-1} L(x_k, u_k) \\ &= \frac{1}{2}\sum_{k=0}^{N-1}(y_k^* y_k - \gamma^2 u_k^* u_k) \end{aligned} \tag{6.16}$$

if the signal sequence $u_0, \cdots, u_{N-1}$ achieves an extremum of J, then there exists a solution to the two point boundary value problem:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \\ \lambda_k &= A^T\lambda_{k+1} + C^T y_k \\ u_k &= \frac{1}{\gamma^2}\left(B^T\lambda_{k+1} + D^T y_k\right) \\ 0 &\leq k \leq N-1 \end{aligned} \tag{6.17}$$

with boundary conditions:

$$\begin{aligned} x_0 &\quad \text{given} \\ \lambda_N &= 0. \end{aligned} \tag{6.18}$$

We can write the above power iteration for this case explicitly. For simplicity, we will assume N=3, although the general case will be obvious from this. To compute Φ in (6.14), note that

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \end{bmatrix} x_0 + \begin{bmatrix} D & 0 & 0 \\ CB & D & 0 \\ CAB & CB & D \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \end{bmatrix}$$

which we will write as

$$y = b + Mu. \tag{6.19}$$

Our optimization problem is equivalent to

$$\max_u J(u, x_0)$$

subject to (6.19) with

$$J(u, x_0) = (\|y\|^2 - \gamma^2\|u\|^2).$$

At this point we can easily determine directly from (6.19) that the global optimum must satisfy

$$u = \frac{1}{\gamma^2} M^*(Mu + b) \tag{6.20}$$

if it exists (if $\gamma > \bar{\sigma}(M)$). If $\gamma < \bar{\sigma}(M)$ then the supremum is $J = \infty$.

For the power iteration the corresponding adjoint-system is also linear time invariant:

$$\begin{bmatrix} u_0 \\ u_1 \\ u_2 \end{bmatrix} = \frac{1}{\gamma^2} \begin{bmatrix} D^T & B^T C^T & B^T A^T C^T \\ 0 & D^T & B^T C^T \\ 0 & 0 & D^T \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix}$$

or

$$u = \frac{1}{\gamma^2} M^* y.$$

In this case the algorithm reduces to a standard power iteration alternating multiplication by M and M^*, or in other words

$$u^{(i+1)} = \frac{1}{\gamma^2} M^*(Mu^{(i)} + b). \tag{6.21}$$

If $\gamma > \bar{\sigma}(M)$ then this iteration itself is just a stable linear discrete time system which will exponentially converge to an equilibrium which is the globally optimum u satisfying (6.20).

The advantage of viewing this problem as a power iteration is that the alternating multiplication by M and M^* can be done by simulation without the need to form the matrices explicitly. It is one of the fastest general approaches to solving this problem, especially for large N.

6.5 Continuous Time

The discussion presented in the previous sections can be easily extended to continuous time. The equilibrium conditions will be derived from the continuous time version of Theorem 10. In this case, optimizing $\|y\|^2 - \gamma^2\|u\|^2$ is equivalent to optimizing the performance index

$$J = \int_{t_i}^{t_f} L(x, u)dt =: \frac{1}{2}(\|y\|^2 - \gamma^2\|u\|^2), \tag{6.22}$$

for the system satisfying the differential equation

$$\dot{x} = f(x, u) \tag{6.23}$$

where

$$y = g(x, u)$$

with given initial conditions x_0.

A signal u can achieve the worst case value of the performance index J only if there exists a signal λ, satisfying

$$-\dot{\lambda} = \left(\frac{\partial f}{\partial x}\right)^T \lambda + \left(\frac{\partial g}{\partial x}\right)^T y \tag{6.24}$$

with final state conditions

$$\lambda(t_f) = 0 \tag{6.25}$$

and satisfying the following alignment conditions

$$u = \frac{1}{\gamma^2}\left(\left(\frac{\partial f}{\partial u}\right)^T \lambda + \left(\frac{\partial g}{\partial u}\right)^T y\right). \tag{6.26}$$

It can readily be seen that the algorithm in Section 6.3 can be extended to the continuous time case. The main difficulty added by considering the system in continuous time is the numerical integration of the differential equations. Note that if these equations are integrated using Euler integration algorithm, the continuous time equations reduce to the discrete time ones for the corresponding discretized system.

6.6 Computational Results

Due to the nature of the results presented here, the success of the algorithm in the nonlinear case can only be determined by gathering experience from many different problems and making computational comparisons with other approaches. To evaluate the proposed power algorithm convergence properties when specialized to linear systems we performed analysis for the linear H_∞ disturbance rejection problem. Once convinced that the algorithm quickly converges to the correct answer in the linear case, we proceeded and tested it on nonlinear examples generated with the "converse Hamilton-Jacobi-Bellman" method (CoHJB) as suggested in [7].

Linear H_∞

The first problem considered in this section is the linear H_∞ disturbance rejection problem, where the worst case disturbance u that maximizes the transfer function $\|T_{uy}\|_\infty$ norm of a linear system is sought. From [9] we know that for a linear system

$$\begin{aligned} \dot{x} &= Ax + Bu \quad x(0) = x_0 \\ y &= Cx \end{aligned}$$

the worst case solution is

$$u_{opt} = \frac{1}{\gamma^2} B^T X$$

where X satisfies following Riccati equation

$$A^T X + XA + \frac{1}{\gamma^2} XBB^T X + C^T C = 0. \tag{6.27}$$

We tried to find the worst case disturbance, using our proposed power algorithm, for the system

$$\begin{aligned} \dot{x} &= \begin{bmatrix} -3 & -2 \\ 1 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u \qquad x_0 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ y &= \begin{bmatrix} 0 & 1 \end{bmatrix} x. \end{aligned}$$

Formulated in correspondence to our framework we want to find

$$\max_u J(u, x_0)$$

where the performance objective is

$$J =: \|y\|^2 - \gamma^2 \|u\|^2$$

and we have chosen $\gamma = 1$.

We compared the solution to this linear constrained optimization problem obtained by our power-type algorithm, that converged after 3 iterations, to the *a priori* known worst case solution u_{opt} obtained by solving the corresponding Riccati equation. The evolution of our algorithm solution compared to the H_∞ worst case solution is given in Figure 6.1. We denote a result of the ith iteration $u^{(i)}$.

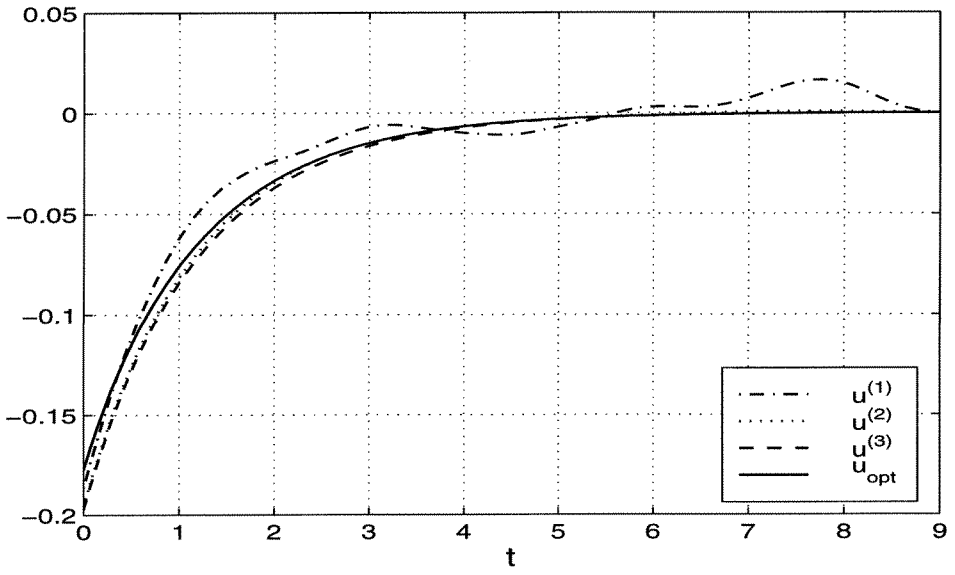

Figure 6.1: Comparison of a linear H_∞ disturbance rejection problem solution obtained by the power-type algorithm and the a priori known optimal solution u_{opt}.

Nonlinear Examples

To generate nonlinear test examples we used the CoHJB method. Starting with the cost and optimal value function V, CoHJB solves HJB PDEs "backwards" to produce nonlinear dynamics and optimal controllers and disturbances. It is computationally tractable and can generate essentially all possible nonlinear optimal control problems. Actual design methods, which must start with the cost and dynamics without knowledge of V can then be studied knowing the optimal control and disturbance. In the rest of this section we will briefly introduce the CoHJB method.

Consider the nonlinear system, with $f(0) = 0$

$$\dot{x}(t) = f(x(t)) + g(x(t))u(x) \tag{6.28}$$

and a performance objective

$$\sup_u J(u(x(t))) \tag{6.29}$$

$$J(u) = \int_0^\infty (q(x(t)) - u^T(t)u(t))dt \tag{6.30}$$

where the state x is available for the control u.

The term *converse* optimal control as introduced in [7] is concerned with the problem defined as follows:

Given a performance defined as (6.30) and a storage function $V : \mathbf{R^n} \rightarrow \mathbf{R}^+$, find a class of nonlinear systems such that the optimal control problem (6.29) has this as its solution.

The converse problem is characterized by the Hamilton-Jacobi-Bellman (HJB) equation

$$V_x f + \frac{1}{4} V_x g g^T V_x^T + q = 0, \tag{6.31}$$

and requires only solving the equation (6.31) as an algebraic equation in the unknowns f, g with V given. The converse problem helps to construct an array of examples which have known optimal solution. Note that almost any nonlinear optimal control problem of the type described above can be generated with this method.

2-D Oscillator

In the following, we will consider a 2-D oscillator system

$$\begin{aligned} \dot{x}_1 &= x_2 & x_1(0) &= 1 \\ \dot{x}_2 &= f(x) + g(x)u & x_2(0) &= 1 \end{aligned}$$

and suppose that $J = \|x_2\|^2 - \|u\|^2$ and $V = x_1^2 + x_2^2$. The HJB is

$$2x_1x_2 + 2x_2 f(x) + \frac{1}{4}(2x_2)^2 g^2(x) = 0.$$

Therefore

$$f(x) = -x_1 - \frac{1}{2}\ x_2(1 + g(x)^2)$$

and the optimal solution is $u_{opt} = g(x)x_2$.

We have chosen $g(x) = x_1$ and compared the solution to this nonlinear constrained optimization problem obtained by our power-type algorithm, that converged after 6 iterations, denoted $u^{(6)}$, and the a priori known optimal solution $u_{opt} = x_1x_2$. The result is given in Figure 6.2.

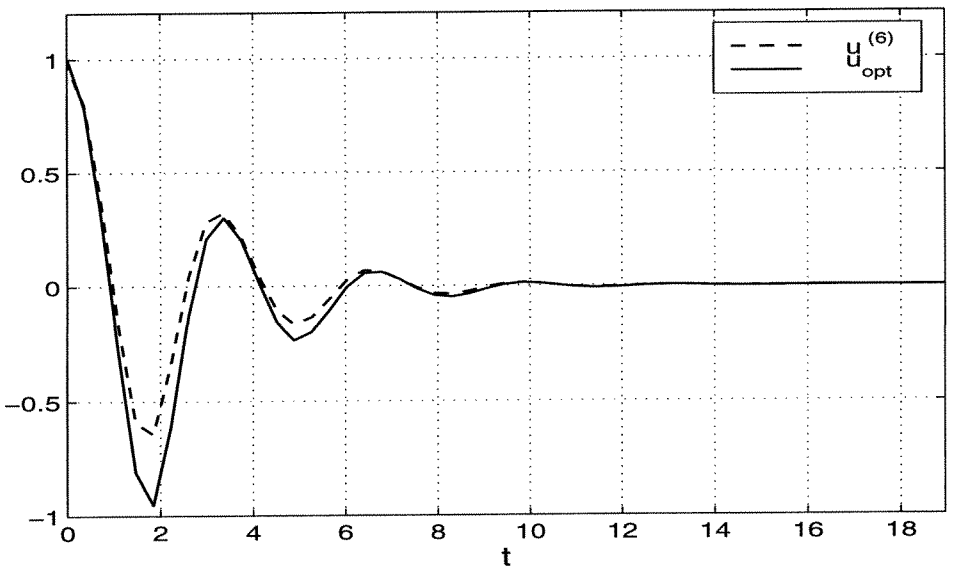

Figure 6.2: Comparison of the 2-D oscillator disturbance rejection problem solution obtained by the power-type algorithm and the *a priori* known optimal solution u_{opt}

Nonlinear H_∞

As a less trivial example we considered a nonlinear system

$$\begin{aligned} \dot{x}_1 &= 2x_2 + (2x_1 - 1)x_1 & x_1(0) &= 1 \\ \dot{x}_2 &= -2x_1 - x_2 - 2x_1^2 + u & x_2(0) &= 1 \\ y &= \begin{bmatrix} \sqrt{2}\ x_1 \\ x_2 \end{bmatrix} \end{aligned}$$

and suppose that $J = \|y\|^2 - \|u\|^2$ and $V = x_1^2 + x_2^2$. In this case the optimal solution is

$$u_{opt} = x_2.$$

We compared the solution to this nonlinear constrained optimization problem obtained by our power-type algorithm, that converged after 8 iterations, denoted $u^{(8)}$, and the a priori known optimal solution $u_{opt} = x_2$. The result is given in Figure 6.3.

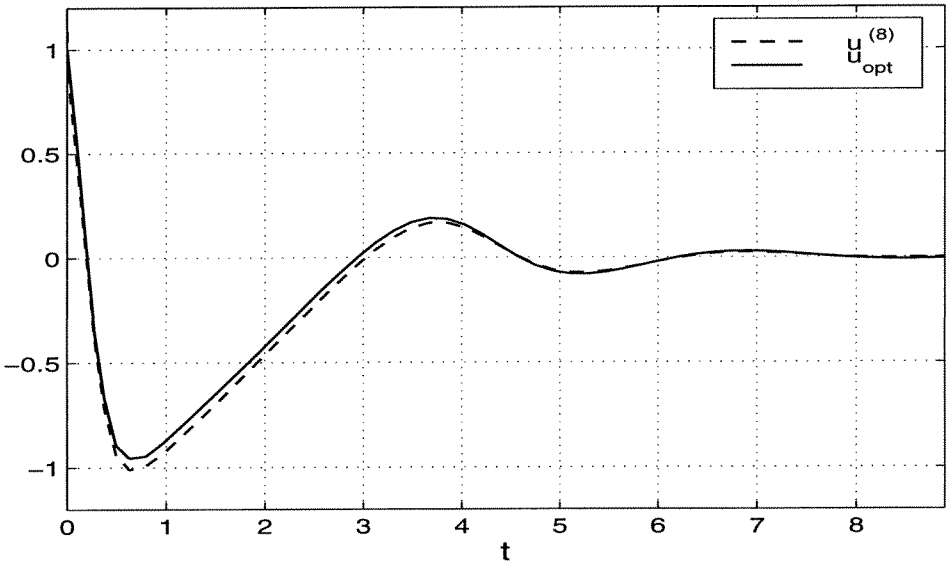

Figure 6.3: Comparison of the nonlinear H_∞ disturbance rejection problem solution obtained by the power-type algorithm and the *a priori* known optimal solution u_{opt}

6.7 Conclusion

In this chapter we showed how a power algorithm can be used to compute a necessary condition for disturbance rejection of discrete and continuous time nonlinear systems by searching for solutions to Euler-Lagrange equations. In the case where the system is linear we showed how the algorithm reduces to a well studied algorithm for the lower bound of μ and that the algorithm is guaranteed to converge to the global optimum. It is important to note that the worst case disturbances obtained by our proposed power algorithm are very close to the worst case disturbances a priori given by other methods, meaning that for the general case of a system with a non-optimal controller this algorithm can provide us with knowledge of the worst case disturbance.

A deeper investigation of the numerical properties of our algorithm is needed and future research will concentrate on this.

Part II

Nonlinear System Model Reduction

Chapter 7

Overview of Nonlinear Systems Model Reduction

Rotating stall in jet engine compressor systems is an instability causing a sudden drop in performance. Feedback control is necessary to avoid the development of rotating stall and the most preferred approach to control design is to use low order models. The three state nonlinear model of Moore and Greitzer (MG3), a Galerkin truncation onto the first Fourier mode of the full Moore-Greitzer model developed in [11], is one of the simple models that adequately describe the basic dynamics of rotating stall.

The classical approach to model reduction of nonlinear systems using the Galerkin method and the Karhunen-Loeve Expansion (KLE) attempts to find an approximate solution of a PDE in the form of a truncated series expansion given by

$$\hat{u}(x,t) = \sum_{n=1}^{N} a_n(t)\varphi_n(x),$$

where the mode functions $\varphi_n(x)$ are based on empirical data and are generated by the standard KLE methods. For systems with rotational (periodic) symmetry, the mode functions are Fourier modes and the order of the reduced model determined by reasonable criteria for the truncation point is not small. For a PDE with a traveling wave as a solution, normally this approach will not give satisfactory results.

The disadvantage of making a Galerkin projection onto a non-propagating function with fixed spatial shape is that it does not properly describe how the stall cell propagates and evolves in simulations. One usually observes that the stall cell quickly develops a square like spatial structure. There are some recent results [29] in modeling a deep stall cell phenomena leading to the conclusion

that stall cell is a rotating square wave. To capture this behavior with non-propagating modes of fixed spatial shape, one needs to include many modes. A remedy for this is to try to capture the dynamics with a family of propagating curves.

In Chapter 9, we propose a new computationally efficient modeling method that captures existing translation symmetry in a compression system (and more generally for systems with a rotational symmetry) by finding an approximate solution of the governing PDE in the form of a truncated series expansion given by

$$\hat{u}(x,t) = \sum_{n=1}^{N} a_n(t)\varphi_n(x + d(t)).$$

To generate an optimal set of basis functions $\varphi_n(x)$, prior to performing KLE we process the available data set using a "centering" procedure which involves giving an appropriate definition of the center of a wave and moving it to a standard position. The eigenvalues of the covariance matrix of the "centered" data decay rapidly and we obtain a low order approximate system of ODEs. This approach has been shown to be efficient in linear and nonlinear scalar wave equations. The method may be viewed as a way of implementing the KLE on the space of solutions of the given PDE modulo a given symmetry group. Viewed this way, the methodology is quite general and therefore should be useful in a variety of problems.

