Contributions to Model Approximation

Hardy Bonatua Siahaan

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy

Department of Engineering Cybernetics
Norwegian University of Science and Technology

2007

Abstract

This dissertation focuses on the approximation problem of models in the form of linear operators and in the form of polynomial dynamical systems. For approximation of linear operators, Schmidt and Mirsky have shown the existence of an optimal approximant which minimizes the induced Euclidean norm distance between the original operator and all possible lower rank approximant. This result is regarded as an important step in the development of model approximation for dynamical systems. In this thesis, a possibility of extending the result of Schmidt and Mirsky to a general induced norm is discussed. For approximation of dynamical systems, three computational schemes are introduced for several classes of polynomial nonlinear systems. The main contribution of this thesis lies on these three schemes.

The first computational scheme is heuristic in nature. The second one is derived based on a reachability approach. These two schemes are mainly to compute a reduced order model for a certain class of polynomial nonlinear systems such that the error model is finite gain $\mathcal{L}_2$ stable. The third scheme is an approach to generalize the balanced truncation method of linear systems to a class of polynomial nonlinear systems. The three schemes utilize the power of sum of squares programming which is amenable to computer solution.

Acknowledgments

I would like to take the opportunity to express my gratitude to Bjarne A. Foss and Ole M. Aamo for supervising me with such a stimulating atmosphere. I am also very grateful for their constant support and encouragement and for giving me the freedom to pursue my own research. I thank Torkel Glad from Linköping University, Anton Shiriaev from Umeå University and Lars Imsland from SINTEF for being the members of my PhD committee.

It is also a good opportunity to thank Sigurd Skogestad who directed me to apply for the PhD position at the department. I am indebted to Bjarne and Sigurd and to John-Morten Godhavn from Statoil for allowing me to join the PhD program in February 2003. I have also benefited from our discussions on slug flows. I want to acknowledge the support of my research from the Norwegian Research Council, Statoil and the Gas Technology Center NTNU.

My gratitude extends to Kristin Y. Pettersen for giving me the chance to practise giving lectures for a full semester course at the department; a very enjoyable experience interacting with the students participating in the course. I owe to Siep Weiland from TU Eindhoven and Anton Stoorvogel from University of Twente for our research cooperation in the Netherlands. I would also like to thank John Doyle for letting me stay for six months in 2005 at CDS-Caltech and to the members of CDS, especially Stephen Prajna and Naoki Yamamoto for all the helps during my stay at Caltech. I am also grateful to IRIS at Bergen, the research institute where I am working now, which gives me a stimulating and encouraging environment to finish writing this thesis.

My appreciations also go to all the staffs of the Department of Engineering Cybernetics; to my fellow PhD students for all discussions, having lunch together, coffee breaks and social activities; to the technical staffs for the helps with all kind of computer and lab problems; to the administrative staffs for all the helps with paperworks; thank you all. Special thanks go to

the Anglican community in Trondheim for the constant support throughout my years in Trondheim and to the Indonesian community in Trondheim, especially the members of Indonesian clubs such as MCC, KT and PPI; you are all that makes me feel like home during my stay in Trondheim.

Finally, I would like to thank my family for their support and unconditional love; to my wife Lasma, my daughter Abigail and my son Jonah for their patience and understanding; to my parents, sisters and brother for their encouragement.

Bergen, December 2007.

Contents

Chapter 1

Introduction

1.1 Background

Model approximation plays an important role in various applications where reducing complexity of a model is needed. This can be formulated differently depending on

- measure of complexity
- measure of accuracy

with the goal of finding a less complex model while keeping the accuracy of the less complex model as close as possible to the original one. There are many ways to quantify complexity of a model. For example, number of grids in power distribution networks, simulation time in computer systems, number of states in differential equations, memory capacity in circuits and number of equations in process optimization can be regarded as measures of complexity. As with accuracy of the less complex model, we want to keep some properties of the original model. For example, small error between a certain variable in the original model and the less complex model for any admissible condition can be a measure of accuracy. Additionally, we may want the less complex model to preserve certain structure from the original system.

This thesis considers models which are represented by input-output relation with intrinsic information given by state space representation. Here, complexity of the models is measured by the number of states. As the number

of states for a particular model is usually increasing in parallel with accuracy, we often encounter that we are only interested in some properties of input-output behavior rather than all information of the states. In this case reducing the number of states while preserving certain properties is essential because of a number of reasons, such as less complex model is easier to analyze and synthesize and is cheaper from computational perspective.

An important step in the development of model approximation in this direction can be traced back to the well known result by Schmidt and Mirsky [3]. It was shown that the minimum misfit between a matrix operator and all possible lower rank approximators in induced Euclidean norm is exactly equal to one of the singular values of the matrix. In this case there exists a lower rank optimal approximator to the given matrix which is viewed as a linear operator.

Inspired by this, Adamjan, Arov and Krein generalized the idea to linear dynamical systems [2]. A linear system induces an operator, Hankel operator, as a map from past inputs to future outputs. This operator is compact and has finite rank. Similarly like the result by Schmidt and Mirsky for linear operator, Adamjan, Arov and Krein have shown that the minimum misfit between a Hankel operator and all possible lower rank approximators in Hankel norm is exactly equal to one of Hankel singular values of the operator. The existence of a lower rank optimal approximator is also guaranteed in this case. Later on Glover shows a systematic way of computing the optimal Hankel approximator [11].

Moore introduced a balanced truncation scheme as another means of model reduction of linear dynamical systems [25]. Moore has shown that combination of reachability and observability is needed to obtain a reduced order model for a linear system. This can be achieved by computing a similarity transformation such that the transformed system has a balanced representation of the reachability and observability gramians. The reduced order model is obtained by truncating the transformed system to remove the least reachable and observable part simultaneously. The result from this scheme can be shown to be related with the result of Adamjan, Arov and Krein. Subsequently many schemes for model reduction of linear system were introduced, c.f. [13], [17], [21], [24], [26], [27], [28], [45].

While model reduction in linear systems is more or less well developed, model reduction for nonlinear systems lacks efficient strategies in its development and implementation. Scherpen [38] has introduced balancing method for a class of nonlinear systems where nonlinear version of the gramians need to

be balanced. The drawback of this method is on the computation of the gramians which is very difficult in general. Another approach which relies heavily on snapshots of data is given by the proper orthogonal decomposition (POD) method where a subspace generated by the data is constructed and a reduced order model is obtained by projecting the original model to the subspace [34]. Within this direction the authors in [20] have introduced an empirical approach for truncating nonlinear systems. A systematic scheme for constructing a reduced order model for polynomial nonlinear systems has been introduced in [33]. The method is easy to implement computationally due to sum of squares programming [30] and shows a promising direction in efficient computation for model reduction of polynomial nonlinear systems.