7.1 Galerkin Projection

The Galerkin method is a discretization scheme for PDEs based on the separation of variables approach which attempts to find an approximate solution in the form of a truncated series expansion given by

$$\hat{u}(x,t) = \sum_{n=1}^{N} a_n(t)\varphi_n(x) + \bar{u}(x), \tag{7.1}$$

where the $\varphi_n(x)$ are known as trial functions and

$$\bar{u}(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T u(x,t)dt.$$

In this way the original infinite dimensional system is approximated by an N dimensional system. We assume that u belongs to a Hilbert space $L_2([0, 2\pi])$, the linear, infinite-dimensional, inner product space of square integrable (complex-valued) functions with inner product

$$\langle f, g \rangle = \int_0^{2\pi} f(x) g^*(x) dx. \tag{7.2}$$

Suppose we have a system governed by the PDE

$$\frac{\partial u}{\partial t} = D(u) \quad u : [0, 2\pi] \times (0, \infty) \to \mathbf{R}$$

with appropriate boundary conditions and initial conditions, where $D(\cdot)$ is a nonlinear operator that may involve spatial derivatives and/or integrals. To be sure that the original PDE is satisfied as closely as possible by (7.1) we choose time dependent coefficients $a_n(t)$ in such a way that they minimize, with respect to a suitable norm, the residual error produced by using (7.1) instead of the exact solution. When the set of trial functions $\{\varphi_n\}$ is orthonormal it is equivalent to forcing the residual

$$r(x, t) = \frac{\partial \hat{u}}{\partial t} - D(\hat{u}) \tag{7.3}$$

to be orthogonal to a chosen number of trial functions, i.e.

$$\langle r(x, t), \varphi_n(x) \rangle = 0 \quad n = 1, \ldots, N.$$

Substituting (7.1) into (7.3) yields,

$$r(x, t) = \sum_{n=1}^{N} \dot{a}_n(t) \varphi_n(x) - D\left(\sum_{n=1}^{N} a_n(t) \varphi_n(x) + \bar{u}(x) \right).$$

Applying the orthogonality condition and using the orthonormality property of the set of trial functions results in a reduced order model which is a system of N ordinary differential equations

$$\dot{a}_i(t) = \langle D\left(\sum_{n=1}^{N} a_n(t) \varphi_n(x) + \bar{u}(x) \right), \varphi_i(x) \rangle \quad i = 1, \ldots, N. \tag{7.4}$$

The initial conditions for the resulting system of ODEs are determined by a second application of the Galerkin approach. We force the residual of the

initial conditions $r_0(x) = u(x,0) - \hat{u}(x,0)$ to also be orthogonal to the first N basis functions and we obtain a system of N linear equations

$$a_i(0) = \langle u(x,0) - \bar{u}(x), \varphi_i(x) \rangle \quad i = 1, \ldots, N. \tag{7.5}$$

Note that to be able to solve the system of ODEs (7.4) and (7.5) one only needs to select the set of trial functions $\{\varphi_n\}$ and the initial conditions of the original system $u(x,0)$. Any complete set of trial functions will suffice, but we focus on those generated by the KLE.

7.2 Karhunen-Loeve Expansion

The Karhunen-Loeve Expansion (KLE) is a well known procedure for extracting a basis for a modal decomposition from an ensemble of signals, such as data measured in the course of an experiment. Its mathematical properties, especially optimality suggest that it is the preferred basis to use in many applications. Karhunen-Loeve expansion provides the most efficient way of capturing the dominant components of an infinite-dimensional process with surprisingly few modes.

In other disciplines the same procedure is known as: proper orthogonal decomposition, principal component analysis, and singular value decomposition, and the basis functions obtained are called: empirical eigenfunctions, empirical basis functions, and empirical orthogonal functions. The KLE was introduced in the context of turbulence by Lumley [28] in the late sixties to analyze experimental data. The aim was to extract dominant features and trends, which are typically patterns in space and time.

The fundamental idea behind KLE is very pragmatic. Suppose we have an ensemble $\{u^{(k)}\}$ of scalar fields, each being a function $u^{(k)} = u^{(k)}(x)$ defined on the domain $0 \leq x \leq 2\pi$. To find a good representation of members of $\{u^{(k)}\}$, we will need to project each $u^{(k)}$ onto candidate basis functions, so we assume that the $u's$ belong to a Hilbert space $L_2([0, 2\pi])$.

We want to find a basis $\{\varphi_n\}$ for L_2 that is optimal for the given data set in the sense that the finite-dimensional representation of the form

$$\hat{u}(x,t) = \sum_{n=1}^{N} a_n(t)\varphi_n(x) + \bar{u}(x) \tag{7.6}$$

describes typical members of the ensemble better than representations of the same dimension in *any* other basis. To simplify notation we introduce the variation of $u(x,t)$ from the mean $\bar{u}(x)$ and denote it as $v(x,t)$. The notion of typical implies the use of time average over an ensemble $\{u^{(k)}\}$ and optimality is equivalent to maximizing the averaged normalized projection of $v(x,t)$ onto $\{\varphi_n\}$.

In general, the existence of the expansion (7.6) is guaranteed under certain conditions by the Karhunen-Loeve expansion theorem, that also provides us with a method for constructing the orthonormal set of functions $\{\varphi_n\}$ and the uncorrelated set of coefficients $\{a_n\}$.

In order to proceed, we need to present some definitions and results from the theory of second order stochastic processes. Concepts, like expectation (ensemble average) will retain their usual meaning, i.e., $E[X] = \int_\Omega X(\omega)P(d\omega)$, and can be computed with some further assumptions.

Definition 5 *A* **second order random variable** X *is one which satisfies* $E[|X|^2] < \infty$. *A second order stochastic process* $\{X_t\}$ *is a one parameter family of second order random variables.*

We can define a two parameter second order stochastic process $\{X_{t,x}\}$ to be a two parameter family of second order random variables. From now on we will assume that both $u(x,t)$ and $v(x,t)$ are second order processes. This assumption is necessary for the mean, correlation function, and covariance function to be defined. Because we are interested in measuring the spatial variation only, we would like the spatial correlation function to be time independent. For that reason, we assume that $u(x,t)$ and $v(x,t)$ are time stationary. With all these assumptions we can write the spatial correlation function as

$$E[v(x,t)v(y,t)] = \mathcal{R}_v(x,y) \tag{7.7}$$

a function independent of time. We have to add the assumption of continuity for second order processes.

Definition 6 *A second order process* $\{X_t\}$ *is* **continuous in quadratic mean** *(q. m. continuous) if for every* $t \in T$ $E[\|X_{t+h} - X_t\|^2] \to 0$ *as* $h \to 0$.

We will assume that $u(x,t)$ and $v(x,t)$ are q. m. continuous in their spatial argument, and therefore the two-point spatial correlation function $\mathcal{R}_v(x,y)$ is continuous in its arguments.

Finally, we can state the Karhunen-Loeve expansion theorem which allows one to express the continuum of random variables by a countable number of orthonormal random variables as presented in [45].

Theorem 11 *Karhunen-Loeve Theorem ([45])*
Let $\{X_t, t \in [a,b]\}$ *be a q. m. continuous second order process with covariance function* $R(t,s)$.

If $\{\varphi_n\}$ *are the orthonormal eigenfunctions of the integral operator with kernel* $R(\cdot,\cdot)$, *and* $\{\lambda_n\}$ *the corresponding eigenvalues,*

$$\int_a^b R(t,s)\varphi_n(s)ds = \lambda_n\varphi_n(t) \quad t \in [a,b] \tag{7.8}$$

then

$$X(t,s) = \lim_{N\to\infty} \sum_{n=1}^{N} \sqrt{\lambda_n} a_n(s)\varphi_n(t) \quad uniformly\ for \quad t \in [a,b] \tag{7.9}$$

where the limit is taken in the q.m. sense and the $\{a_n\}$ *satisfy*

$$a_n(s) = (\sqrt{\lambda_n})^{-1} \int_a^b \varphi_n(t)X(s,t)dt$$

and

$$E[a_m a_n] = \sigma_{mn}.$$

Conversely, if $X(t,s)$ *has an expansion of the form 7.9 with*

$$\int_a^b \varphi_m(x)\varphi_n(x)dx = \sigma_{nm}$$

and

$$E[a_m a_n] = \sigma_{mn}$$

then $\{\varphi_n\}$ *and* $\{\lambda_n\}$ *must be eigenfunctions and eigenvalues respectively of the integral operator with kernel* $R(\cdot,\cdot)$, *i.e., satisfy equation (7.8).*

In order to apply this result to our process $\{v_{x,t}\}$, we invoke the previously stated assumption of time invariance and get the one parameter process $v_x(t)$

with two point spatial covariance function $R(x,y)$. Then we can expand $v(x,t)$ using the Karhunen-Loeve theorem,

$$v(x,t) = \sum_{n=1}^{\infty} \sqrt{\lambda_n} a_n(t) \varphi_n(x)$$

where the limit is in the q. m. sense and

$$a_n(t) = (\sqrt{\lambda_n})^{-1} \int_{\mathcal{D}} \varphi_n(x) v(x,t) dx$$

and

$$E[a_m(t) a_n(t)] = \delta_m^n.$$

The orthonormal basis functions $\{\varphi_n(x)\}$ are found via the integral equation

$$\int_{\mathcal{D}} R_v(x,y) \varphi_n(y) dy = \lambda_n \varphi_n(x), \quad x \in \mathcal{D}.$$

To gain better understanding of the KLE method we will consider the case when we want to find the best approximation to the ensemble members using a single function. We are actually trying to perform the following maximization procedure:

$$\max_{\varphi \in L_2([0,2\pi])} \frac{E(|\langle v(x,t), \varphi(x) \rangle|^2)}{\|\varphi(x)\|^2}$$

where $|\cdot|$ denotes the absolute value and $\|\cdot\|$ is the L_2 norm

$$\|f(x)\| = \langle f(x), f(x) \rangle^{1/2}$$

and $E(\cdot)$ stands for time averaging, i.e.,

$$E(f(t)) = \frac{1}{T} \lim_{T \to \infty} \int_0^T f(t) dt.$$

The functional corresponding to this constrained variational problem is

$$J[\varphi(x)] = E(|\langle v(x,t), \varphi(x) \rangle|^2) - \lambda(\|\varphi(x)\|^2 - 1) \tag{7.10}$$

and a necessary condition for an extrema is that the functional derivative vanishes for all variations $\varphi(x) + \delta\psi(x) \in L_2([0,2\pi]), \delta \in \mathbf{R}$ is

$$\frac{d}{d\delta} J[\varphi + \delta\psi]|_{\delta=0} = 0.$$

From (7.10) we have

$$\begin{aligned}
&\frac{d}{d\delta} J[\varphi + \delta\psi]|_{\delta=0} \\
&= \frac{d}{d\delta}[E(\langle v, \varphi + \delta\psi\rangle\langle\varphi + \delta\psi, v\rangle) - \lambda\langle\varphi + \delta\psi, \varphi + \delta\psi\rangle]|_{\delta=0} \\
&= 2Re[E(\langle v, \psi\rangle\langle\varphi, v\rangle) - \lambda\langle\varphi, \psi\rangle] \\
&= \int_0^{2\pi}\left[\int_0^{2\pi} E(v(x,t)v^*(x',t))\varphi(x')dx' - \lambda\varphi(x)\right]\psi^*(x)dx \\
&= 0
\end{aligned}$$

where we have used the commutativity of averaging over time and integration over x. Since $\psi(x)$ is an arbitrary variation, our condition reduces to

$$\begin{aligned}
\int_0^1 R(x,x')\varphi(x')dx' &= \lambda\varphi(x) \qquad (7.11)\\
R(x,x') &\doteq E(v(x,t)v^*(x',t)).
\end{aligned}$$

In the general case the optimal basis is given by the eigenfunctions $\{\varphi_n\}$ of the integral equation (7.8) whose kernel is the averaged autocorrelation function and in the rest of the chapter they will be called empirical eigenfunctions. If we define the mean energy projection as $E[|\langle u, \varphi_n\rangle|^2]$, then the eigenvalues $\{\lambda_n\}$ corresponding to the empirical eigenfunctions may be interpreted as the "the mean energy of the process $u(x,t)$ projected on the φ_n axis in function space."

It has been shown in [22] that almost every member of the original ensemble can be reconstructed as a linear combination of empirical eigenfunctions having strictly positive eigenvalues. And not only that, but every empirical eigenfunction can be expressed as a linear combination of observations, implying that any property of the ensemble members that is preserved under linear combinations is inherited by all functions spanned by the empirical basis functions.

Optimality

Suppose that we have a stationary random field $v(x,t)$ in $L_2([0,2\pi])$ and that $\{\varphi_n\}$ is the set of orthonormal empirical eigenfunctions obtained from

time averages of $v(x,t)$. Let

$$v(x,t) = \sum_{n=1}^{\infty} a_n(t)\varphi_n(x),$$

be the decomposition with respect to this basis. Assume that the eigenvalues $\{\lambda_n\}$ corresponding to $\{\varphi_n\}$ have been decreasingly ordered so that $\lambda_{i+1} > \lambda_i$ $(\forall i)$. It can be shown that if $\{\psi_n\}$ is some arbitrary set of orthonormal basis functions in which we expand $v(x,t)$ then for any value of N

$$\sum_{n=1}^{N} E[|\langle \varphi_n, v\rangle|^2] = \sum_{n=1}^{N} \lambda_n \geq \sum_{n=1}^{N} E[|\langle \psi_n, v\rangle|^2].$$

Therefore, for a given number of modes N the projection on the subspace used for modeling the flow will on average contain the most energy possible compared to all other linear decompositions.

Method of Snapshots

The main goal of this type of data analysis is to generate optimal basis functions for Galerkin representations of PDEs. We consider a linearly independent set of snapshot data samples $\{u^{(1)}, u^{(2)}, \ldots, u^{(M)}\}$ which is either the result of a performed physical experiment or generated as the numerical solution to a scalar nonlinear PDE. The averaged snapshot is computed as

$$\bar{u} = \frac{1}{M}\sum_{k=1}^{M} u^{(k)}$$

and the mean adjusted snapshots are given by

$$v^{(k)} = u^{(k)} - \bar{u}.$$

For computational purposes, we discretize the spatial domain, which usually leads to a very large spatial correlation matrix and determining the corresponding eigenvalue decomposition is extremely costly. Assuming that $u(x,t)$ is an ergodic process, meaning that time averages equal ensemble averages for each fixed value of x, we can represent the averaged spatial correlation function as

$$R(x,y) = \lim_{T\to\infty} \frac{1}{T}\int_0^T v(x,t)v(y,t)dt.$$

The empirical data set is limited to a finite number of snapshots (say M samples). Thus, we use the following approximation of the averaged correlation function

$$\hat{R}_M(x,y) = \frac{1}{M}\sum_{i=1}^{M} v^{(i)}(x)v^{(i)}(y). \tag{7.12}$$

We call $\hat{R}_M(x,y)$ the empirically determined spatial correlation function. Because computing the empirical basis functions is based on the integral operator with kernel $R(x,y)$, we examine the integral operator of the approximation $\hat{R}_M(x,y)$:

$$\begin{aligned}\int_0^{2\pi} \hat{R}_M(x,y)\varphi(y)dy &= \int_0^{2\pi} \frac{1}{M}\sum_{i=1}^{M} v^{(i)}(x)v^{(i)}(y)\varphi(y)dy \\ &= \frac{1}{M}\sum_{i=1}^{M} v^{(i)}(x)\int_0^{2\pi} v^{(i)}(y)\varphi(y)dy \\ &= \sum_{i=1}^{M} \alpha_i v^{(i)}(x).\end{aligned}$$

Thus, since the eigenfunctions of the integral operator with kernel $\hat{R}_M(x,y)$ must satisfy

$$\sum_{i=1}^{M} \alpha_i v^{(i)}(x) = \lambda\varphi(x)$$

the empirical eigenfunctions can be written as

$$\varphi(x) = \sum_{i=1}^{M} b_i v^{(i)}(x) \tag{7.13}$$

for some constants $b_1,\ldots,b_M$. One can conclude that the empirical eigenfunctions of the integral operator with kernel $\hat{R}_M(x,y)$ are linear combinations of the data snapshots $v^{(1)},\ldots,v^{(M)}$. This justifies our previous statement that any property of the ensemble members that is preserved under linear combination is inherited by all functions spanned by the empirical basis functions.

Substituting equations (7.12) and (7.13) into (7.11) yields

$$\begin{aligned}\int_0^{2\pi} \frac{1}{M}\sum_{i=1}^{M} v^{(i)}(x)v^{(i)}(y)\sum_{j=1}^{M} b_j v^{(j)}(y)dy &= \lambda\sum_{j=1}^{M} b_j v^{(j)}(x) \\ \frac{1}{M}\sum_{i=1}^{M} v^{(i)}(x)\sum_{j=1}^{M} b_j \int_0^{2\pi} v^{(i)}(y)v^{(j)}(y)dy &= \lambda\sum_{j=1}^{M} b_j v^{(j)}(x)\end{aligned}$$

which can be written as

$$(Cb)^T V = \lambda b^T V,$$

where

$$\begin{aligned} (C)_{ij} &= \frac{1}{M}\int_0^{2\pi} v^{(i)}(x)v^{(j)}(x)dx \quad i,j = 1,\ldots,M \\ b &= [b_1,\ldots,b_M]^T \\ V &= [v^{(1)}(x),\ldots,v^{(M)}(x)]^T. \end{aligned} \tag{7.14}$$

If we assume that data snapshots are linearly independent, and therefore $(Cb)^T V = \lambda b^T V$ if and only if

$$Cb = \lambda b. \tag{7.15}$$

Thus, once we find the eigenvectors f^n and corresponding eigenvalues of the $M \times M$ matrix C, the empirical eigenfunctions are computed as linear combination of the used data snapshots by

$$\varphi_n(x) = \sum_{k=1}^{M} f_k^n v^{(k)}(x) \qquad n = 1,2,\ldots,M. \tag{7.16}$$

This approach is known as the method of snapshots and was introduced by Sirovich in [26].

Method of Snapshots and Data Matrix Singular Value Decomposition

Now, suppose we have a linearly independent set of snapshot data samples $\{u^{(1)}, u^{(2)}, \ldots, u^{(M)}\}$ that are the result of performing a physical experiment. For measurement purposes, we have discretized the spatial domain, so assume that each data snapshot consists of J data samples, i.e.,

$$u^{(i)} = [\ u^{(i)}(x_1) \quad \ldots \quad u^{(i)}(x_J)\ ].$$

Using all the mean adjusted data snapshots we build a matrix

$$V = \begin{bmatrix} v^1(x_1) & \ldots & v^1(x_J) \\ v^2(x_1) & \ldots & v^2(x_J) \\ \vdots & & \vdots \\ v^M(x_1) & \ldots & v^M(x_J) \end{bmatrix}$$

in such a way that each row is one data snapshot, and each column corresponds to the measurement at the same spatial point. Then, if we approximate the integral with a sum, the averaged correlation matrix given in (7.14) can be written as

$$C = \alpha V V^T,$$

where α is a real constant. Finding the eigenvalue decomposition (7.15) is equivalent to finding the singular value decomposition of the data matrix V.

Definition 7 *Singular Value Decomposition*
Any $m \times n$ matrix A can be factored into

$$A = Q_1 \Sigma Q_2^T \qquad Q_1^T Q_1 = I, \; Q_2^T Q_2 = I \tag{7.17}$$

The diagonal (but rectangular) matrix Σ is nonnegative and its positive entries $\sigma_1, \ldots, \sigma_r$, are the square roots of the nonzero eigenvalues of both AA^T and $A^T A$. They are called the **singular values** *of A. The columns of Q_1 are eigenvectors of AA^T, and the columns of Q_2 are eigenvectors of $A^T A$.*

If we find the singular value decomposition for the data matrix V

$$V = F \Sigma \Phi^T \qquad F^T F = I, \; \Phi^T \Phi = I$$

then the eigenvectors f^n of the matrix C defined in Section 7.2 are the scaled columns of the matrix F and the empirical eigenfunctions φ_n from (7.16) are the scaled columns of Φ. Thus empirical eigenfunctions span the row space of the data matrix V.

Thus, computing the KLE from the available data snapshots is equivalent to computing the singular value decomposition of the data matrix, and developing the reduced order model is equivalent to reducing the data matrix.

7.3 Symmetry and Karhunen-Loeve Expansion

Physical systems may exhibit various types of both *continuous* and *discrete* symmetries. To characterize the relation between underlying symmetries and subspaces spanned by the empirical eigenfunctions we need the notions of equivariant dynamical systems and invariant sets. The following definition is taken from [22].

Definition 8 *Equivariance*
Let

$$\dot{u} = f(u) \tag{7.18}$$

be an n-dimensional system of ODEs and Γ *be a symmetry group acting on the phase space* $\mathbf{R}^n$*, where the elements* γ *of* Γ *are just* $n \times n$ *matrices. The equation (7.18) is said to be* **equivariant** *under* Γ *if for every* $\gamma \in \Gamma$ *the equation*

$$\gamma f(u) = f(\gamma u)$$

holds.

This implies that if u is a solution of (7.18), then so is $\gamma u(t)$.

A set $\mathcal{S}$ is invariant for the flow Φ_t generated by an equation (7.18) if, when the initial condition $u(0)$ lies in $\mathcal{S}$, then so does the solution $u(t) = \Phi_t(u(0))$, for all t. Now, we can define an attracting set $\mathcal{A}$ as an invariant set which attracts all solutions starting in some open neighborhood of $\mathcal{A}$, and we say an orbit is dense in $\mathcal{S}$ if that orbit goes arbitrarily close to every point in $\mathcal{S}$.

Definition 9 *An* **attractor** *is an attracting set which contains a dense orbit.*

The requirement of a dense orbit guarantees that almost all solutions in an attractor display the typical behavior of that attractor. The attractor is ergodic if time averages and averages over the part of phase space containing the attractor coincide.

While a physical system or its dynamical system model may well admit a symmetry group, one can not expect ensembles of observations to share the full underlying symmetry group. A simple example of this would be a system with several distinct attractors. Then the time average of a single solution will reproduce just one of these attractors and empirical eigenfunctions generated by time averaging data snapshots obtained in one experimental run have less symmetry than the original problem.

Adopting the same philosophy for the KLE concept leads to the conclusion that if φ_n and λ_n are the empirical eigenvectors and corresponding eigenvalues generated from a set of experiments $\{u^{(k)}\}$ of a dynamical system equivariant under a group Γ then a necessary condition for the system generating $\{u^{(k)}\}$ to

be ergodic is that each of the finite dimensional eigenspaces corresponding to a given empirical eigenvalue be invariant under Γ. This can easily be checked experimentally. An alternative approach is to assume that a system is ergodic and use its known symmetries to increase the size of the ensemble, generating a symmetric data set $\{\gamma u^{(k)}\}$ from the available measured ensemble $\{u^{(k)}\}$. This approach has been advocated by Sirovich in [26].

Because of the nature of a stall cell phenomena, we are interested in rotational symmetry, called homogeneity in the turbulence literature. In this case the averaged two point correlation function $R(x,y)$ is homogeneous, meaning that it depends only on the difference of the two coordinates. In the case of a finite domain $\mathcal{D} = [0, 2\pi]$, we may develop a homogeneous $R(x,y)$ in a Fourier series representation

$$R(x,y) = R(x-y) = \sum a_k e^{ik(x-y)} \tag{7.19}$$

which implies that $\{e^{ikx}\}$ are exactly the eigenfunctions of the integral equation (7.11). Conversely, if the eigenfunctions are Fourier modes then equation (7.19) holds, leading to the conclusion that $R(x,y)$ is homogeneous. Thus, homogeneity completely determines the form of the empirical eigenfunctions, whereas ordering of the eigenvalues depends on the Fourier spectrum of the data involved.

We will present a very simple but illustrative example that justifies that KLE is an inadequate method for model reduction when a system has rotational symmetry.

Assume that we have just one impulse rotating around the annulus of the compressor rig with an angular speed equal to the sampling rate. In that case, at every sample time the impulse will be measured by the next sensor, and the data matrix is

$$U = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = I_M.$$

The averaged covariance matrix is $C = \alpha I_M$ and all the eigenvalues are equal

$$\lambda_i = \alpha, \;\; i = 1, \dots, M.$$

Thus it is impossible to determine a truncation point and obtain a reduced order model. The same argument holds even when the rotating speed of an impulse is not equal to the sampling rate.

In the case of a stall cell where we have a square like spatial structure the disadvantage of making a Galerkin truncation of corresponding ensemble of data onto Fourier modes is that to capture how the stall cell propagates and evolves one needs to include a large number of modes. A natural remedy for this is to try to capture the dynamics with a family of propagating curves. We will concentrate on that approach in Chapter 9. In the following chapter we will compare the balanced truncation method, and the Galerkin method for system model reduction.

Chapter 8

Model Reduction Technique Comparison

In control system design simple linear models of systems and controllers are preferred, since they are much easier to perform analysis and synthesis on. When a system is infinite dimensional, the model approximation becomes essential. In this chapter we consider the problem of reducing the order of a dynamical system.

There are many ways to reduce the order of a linear system, but we will present only one of them: the balanced truncation method which is very simple and performs fairly well. We will present the theory of a linear system balanced realization, based on linear system observability and controllability theory, and widely used in the robust control. More details of this theory can be found in [52]. Balanced truncation model reduction was introduced by Moore in [30] by applying principal component analysis to responses of linear system models. We will give a brief overview of his approach to model reduction.

To better understand model reduction of nonlinear systems based on the Galerkin projection and KLE, this methodology , and the balanced truncation method are compared by applying them to the same system driven by a linear PDE.

8.1 Linear System Model Reduction by Balanced Truncation

Finding reduced order models for linear systems is based on the input-output characteristics of a system. Given a full order model of a linear dynamical system in the form of a transfer function $G(s)$, we want to find a lower

order model, say an rth order model G_r, such that G and G_r are close in the following sense

$$\inf_{deg(G_r)\leq r} \|G - G_r\|_\infty,$$

where

$$\|G\|_\infty \doteq ess\sup_\omega \overline{\sigma}\{G(j\omega\ )\}.$$

There are infinitely many different state space realizations for a given transfer matrix, but some have proven to be useful in control engineering. In this subsection we will show the effectiveness of an approach based on balanced realizations.

Observability operators

Consider an LTI system with zero input and non-zero initial state

$$\begin{aligned} \dot{x}(t) &= Ax(t), \quad x(0) = x_0 \in \mathbf{C}^n \\ y(t) &= Cx(t) \end{aligned}$$

where A is a Hurwitz matrix. The solution to this system is

$$y(t) = Ce^{At}x_0, \quad t \geq 0. \tag{8.1}$$

Define the observability operator $\phi_o : \mathbf{C}^n \to L_2[0,\infty)$ by

$$x_0 \to Ce^{At}x_0.$$

For all initial conditions x_0 and time t, since A is Hurwitz, there exist positive constants k and α such that

$$\|y\|_2 = \|\phi_o x_0\| \leq \frac{k}{\alpha}\|x_0\|_2,$$

meaning that ϕ_0 is a bounded operator. Then the energy of $y = \phi_o x_0$ is given by

$$\begin{aligned} \|y\|_2^2 &= \langle \phi_o x_0, \phi_o x_0 \rangle \\ &= \langle x_0, \phi_o^* \phi_o x_0 \rangle \\ &= x_0^* Y_o x_0 \end{aligned}$$

where Y_o is the *observability gramian* defined as

$$Y_o \doteq \phi_o^* \phi_o \doteq \int_0^\infty e^{A^* t} C^* C e^{At} dt.$$

The observability gramian is the unique positive semi-definite matrix solution to the Lyapunov equation

$$A^* Y_o + Y_o A + C^* C = 0$$

that determines how much energy in the output y is given by an initial state x_0, and its eigenvalue decomposition provides a way to assess the relative observability of various directions in the state space.

Let

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0,$$

be the eigenvalues of Y_o and

$$v_1, \ldots, v_n$$

corresponding unit eigenvectors. Then the eigenvalues, σ_k, give a notion of "observability" of a particular direction in state space. If $\sigma_k >> \sigma_j$ then the output energy resulting from initial state v_j is much smaller then the energy observed when the initial condition is v_k. We say that state v_k is "more observable" than state v_j.

Controllability operators

Here we will present the dual idea of that pursued in the previous subsection. This time consider an LTI system with non-zero input

$$\dot{x}(t) = Ax(t) + Bw(t), \quad x(-\infty) = 0 \tag{8.2}$$

where A is a Hurwitz matrix. The vector $x(0)$ is the response of this system to an input function $w \in L_2[-\infty, 0)$ given as

$$x(0) = \int_{-\infty}^0 e^{-At} B w(t) dt. \tag{8.3}$$

Define the controllability operator $\phi_c : L_2(-\infty, 0] \to \mathbf{C}^n$ by

$$w \to \int_{-\infty}^0 e^{-At} B w(t) dt.$$

For a given unit norm state $x_0 \in \mathbf{C}^n$ the smallest norm $w \in L_2(-\infty, 0]$ that solves

$$\phi_c w = x_0$$

is given by

$$w_{opt} = \phi_c^* Y_c^{-1} x_0.$$

The energy of w_{opt} is given by

$$\begin{aligned} \|w_{opt}\|_2^2 &= \langle \phi_c^* Y_c^{-1} x_0, \phi_c^* Y_c^{-1} x_0 \rangle \\ &= \langle Y_c^{-1} x_0, \phi_c \phi_c^* Y_c^{-1} x_0 \rangle \\ &= x_0^* Y_c^{-1} x_0 \end{aligned}$$

where Y_c is called the *controllability gramian* defined as

$$Y_c = \phi_c \phi_c^* = \int_0^\infty e^{-At} B B^* e^{-A*t} dt.$$

The controllability gramian is the unique positive semi-definite matrix solution to the Lyapunov equation

$$A Y_c + Y_c A^* + B B^* = 0$$

that determines all the possible final states $x_0 = \phi_c w$ that can result, given an input $\|w\|_2 = 1$. An eigenvalue decomposition of the controllability gramian provides a way to rank the relative controllability of various directions in the state space.

Let

$$\delta_1 \geq \delta_2 \geq \ldots \geq \delta_n \geq 0,$$

be the eigenvalues of Y_o and

$$r_1, \ldots, r_n$$

corresponding unit norm eigenvectors. Then the eigenvalues δ_k give us a notion of "controllability" of a particular direction in state space. If $\delta_k >> \delta_j$ then state direction r_k is "more controllable" than state direction r_j.

Balanced Realization and Model Reduction by Balanced Truncation

In the previous two subsections we have provided a method (in terms of gramians) to rank directions in state space by controllability and observability. In general we are considering the system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bw(t), \quad x(0) = x_0 \\ y(t) &= Cx(t) \end{aligned} \tag{8.4}$$

for $t \geq 0$. The goal is to find a natural basis for this state space that determines which states dominate the system's behavior. We have concluded that states corresponding to small eigenvalues of the observability gramian are weakly observable, but they can be strongly controllable, meaning that they do contribute to the input-output behavior of the system. In the same manner states corresponding to small eigenvalues of the controllability gramian are weakly controllable, but they can be strongly observable, meaning that they also contribute to the input-output behavior of the system. Intuitively, we would like to find a realization of the system, such that strongly controllable states are strongly observable and vice versa. Assume that we change the basis of state space for the system (8.4). A transformed realization

$$\begin{aligned} \hat{A} &= TAT^{-1} \\ \hat{B} &= TB \\ \hat{C} &= CT^{-1}, \end{aligned}$$

is constructed, where T is a similarity transformation. The controllability and observability gramians associated with this new realization are

$$\begin{aligned} \hat{Y}_c &= TY_cT^* \\ \hat{Y}_o &= (T^{-1})^*Y_oT^{-1} \end{aligned}$$

and the eigenvalues of their product $\hat{Y}_c\hat{Y}_o = TY_cY_oT^{-1}$ are invariant under this state transformation. In the case of a minimal realization a similarity transformation T which gives the eigenvector decomposition

$$Y_cY_o = T^{-1}\Lambda T,$$

can be chosen such that $\hat{Y}_c = \hat{Y}_o = \Sigma$, where $\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$ and $\Sigma^2 = \Lambda$. This new realization, having equal controllability and observability gramians is called a balanced realization. The decreasingly ordered $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$, are called the Hankel singular values of the system.

Given a minimal realization of a system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bw(t) \\ y(t) &= Cx(t) + Dw(t), \end{aligned}$$

a balanced realization can be obtained by the following procedure:

- Compute the controllability and observability gramians $Y_c > 0, Y_o > 0$.
- Find a matrix R such that $Y_c = R^*R$.
- Diagonalize RY_oR^* to get $RY_oR^* = U\Sigma^2U^*$.
- Let $T^{-1} = R^*U\Sigma^{-\frac{1}{2}}$. Then

$$\begin{aligned} TY_cT^* &= (T^*)^{-1}Y_oT^{-1} = \Sigma \\ &= diag(\sigma_1, \sigma_2, \ldots, \sigma_n) \end{aligned}$$

and

$$\begin{aligned} \dot{\hat{x}}(t) &= TAT^{-1}\hat{x}(t) + TBw(t) \\ y(t) &= CT^{-1}\hat{x}(t) + Dw(t) \end{aligned}$$

is balanced.

More generally, if the realization of a system is not minimal, then there is a similarity transformation such that both controllability and observability gramians are diagonal and the controllable and observable subsystem is balanced.

Assume that we do have a balanced realization of a system and that the Hankel singular values of the system are decreasingly ordered

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n.$$

Suppose $\sigma_r >> \sigma_{r+1}$. Then the states corresponding to the singular values $\sigma_{r+1}, \ldots, \sigma_n$ are less controllable and observable than states corresponding to $\sigma_1, \ldots, \sigma_r$. Thus, by truncating those less controllable and less observable states we will not loose much information about the system. This is stated formally in the following theorem.

Theorem 12 *[52] Consider a stable system G(s) given by a balanced realization*

$$\begin{aligned} \dot{x}_1(t) &= A_{11}x_1(t) + A_{12}x_2(t) + B_1w(t), \quad x_1(0) = x_{1_0} \\ \dot{x}_2(t) &= A_{21}x_1(t) + A_{22}x_2(t) + B_2w(t), \quad x_2(0) = x_{2_0} \\ y(t) &= C_1x_1(t) + C_2x_2(t) + Dw(t) \end{aligned}$$

with gramians $\Sigma = diag(\Sigma_1, \Sigma_2)$

$$\Sigma_1 = diag(\sigma_1 I_{s_1}, \sigma_2 I_{s_2}, \ldots, \sigma_r I_{s_r})$$
$$\Sigma_2 = diag(\sigma_{r+1} I_{s_{r+1}}, \sigma_{r+2} I_{s_{r+2}}, \ldots, \sigma_N I_{s_N})$$

and

$$\sigma_1 > \sigma_2 > \cdots > \sigma_r > \sigma_{r+1} > \cdots > \sigma_N$$

where σ_i *has multiplicity* s_i, $i = 1, 2, \ldots, N$ *and* $s_1 + s_2 + \ldots + s_N = n$. *Then the truncated system* $G_r(s)$ *given by*

$$\begin{aligned} \dot{x}_r(t) &= A_{11}x_r(t) + B_1w(t) \\ y(t) &= C_1x_r(t) + Dw(t) \end{aligned}$$

is balanced and asymptotically stable. Furthermore

$$\|G(s) - G_r(s)\|_\infty \leq 2 \sum_{i=r+1}^{N} \sigma_i. \tag{8.5}$$

Balanced truncation of a linear system is based on the idea of preserving the states that contribute to the behavior of the system the most, and that concept is indirectly based on the energy distribution among the states.

Early balancing work was motivated by the principal component analysis theory that is strongly linked to the KLE method, and in the next section we will discuss the connection between these two methods.

8.2 Principal Component Analysis in Linear Systems

In this section we will present the principal component analysis approach to linear system model reduction that resulted in the development of the balanced truncation method. The original work is presented in [30].

Consider a plant with inputs u and outputs y operating at the equilibrium point (x_0, y_0). The corresponding model

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t), \end{aligned} \tag{8.6}$$

when started at rest ($x(0) = 0$), simulates exactly the small signal input-output characteristics of the plant, with a coordinate system translated to (x_0, y_0).

Definition 10 *Let $F : \mathbf{R} \to \mathbf{R}^{n\times m}$ be a piecewise continuous map represented in matrix form by $F(t)$. The* **gramian**

$$W^2 \doteq \int_{t_1}^{t_2} F(t)F^T(t)dt$$

is a positive semidefinite matrix with a set of nonnegative real eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_n^2 \geq 0$ and corresponding mutually orthogonal unit eigenvectors $v_1, v_2, \ldots, v_n$. The map F may be represented with $v_1, v_2, \ldots, v_n$ used as orthonormal basis vectors for $\mathbf{R}^n$, i.e.,

$$F(t) = v_1 f_1^T(t) + v_2 f_2^T(t) + \ \ldots \ v_n f_n^T(t)$$

where $f_i^T(t) \doteq v_i^T F(t)$ for $1 \leq i \leq n$. We refer to v_i as a **component vector**, *σ_i as a* **component magnitude**, *and $v_i f_i^t(t)$ as a* **principal component**.

Let $F(t)$ be the impulse response matrix of a linear time invariant system (8.6), and let σ_i and v_i be component magnitudes and vectors of $F(t)$ over $[0, T]$. Define

$$\begin{aligned} \Sigma &\doteq diag(\sigma_1, \sigma_2, \ldots, \sigma_n) \\ V &\doteq (v_1, v_2, \ldots, v_n). \end{aligned}$$

The following proposition shows that for Ω representing the class of all inputs $u(t)$ which are piecewise continuous on $[0, T]$ and satisfy the norm bound $(\int_0^T \|u(t)\|^2 dt)^{\frac{1}{2}} \leq 1$, the set $\hat{S} = \{y : \ y = V\Sigma p, \|p\| = 1\}$ corresponds to the surface of $S = \{y : y = \int_0^t F(t-\tau)u(\tau)d\tau, \ t \leq T, \ u \in \Omega\}$.

Proposition 2 *[30] The set S can be characterized as follows*

$$S = \{y : y = \alpha\hat{z}, \ \hat{z} \in \hat{S}, \ 0 \leq \alpha \leq 1\}.$$

It can be shown, too, that the component vectors corresponding to the nonzero principal components of $e^{At}B$ span the controllable subspace, and the component vectors corresponding to the nonzero principal components of $e^{A^T t}C^T$ span the unobservable subspace. A main problem that we want to deal with is the fact that the internal responses $e^{At}B$, $e^{A^T t}C^T$ depend on the internal coordinate system. The existence of small components in $e^{At}B$ or $e^{A^T t}C^T$ implies nothing about their importance with respect to input-output properties of the model. A special coordinate system where input-output properties are reflected by internal principal components can be derived.

The discrete time equivalent of system (8.6) obtained by sampling and holding inputs every t_s seconds is

$$\begin{aligned} x^d_{k+1} &= Fx^d_k + Gu^d_k \\ y^d_k &= Cx^d_k \end{aligned}$$

where

$$\begin{aligned} F &= e^{At_s} \\ G &= \int_0^{t_s} e^{A(t_s-\tau)}Bd\tau. \end{aligned}$$

We assume that system responses are analyzed over an interval $[0, T_a]$, and that $N = \frac{T_a}{t_s}$ is an integer. The extended controllability and observability matrices corresponding to (F, G, C) are:

$$\begin{aligned} Q_c(t_s) &= [\ G \quad FG \quad \ldots \quad F^NG\ ] \\ Q_o(t_s) &= \begin{bmatrix} C \\ CF \\ \vdots \\ CF^N \end{bmatrix} \end{aligned}$$

The matrix $Q_c(t_s)$ is a data matrix closely related to $e^{At}B$, and $Q_o(t_s)$ is formed by sampling Ce^{At} every t_s seconds over the time interval $[0, T_a]$. We adopt the following notation.