The fact that there are only a few results for nonlinear systems is due to several reasons such as the difficulty of the problem and computational aspects. Therefore a strategy which is easy to be implemented computationally will give benefit to the community. The results in this dissertation are mainly to serve this purpose. In particular, the power of sum of squares programming for model reduction of polynomial nonlinear systems is exploited.

1.2 Contributions

This dissertation contributes toward the understanding of approximation problems for models in the form of linear operators and dynamical systems. For model approximation of linear operators, a possibility of extending the result of Schmidt and Mirsky to a general induced norm is shown. For model approximation of dynamical systems which is the focus of this thesis, the emphasis is on developing computational schemes to get a reduced order model for continuous-time polynomial nonlinear systems. The schemes employ sum of squares (SOS) programming which is the LMI (linear matrix inequality) version of polynomial inequality. In particular, $\mathcal{L}_2$-gain model reduction and approximate balanced truncation for a certain class of polynomial nonlinear systems are presented. These results are generalization of the $\mathcal{H}_\infty$-model reduction and balanced truncation for linear systems, respectively. The table below shows indication of how the methods for polynomial nonlinear systems in this thesis are compared to those for linear systems.

Linear Systems (LMI)	Polynomial Nonlinear Systems (SOS)
Balanced Truncation	Approximate Balanced Truncation
$\mathcal{H}_\infty$	$\mathcal{L}_2$-gain

In summary, the contributions of this dissertation are as follows.

1.2.1 Main Contribution

Optimal Approximation of Linear Operators from a Singular Value Decomposition Approach

The purpose of this part is to propose a definition of a set of singular values and a singular value decomposition associated with a linear operator defined on arbitrary normed linear spaces. This generalizes the usual notion of singular values and singular value decompositions to operators defined on spaces equipped with the p-norm, where p is arbitrary. Basic properties of these generalized singular values are derived and the problem of optimal rank approximation of linear operators is investigated in this context. We give sufficient conditions for the existence of optimal rank approximants in the p-induced norm and discuss an application of this concept for the identification of dynamical systems from data.

$\mathcal{L}_2$ Gain Approximation of Nonlinear Systems: a Heuristic Approach

This part considers a computational mechanism to approximate a restricted class of polynomial nonlinear systems with reduced order models. First a necessary condition for a nonlinear system to have a reduced order model is given. Then a heuristic approach which utilizes the necessary condition is established. The method computes a reduced order model for the nonlinear system such that the error model is finite gain $\mathcal{L}_2$ stable.

Reachability-Based Approach for $\mathcal{L}_2$ Gain Approximation of Polynomial Systems

This part considers another computational mechanism to approximate a class of polynomial nonlinear systems with reduced order models. The approach is based on estimate of the reachability set and a finite gain $\mathcal{L}_2$ stability condition. The approach benefits from the use of sum of squares programming where the computation is rendered tractable.

Approximate Balanced Truncation of Polynomial Nonlinear Systems

The approach of model reduction for polynomial nonlinear systems in this part is based on balancing generalized gramians of the system and truncate the system based on the balanced generalized gramians. The approach utilizes sum of squares programming for its computational purposes.

1.2.2 Additional Contribution

In addition to the topic of model approximation, another topic of research was conducted during the period of PhD work. This work is not related to the topic of model approximation, but more on stabilization of multiphase flow in the riser pipeline experiencing slugs. The aim of stabilization is to suppress oscillation which occurs in the riser. One of the author's main contribution in this area is as follow.

Suppressing Riser-Based Slugging in Multiphase Flow by State Feedback

This part proposes a state feedback design method for attenuating severe slugging in multiphase flow pipeline systems. The feedback is designed based on the input output linearization method, and incorporates the saturation effect on the input. The designed feedback can suppress the slugging phenomena provided some sufficient conditions are satisfied. Finally, checking the conditions leads to the selection of the variable which is more relevant to be controlled.

1.3 List of Papers

The following is the list of publications related to the main contribution of model approximation.

1. H.B. Siahaan, O.M. Aamo and B.A. Foss. Approximate Balanced Truncation of a Class of Polynomial Nonlinear Systems, submitted to *Automatica*.

2. H.B. Siahaan. A Balancing Approach to Model Reduction of Polynomial Nonlinear Systems, accepted for publication in *Proceedings 17th IFAC World Congress*, Seoul, Korea, 2008.

3. H.B. Siahaan, O.M. Aamo and B.A. Foss. Reachability-Based Approach for $\mathcal{L}_2$ Gain Approximation of Polynomial Systems, *Proceedings European Control Conference 2007*, Kos, Greece, pp. 1119-1125, 2007.

4. H.B. Siahaan, $\mathcal{L}_2$ Gain Approximation of Nonlinear Systems: a Heuristic Approach, *Proceedings 45th IEEE Conference on Decision and Control*, San Diego, USA, pp. 3696-3701, 2006.

5. H.B. Siahaan, S. Weiland and A.A. Stoorvogel, Optimal Approximation of Linear Operators: a Singular Value Decomposition Approach, *Proceedings 15th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, Notre Dame, USA, 2002.

The following is the list of publications related to the additional contribution in riser-based slugging in multiphase flow.

1. H.B. Siahaan, O.M. Aamo and B.A. Foss. Suppressing Riser-Based Slugging in Multiphase Flow with State Feedback, *Proceedings 44th IEEE Conference on Decision and Control and European Control Conference 2005*, Seville, Spain, 2005.

2. O.M. Aamo, G.O. Eikrem, H.B. Siahaan and B.A. Foss, Observer Design for Multiphase Flow in Vertical Risers with Gas-Lift - Theory and Experiments, *Journal of Process Control*, Vol. 15, Issue 3, 2005.

3. O.M. Aamo, G.O. Eikrem and H.B. Siahaan, B.A. Foss, Observer Design for Gas Lifted Oil Wells, *Proceedings 2004 American Control Conference*, Boston, USA, 2004.

4. G.O. Eikrem, O.M. Aamo, H.B. Siahaan and B.A. Foss, Anti-Slug Control of Gas Lift Well - Experimental Results, *Proceedings IFAC Nonlinear Control Conference (NOLCOS) 2004*, Stuttgart, Germany, 2004.

1.4 Structure

The dissertation is structured as follows. Chapter 2 contains a selection of results from literature study which is important for the rest of the thesis.

Chapter 3 covers approximation of linear operator from a singular value decomposition approach. Model reduction for polynomial systems and several computational schemes using sum of squares programming are described in chapter 4-6. Chapter 7 consists of additional contribution on stabilization of multiphase flow in the riser pipeline. Chapter 8 contains concluding remarks of the thesis.