1. $V_c = [v_{c_1} \ldots v_{c_n}]$, $\Sigma_c = diag(\sigma_{c_1} \ldots \sigma_{c_n})$ where v_{c_i} and σ_{c_i} are the ith component vector and magnitude of $e^{At}B$.

2. $V_o = [v_{o_1} \dots v_{o_n}]$, $\Sigma_o = diag(\sigma_{o_1} \dots \sigma_{o_n})$ where v_{o_i} and σ_{o_i} are the ith component vector and magnitude of $e^{A^T t} C^T$.

3. $V_c^d(t_s)\Sigma_c^d(t_s)(U_c^d)^T(t_s) = Q_c(t_s)$ (SVD).

4. $V_o^d(t_s)\Sigma_o^d(t_s)(U_o^d)^T(t_s) = Q_o(t_s)$ (SVD).

The following proposition shows that the left singular vectors and corresponding singular values of the matrices $Q_c(t_s)$ and $Q_o(t_s)$ computed via sampled data converge to the principal vectors and component magnitudes of $e^{At}B$ and $e^{A^T t}C^T$, respectively.

Proposition 3 *([30]) The singular values satisfy*

$$\lim_{t_s \to 0} \frac{1}{\sqrt{t_s}} \Sigma_c^d(t_s) = \Sigma_c \qquad \lim_{t_s \to 0} \frac{1}{\sqrt{t_s}} \Sigma_o^d(t_s) = \Sigma_0.$$

If the diagonal elements of $\Sigma_c(t_s)$, $\Sigma_o(t_s)$ are distinct, then

$$\lim_{t_s \to 0} V_c^d(t_s) = V_c \qquad \lim_{t_s \to 0} V_o^d(t_s) = V_o.$$

Moore showed that there is a coordinate transformation T such that model obtained has balanced internal dynamics, meaning that the controllability and observability gramians defined in Section 8.1 are equal. The transformation matrix T is a function of V_c, Σ_c, V_o, and Σ_o, and is computed using their sampled data equivalents V_c^d, Σ_c^d, V_o^d, and Σ_o^d.

Assume that the system is asymptotically stable and that (A, B, C) is internally balanced over $[0, \infty)$, i.e.,

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

or, equivalently,

$$Y_c = Y_o = \Sigma = diag(\sigma_1, \dots, \sigma_n).$$

The diagonal entries of Σ are decreasingly ordered.

Consider the case when $\sigma_{k+1} \ll \sigma_k$. The basic model reduction idea is to partition the state variables of the internally balanced model in accordance with the partitioning of Σ to obtain

$$\begin{aligned} \dot{x}_1(t) &= A_{11}x_1(t) + A_{12}x_2(t) + B_1 u(t) \\ \dot{x}_2(t) &= A_{21}x_1(t) + A_{22}x_2(t) + B_2 u(t) \\ y(t) &= C_1 x_1(t) + C_2 x_2(t) + Du(t). \end{aligned}$$

The subspace spanned by the first k states is a working approximation of the controllable and observable space of the original system (8.6). The resulting lower order model (A_{11}, B_1, C_1) is asymptotically stable and internally balanced.

The main idea underlying this model reduction method is to eliminate any weak subsystem which contributes little to the impulse response matrix. This implicitly defines the meaning of a dominant subsystem whose impulse response matrix is close to that of the full model.

8.3 Galerkin Projection onto Linearly Independent Set of Functions

In Section 7.1 we briefly reviewed the Galerkin discretization scheme for PDEs based on the separation of variables approach which attempts to find an approximate solution in the form of a truncated series expansion given by

$$\hat{u}(x,t) = \sum_{n=1}^{N} a_n(t)\varphi_n(x) + \bar{u}(x), \tag{8.7}$$

where the set of trial functions $\{\varphi_n(x)\}$ is orthonormal and

$$\bar{u}(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T u(x,t)dt. \tag{8.8}$$

To develop the mathematical machinery we need for comparison of different model reduction techniques, in this section we will show how one can perform Galerkin projection onto a linearly independent set of trial functions as explained in more detail in [21].

Suppose again, a system is governed by the PDE

$$\frac{\partial u}{\partial t} = D(u) \quad u : [0, 2\pi] \times (0, \infty) \to \mathbf{R}$$

with given initial, and boundary conditions, where $D(\cdot)$ is a linear (spatial) differential operator. At each instant t, we assume that the coefficients $a_n(t)$ are known and we seek the values for the N independent quantities $\dot{a}_n(t)$ that minimize $\langle r(x,t), r(x,t)\rangle$ where $r(x,t)$ is the residual error produced by using the approximate solution (8.7) instead of the exact solution. The resulting system of N ordinary differential equations

$$\frac{d}{dt}\langle \hat{u}, \varphi_i\rangle = \langle D(\hat{u}), \varphi_i(x)\rangle \tag{8.9}$$

or

$$\begin{aligned}\sum_{n=1}^{N}\dot{a}_n(t)\langle\varphi_n(x),\varphi_i(x)\rangle &= \langle D\left(\sum_{n=1}^{N}a_n(t)\varphi_n(x)+\bar{u}(x)\right),\varphi_i(x)\rangle \\ &= \langle\sum_{n=1}^{N}a_n(t)D(\varphi_n(x)),\varphi_i(x)\rangle + \\ &\quad \langle D(\bar{u}(x)),\varphi_i(x)\rangle \\ &\qquad\qquad i=1,\ldots,N.\end{aligned}$$

is the obtained reduced order model. This model can be written as

$$F\dot{a}(t) = Aa(t) + B, \tag{8.10}$$

where

$$\begin{aligned}a(t) &= \begin{bmatrix} a_1(t) \\ \vdots \\ a_N(t)\end{bmatrix} \\ F &= \begin{bmatrix} \langle\varphi_1,\varphi_1\rangle & \cdots & \langle\varphi_1,\varphi_N\rangle \\ \vdots & & \vdots \\ \langle\varphi_N,\varphi_1\rangle & \cdots & \langle\varphi_N,\varphi_N\rangle\end{bmatrix} \\ A &= \begin{bmatrix} \langle\varphi_1(x),D(\varphi_1(x))\rangle & \ldots & \langle\varphi_1(x),D(\varphi_N(x))\rangle \\ \vdots & & \vdots \\ \langle\varphi_N(x),D(\varphi_1(x))\rangle & \ldots & \langle\varphi_N(x),D(\varphi_N(x))\rangle\end{bmatrix}. \\ B &= \begin{bmatrix} \langle\varphi_1(x),D(\bar{u}(x))\rangle \\ \vdots \\ \langle\varphi_N(x),D(\bar{u}(x))\rangle\end{bmatrix}.\end{aligned} \tag{8.11}$$

Since trial functions are linearly independent, the matrix F has full rank and the reduced order model (8.10) can be put in the following form

$$\dot{a}(t) = F^{-1}Aa(t) + F^{-1}B. \tag{8.12}$$

When a set of trial functions is orthonormal, $F = I_N$, then same result as in Section 7.1 is obtained.

The initial conditions for the resulting system of ODEs are again determined by a second application of the Galerkin approach. We choose them

in such a way that they minimize the norm of the initial conditions residual $r_0(x) = u(x,0) - \hat{u}(x,0)$ and we obtain a system of N linear equations

$$0 = \frac{d}{da_i(0)} \quad \langle \quad u(x,0) - \sum_{n=1}^{N} a_n(0)\varphi_n(x) - \bar{u}(x),$$
$$u(x,0) - \sum_{n=1}^{N} a_n(0)\varphi_n(x) - \bar{u}(x)\rangle$$

or

$$\begin{aligned} \sum_{n=1}^{N} a_n(0)\langle \varphi_i(x), \varphi_n(x)\rangle &= \langle u(x,0) - \bar{u}(x), \varphi_i(x)\rangle \\ &= \langle v(x,0), \varphi_i(x)\rangle \quad i = 1, \ldots, N. \end{aligned}$$

This is equivalent to

$$a(0) = F^{-1}V_0 \tag{8.13}$$

where

$$V_0 = \begin{bmatrix} \langle \varphi_1(x), v(x,0)\rangle \\ \vdots \\ \langle \varphi_N(x), v(x,0)\rangle \end{bmatrix}. \tag{8.14}$$

It is important to note that to solve the system of ODEs (8.12) and (8.13) one only needs to select the set of trial functions $\{\varphi_n\}$ and the initial conditions of the original system $u(x,0)$.

8.4 Balancing and Galerkin Commute

To better understand model reduction of a nonlinear system based on Galerkin projection and KLE, we will compare this methodology and the balanced truncation model reduction method by applying them to the same system. This could lead to an extension of the widely accepted concept of model reduction by the balanced truncation method for linear systems to model reduction for nonlinear systems.

Suppose we have a system governed by the PDE

$$\frac{\partial u}{\partial t} = D(u) \quad u : [0, 2\pi] \times (0, \infty) \to \mathbf{R} \tag{8.15}$$

with given initial and boundary conditions, where $D(\cdot)$ is a linear (spatial) differential operator. Determining a set of trial functions $\{\varphi_n\}$ by KLE applied to the available data snapshots, and performing a Galerkin projection

onto $\{\varphi_n\}$ will give an approximate solution in the form of a truncated series expansion

$$\hat{u}(x,t) = \sum_{n=1}^{N} a_n(t)\varphi_n(x) + \bar{u}(x), \quad (\forall i,j)\ \langle\varphi_i,\varphi_j\rangle = \sigma_{ij}. \tag{8.16}$$

Thus, the original infinite dimensional system (8.15) is approximated by a system of N ODEs

$$\dot{a}(t) = Aa(t) + B, \quad a(0) = a_0 \tag{8.17}$$

where

$$\begin{aligned} A_{ij} &= \langle\varphi_i(x), D(\varphi_j(x))\rangle \\ B_i &= \langle\varphi_i(x), D(\bar{u}(x))\rangle \\ a_i(0) &= \langle\varphi_i(x), u(x,0) - \bar{u}(x)\rangle \\ i,j &= 1,\ldots,N. \end{aligned}$$

The balanced realization of the linear system is based on its input-output properties and we need matrices A, B and C from any state representation of a system to obtain a balanced realization of the system. System (8.17) can be viewed as a state space model if we assume that the input $w(t)$ is constant and equal to 1, but there is no equation defining the output behavior of the system. Since for us a reduced order model is a model describing the time evolution of time dependent coefficients $\{a_n(t)\}$ we will define them as outputs of the system,

$$y(t) = I_N a(t). \tag{8.18}$$

Then we can write equations (8.17) and (8.18) in state space form as

$$\begin{aligned} \dot{a}(t) &= Aa(t) + Bw(t), \quad a(0) = a_0 \\ y(t) &= I_N a(t), \end{aligned} \tag{8.19}$$

where

$$w(t) = 1, \quad t \geq 1. \tag{8.20}$$

Next we will change the basis of state space by

$$x(t) = Ta(t) \tag{8.21}$$

to obtain a balanced realization of system (8.19)

$$\begin{aligned} \dot{x}(t) &= TAT^{-1}x(t) + TBw(t) \\ y(t) &= I_N T^{-1} a(t). \end{aligned} \tag{8.22}$$

An approximate solution (8.16) can then be represented in terms of new state space variables in the following way

$$\begin{aligned} \hat{u}(x,t) &= \bar{u}(x) + \begin{bmatrix} a_1(t) & \dots & a_N(t) \end{bmatrix} \begin{bmatrix} \varphi_1(x) \\ \vdots \\ \varphi_N(x) \end{bmatrix} \\ &= \bar{u}(x) + a^T(t)\Psi(x) \\ &= \bar{u}(x) + (T^{-1}x(t))^T \Psi(x) \\ &= \bar{u}(x) + x^T(t)(T^{-T}\Psi(x)) \\ &= \bar{u}(x) + x^T(t)\Phi(x) \end{aligned} \tag{8.23}$$

where

$$\begin{aligned} \Phi(x) &= T^{-T} \begin{bmatrix} \varphi_1(x) \\ \vdots \\ \varphi_N(x) \end{bmatrix} \\ &= \begin{bmatrix} \phi_1(x) \\ \vdots \\ \phi_N(x) \end{bmatrix}, \quad \forall(i,j)\ \langle \phi_i(x), \phi_j(x) \rangle \neq \sigma_{ij} \end{aligned} \tag{8.24}$$

are new trial functions.

It would be nice to know that if we have initially chosen to find an approximate solution of a PDE (8.15) using a set of linearly independent functions $\{\phi_n\}$ in the form of

$$\hat{u}(x,t) = \bar{u}(x) + \sum_{n=1}^{N} c_n(t)\phi_n(x) \tag{8.25}$$

then will we get the same system of ODEs governing the time dependent coefficients evolution.

We have the original system (8.15) governed by the PDE and a given initial and boundary conditions. Using Galerkin projection onto $\{\phi_n\}$ as explained

in the Section 8.3 will determine a reduced order model of (8.15)

$$\begin{aligned} \dot{c}(t) &= F^{-1}\hat{A}c(t) + F^{-1}\hat{B} \\ c(0) &= F^{-1}U \end{aligned} \tag{8.26}$$

where

$$\begin{aligned} F_{ij} &= \langle \phi_i(x), \phi_j(x) \rangle \\ \hat{A}_{ij} &= \langle \phi_i(x), D(\phi_j(x)) \rangle \\ \hat{B}_i &= \langle \phi_i(x), D(\bar{u}(x)) \rangle \\ U_i &= \langle \phi_i, u(x,0) - \bar{u}(x) \rangle \\ i,j &= 1, \ldots, N. \end{aligned}$$

To be able to determine if $\{x_n\}$ and $\{c_n\}$ are the same we have to compare systems (8.19) and (8.26). Let's denote $L = T^{-T}$. Then

$$\begin{aligned} F_{ij} &= \langle \phi_i(x), \phi_j(x) \rangle \\ &= \langle \sum_{k=1}^{N} L_{ik}\varphi_k(x), \sum_{m=1}^{N} L_{jm}\varphi_m(x) \rangle \\ &= \sum_{k=1}^{N}\sum_{m=1}^{N} L_{ik}L_{jm}\langle \varphi_k(x), \varphi_m(x) \rangle \\ &= \sum_{k=1}^{N} L_{ik}L_{jk} \\ &= (LL^T)_{ij} \\ &= (T^{-T}T^{-1})_{ij}, \end{aligned} \tag{8.27}$$

$$\begin{aligned} \hat{A}_{ij} &= \langle \phi_i(x), D(\phi_j(x)) \rangle \\ &= \langle \sum_{k=1}^{N} L_{ik}\varphi_k(x), \sum_{m=1}^{N} L_{jm}D(\varphi_m(x)) \rangle \\ &= \sum_{k=1}^{N}\sum_{m=1}^{N} L_{ik}L_{jm}\langle \varphi_k(x), D(\varphi_m(x)) \rangle \\ &= (LAL^T)_{ij} \\ &= (T^{-T}AT^{-1})_{ij} \end{aligned} \tag{8.28}$$

and

$$\begin{aligned}
\hat{B}_i &= \langle \phi_i(x), D(\bar{u}(x)(x)) \rangle \\
&= \langle \sum_{k=1}^{N} L_{ik}\varphi_k(x), D(\bar{u}(x)) \rangle \\
&= \sum_{k=1}^{N} L_{ik} \langle \varphi_k(x), D(\bar{u}(x)) \rangle \\
&= \sum_{k=1}^{N} L_{ik} B_k \\
&= (T^{-T}B)_i. \qquad (8.29)
\end{aligned}$$

Substituting (8.27), (8.28) and (8.29) back into (8.26) yields

$$\begin{aligned}
\dot{c}(t) &= F^{-1}(\hat{A}c(t) + \hat{B}) \\
&= (T^{-T}T^{-1})^{-1}(T^{-T}AT^{-1}c(t) + T^{-T}B) \\
&= TAT^{-1}c(t) + TB \qquad (8.30)
\end{aligned}$$

and

$$\begin{aligned}
c(0) &= F^{-1}T^{-T}a(0) \\
&= (T^{-T}T^{-1})^{-1}T^{-T}a(0) \\
&= x(0). \qquad (8.31)
\end{aligned}$$

Thus we can conclude that

$$x(t) = c(t). \qquad (8.32)$$

It is important to say that the fact that D is a linear spatial differential operator enabled us to derive all the formulas (8.27) - (8.31). We have shown that in this particular case we can think of balancing the reduced order model obtained by Galerkin projection and consisting of N ODEs as just changing a set of trial functions used for the truncated series expansion. In other words the balancing transformation and Galerkin projection commute. We can either change the set of trial functions or we can change state space coordinates, using the same transformation.

8.5 Balanced Truncation and Galerkin Do Not Commute

Given a balanced realization of the system (8.19)

$$\begin{aligned} \dot{x}(t) &= \hat{A}x(t) + \hat{B}w(t), \quad x(0) = Ta(0) \\ y(t) &= \hat{C}a(t), \end{aligned} \tag{8.33}$$

where

$$\begin{aligned} \hat{A} &= TAT^{-1} \\ \hat{B} &= TB \\ \hat{C} &= I_N T^{-1}, \end{aligned}$$

an approximate solution (8.16) can then be represented in terms of new state space variables in the following way

$$\hat{u}(x,t) = \bar{u}(x) + x^T(t)\Phi(x). \tag{8.34}$$

In the previous section we have investigated what balancing does to a reduced order model of a linear system. Here we will go even further and try to understand what happens once we truncate the balanced realization of the reduced order system. State truncation is performed by a simple matrix multiplication. Assume we want to keep just the first M states, then

$$\begin{aligned} z(t) &= \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} \\ &= \begin{bmatrix} I_M & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ \vdots \\ x_N(t) \end{bmatrix} \\ &= Rx(t). \end{aligned}$$

Then the truncated system is given by

$$\begin{aligned} \dot{z}(t) &= R\hat{A}R^T z(t) + R\hat{B}, \quad z(0) = Rx(0) \\ \tilde{y} &= \hat{C}R^T z \end{aligned}$$

and an approximate solution represented in terms of the truncated system output variables is

$$\begin{aligned}\hat{u}^t(t,x) &= \bar{u}(x) + \tilde{y}^T(t)\Psi(x) \\ &= \bar{u}(x) + z^T(t)R\hat{C}^T\Psi(x) \\ &= \bar{u}(x) + z^T(t)(R\Phi(x)) \\ &= \bar{u}(x) + z^T(t)\Phi^t(x)\end{aligned}$$

where

$$\Phi^t(x) = \begin{bmatrix} \phi_1(x) \\ \vdots \\ \phi_M(x) \end{bmatrix} \tag{8.35}$$

are the first M trial functions obtained after balancing the original system, as shown in (8.24). Having this reduced set of trial functions we would be interested in seeing what reduced order model would be obtained if we use Galerkin projection onto $\{(\phi^t)^n\}$ as explained in Section 8.3 and obtain an approximate solution

$$\hat{u}(x,t) = \bar{u}(x) + \sum_{n=1}^{M} d_n(t)\phi_n(x). \tag{8.36}$$

The reduced order model is

$$\dot{d}(t) = (F^M)^{-1}(A^M d(t) + B^M), \quad d(0) = (F^M)^{-1}V^t \tag{8.37}$$

where

$$\begin{aligned} F_{ij}^M &= \langle \phi_i(x), \phi_j(x) \rangle \\ A_{ij}^M &= \langle \phi_i(x), D(\phi_j(x)) \rangle \\ B_i^M &= \langle \phi_i(x), D(\bar{u}(x)) \rangle \\ V_i^t &= \langle \phi_i, u(x,0) - \bar{u}(x) \rangle \\ i,j &= 1, \ldots, M. \end{aligned}$$

Let's denote $L = T^{-T}$. Then

$$\begin{aligned} F_{ij}^M &= \langle \phi_i(x), \phi_j(x) \rangle \\ &= \langle \sum_{k=1}^{M} L_{ik}\varphi_k(x), \sum_{m=1}^{M} L_{jm}\varphi_m(x) \rangle \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{N} L_{ik} L_{jk} \\
&= (LL^T)_{ij} \\
&= (T^{-T} T^{-1})_{ij}, \\
F^M &= R(T^{-T} T^{-1}) R^T
\end{aligned}
$$

$$
\begin{aligned}
A_{ij}^M &= \langle \phi_i(x), D(\phi_j(x)) \rangle \\
&= \langle \sum_{k=1}^{M} L_{ik} \varphi_k(x), \sum_{m=1}^{M} L_{jm} D(\varphi_m(x)) \rangle \\
&= (LAL^T)_{ij} \\
&= (T^{-T} A T^{-1})_{ij} \\
A^M &= R(T^{-T} A T^{-1}) R^T,
\end{aligned}
\tag{8.38}
$$

$$
\begin{aligned}
B_i^M &= \langle \phi_i(x), D(\bar{u}(x)(x)) \rangle \\
&= \langle \sum_{k=1}^{N} L_{ik} \varphi_k(x), D(\bar{u}(x)) \rangle \\
&= \sum_{k=1}^{M} L_{ik} B_k \\
&= (T^{-T} B)_i \\
B^M &= R T^{-T} B,
\end{aligned}
$$

and

$$
\begin{aligned}
(V^t)_i &= \langle \phi_i(x), v(x,0) \rangle \\
&= \sum_{k=1}^{M} L_{ik} \langle \varphi_k(x), v(x,0) \rangle \\
B^M &= R T^{-T} a(0)
\end{aligned}
$$

and the reduced order model can be written as

$$
\begin{aligned}
\dot{d}(t) &= \left(R T^{-T} T^{-1} R^T\right)^{-1} \left(R(T^{-T} A T^{-1}) R^T d(t) + R T^{-T} B\right) \\
d(0) &= \left(R T^{-T} T^{-1} R^T\right)^{-1} R T^{-T} a(0).
\end{aligned}
\tag{8.39}
$$

In the case when $TT^T = I_N$ and because $RR^T = I_M$, system (8.39) is equivalent to the system

$$\begin{aligned} \dot{d}(t) &= R(TAT^{-1})R^T d(t) + RTB \\ &= R\hat{A}R^T d(t) + R\hat{B} \\ d(0) &= RTa(0) \\ &= Rx(0) \end{aligned}$$

and we can claim that

$$d(t) = z(t). \tag{8.40}$$

Thus, as long as the initial change of state space basis T is an orthogonal matrix, we have that $d(t) = z(t)$. Since a set of trial functions obtained by the KLE method is an orthonormal set, if we have an orthogonal state space transformation matrix T, the new set of trial functions $\{\psi_n(x)\}$, corresponding to transformed state space is also an orthonormal set of functions.

In general, we don't have $TT^T = I_N$ meaning that equality (8.40) does not hold always, and we can not represent model reduction by balanced truncation just as obtaining the reduced order model by Galerkin projection onto truncated set of trial functions corresponding to a balanced realization of a reduced order model. In other words, balanced truncation and Galerkin do not commute.

8.6 Conclusion

To better understand model reduction of a nonlinear systems based on Galerkin projection and KLE, in this chapter we compared this methodology and the balanced truncation method by applying them to the same system driven by a linear PDE.

We have shown that in this particular case we can think of balancing the reduced order model obtained by Galerkin projection and consisting of N ODEs as just changing the set of trial functions used for the truncated series expansion. Thus, one can conclude that the balancing transformation and the Galerkin method commute.

We demonstrated that only when we have an orthogonal state space transformation matrix T, balanced truncation and the Galerkin method commute.

Unfortunately, balancing is in general not an orthogonal transformation.

The most important step in extending the balanced truncation methodology to nonlinear systems will be in determining how to perform balancing of the system. The work presented in this chapter and a review of Moore's early development of the balancing method do suggest how we might generalize the method of balanced realization to nonlinear systems. It was shown by Moore that for linear systems the gramian computed via sampled data converges to the actual gramian as the sample rate tends to zero. To perform balancing for nonlinear systems, we need some method for computing gramians based on data.

Chapter 9

Centering and Karhunen-Loeve Expansion

9.1 Centering

Starting with this section and throughout this chapter we will concentrate on systems with rotational symmetry. Thus we will consider systems governed by the following type of PDE

$$\frac{\partial u}{\partial t} + \omega \frac{\partial u}{\partial x} = D(u) \quad u : [0, 2\pi] \times (0, \infty) \to \mathbf{R} \tag{9.1}$$

with periodic boundary conditions,

$$u(0, t) = u(2\pi, t)$$

where $D(\cdot)$ is a nonlinear operator that may involve spatial derivatives. In general these PDEs have a traveling (rotating) wave solution and we would like to obtain as few modes as necessary to accurately approximate the shape of the propagating wave. To accomplish this, we have to separate the movement of a solution $u(x, t)$ from the evolution of a wave shape.

First, we define a center of each member of an available ensemble $\{u^{(k)}\}$.

Definition 11 *Let $f(x)$ be a periodic function defined on $[0, 2\pi]$ with a period 2π,*

$$f(x) = f(x + 2\pi).$$

Define the center C of a wave $f(x)$ as

$$\int_0^C f(x)^2 dx = \int_C^{2\pi} f(x)^2 dx.$$

In order to extract the propagating wave we position all data snapshots so that their centers are at the same point. For simplicity and without loss of generality, we have chosen to place data snapshots centers at π so that

$$\int_0^{\pi} f(x-d)^2 dx = \int_{\pi}^{2\pi} f(x-d)^2 dx.$$

We call this procedure the centering of a wave. It is performed using the following iterative procedure.

1. Start with snapshot $u^t(x) = u(t,x)$
2. Compute a center C of the wave $u^t(x)$
3. Shift the wave $u^t(x)$ to the right by $|\pi - C|$
4. Find the center of $u^t(x)$
5. If converged ($C = \pi$), then stop, else go to 3

Suppose that we have an ensemble $\{u^{(k)}\}$ of scalar fields, each being a function $u^{(k)} = u^{(k)}(x)$ defined on the domain $0 \leq x \leq 2\pi$. To find a good representation of the members of $\{u^{(k)}\}$, we center each member of an ensemble to obtain a centered data ensemble $\{(u^c)^{(k)}\}$ and then project each $(u^c)^{(k)}$ onto candidate basis functions. Because we assumed that the u's belong to a Hilbert space $L^2([0,2\pi])$, this also holds for the u^c's.

Performing KLE on the centered data set, we find a basis $\{\varphi_n\}$ for L^2 that gives a finite-dimensional centered data representation of the form

$$\hat{u}^c(x,t) = \sum_{n=1}^{N} a_n(t)\varphi_n(x) + \bar{u}^c(x).$$

The original ensemble is then approximated as

$$\begin{aligned} \hat{u}(x,t) &= \hat{u}^c(t, x + d(t)) \\ &= \sum_{n=1}^{N} a_n(t)\varphi(x + d(t)) + \bar{u}^c(x + d(t)). \end{aligned} \tag{9.2}$$

9.2 Centering and Method of Characteristics

To show how centering works and how it relates to the standard method of characteristics, we will consider a simple first order non dimensional wave equation

$$\begin{aligned} \frac{\partial u}{\partial t} + \omega \frac{\partial u}{\partial x} &= 0 \\ u(x,0) &= g(x), \quad x \in [0, 2\pi] \quad g(0) = g(2\pi) \\ u(0,t) &= u(2\pi, t), \quad t \geq 0. \end{aligned} \tag{9.3}$$

Assuming that x depends on time t, the first derivative of $u(x,t)$ with respect to time t is

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x}\frac{dx}{dt}, \tag{9.4}$$

and along characteristics, where $\frac{dx}{dt} = \omega$ holds, we can write equation (9.3) as

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \omega \frac{\partial u}{\partial x} = 0. \tag{9.5}$$

Thus, we conclude that a solution to equation (9.5) does not change with time along the characteristics, and for every point (x_0, t_0) in the (x,t) plane the solution is equal to the initial condition at the point on the x-axis that lies on the line with slope ω passing through the point (x_0, t_0), i.e.,

$$u(x,t) = g(x - \omega t). \tag{9.6}$$

Now, suppose that we have an ensemble $\{u^{(k)}\}$ of data snapshots shown in Figure 9.2 obtained by simulation of this simple equation for $\omega = 4$ and $g(x)$ shown in the Figure 9.1, and we center them to obtain the centered data snapshots shown in Figure 9.3.

We extract the mean centered snapshot shown in Figure 9.4 which is just the initial condition rotated by some angle, i.e.

$$g(x) = u(x,0) = \bar{u}^c(x + d_0)$$

and an approximate solution of the equation (9.3)

$$\hat{u}(x,t) = \bar{u}^c(x + d(t))$$

is equal to the exact solution. Thus, by centering, we have extracted the wave shape $g(x)$ and the speed of rotation $-\dot{d}(t)$, where $d(t)$ is shown in Figure 9.5, and that is all we need to be able to represent the solution to equation (9.3).

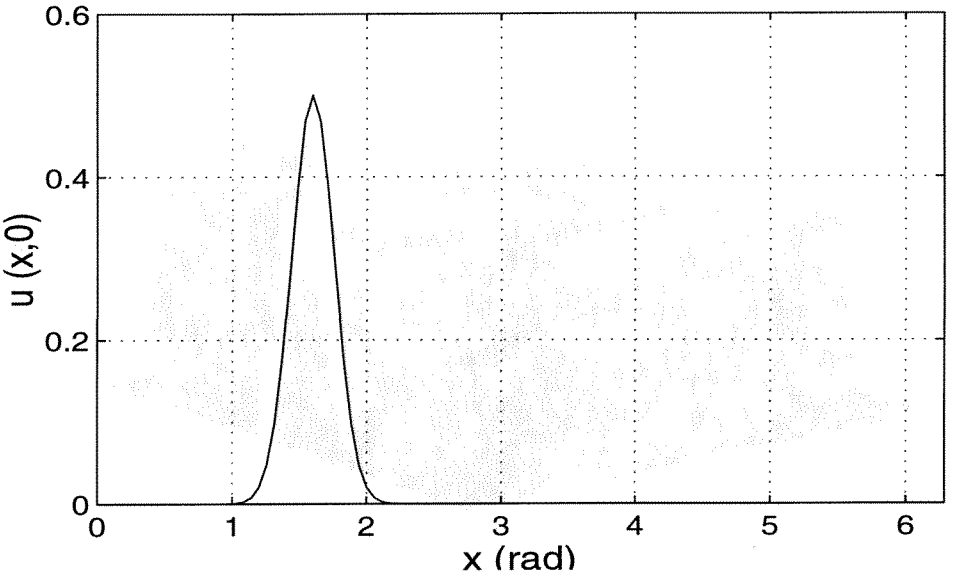

Figure 9.1: Initial condition for linear wave equation.

9.3 Optimality of the KLE in the Centered Space

We want to show that the KLE method provides an expansion of $u^c(x,t)$ which is optimal for modeling or reconstructing the original solution. The eigenvalues $\{\lambda_n\}$ corresponding to the eigenfunctions $\{\varphi_n(x)\}$ of the integral operator with kernel $\mathcal{R}_{v^c}(\cdot,\cdot)$, where

$$\mathcal{R}_{v^c}(x,y) = E[v^c(x,t)v^c(y,t)],$$

may be interpreted as the mean energy of $v^c(x,t)$ projected onto the φ_n axis in the centered function space. We have already defined mean energy projection in (7.2) as $E[|\langle v^c(x,t),\varphi_n(x)\rangle|^2]$. Then

$$\begin{aligned}
E[|\langle v^c(x,t),\varphi_i(x)\rangle|^2] &= E[\int_0^{2\pi} v^c(x,t)\varphi_i(x)dx \int_0^{2\pi} v^c(y,t)\varphi_i(y)dy] \\
&= E[\int_0^{2\pi}\int_0^{2\pi} v^c(x,t)\varphi_i(x)v^c(y,t)\varphi_i(y)dxdy] \\
&= \int_0^{2\pi}\int_0^{2\pi} \varphi_i(x)E[v^c(x,t)v^c(y,t)]\varphi_i(y)dxdy \\
&= \int_0^{2\pi} \varphi_i(x)\int_0^{2\pi} \mathcal{R}_{v^c}(x,y)\varphi_i(y)dydx
\end{aligned}$$

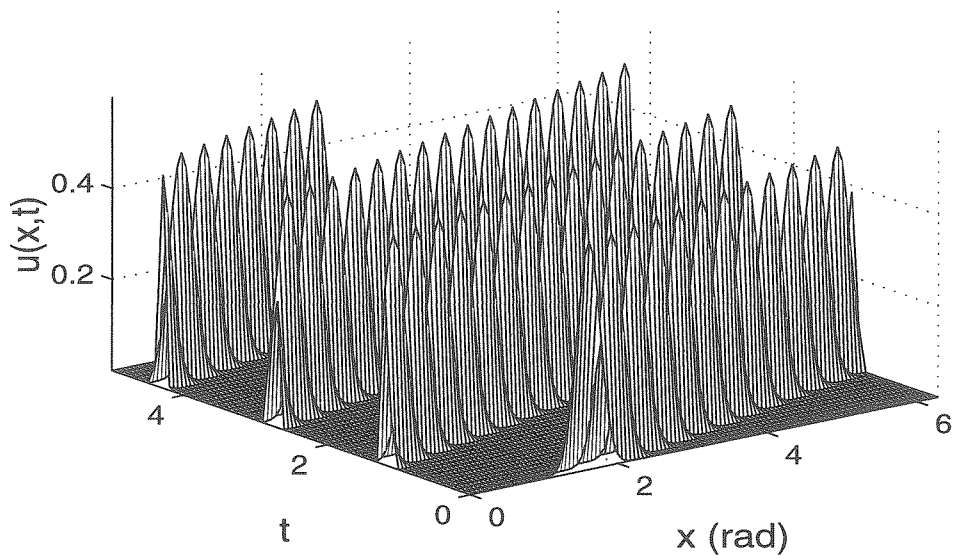

Figure 9.2: Data snapshots used for linear wave equation.

$$= \lambda_i \int_0^{2\pi} \varphi_i^2(x)dx$$
$$= \lambda_i.$$

Assuming that the eigenvalues $\{\lambda_n\}$ corresponding to $\{\varphi_n\}$ have been decreasingly ordered so that $\lambda_{i+1} > \lambda_i$ $(\forall i)$, it can be shown that if $\{\psi_n\}$ is an arbitrary set of orthonormal basis functions in which we expand $v^c(x,t)$ then for any value of N

$$\sum_{n=1}^{N} E[|\langle v^c(x,t), \varphi_n(x)\rangle|^2] = \sum_{n=1}^{N} \lambda_n$$
$$\geq \sum_{n=1}^{N} E[|\langle v^c(x,t), \psi_n(x)\rangle|^2].$$

Therefore, for a given number of modes N the projection on the subspace of a centered space spanned by $\{\varphi_n\}$ will contain the most average energy possible compared to all other linear decompositions.

9.4 A Reduced Order Model

Suppose we have a system governed by the PDE (9.1). The original attempt was to find an approximate solution in the form of a truncated series expansion

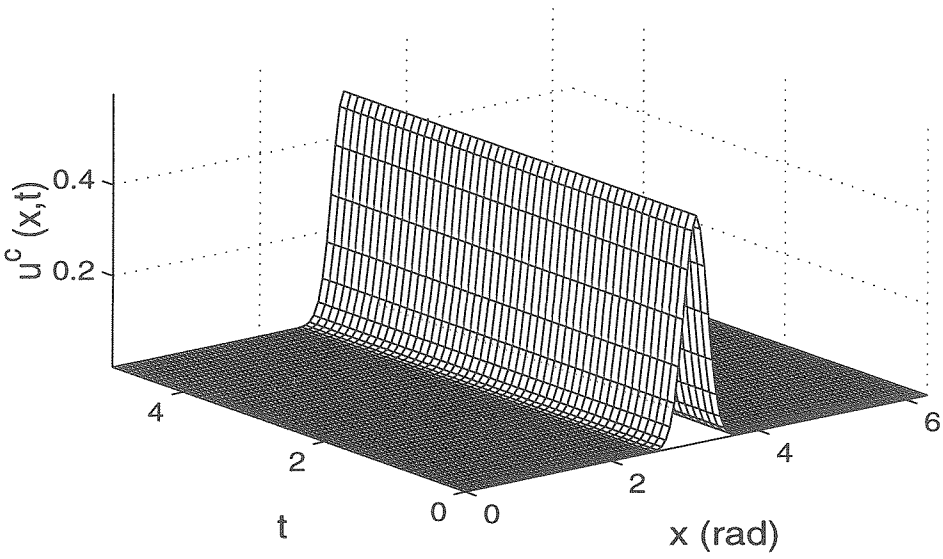

Figure 9.3: Centered data snapshots for linear wave equation.

given by

$$\hat{u}(x,t) = \sum_{n=1}^{N} a_n(t)\varphi_n(x + d(t)) + \bar{u}^c(x + d(t)), \tag{9.7}$$

where the $\varphi_n(x)$ are trial functions obtained after performing KLE on the centered data ensemble $\{(u^c)^{(k)}\}$. This way the original infinite dimensional system is again approximated by an N dimensional system.