1.5 Notations

The following notations are fairly standard and will be used throughout the thesis. The set of real numbers is denoted by $\mathbb{R}$. The collection of all real matrices of size $n \times m$ is denoted by $\mathbb{R}^{n\times n}$. The superscript ^{T} stands for real valued matrix transposition. The superscript * stands for complex valued matrix transposition. The notation I_p means the identity matrix of dimension $p \times p$. The set of symmetric matrices in $\mathbb{R}^{n\times n}$ is denoted by S_n. The matrix inequality $W \succ 0\,(\prec 0)$ means that W is a positive (negative) definite symmetric matrix while $W \succeq 0(\preceq 0)$ means that W is a positive (negative) semidefinite symmetric matrix. The p-norm of the vector $x \in \mathbb{R}^n$ is given by

$$\|x\|_p := \begin{cases} \left(\Sigma_{i=1}^n |x_i|^p\right)^{1/p} & \text{if } p < \infty \\ \max_{i=1,\ldots,n} |x_i| & \text{if } p = \infty \end{cases}.$$

For $p = 2$ we sometimes drop the index and write $\|\cdot\|$ to refer to the Euclidean norm of the vector involved. The p-norm space is a linear vector space equipped with p-norm. The induced p-norm of a matrix $M \in \mathbb{R}^{n\times n}$ is

$$\|M\|_{p-ind} := \sup_{0\neq x\in\mathbb{R}^n} \frac{\|Mx\|_p}{\|x\|_p}.$$

$\mathcal{L}_p[t_o, t_f]$ means the vector space of function $v : [t_o, t_f] \to \mathbb{R}^q$ that satisfies

$$\begin{aligned} \|v\|_{\mathcal{L}_p[t_o,t_f]} &:= \left(\int_{t_f}^{t_o} \|v(t)\|^p\, dt\right)^{1/p} < \infty \quad \text{if } p \in [1,\infty), \\ \|v\|_{\mathcal{L}_\infty[t_o,t_f]} &:= \max_{t\in[t_o,t_f]} \|v(t)\| < \infty \qquad\quad \text{if } p = \infty. \end{aligned}$$

The set of polynomial in x with real coefficient is denoted by $\mathbb{R}[x]$. For polynomial $p \in \mathbb{R}[x]$ we define $\mu_{\min}(p(x))$ and $\mu_{\max}(p(x))$ as the minimum and maximum degree, respectively, of its monomials. For example $p(x) = x_2 + x_1x_2 + x_1^3x_2$ will give $\mu_{\min}(p(x)) = 1$ from the monomial x_2 and $\mu_{\max}(p(x)) = 4$ from the monomial $x_1^3x_2$. The set of matrices of size

$n \times m$ whose entries are polynomial in x with real coefficient is denoted by $\mathbb{R}^{n\times m}[x]$. A scalar function $w(x)$ is said to be positive definite if $w(0) = 0$ and $w(x) > 0$ for all $x \neq 0$. The matrix inequality $W(x) \succ 0 \, (\prec 0)$ means that W is a positive (negative) definite symmetric matrix for all x while the matrix inequality $W(x) \succeq 0 (\preceq 0)$ means that W is a positive (negative) semidefinite matrix for all x. For any matrix Γ we denote N_Γ as the full rank matrix satisfying $\operatorname{Im} N_\Gamma = \ker \Gamma$. A linear system $\hat{G}$ with realization $\{\hat{A}, \hat{B}, \hat{C}, \hat{D}\}$ of order n can be written in a state space form

$$\begin{aligned} \dot{\hat{x}} &= \hat{A}\hat{x} + \hat{B}\hat{u}, \\ \hat{y} &= \hat{C}\hat{x} + \hat{D}\hat{u}, \end{aligned}$$

where $\hat{x}$ is the state, $\hat{u}$ is the input and $\hat{y}$ is the output. The system $\hat{G}$ is said to be controllable if the controllability matrix

$$\begin{bmatrix} \hat{C} \\ \hat{C}\hat{A} \\ \vdots \\ \hat{C}\hat{A}^{n-1} \end{bmatrix}$$

has full rank. The system $\hat{G}$ is said to be observable if the observability matrix

$$\begin{bmatrix} \hat{B} & \hat{A}\hat{B} & \dots & \hat{A}^{n-1}\hat{B} \end{bmatrix}$$

has full rank. The square matrix $\hat{A}$ is Hurwitz if all of its eigenvalues have strictly negative real part. The ball in $\mathbb{R}^n$ is denoted by

$$B_r = \left\{ x \in \mathbb{R}^n | \ \|x\|^2 \leq r \right\}.$$

For a scalar function $\xi(x) = \xi(x_1, \dots, x_n)$, the function is said to belong to class C^k for some positive integer k if all of its partial derivative of order $\leq k$ with respect to $x_1, \dots, x_n$ exist and continuous. A matrix-valued function belongs to class C^k if each of its element belongs to class C^k.

Chapter 2

Preliminaries

2.1 Introduction

This chapter discusses some well known results on model approximation which are relevant for the thesis (for a survey on model reduction techniques, see [3]). We will begin with the problem of approximation of linear operators with lower rank linear operators. For the case when the operators work in the standard Euclidean space, Schmidt and Mirsky [3] have shown that there exists an optimal approximant which minimizes the distance between the original operator and all lower rank approximants. Here, the distance is quantified by the standard Euclidean induced norm.

The result of Schmidt and Mirsky is more or less seen as the inspiration to generalize the problem to linear systems where we seek to approximate a stable linear dynamical system with another lower order stable linear dynamical system. Based on the same spirit, Adamjan, Arov and Krein [2] have shown the existence of a lower order optimal approximant to the linear system. In this case the linear system and the lower order optimal approximant are viewed as operators with special structure. Though these operators, the Hankel operators, operate in infinite dimensional space they have finite rank. The Hankel operator of the optimal approximant has a lower rank than that of the original linear system. The optimal operator itself minimizes the distance quantified by the Hankel norm.

The Hankel operator of a linear system is closely related to the reachability and observability gramians of the system. The gramians are fundamental for another means of obtaining a reduced order model, known as balanced

truncation [25]. The method first computes the gramians from Lyapunov equations and then compute a transformation such that the reachability and observability gramians are balanced. The reduced order model is obtained by truncating the transformed system such that the least reachable and observable part is removed.

Another approach for model reduction of linear systems is by computing a generalized version of the gramians. The generalized gramians are computed from Lyapunov inequalities instead of equations. It can be further shown that for a stable linear system there exists a lower order stable system such that the approximation error in $\mathcal{H}_\infty$-norm is bounded by any given constant provided a certain coupling constraint between the generalized reachability and observability gramians is satisfied [9].

2.2 Approximation of Linear Operators

We consider the problem of approximating a full rank matrix $M \in \mathbb{R}^{n\times n}$ by another matrix $M' \in \mathbb{R}^{n\times n}$ of lower rank where all the matrices are viewed as maps from the p-norm space to the p-norm space. To be precise the problem is given by

$$OPT - p : \min_{rank(M')\leq k<n} \left\|M - M'\right\|_{p-ind}.$$

For general p, the problem $OPT - p$ is discussed in more detail in the next chapter. For the special case of $p = 2$ the problem has been solved by Schmidt and Mirsky in [3]. Indeed, if

$$M = USV^T$$

is the singular value decomposition (SVD) of the matrix M where $U = \begin{bmatrix} u_1 & \dots & u_n \end{bmatrix} \in \mathbb{R}^{n\times n}, V = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $S = diag(\sigma_i) \in \mathbb{R}^{n\times n}$ with $\sigma_1 \geq \dots \geq \sigma_n > 0$, then the solution M'_k which satisfies

$$\left\|M - M'_k\right\|_{2-ind} = \min_{rank(M')\leq k<n} \left\|M - M'\right\|_{2-ind}$$

is given by

$$M'_k = \sigma_1 u_1 v_1^T + \dots + \sigma_k u_k v_k^T$$

with

$$\left\|M - M'_k\right\|_{2-ind} = \sigma_{k+1}.$$

2.3 Reachability and Observability Maps

In this part we review the reachability and observability maps of linear systems. These maps are of fundamental importance for model approximation. We consider an LTI system G of the form

$$\dot{x} = Ax + Bu \tag{2.1a}$$
$$y = Cx \tag{2.1b}$$

where $x \in \mathbb{R}^n, u \in \mathbb{R}^{n_u}, y \in \mathbb{R}^{n_y}, A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times n_u}, C \in \mathbb{R}^{n_y\times n}$. The order of the system is given by n. Throughout the thesis we assume that (2.1) is a minimal representation (observable and controllable) and A is Hurwitz. The materials in this part are based on generalization of the reachability and observability maps from [43] on $\mathcal{L}_p$-space.

Indeed, for an initial condition $x(t_0)$ at time t_0, the solution to (2.1) at time t can be written as

$$x(t) = e^{A(t-t_o)}x(t_o) + \int_{t_o}^{t} e^{A(t-\tau)}Bu(\tau)\,d\tau.$$

Suppose that the initial condition is set to zero ($x(t_o) = 0$). For the class of input $u \in \mathcal{L}_p[t_o, t_f]$ the set of solutions given by

$$\mathcal{R}_p([t_o, t_f]) = \left\{ x \in \mathbb{R}^n |\ x = \varphi_{[t_o,t_f]}(u) = \int_{t_o}^{t_f} e^{A(t_f-\tau)}Bu(\tau)\,d\tau, u \in \mathcal{L}_p[t_o, t_f] \right\}$$

is a set containing all states at time t_f reachable from the origin. The reachability map $\varphi_{[t_o,t_f]}$ is a linear map from $\mathcal{L}_p[t_o, t_f]$ to $\mathbb{R}^n$.

Next we focus at the event of no input to the system ($u = 0$), that is

$$\dot{x} = Ax,$$
$$y = Cx.$$

For an initial condition set to $x(t_o) = x$ the output will be

$$y(t) = Cx(t) = Ce^{A(t-t_o)}x(t_o) = Ce^{A(t-t_o)}x$$

and the set containing all type of this output is given by

$$\mathcal{O}_p(x, [t_o, t_f]) = \left\{ y \in \mathcal{L}_p[t_o, t_f] \mid y = (\phi_{t_o}(x))(t) = Ce^{A(t-t_o)}x, t_o \le t \le t_f \right\}.$$

The observability map ϕ_{t_o} is a linear map from $\mathbb{R}^n$ to $\mathcal{L}_p[t_o, t_f]$.

2.4 Hankel Operator

In this section we will discuss the relation between the observability and reachability maps with an operator known as Hankel operator. Furthermore we will see that we can define the quality of an approximant based on this operator. This section is largely based on [9] and [10].

Consider an LTI system G with impulse response given by $g(t)$. The output of the system with respect to the response $u(t)$ satisfies the convolution given by

$$y(t) = \int_{-\infty}^{\infty} g(t-\tau)\, u(\tau)\, d\tau.$$

Now let us consider the following input

$$\hat{u}(t) = P_{-}u(t) = \begin{cases} u(t), & \text{if } t \leq 0 \\ 0, & \text{if } t > 0 \end{cases}.$$

Then the output with respect to this type of input is given by

$$\bar{y}(t) = \int_{-\infty}^{\infty} g(t-\tau)\, \hat{u}(\tau)\, d\tau = \int_{-\infty}^{0} g(t-\tau)\, \hat{u}(\tau)\, d\tau.$$

We are interested in the output

$$\hat{y}(t) = P_{+}\bar{y}(t) = \begin{cases} 0, & \text{if } t < 0 \\ \bar{y}(t), & \text{if } t \geq 0 \end{cases}.$$

The Hankel operator of the system G is a map $\Gamma_G : \mathcal{L}_p(-\infty, 0] \to \mathcal{L}_p[0, \infty)$ which satisfies $\hat{y} = \Gamma_G \hat{u}$. Here, the operator maps past inputs to future outputs. The induced norm of the operator is defined by

$$\|\Gamma_G\|_{H,p} := \sup_{\hat{u} \in \mathcal{L}_p(-\infty,0]} \|\Gamma_G \hat{u}\|_{\mathcal{L}_p[0,\infty)}$$

which is referred to as Hankel norm.

For the system $G = (A, B, C)$ in (2.1) the impulse response is given by $g(t) = Ce^{At}B$ with Hankel operator given by

$$\hat{y}(t) = (\Gamma_G \hat{u})(t) = \begin{cases} 0, & \text{if } t < 0 \\ \int_{-\infty}^{0} Ce^{A(t-\tau)}B\hat{u}(\tau)\, d\tau, & \text{if } t \geq 0 \end{cases}$$

We can decompose the operator into two parts

$$\hat{y}(t) = (\phi_0(x_0))(t) = Ce^{At}x_0,\ t \geq 0,$$
$$x_0 = \varphi_{(-\infty,0]}(\hat{u}) = \int_{-\infty}^{0} e^{-A\tau}B\hat{u}(\tau)\, d\tau,$$

which shows that

$$\hat{y} = \Gamma_G \hat{u} = \phi_0 \left(\varphi_{(-\infty,0]} \left(\hat{u} \right) \right).$$

Hence the Hankel operator is given by

$$\Gamma_G = \phi_0 \varphi_{(-\infty,0]}.$$

Furthermore, we can even determine the upper bound of the rank of Γ_G. Here, the rank of Γ_G refers to the maximum number of linearly independent outputs of the operator.