To verify that the original PDE is satisfied as closely as possible by (9.7) we choose time dependent coefficients $a_n(t)$ so that the residual error produced by using (9.7) instead of the exact solution is minimized. At any time t we want the residual

$$r(t,x) = \frac{\partial \hat{u}(x,t)}{\partial t} + \omega \frac{\partial \hat{u}(x,t)}{\partial x} - D(\hat{u}(x,t)) \tag{9.8}$$

to be orthogonal to a chosen set of trial functions, i.e.,

$$\langle r(t,x), \varphi_i(x + d(t))\rangle = 0 \quad i = 1, \ldots, N.$$

Substituting (9.7) into (9.8) yields,

$$r(t,x) \;=\; \sum_{n=1}^{N} \dot{a}_n(t)\varphi_n(x + d(t)) +$$

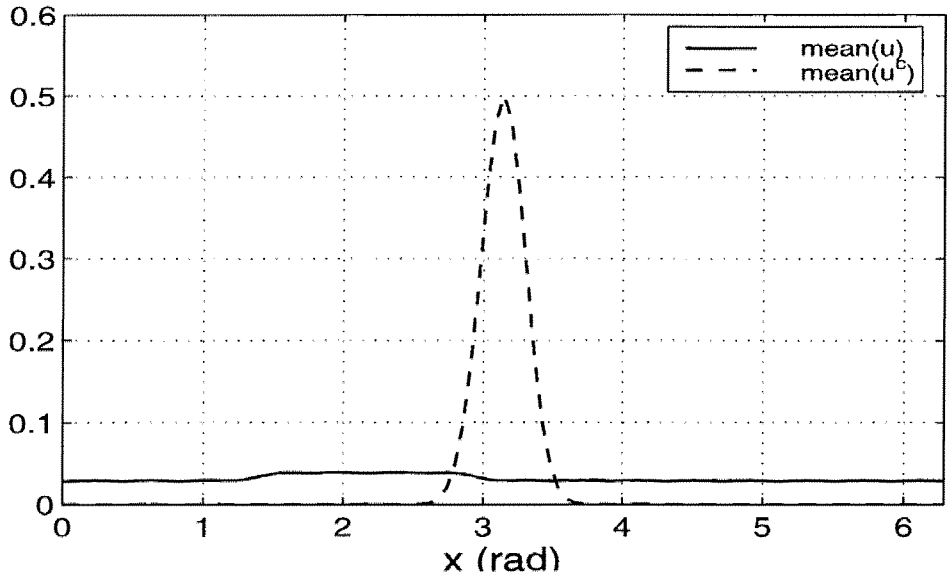

Figure 9.4: Mean data snapshots for linear wave equation.

$$
\begin{aligned}
&(\dot{d}+\omega)\bar{u}^{c\prime}(x+d(t))+\\
&(\dot{d}+\omega)\sum_{n=1}^{N}a_n(t)\varphi_n'(x+d(t))-\\
&D\left(\sum_{n=1}^{N}a_n(t)\varphi_n(x+d(t))+\bar{u}^c(x+d(t))\right).
\end{aligned}
$$

Applying the orthogonality condition and using the orthonormality property of the set of trial functions results in a reduced order model which is a system of N ordinary differential equations

$$
\begin{aligned}
\dot{a}_i(t) \;=\; & (\dot{d}+\omega)\sum_{n=1}^{N}a_n(t)\langle\varphi_i(x+d(t)),\varphi_n'(x+d(t))\rangle+\\
& (\dot{d}+\omega)\langle\bar{u}^{c\prime}(x+d(t)),\varphi_i(x+d(t))\rangle+\\
& \left\langle D\left(\sum_{n=1}^{N}a_n(t)\varphi_n(x+d(t))+\bar{u}^c(x+d(t))\right),\varphi_i(x+d(t))\right\rangle. \qquad (9.9)
\end{aligned}
$$

The initial conditions for the resulting system of ODEs are determined by a second application of the Galerkin approach. We force the residual of the initial conditions $r_0(x) = u(x,0) - \hat{u}(x,0)$ to be orthogonal to the first N basis functions and we obtain a system of N linear equations.

$$
a_i(0) = \langle u(0,x) - \bar{u}^c(x+d(0)),\varphi_i(x+d(0))\rangle. \qquad (9.10)
$$

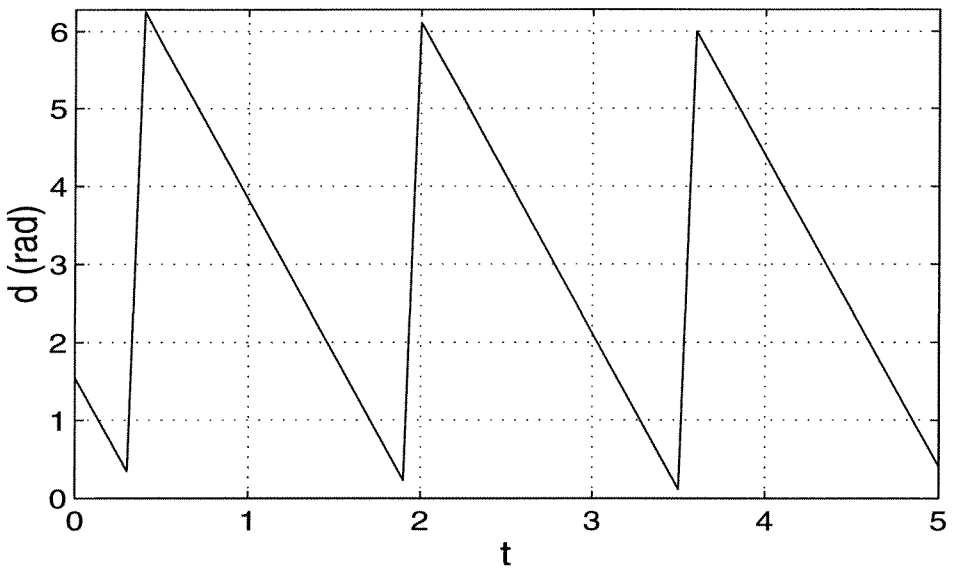

Figure 9.5: d(t) for linear wave equation.

Centering separates the evolution of the wave shape and movement of the wave. The system of ODEs (9.9) and (9.10) model the the evolution of the wave shape. We assume that propagation of the wave can be represented by the movement of its center, and a single ODE modeling movement of the wave center can be extracted from $d(t)$ obtained by centering. In the case when $d(t)$ depends linearly on time, and that holds for all the examples considered in this thesis, waves rotate with a constant speed and an ODE modeling $d(t)$ is

$$d(t) = d_0 + \dot{d}(t)t \tag{9.11}$$

where

$$\dot{d}(t) = -\omega.$$

Note that to solve the system of ODEs (9.9), (9.10), and (9.11) one needs to select the set of trial functions $\{\varphi_n\}$, the initial conditions of the original system $u(x, 0)$, and the initial condition for $d(t)$. We choose trial functions generated by the KLE performed on the "centered" data snapshots, and propose determining d_0 by centering the first data snapshot.

9.5 Computational Results

Nonlinear Wave Equation

To demonstrate the method's performance we simulated the following non dimensional PDE

$$\frac{\partial u(t,x)}{\partial t} = 4\frac{\partial u(t,x)}{\partial x} + u(t,x)^2 \tag{9.12}$$

and performed standard model reduction using Galerkin projection on KL modes and model reduction via centering using only the first two centered KL modes.

Figure 9.6 shows the data snapshots obtained from the PDE simulation, and Figure 9.7 shows these data snapshots centered.

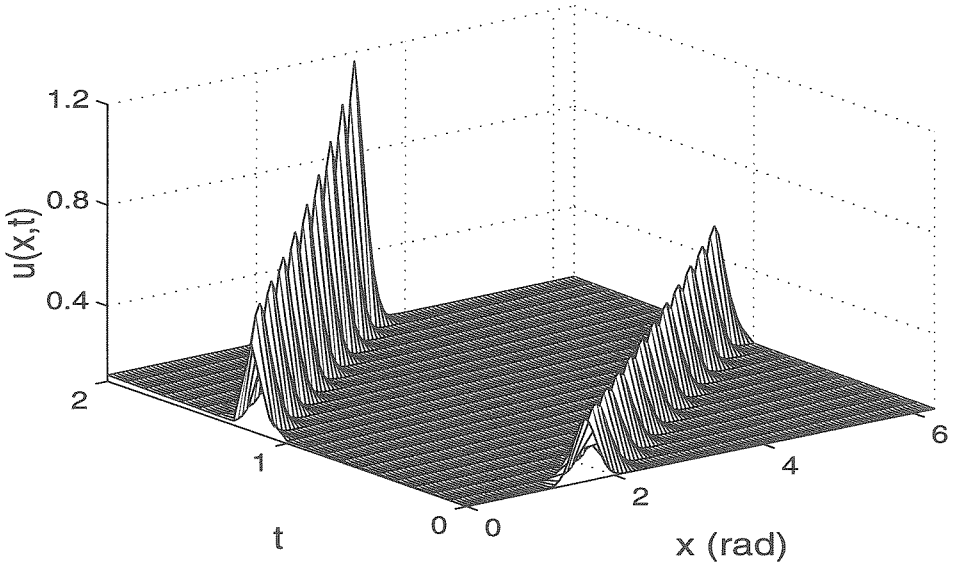

Figure 9.6: Used data snapshots for nonlinear wave equation

To see if we are obtaining more information from the available data by centering it, we compare the mean snapshot of the original data ensemble and of the centered data ensemble. They are shown in Figure 9.8. We can immediately observe that the wave shape is captured rather well in the shape of a mean snapshot of the centered data, whereas the mean snapshot of the original data ensemble does not contain much information. Figure 9.9 shows a comparison of the first two KL and centered KL (CKL) modes. From the corresponding eigenvalues we see that the first two centered modes contain

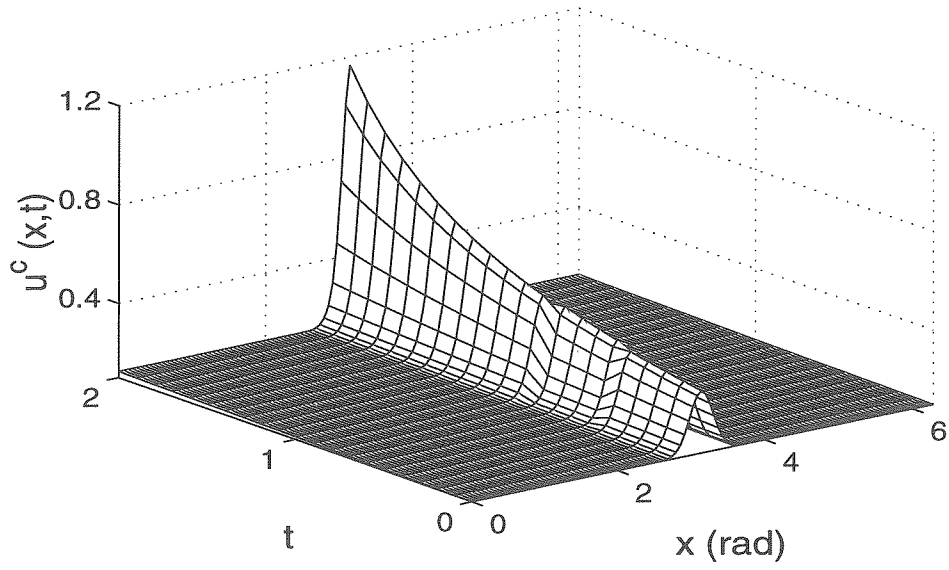

Figure 9.7: Centered data snapshots for nonlinear wave equation

more than 99% of the energy of the data ensemble. The first two KL modes contain just slightly more than 31% of the original data ensemble energy.

Figure 9.10 shows the evolution of the first two time dependent modal coefficients $a_n(t)$ of a truncated series expansion using centered KL modes.

Figure 9.11 shows one of the used data snapshots and its reconstructions $ru(x,t)$ using the first five KL modes and its reconstructions $ru^c(x,t)$ using only the first two centered KL modes. It is obvious that we are outperforming the classical method by the use of centering.

Centering introduces a significant improvement compared to the classical technique.

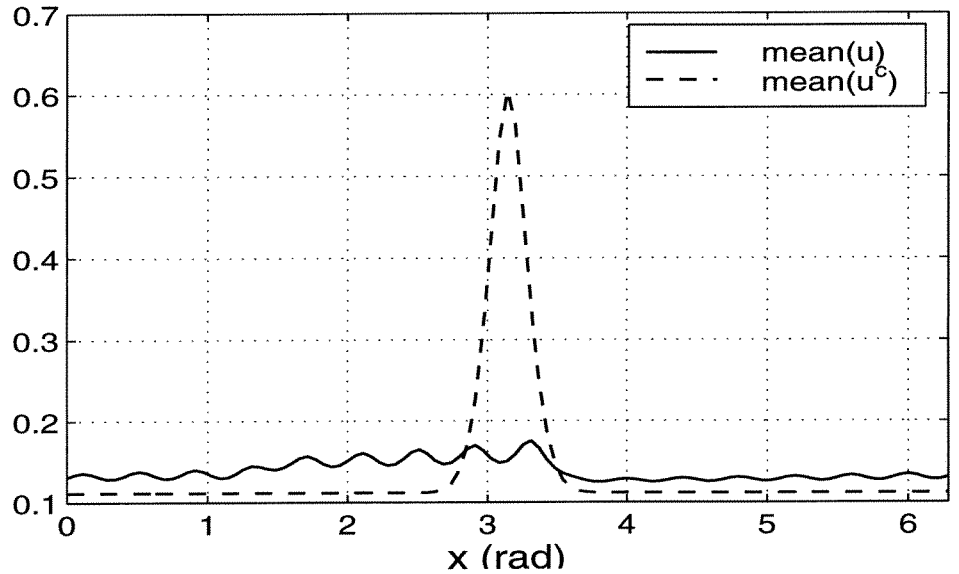

Figure 9.8: Mean data of original data snapshots and centered data snapshots

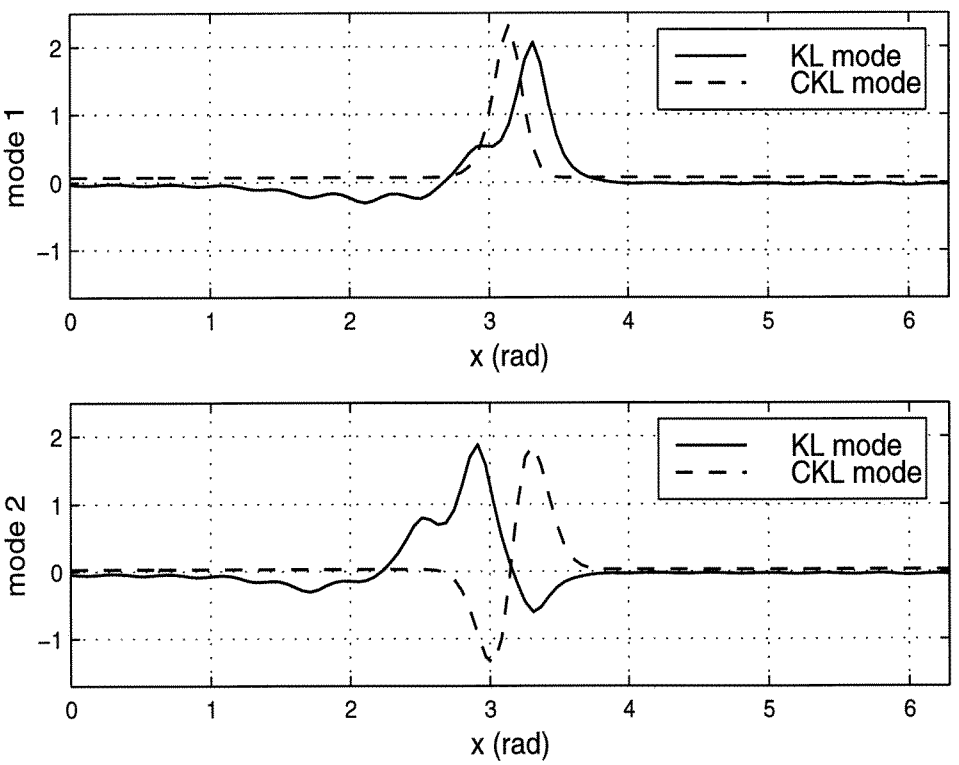

Figure 9.9: Percentage of data energy captured: first KL mode 17.21% and first CKL mode 96.79% ; second KL mode 13.91% and second CKL mode 2.86% .

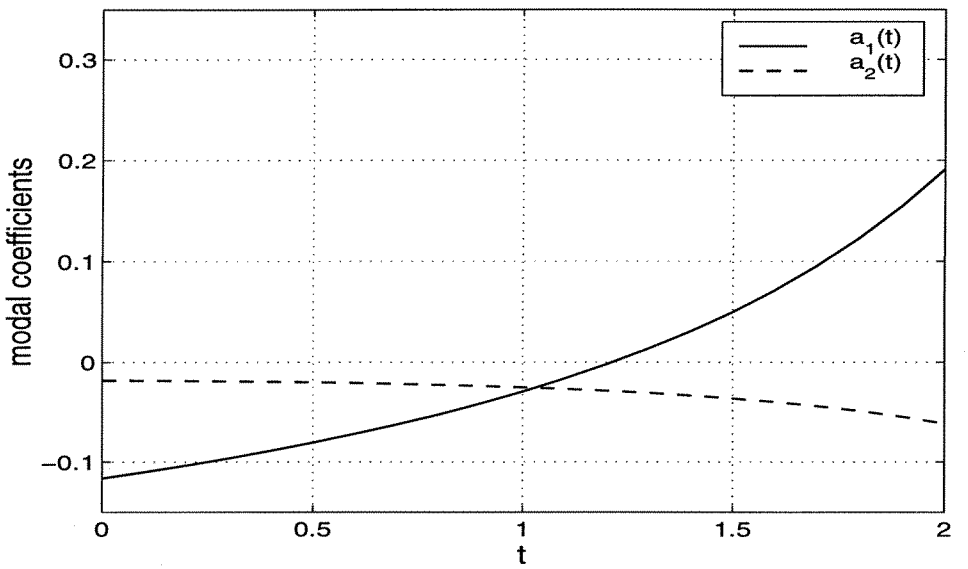

Figure 9.10: Evolution of time dependent coefficients.

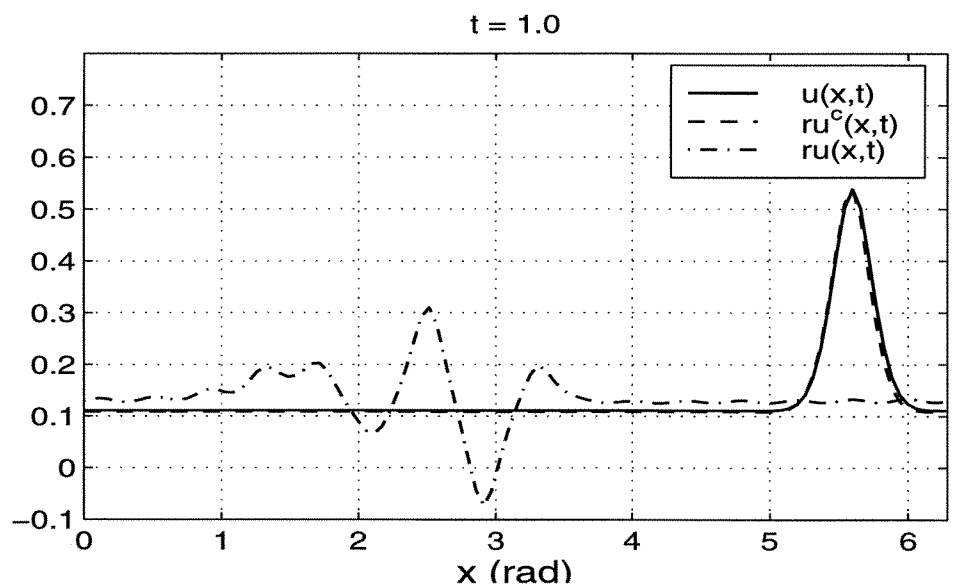

Figure 9.11: Original snapshot of u(t,x) and its reconstructions by 5 KL modes $ru(t,x)$ and by 2 centered KL modes $ru^c(t,x)$.

Deep stall cell phenomena

Deep stall cell is one of the stalled flow regimes in compression systems, and is essentially two dimensional. This was observed experimentally and summarized in the book by Pampreen [39]. In [29] a large–scale theory that predicts the crucial features of deep stall cell dynamics was introduced. The analysis, based on averaging of the incompressible Navier-Stokes equation, confirmed that in the deep stall cell regime the flow in the compressor away from the hub and casing is indeed two dimensional. It was hypothesized that the stall instability is a consequence of a competition between flow acceleration due to blade forces and moment redistribution due to velocity fluctuation.

The rotating stall is treated as a large-scale phenomenon, and is a feature of the average flow. The averaging volume is extended over rotor and stator rows and many blades, and the role of velocity fluctuation in deep stall instability is emphasized.