Proposition 2.1 *[9] For the system G of order n then rank $(\Gamma_G) \leq n$.*

Let us consider the following $\mathcal{L}_p$-induced norm of the system $G : \mathcal{L}_p(-\infty, \infty) \rightarrow \mathcal{L}_p(-\infty, \infty)$

$$\|G\|_{\mathcal{L}_p - ind} := \sup_{u \in \mathcal{L}_p(-\infty,\infty)} \|Gu\|_{\mathcal{L}_p(-\infty,\infty)}.$$

The following proposition shows that the distance of the original system from any approximant of order n_r is bounded below by the minimum value of $\|\Gamma_G - \Gamma_{G'}\|_{H,p}$ over all possible Hankel operators $\Gamma_{G'}$ of rank not larger than n_r.

Proposition 2.2 *For the system G of order n, it follows that*

$$\|G - G_r\|_{\mathcal{L}_p - ind} \geq \min_{rank(\Gamma_{G'}) \leq n_r} \|\Gamma_G - \Gamma_{G'}\|_{H,p}$$

for all system G_r of order $n_r < n$.

Proof. *For any system Ξ of order n it follows that*

$$\begin{aligned} \|\Xi\|_{\mathcal{L}_p - ind} &\geq \sup_{\hat{u} \in \mathcal{L}_p(-\infty,0]} \|\Xi \hat{u}\|_{\mathcal{L}_p(-\infty,\infty)} \geq \sup_{\hat{u} \in \mathcal{L}_p(-\infty,0]} \|P_+ \Xi \hat{u}\|_{\mathcal{L}_p(-\infty,\infty)} \\ &= \sup_{\hat{u} \in \mathcal{L}_p(-\infty,0]} \|P_+ \Xi \hat{u}\|_{\mathcal{L}_p[0,\infty)} = \sup_{\hat{u} \in \mathcal{L}_p(-\infty,0]} \|\Gamma_\Xi \hat{u}\|_{\mathcal{L}_p[0,\infty)} = \|\Gamma_\Xi\|_{H,p}. \end{aligned}$$

Similarly,

$$\|G - G_r\|_{\mathcal{L}_p - ind} \geq \|\Gamma_{G - G_r}\|_{H,p} = \|\Gamma_G - \Gamma_{G_r}\|_{H,p}.$$

By Proposition 2.1, rank$(\Gamma_{G_r}) \leq n_r$ and

$$\|\Gamma_G - \Gamma_{G_r}\|_{H,p} \geq \min_{rank(\Gamma_{G'}) \leq n_r} \|\Gamma_G - \Gamma_{G'}\|_{H,p}.$$

■

This proposition shows the limit of how good an approximant G_r can be where it is shown that the approximation error can not be better than the lower bound given by the minimum value of $\|\Gamma_G - \Gamma_{G'}\|_{H,p}$.

2.5 Special Case for $p = 2$

For $p = 2$ all the operators involved are working on Hilbert space which has nice properties to be exploited.

2.5.1 Gramians

In view of the fact that a Hilbert space is equipped with an inner product we can obtain the adjoint of reachability and observability maps and show that they are related to the reachability and observability gramians defined as follows.

Definition 2.1 *[43] The reachability and observability gramians of (2.1) are given, respectively by*

$$Y_{[t_o,t_f]} = \int_{t_o}^{t_f} e^{A(t_f-t)} B B^T e^{A^T(t_f-t)} dt,$$
$$X_{[t_o,t_f]} = \int_{t_o}^{t_f} e^{A^T(t-t_o)} C^T C e^{A(t-t_o)} dt.$$

The adjoint reachability map $\varphi^*_{t_f}$ satisfying

$$\left\langle \upsilon, \varphi^*_{t_f}(z) \right\rangle_{\mathcal{L}_2[t_o,t_f]} = \left\langle \varphi_{[t_o,t_f]}(\upsilon), z \right\rangle_{\mathbb{R}^n}$$

is given by

$$\left(\varphi^*_{t_f}(z)\right)(\tau) = B^T e^{A^T(t_f-\tau)} z.$$

Then we have the following.

Proposition 2.3 *[43] The reachability gramian is given by* $Y_{[t_o,t_f]} = \varphi_{[t_o,t_f]} \varphi^*_{t_f}$.

Proposition 2.4 *[43] Given x, suppose that z is any solution of $Y_{[t_o,t_f]}z = x$. Let $u_{opt} = \varphi^*_{t_f}(z)$. Then*

$$\int_{t_o}^{t_f} \|u_{opt}(t)\|^2 dt \leq \int_{t_o}^{t_f} \|u(t)\|^2 dt$$

for all $u \in \mathcal{L}_2[t_o, t_f]$ such that $x = \varphi_{[t_o,t_f]}(u)$.

The result shows that u_{opt} is the minimum energy of input to reach state x at time t_f from the origin. In this case we have

$$\begin{aligned}\int_{t_o}^{t_f} \|u_{opt}(t)\|^2 dt &= \langle u_{opt}, u_{opt}\rangle_{\mathcal{L}_2[t_o,t_f]} = \left\langle \varphi^*_{t_f}(z), \varphi^*_{t_f}(z)\right\rangle_{\mathcal{L}_2[t_o,t_f]} \\ &= \left\langle \varphi_{[t_o,t_f]}\varphi^*_{t_f} Y^\dagger_{[t_o,t_f]}x, Y^\dagger_{[t_o,t_f]}x\right\rangle_{\mathbb{R}^n} = x^T Y^\dagger_{[t_o,t_f]}x.\end{aligned}$$

Remark 2.1 *When $Y_{[t_o,t_f]} = \varphi_{[t_o,t_f]}\varphi^*_{t_f}$ is invertible we have $Y^\dagger_{[t_o,t_f]} = Y^{-1}_{[t_o,t_f]}$.*

Here, the reachability gramian determines the states which are hard to reach by looking at the minimum energy $x^T Y^\dagger_{[t_o,t_f]}x$ required to reach the state x from the origin. Throughout the thesis, whenever referred to, the term reachability gramian is for $t_0 = -\infty$ and $t_f = 0$. The reachability gramian can then be computed by solving the following equation.

Proposition 2.5 *[9] For $t_0 = -\infty$ and $t_f = 0$ the reachability gramian $Y = Y_{[-\infty,0)}$ is the solution to the Lyapunov equation*

$$AY + YA^T + BB^T = 0. \tag{2.2}$$

Furthermore, $Y \succ 0$.

The adjoint observability map $\phi^*_{[t_o,t_f]}$ satisfying

$$\left\langle \phi^*_{[t_o,t_f]}(\upsilon), z\right\rangle_{\mathbb{R}^n} = \langle \upsilon, \phi_{t_o}(z)\rangle_{\mathcal{L}_2[t_o,t_f]}$$

is given by

$$\phi^*_{[t_o,t_f]}(\upsilon) = \int_{t_o}^{t_f} e^{A^T(\tau - t_o)} C' \upsilon(\tau)\, d\tau.$$

Then we have the following.

Proposition 2.6 *[43] The observability gramian is given by* $X_{[t_o,t_f]} = \phi^*_{[t_o,t_f]}\phi_{t_o}$.

For initial condition $x(t_o) = x$ we have

$$\int_{t_o}^{t_f} \|y(t)\|^2 dt = \langle \phi_{t_o}(x), \phi_{t_o}(x) \rangle_{\mathcal{L}_2[t_o,t_f]} = \left\langle x, \phi^*_{[t_o,t_f]}(\phi_{t_o}(x)) \right\rangle_{\mathbb{R}^n}$$
$$= x^T X_{[t_o,t_f]} x.$$

Here, the observability gramian determines the states which are difficult to observe by looking at the output energy $x^T X_{[t_o,t_f]} x$ produced when the initial state is at x. Throughout the thesis, whenever referred to, the term observability gramian is for $t_0 = 0$ and $t_f = \infty$. The observability gramian can then be computed by solving the following equation.

Proposition 2.7 *[9] For* $t_0 = 0$ *and* $t_f = \infty$ *the observability gramian* $X = X_{[0,\infty)}$ *is the solution to the Lyapunov equation*

$$A^T X + XA + C^T C = 0. \tag{2.3}$$

Furthermore, $X \succ 0$.

The existence and uniqueness of positive definite solutions Y and X to (2.2) and (2.3), respectively are guaranteed by the asymptotic stability and minimality (observability and controllability) of the system.

2.5.2 Hankel Singular Values

The relation of Hankel operator with the gramians is described as follows. The Hankel norm of the system $G = \{A, B, C\}$ in (2.1) for $p = 2$ is given by the square root of the largest eigenvalue of cross gramian XY, that is

$$\|\Gamma_G\|_{H,2} = \sqrt{\lambda_{\max}(XY)}.$$

In fact all eigenvalues of XY are equal to all nonzero eigenvalues of finite rank operator Γ_G. The square roots of these eigenvalues denoted by

$$\sigma_1^H \geq \ldots \geq \sigma_n^H > 0$$

are called the Hankel singular values of the system G.

Furthermore, by Proposition 2.2 for $p = 2$, any approximant G_r of order $n_r < n$ for the system G of order n satisfies

$$\|G - G_r\|_{\mathcal{L}_2-ind} \geq \min_{rank(\Gamma_{G'})\leq n_r} \|\Gamma_G - \Gamma_{G'}\|_{H,2}$$

The exact value of the lower bound is given by

$$\min_{rank(\Gamma_{G'})\leq n_r} \|\Gamma_G - \Gamma_{G'}\|_{H,2} = \sigma^H_{n_r+1}$$

which is due to Adamjan, Arov and Krein [2]. This is a generalization of the matrix rank minimization due to Schmidt and Mirsky (see Section 2.2) for linear dynamical systems.

2.5.3 Balanced Truncation Approach

Model reduction by balancing approach was introduced for the first time in [25]. The method is performed in two steps. The first step is by transforming the original system (2.1) such that the new system has balanced representation of the gramians. The transformation is constructed from the gramians Y and X obtained from Proposition 2.5 and 2.7, respectively. For the transformed system we have the following well-known result.

Theorem 2.1 *[9] Consider $G = \{A, B, C\}$ in (2.1). There exists a nonsingular matrix T such that the gramians of the transformed system $\{\hat{A}, \hat{B}, \hat{C}\} = \{TAT^{-1}, TB, CT^{-1}\}$ are identical. In this case the positive definite matrix $\Sigma = diag\left(\sigma^H_i\right)$ with ordered Hankel singular values $\sigma^H_1 \geq \ldots \geq \sigma^H_n > 0$ is the balanced reachability and observability gramian for the transformed system and satisfies*

$$\hat{A}\Sigma + \Sigma\hat{A}^T + \hat{B}\hat{B}^T = 0,$$
$$\hat{A}^T\Sigma + \Sigma\hat{A} + \hat{C}^T\hat{C} = 0.$$

The second step of the approach is by truncating the transformed system such that the least important structure is removed. The least important ones are related to the lower values of the Hankel singular values. In this case we truncate the transformed system at order $n_r < n$ where there is a gap between the adjacent Hankel singular values $\sigma^H_{n_r} > \sigma^H_{n_r+1}$. The truncation can be done as follows. Suppose we partition the transformed system conformally into

$$\hat{A} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix},\ \hat{B} = \begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix},\ \hat{C} = \begin{bmatrix} \hat{C}_1 & \hat{C}_2 \end{bmatrix}.$$

The reduced order model is then given by representation $G_r = \{\hat{A}_{11}, \hat{B}_1, \hat{C}_1\}$ of order n_r. The reduced order model G_r has balanced gramians and preserves asymptotic stability of the original model, that is $\hat{A}_{11}$ being Hurwitz. Furthermore, the error model is guaranteed to satisfy

$$\sigma_{n_r+1}^H \leq \|G - G_r\|_{\mathcal{L}_2-ind} \leq 2\left(\sigma_1^H + \ldots + \sigma_{n_r}^H\right).$$

2.5.4 $\mathcal{H}_\infty$ Model Reduction

Another means of quantifying the quality of a reduced model is defined by the $\mathcal{H}_\infty$-norm, which is equivalent to the $\mathcal{L}_2$-induced norm. In this case, the existence of reduced order model for a linear system is given by the following.

Proposition 2.8 *[9] Given G with a realization $\{A, B, C, D\}$ where $A \in \mathbb{R}^{n\times n}$ is Hurwitz, there exists G_r of order $n_r < n$ with a realization $\{A_r, B_r, C_r, D_r\}$ where $A_r \in \mathbb{R}^{n_r\times n_r}$ is Hurwitz such that $\|G - G_r\|_{\mathcal{H}_\infty} := \|G - G_r\|_{\mathcal{L}_2-ind} < \hat{\epsilon}$ if and only if there exist generalized gramians $X \succ 0$ and $Y \succ 0$ satisfying*

- $AY + YA^T + BB^T \preceq 0$,
- $A^TX + XA + C^TC \preceq 0$,
- $X - \hat{\epsilon}^2Y^{-1} \succeq 0$, $rank\left(X - \hat{\epsilon}^2Y^{-1}\right) \leq n_r$.

Here, the generalized gramians X and Y are coupled with rank condition while the gramians X and Y in the balanced truncation approach are independent of each other.