In this section we will apply model reduction by centering and KLE to the equation modeling the unsteady axial flow in the compression system introduced in [29]

$$\frac{\partial u}{\partial t} + \omega \frac{\partial u}{\partial x} = f(u) - \langle f(u) \rangle + \gamma \frac{\partial^2 u}{\partial x^2}, \tag{9.13}$$

where u is the axial velocity, t is time, x is angular variable, ω is the velocity of stall cell rotation, and the compression system characteristic function

$$f(u) = f_0 + H\left(1 + \frac{3}{2}\left(\frac{u}{W} - 1\right) - \frac{1}{2}\left(\frac{u}{W} - 1\right)^3\right), \tag{9.14}$$

is shown in Figure 9.12 and $\langle f(u) \rangle$ is the annulus average of the characteristic function

$$\langle f(u) \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(u) dx. \tag{9.15}$$

This is a non dimensional equation.

The model is not valid near the casing due to the no-slip condition. This is a reaction-diffusion type equation, with cubic nonlinearity. The reaction term $f(u) - \langle f(u) \rangle$ is caused by combined effects of pressure equalization at the plenum and blade forces. The diffusion term is caused by the inviscid process of turbulent momentum transport via Reynolds stresses.

The steady state, nonuniform solutions of the equation (9.13) are the stall cells that rotate around the annulus with the average velocity being one half of

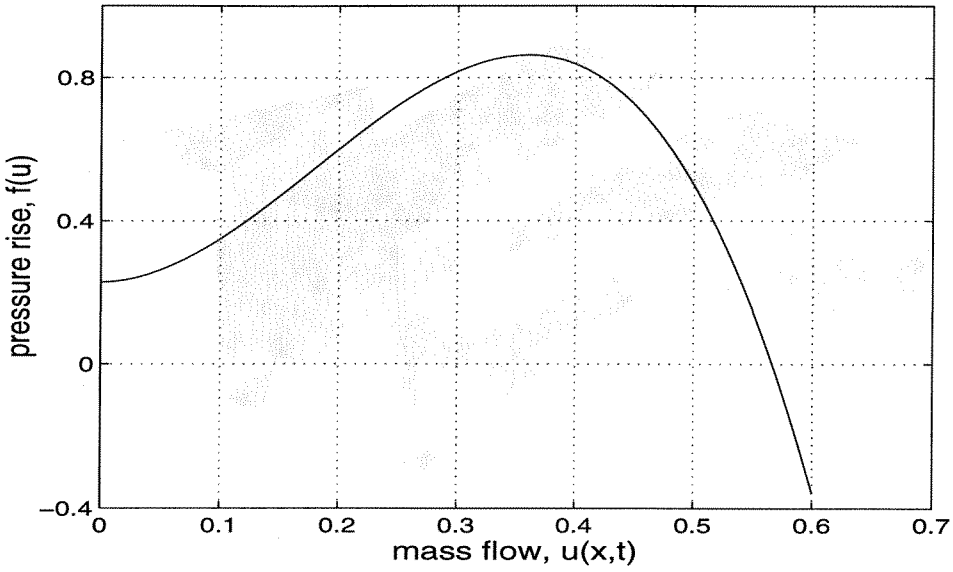

Figure 9.12: Characteristic function f used in simulation.

the rotor velocity. The numerical existence of such solutions has been shown in [29].

In this section we will show the results of simulation of the equation (9.13) carried out with $\omega = 0.5$, $f_0 = 0.23$, $H = 0.32$, $W = 0.18$, and $\gamma = 0.01$. The evolution of a small sinusoidal disturbance superimposed on the uniform flow for the mean flow $\Phi = 0.3$ is shown in the Figure 9.13, and this data centered is shown in the Figure 9.14.

Figure 9.15 shows a comparison of the mean snapshot of the original data ensemble and the mean snapshot of the centered data ensemble. The square wave shape appears immediately in the shape of a mean snapshot of the centered data ensemble, whereas the mean snapshot of the original data ensemble gives no helpful information about the shape of the rotating wave.

Figure 9.16 shows a comparison of the first two standard and centered KL modes. From the corresponding eigenvalues we see that the first two centered modes contain more than 96% of the energy of the data ensemble. The first two KL modes contain a bit more than 67% of the original data ensemble energy. Because the square wave develops rather quickly, most of the data snapshots are just rotated versions of a square shape, meaning that even though the system does not exhibit strict $SO(2)$ symmetry, the KL modes obtained are

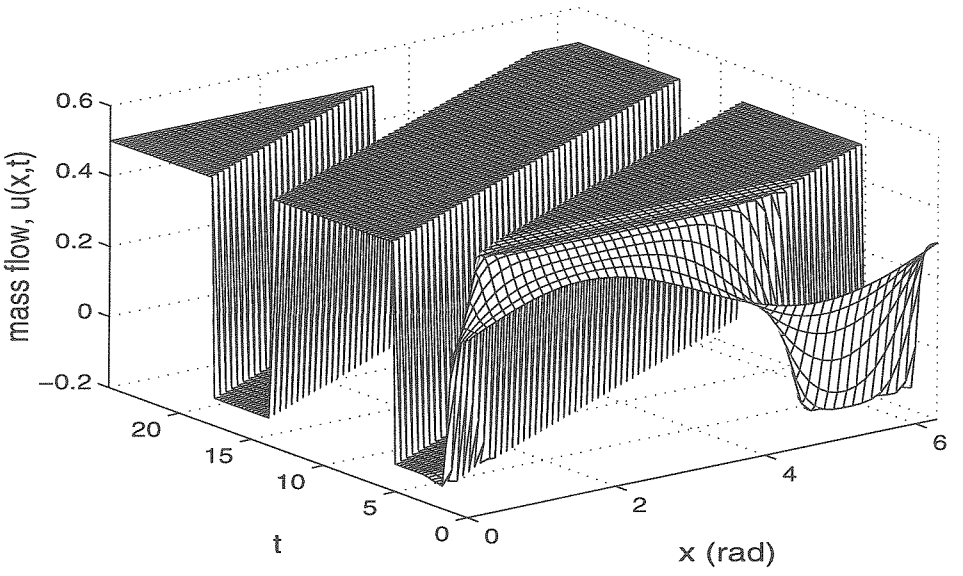

Figure 9.13: Stall cell evolution, $\Phi = 0.3$.

just Fourier modes.

Figure 9.17 shows the evolution of the first two time dependent coefficients $a_n(t)$ of a truncated series expansion using centered KL modes. Since we model rotating stall cell as a rotating mean square wave plus a sum of rotating modes, these coefficients present a deviation of the approximate solution of the PDE (9.13) around that mean square wave.

Figure 9.18 shows one of the used data snapshots and its reconstructions $ru(x,t)$ using the first five KL modes and its reconstructions $ru^c(x,t)$ using only the first two centered KL modes. It is clear that we are outperforming the classical method by the use of centering.

Once we extracted centered KL modes, and obtained time varying ODEs that model the deviation of the PDE solution around the mean square wave, we would like to justify our model. Thus, we simulate the original PDE using $u(x,0) = \sum_{i=1}^{2} a_i^0 \varphi_i(x + d_0)$ as an initial condition. In Figure 9.19 we show a new initial condition for the Mezić PDE, denoted $u^n(x,0)$, and an initial condition for the Mezić PDE that we used to calculate centered KL modes.

We project the results of the simulation onto previously extracted centered KL modes to obtain a set of ODEs modeling the wave shape evolution, and we simulate them. We also simulate our time varying ODE using a_1^0, a_2^0, and

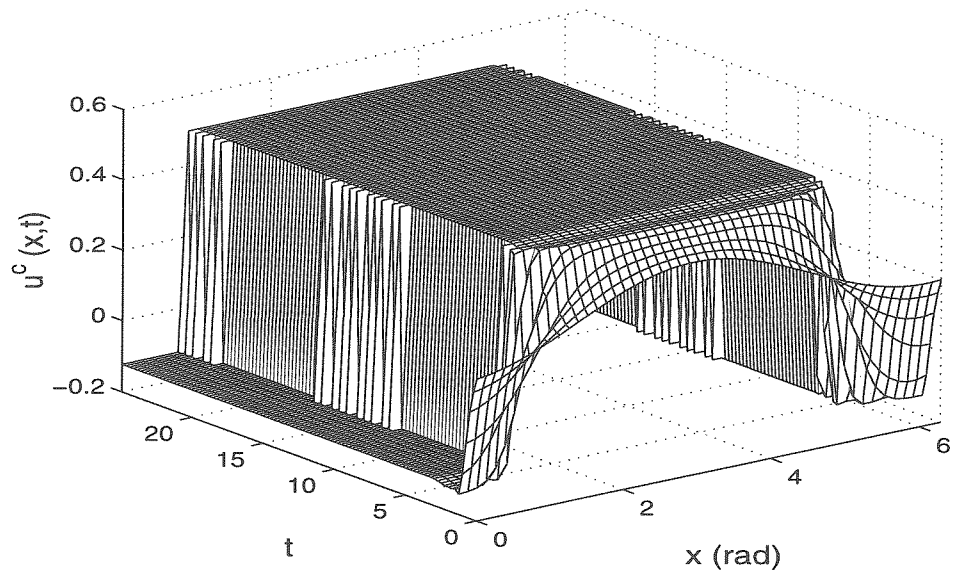

Figure 9.14: Centered stall cell evolution , $\Phi = 0.3$.

d_0 as initial conditions, and denote the computed modal coefficients as $a_i^o(t)$. In Figure 9.20 we compare the obtained modal coefficients. In Figure 9.21 we show approximate solutions when we use both sets of coefficients. The approximate solution using a_i^o's is denoted $ru^o(x,t)$.

Our reduced order model predicts system behavior rather well.

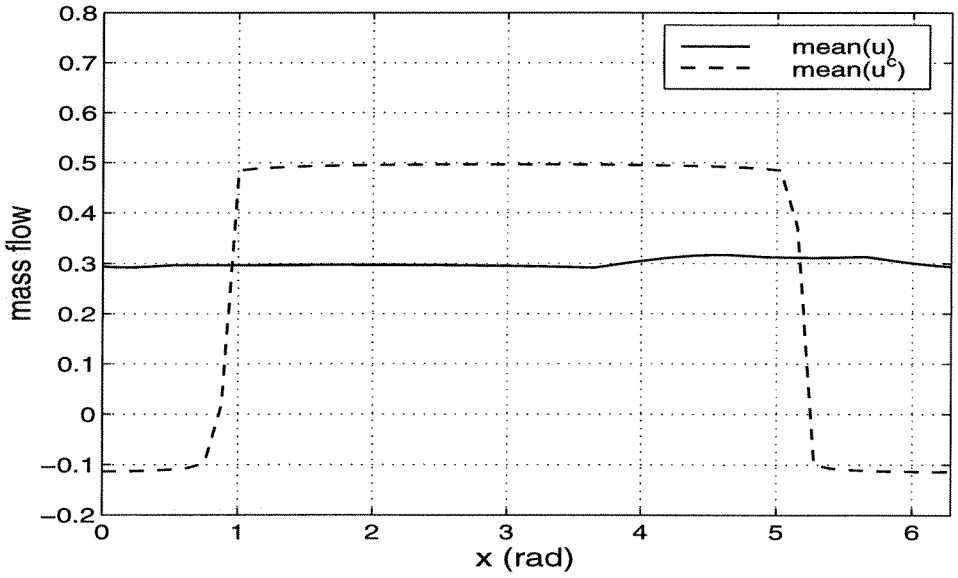

Figure 9.15: Mean data of original data snapshots and centered data snapshots

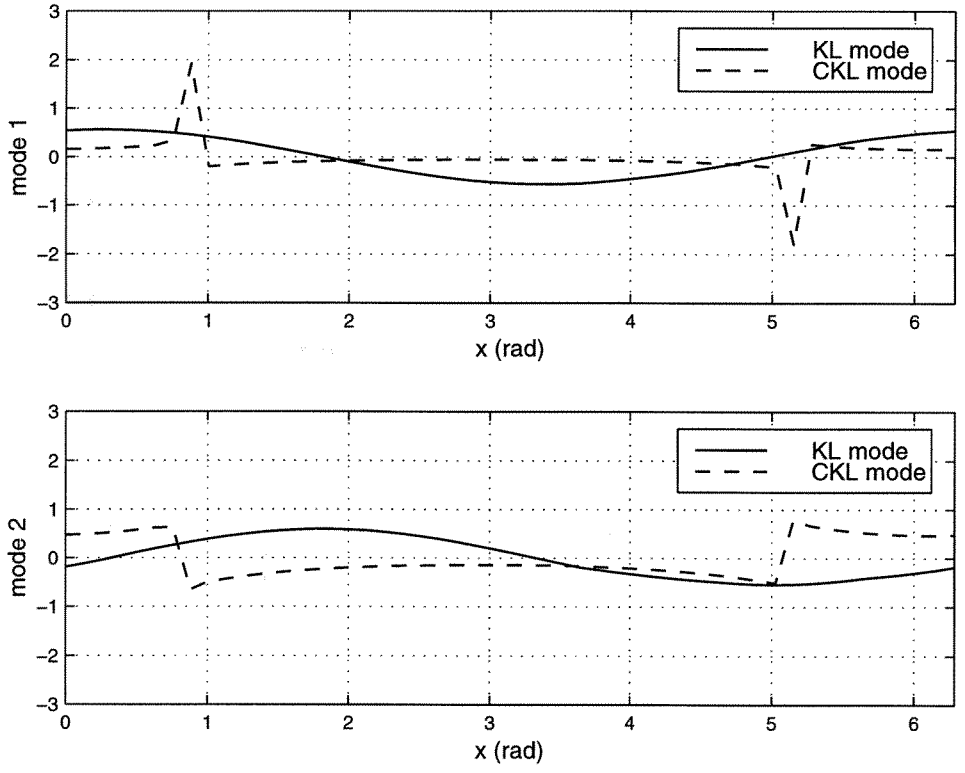

Figure 9.16: Percentage of data energy captured: first KL mode 34.66% and first centered KL mode 61.71% ; second KL mode 32.43% and second centered KL mode 34.96% .

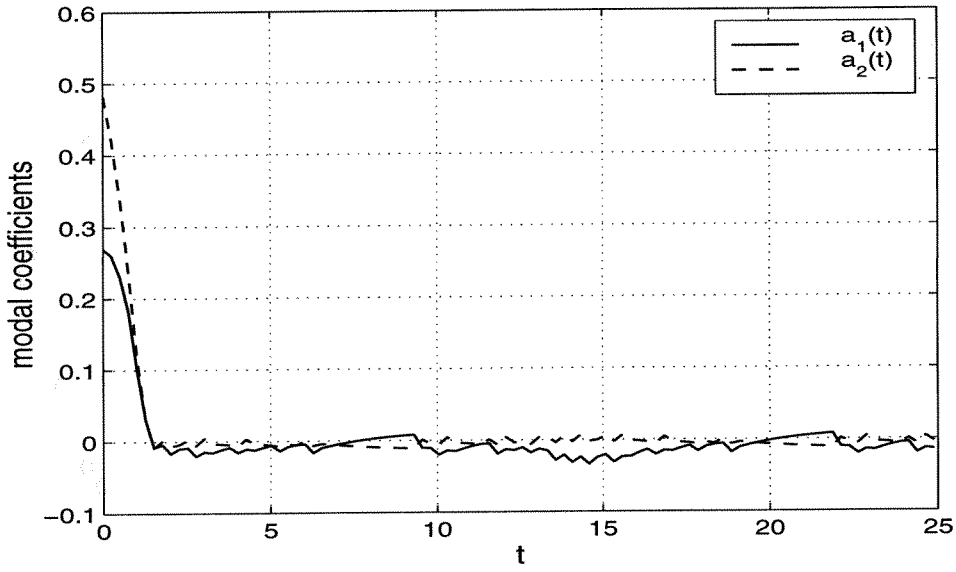

Figure 9.17: Evolution of time dependent coefficients.

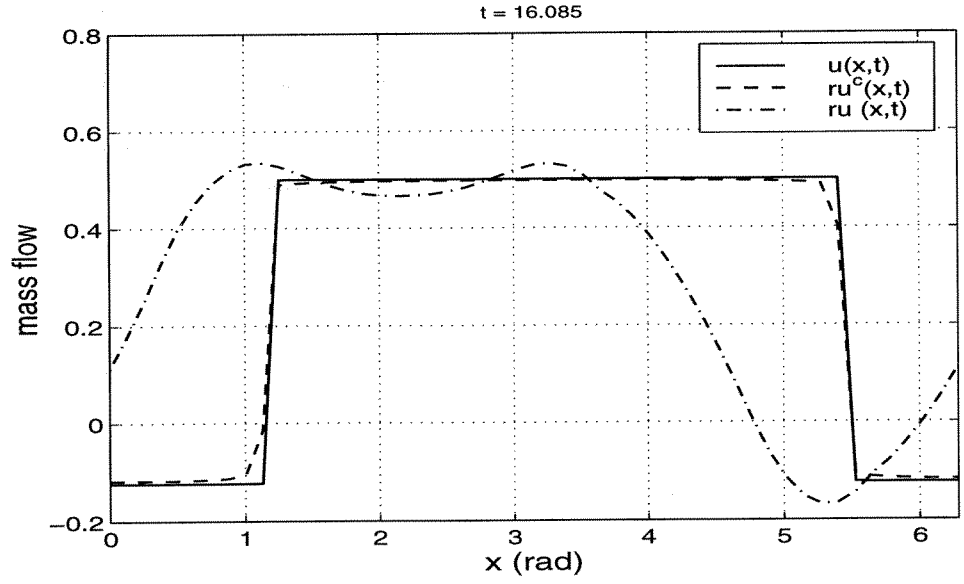

Figure 9.18: Original snapshot of u(t,x) and its reconstructions by 5 KL modes $ru(t,x)$ and by 2 centered KL modes $ru^c(t,x)$.

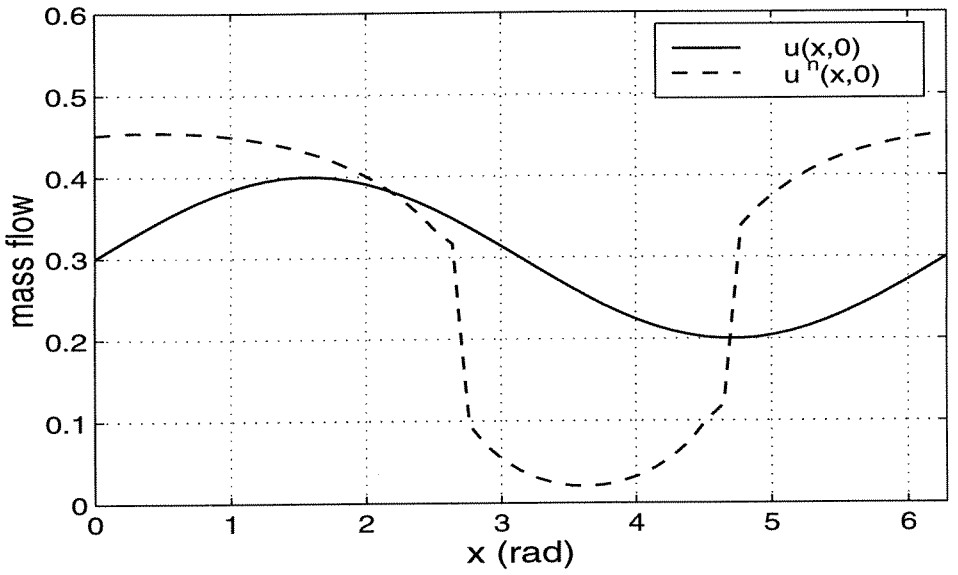

Figure 9.19: Initial conditions for Mezic PDE simulation.