2.6 Sum of Squares

We define a polynomial in the form

$$p(x) = \sum_i p_i^2(x)$$

as a sum of squares (SOS) polynomial when $p_i(x)$ are polynomials. It is obvious that any polynomial which can be expressed as an SOS of other polynomials is nonnegative everywhere. One way to express an SOS equivalently is by

$$p(x) = z^T(x)Mz(x)$$

where M is a positive semidefinite symmetric matrix and $z(x)$ is a monomial of degree less than or equal to half of the degree of $p(x)$. For the same monomial $z(x)$ it might be possible to have similar representation with different M with M being not positive semidefinite. Thus if the intersection of $\left\{M \in S_n | p(x) = z^T(x) M z(x)\right\}$ with $\{M \in S_n | M \succeq 0\}$ is not empty then $p(x) = z^T(x) M z(x)$ is an SOS. Within this direction, in [30] the author has showed that determining whether a polynomial can be expressed as an SOS is an LMI problem. Hence the problem of testing whether a polynomial is sum of squares becomes relatively easy as it can be computed using semidefinite programming. In view of the fact that verifying nonnegativity of a polynomial is very difficult, throughout the thesis, we will relax most polynomial inequalities by replacing nonnegativity with SOS condition.

Chapter 3

Optimal Approximation of Linear Operators: a Singular Value Decomposition Approach

3.1 Introduction

Singular values and singular value decompositions are among the most important tools in linear algebra that have played a key role in systems analysis, control system design, model reduction, data compression, perturbation theory, signal analysis and many applications in numerical linear algebra. Unlike eigenvalues and eigenvalue decompositions, singular values and singular value decompositions provide structural information on the spacial distribution of mutually orthogonal amplification directions in the domain and co-domain of a linear map. As such, the singular value decomposition defines a numerically well conditioned basis for both the domain and the co-domain of a linear operator and is, in fact, the core numerical tool to implement basic algebraic concepts such as rank, null space, range, orthogonal complements, etc.

A basic algebraic treatment of singular values and their applications can be found in the standard works [12], [40]. In short, every matrix $M \in \mathbb{C}^{m\times n}$ admits a decomposition of the form

$$M = Y\Sigma X^* \tag{3.1}$$

where $X \in \mathbb{C}^{n\times n}$ and $Y \in \mathbb{C}^{m\times m}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{m\times n}$ is

a matrix whose diagonal entries $(\Sigma)_{ii} = \sigma_i$, $i = 1, \ldots, \min(m, n)$, and which is zero elsewhere. Here, σ_i are non-negative real numbers, ordered according to $\sigma_1 \geq \ldots \geq \sigma_{\min(m,n)} \geq 0$ and called the *singular values* of M. The column vectors x_i of X and y_i of Y are the right and left *singular vectors* and equation (3.1) is referred to as a *singular value decomposition* of M. From (3.1) it follows that M allows a *diadic expansion* $M = \Sigma_{k=1}^{r} \sigma_i y_i x_i^*$, where $r = \operatorname{rank} M$.

The decomposition (3.1) proves useful for a wide variety of problems. It is the purpose of this chapter to propose a generalization of this traditional notion of a singular value decomposition and to establish a number of its properties. In addition, we consider the approximation problem to find lower rank approximants M' of M which are optimal in that the error $M - M'$ has minimal induced norm when viewed as an operator on arbitrary normed spaces.

The chapter is organized as follows. In section 3.2 we introduce singular values in a general fashion and establish some of its elementary properties. Problem formulations are collected in section 3.3. The main results on optimal rank approximations are given in section 3.4. An application on optimal system identification is discussed in section 3.5.

3.2 Generalized Singular Values

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite dimensional vector spaces over the field of scalars $\mathbb{F}$. Let $n = \dim \mathcal{X}$ and $m = \dim \mathcal{Y}$ and define the p-norm of elements $x \in \mathcal{X}$ as

$$\|x\|_p := \begin{cases} \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} & \text{if } p < \infty \\ \max_{i=1,\ldots,n} |x_i| & \text{if } p = \infty \end{cases}$$

Here, x_i denotes the ith component of x. Let $(\mathcal{X}, \|\cdot\|_p)$ and $(\mathcal{Y}, \|\cdot\|_p)$ be normed linear vector spaces and let $M : \mathcal{X} \to \mathcal{Y}$ be a linear mapping. The *induced p-norm of* M is

$$\|M\|_{p\text{-ind}} := \sup_{0 \neq x \in \mathcal{X}} \frac{\|Mx\|_p}{\|x\|_p}.$$

Throughout, the notation $\mathcal{L} \subseteq \mathcal{X}$ is understood to mean that $\mathcal{L}$ is a *linear subspace* of $\mathcal{X}$. If $\mathcal{L} \subseteq \mathcal{X}$, then $M|_{\mathcal{L}}$ denotes the restriction of M to $\mathcal{L}$, i.e., $M|_{\mathcal{L}} : \mathcal{L} \to \mathcal{Y}$ is defined as $M|_{\mathcal{L}} x = Mx$ for $x \in \mathcal{L}$.

Definition 3.1 *The* p-norm induced singular values *of* M *are the numbers*

$$\sigma_k^{(p)} := \inf_{\substack{\mathcal{L}\subseteq\mathcal{X},\\ \dim\mathcal{L}\geq n-k+1}} \sup_{0\neq x\in\mathcal{L}} \frac{\|Mx\|_p}{\|x\|_p} \tag{3.2}$$

where k *runs from 1 till* n*. The set of these numbers is denoted by* $\sigma^{(p)}(M)$*.*

Note that the induced p-norm singular values are non-negative real numbers. For $k = 1, \ldots, n$ we will also be interested in the arguments of the infimum in (3.2). For this purpose, define

$$\mathbb{L}_k^{(p)} := \{\mathcal{L} \subseteq \mathcal{X} \mid \dim\mathcal{L} \geq n-k+1 \text{ and } \sup_{0\neq x\in\mathcal{L}} \frac{\|Mx\|_p}{\|x\|_p} = \sigma_k^{(p)}\}. \tag{3.3}$$

Note that $\mathbb{L}_k^{(p)}$ is non-empty for all k and all p and that $\mathbb{L}_1^{(p)} = \mathcal{X}$ for all p. Whenever p is understood from the context we omit the superscript (p) and write σ_k, $\sigma(M)$ and $\mathbb{L}_k$. It is easy to see that

$$\begin{aligned}
\sigma_1^{(p)} &= \|M\|_{p\text{-ind}} \\
\sigma_k^{(p)} &= \|M|_{\mathcal{L}_k}\|_{p\text{-ind}} \\
\sigma_n^{(p)} &= \inf_{0\neq x\in\mathcal{X}} \frac{\|Mx\|_p}{\|x\|_p}
\end{aligned}$$

where $\mathcal{L}_k \in \mathbb{L}_k^{(p)}$ and $k = 1, \ldots, n$. Some elementary results pertaining to the p-norm induced singular values are summarized in the following Proposition.