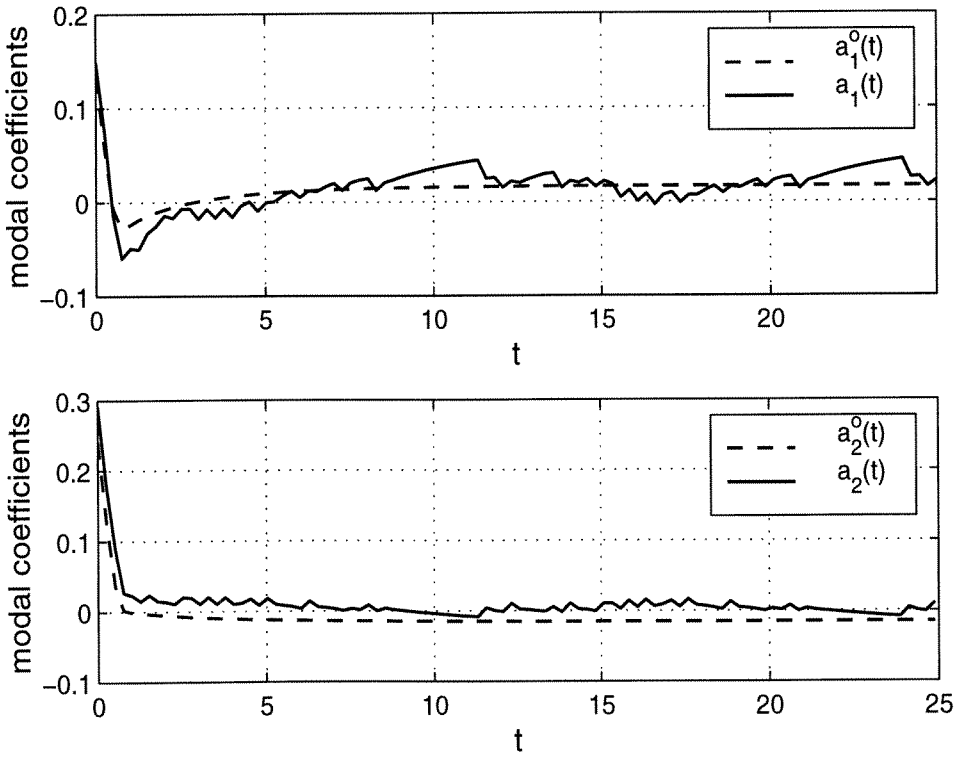

Figure 9.20: Modal coefficients comparison.

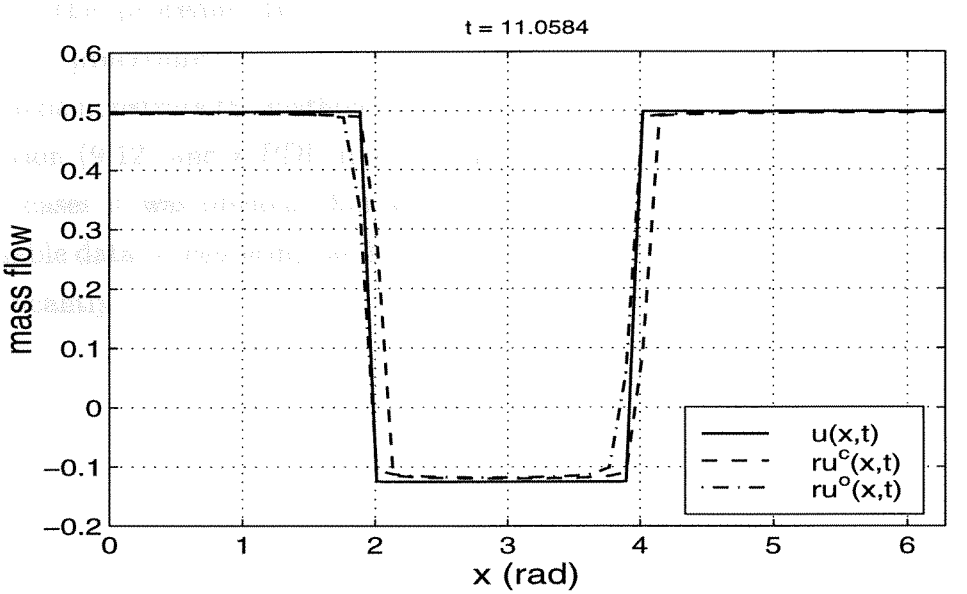

Figure 9.21: Model comparison.

9.6 Conclusion

In this chapter we considered systems governed by PDEs have a traveling (rotating) wave solution and we showed how to obtain as few as possible modes necessary to accurately approximate the shape of a propagating wave. We accomplished this by separating the movement of a solution $u(x,t)$ and the evolution of a wave shape.

In order to extract the propagating wave we position all data snapshots in such a way that their centers lie at the same point. For simplicity and without loss of generality we have chosen to place data snapshots centers at π, and we call this procedure the centering of a wave. It is performed using simple iterative procedure.

To demonstrate the method's performance we applied it both to a nonlinear equation (9.12) and a PDE modeling a deep stall cell phenomena (9.13). In both cases it was obvious that we are obtaining more information from the available data by centering, and that we reduce the order of the models needed significantly.

Chapter 10
Concluding Remarks

In this last chapter we conclude this thesis by summarizing the results presented and by suggesting future lines of work.

10.1 Summary

The development of numerical tools for robustness analysis of linear systems has been successful. In the last two decades these tools have been widely adopted and used for a variety of applications. The main motivation behind the first part of this thesis is to broaden the classes of systems and types of problems that these tools can be used for.

A great deal of interest has arisen with regard to robustness problems involving uncertain parameters that are not only norm bounded, but also constrained to be real. Several new approaches to computing an improved μ lower bound have been presented in Chapter 3. These algorithms have been combined to yield a substantially improved power algorithm.

In Chapter 4 we considered a class of uncertain systems subject to norm bounded structured LTI perturbations. We showed that the worst case $\mathcal{H}_\infty$ gain of a system can be written exactly in terms of the skewed structured singular value. Although, like μ, the skewed structured singular value can not be computed exactly, we discussed an efficient algorithm to compute corresponding upper and lower bounds. The results presented show that the enhanced algorithm developed recently for the structured singular value can be extended to the problem of computing worst case gains under fixed size uncertainty, without significant loss of performance or accuracy.

In chapter 5 we presented how the worst case $\mathcal{H}_2$ norm of an uncertain

system subject to norm bounded structured LTI perturbations can be written exactly in terms of the complex skewed structured singular value. Even though computation of the lower and upper bound for the worst case $\mathcal{H}_2$ performance implies numerical integration over frequency, which is the same as in a frequency by frequency evaluation of the structured singular value (the current practice for robustness analysis in industry), this new approach can be effortlessly integrated into the current robustness analysis tools.

In Chapter 6 we showed how a power algorithm can be used to compute a necessary condition for disturbance rejection of discrete and continuous time nonlinear systems by searching for solutions to Euler-Lagrange equations. The performance index used is different from the one considered in previous work. In the case where the system is linear we showed how the algorithm reduces to a well studied algorithm for the lower bound of μ and that the algorithm is guaranteed to converge to the global optimum. The worst case disturbances obtained by our proposed power algorithm are very close to the worst case disturbances a priori given by other methods, meaning that for the general case of a system with a non-optimal controller this algorithm can provide us with knowledge of the worst case disturbance.

In the second part of this thesis we explored different approaches to the model reduction of system. First, in Chapter 8, we compared the model reduction of a system based on Galerkin projection to the balanced truncation method by applying them to the same system. We have shown that in this particular case, when the system is driven by a linear PDE, we can think of balancing the reduced order model obtained by Galerkin projection and consisting of N ODEs as just changing a set of trial functions used for the truncated series expansion. Thus, the balancing transformation and Galerkin commute.

We demonstrated that when we have an orthogonal state space transformation matrix we can represent model reduction by balanced truncation as just obtaining the reduced order model by Galerkin projection onto a truncated set of trial functions corresponding to a transformed realization of a reduced order model. Only in this case do balanced truncation and Galerkin projection commute.

In Chapter 9 we pursued model reduction of nonlinear systems with rota-

tional symmetry by incorporating symmetry into a KLE based method. We showed how to obtain as few modes as necessary to accurately approximate the shape of a propagating wave. We accomplished this by separating the movement of the wave from the evolution of the wave shape.

In order to extract the propagating wave we position all data snapshots so that their centers are at the same point, and we call this procedure the centering of a wave. It is performed using a simple iterative procedure. To demonstrate the method's performance we applied it both to a nonlinear wave equation and the Mezić PDE which models a deep stall cell phenomena. In both cases we obtained more information from the available data by centering, and we reduced the order of the models needed significantly.

10.2 Future Research

In the first part of this thesis we presented extensions of the standard μ analysis methods to systems with real parametric uncertainties and norm bounded uncertainties. We considered both the problem of computing the worst case $\mathcal{H}_\infty$ norm and the worst case $\mathcal{H}_2$ norm of an uncertain system subject to norm bounded structured LTI perturbations.

First, we showed new approaches to computing an improved mixed μ lower bound. The algorithms described and compared here are by no means optimized, so it is reasonable to expect that this work will lead to even better results. Furthermore, based on this new algorithm we developed an efficient power algorithm for the skewed-μ lower bound. To understand bounds on its performance, and develop improvements we need to gain more experience from testing the algorithm on systems derived from real life engineering applications.

In chapter 5 we wrote the worst case $\mathcal{H}_2$ norm of an uncertain system subject to norm bounded structured LTI perturbations exactly in terms of the complex skewed structured singular value. This solution uses the true definition of the worst case $\mathcal{H}_2$ norm and both upper and lower bounds can be computed, and can be integrated into the current robustness analysis tools. Comparisons with other available methods need to be carried out to establish their relative merits. Still to be resolved is the issue of causality of the perturbations, and how much conservativeness is introduced by not imposing it.

It is also necessary to investigate under which circumstances the perturbation that achieves the worst case norm, as presented in that chapter, is causal.

We also showed how a power algorithm can be used to compute a necessary condition for disturbance rejection of both discrete and continuous time nonlinear systems. For the general case of a system with a non-optimal controller this algorithm can provide us with knowledge of the worst case disturbance. A deeper investigation of the numerical properties of our algorithm is needed and future research should concentrate on this.

As a first step in the direction of extending linear system model reduction techniques to nonlinear systems, the work in the second part of this thesis poses a set of new questions to be answered.

We demonstrated in Chapter 9 that only when we have an orthogonal balancing state space transformation matrix, balanced truncation and Galerkin projection commute. In general, balancing is not an orthogonal transformation, and balanced truncation and Galerkin projection do not commute. Some recent results [27] show that acknowledging this fact might lead to advancement in the research of model reduction for nonlinear system.

In Chapter 9 we pursued a different approach to nonlinear system model reduction, that incorporates an existing symmetry of the system into a KLE based method. First, we considered model reduction for a class of nonlinear systems with rotational symmetry. By separation of the movement of a rotating wave from the evolution of the wave shape we improved the extraction of the shape of the propagating wave, and significantly reduced the order of a model. The search for a generalization of this approach to higher dimensional systems is going to constitute the main thrust in this area of research. Procedures separating complex movements from the shape evolution, equivalent to centering in one-dimensional space, will be sought. This will allow us to extend the scope of model reduction even further.

Bibliography

[1] U. M. Ascher, R. M. Mattheij, and R. D. Russel. *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations.* Prentice Hall, 1988.

[2] S. P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* SIAM, 1994.

[3] R. D. Braatz, P. M. Young, J. C. Doyle, and M. Morari. Computational complexity of μ calculation. *IEEE Transactions on Automatic Control,* 39:1000–1002, 1994.

[4] A. E. Bryson and Y. C. Ho. *Applied Optimal Control.* Hemisphere Publishing, 1975.

[5] J. Chen, M. K. H. Fan, and C. N. Nett. The structured singular value and stability of uncertain polynomials: A missing link. *Control of Systems with Inexact Dynamic Models, ASME,* pages 15–23, 1991.

[6] J. Demmel. The componentwise distance to the nearest singular matrix. *SIAM Journal on Matrix Analysis and Applications,* 13:10–19, 1992.

[7] J. Doyle, J. Primbs, B. Shapiro, and V. Nevistić. Nonlinear games: examples and counterexamples. In *Proceedings of the 35th Conference on Decision and Control,* pages 3915–3919. IEEE, 1996.

[8] J. C. Doyle. Analysis of feedback systems with structured uncertainty. *IEE Proceedings, Part D,* 129(6):242–250, November 1982.

[9] J. C. Doyle, K. Glover, P. Khargonekar, and B. Francis. State space solutions to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems. *IEEE Transactions on Automatic Control,* 34(8):831–847, August 1989.

[10] D. F. Enns. Model reduction with balanced realizations: An error bound and a frequency weighted generalization. In *Proceedings of the 23rd Conference on Decision and Control*, pages 127–132. IEEE, 1984.

[11] F. K. Moore, E. M. Greitzer. A theory of post-stall transients in axial compression systems – Part I: Development of equations. *Journal of Engineering for Gas Turbines and Power*, 108:68–76, 1986.

[12] M. K. H. Fan and A. L. Tits. A measure of worst-case H_∞ performance and of largest acceptable uncertainty. *Systems and Control Letters*, 18(6):409–421, 1992.

[13] M. K. H. Fan, A. L. Tits, and J. C. Doyle. Robustness in the presence of joint parametric uncertainty and unmodeled dynamics. In *Proceedings of the American Control Conference*, pages 1195–1200, 1988.

[14] M. K. H. Fan, A. L. Tits, and J. C. Doyle. Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics. *IEEE Transactions on Automatic Control*, AC-36:25–38, 1991.

[15] E. Feron. Analysis of robust H_2 performance using multiplier theory. *Control and Optimization*, 35(1):160–177, January 1997.

[16] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP Completeness.* W. H. Freeman, New York, 1979.

[17] S. Glavaški, J. E. Marsden, and R. M. Murray. Model Reduction, Centering, and the Karhunen-Loeve Expansion. In *Proceedings of the 37th Conference on Decision and Control*, Tampa, Florida, 1998. IEEE. Submitted.

[18] S. Glavaški and J. E. Tierno. Robustness analysis: Nonlinear and growing. In *Proceedings of the 35th Conference on Decision and Control*, pages 3921–3925. IEEE, 1996.

[19] S. Glavaški and J. E. Tierno. Advances in Worst-case $\mathcal{H}_\infty$ Performance Computation. In *Proceedings of the IEEE International Conference on Control Applications*, Trieste, Italy, 1998. Submitted.

[20] K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their L_∞-error bounds. *International Journal of Control*, 39(6):1115–1193, 1984.

[21] D. Gottlieb and S. A. Orszag. *Numerical Analysis of Spectral Methods: Theory and Application.* SIAM, Philadelphia, Pennsylvania, 1977.

[22] P. Holmes, J. L. Lumley, and G. Berkooz. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry.* Cambridge University Press, 1996.

[23] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, 24:417–441, 1933.

[24] K. Karhunen. *Über* linear Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fenn.*, 37, 1947.

[25] H. B. Keller. *Numerical Methods for Two Point Boundary-Value Problems.* Blaisdell, 1968.

[26] L. Sirovich. Turbulence and the dynamics of coherent structures, Parts I-III. *Quarterly of Applied Mathematics*, XLV(3):561–582, 1987.

[27] S. Lall. Balancing and Karhunen-Loeve Expansion. Personal Communication, 1998.

[28] J. L. Lumley. *Atmospheric Turbulence and Wave Propagation.* Nauka, Moscow, 1967.

[29] I. Mezić. A large-scale theory of axial compression system dynamics. In preparation, 1998.

[30] B. C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, 1981.

[31] M. Morari and E. Zafiriou. *Robust Process Control.* Prentice Hall, New Jersey, 1989.

[32] M. P. Newlin and S. Glavaški. Advances in the computation of the μ lower bound. In *Proceedings of the American Control Conference*, pages 442–446, 1995.

[33] M. P. Newlin and P. M. Young. Mixed μ problem and branch and bound technique. *International Journal of Robust and Nonlinear Control*, 5(6):573–590, 1995.

[34] A. J. Newman. Model reduction via Karhunen-Loeve expansion, Part I: An exposition. Technical report, T.R. 96-32, Institute for Systems Research, Department of Electrical Engineering, University of Maryland, 1996.

[35] A. J. Newman. Model reduction via Karhunen-Loeve expansion, Part II: Some Elementary Examples. Technical report, T.R. 96-33, Institute for Systems Research, Department of Electrical Engineering, University of Maryland, 1996.

[36] A. K. Packard and J. C. Doyle. The complex structured singular value. *Automatica*, 29:71–109, 1993.

[37] F. Paganini. Necessary and sufficient conditions for robust H_2 performance. In *Proceedings of the 34th CDC*, pages 1970–1975, New Orleans, LA, 1995.

[38] F. Paganini and E. Feron. Analysis of robust $\mathcal{H}_2$ performance: Comparisons and examples. In *Proceedings of the 36th Conference on Decision and Control*, Piscataway, NJ, 1997. IEEE.

[39] R. C. Pampreen. *Compressor Surge and Stall.* Concepts ETI, Norwich, Vermont, 1993.

[40] L. Parnebo and L. M. Silverman. Model reduction via balanced state space representation. *IEEE Transactions on Automatic Control*, 27(2):382–387, 1982.

[41] J. E. Tierno. *A Computational Approach to Nonlinear System Analysis.* PhD thesis, California Institute of Technology, 1996.

[42] J. E. Tierno and S. Glavaški. Practical upper and lower bounds for robust $\mathcal{H}_2$ performance under LTI perturbations. In *AIAA Guidance, Navigation, and Control Conference*, Boston, MA, 1998. (Submitted).

[43] J. E. Tierno, R. M. Murray, J.C. Doyle, and I. M. Gregory. Numerically efficient robustness analysis of trajectory tracking for nonlinear systems. *AIAA Journal of Guidance, Control, and Dynamics*, 20(4):640–647, 1997.

[44] J. E. Tierno and P. M. Young. An improved μ lower bound via adaptive power iteration. In *Proceedings of the 31st Conference on Decision and Control*, pages 3181–3186, 1992.

[45] E. Wong. *Stochastic Processes in Information and Dynamical Systems.* McGraw-Hill, 1971.

[46] P. M. Young. The rank one mixed μ problem and "Kharitonov-type" methods. In *Proceedings of the 32nd Conference on Decision and Control*, pages 523–528, 1993.

[47] P. M. Young. *Robustness with Parametric and Dynamic Uncertainty.* PhD thesis, California Institute of Technology, 1993.

[48] P. M. Young and J. C. Doyle. Computation of μ with real and complex uncertainties. In *Proceedings of the 29th Conference on Decision and Control*, pages 1230–1235. IEEE, 1990.

[49] P. M. Young and M. P. Newlin. Algorithms for practical computation of skewed μ. Personal Communication.

[50] P. M. Young, M. P. Newlin, and J. C. Doyle. Practical computation of the mixed μ problem. In *Proceedings of the American Control Conference*, pages 2190–2194, 1992.

[51] G. Zames. On the input-output stability of nonlinear time-varying feedback systems, parts i and ii. *IEEE Transactions on Automatic Control*, 11:228–238 and 465–476, 1966.

[52] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control.* Prentice Hall, 1996.