Proposition 3.1 *For all* $p \in [1, \infty]$ *there holds*

1. $\sigma_1^{(p)} \geq \sigma_2^{(p)} \geq \ldots \geq \sigma_n^{(p)} \geq 0$.
2. $\operatorname{rank}(M) = r < n$ *if and only if* $\sigma_{r+1}^{(p)} = \cdots = \sigma_n^{(p)} = 0$.
3. $\operatorname{rank}(M) = n$ *if and only if* $\sigma_n^{(p)} > 0$.
4. $\sigma_1^{(\infty)} \geq \sigma_1^{(p)}$.

Proof. Fix $p \in [1, \infty]$ and let $\mathbb{S}_k := \{\mathcal{L} \subseteq \mathcal{X} \mid \dim\mathcal{L} \geq n-k+1\}$.

1. Obviously $\sigma_k \geq 0$ for all $k = 1, \ldots, n$. Since $\mathbb{S}_k \subseteq \mathbb{S}_{k+1}$ it is immediate that

$$\begin{aligned}\sigma_k &= \inf_{\mathcal{L}\in\mathbb{S}_k} \sup_{0\neq x\in\mathcal{L}} \frac{\|Mx\|_p}{\|x\|_p} \\ &\geq \inf_{\mathcal{L}\in\mathbb{S}_{k+1}} \sup_{0\neq x\in\mathcal{L}} \frac{\|Mx\|_p}{\|x\|_p} \\ &= \sigma_{k+1}.\end{aligned}$$

2. Suppose $\operatorname{rank}(M) = r < n$ and define $\mathcal{K} := \ker M$. Then $\dim \mathcal{K} = n-r$ so that $\mathcal{K} \in \mathbb{S}_k$ for $k = r+1, \ldots, n$. But then

$$\begin{aligned}\sigma_k &= \inf_{\mathcal{L}\in\mathbb{S}_k} \sup_{0\neq x\in\mathcal{L}} \frac{\|Mx\|_p}{\|x\|_p} \leq \sup_{0\neq x\in\mathcal{K}} \frac{\|Mx\|_p}{\|x\|_p} \\ &= 0\end{aligned}$$

   for $k = r+1, \ldots, n$. Since, $\sigma_k \geq 0$ (statement 1), it follows that $\sigma_{r+1} = \cdots = \sigma_n = 0$.

3. Since M is linear and since $\mathbb{S}_n$ consists of one dimensional subspaces of $\mathcal{X}$ we have that $\sigma_n = \inf_{0\neq x\in\mathcal{X}} \frac{|Mx\|_p}{\|x\|_p}$ is strictly larger than zero if M has rank n.

4. It is shown in the Appendix C.2.16 of [8] that $\|M\|_{\infty\text{-ind}} \geq \|M\|_{p\text{-ind}}$. This is equivalent to $\sigma_1^{(\infty)} \geq \sigma_1^{(p)}$.

■

3.3 Problem Formulations

In this section we consider a number of problems where the p-norm induced singular values play a natural role.

3.3.1 Rank Deficiency

An important application of singular values stems from the numerical difficulty to determine the rank of a matrix M. In particular, for situations where M is near rank deficient, a numerically reliable calculation of $\operatorname{rank}(M)$

is sensitive to errors. Most numerical implementations to determine $\operatorname{rank}(M)$ calculate the *numerical rank*, defined as

$$r' = \operatorname{rank}(M, \varepsilon) := \min_{\|M - M'\|_{p\text{-ind}} \leq \varepsilon} \operatorname{rank}(M')$$

where $\varepsilon > 0$ is an accuracy level. In fact, this problem is a special case of the *optimal rank approximation problem*, which we formulate next.

3.3.2 Optimal Rank Approximation

Let $(\mathcal{X}, \|\cdot\|_p)$ and $(\mathcal{Y}, \|\cdot\|_p)$ be finite dimensional normed linear vector spaces of dimension n and m, respectively, and let $M : \mathcal{X} \to \mathcal{Y}$ be a linear mapping of rank r. Consider the problem of approximating M by a linear map $M' : \mathcal{X} \to \mathcal{Y}$ of rank at most k ($k < r$), such that the p-induced norm

$$\|M - M'\|_{p\text{-ind}}$$

is minimal. We refer to this problem as the *optimal rank approximation problem* and to solutions M' as *optimal rank k approximants*.

3.3.3 Optimal System Identification

Consider the problem to model the (real scalar valued) observed time series $\tilde{w}(t)$, $t = 0, \ldots N$, by an auto-regressive linear model of the form

$$\sum_{i=0}^{n} x_i w(t+i) = 0$$

where $x_i \in \mathbb{R}$ are the model coefficients and $n \geq 0$ is the model order. Let $x = (x_0, \ldots, x_n)^\top$ denote the model coefficient vector and define the *misfit* between model x and the data $\tilde{w}$ by

$$\mu(x, \tilde{w}) := \frac{\|e\|_p}{\|x\|_p}$$

where e is the vector of *residuals* $e(t) = \sum_{i=0}^{n} x_i \tilde{w}(t+i)$, $t = 0, \ldots N - n$. Given $\tilde{w}$, $n \geq 0$ and $\varepsilon \geq 0$, the *identification problem* amounts to finding all model coefficient vectors $x \in \mathbb{R}^{n+1}$ which have a guaranteed misfit in that $\mu(x, \tilde{w}) \leq \varepsilon$, i.e., we wish to characterize all models that can not be distinguished if one tolerates a misfit level ε. Note that this set may be

empty. A model $x^* \in \mathbb{R}^{n+1}$ is said to be *optimal* if it minimizes the misfit $\mu(\cdot, \tilde{w})$.

Note the importance and interpretation of this problem for different values of p. The usual phrasing of this problem is in a stochastic context where the variance of e is to be minimized. This is equivalent of setting $p = 2$. Less conventional is the case where $p = \infty$. Solutions of the identification problem then have guaranteed upperbounds on the *amplitude* of their residuals, which seems of considerable interest for many applications in modeling.

We remark that the assumption that $\tilde{w}(t)$ and x_i are scalar valued has been made to simplify exposition only. Multivariable generalizations of this identification problem can be incorporated in a straightforward way. See Section 3.5 below.

3.4 Optimal Rank Approximation

If $p = 2$, the optimal rank optimization problem is well understood and has a simple solution. Indeed, let (3.1) be a singular value decomposition of M and, in the notation of Section 3.1, set $M_k := \sum_{i=1}^{k} \sigma_i x_i y_i^*$. Then $\operatorname{rank} M_k \leq k$ and

$$\min_{\operatorname{rank}(M') \leq k} \|M - M'\|_{\text{2-ind}} = \|M - M_k\|_{\text{2-ind}} = \sigma_{k+1}^{(2)}$$

which shows that M_k is an optimal rank k approximant of M. In particular, any truncation of the diadic expansion of M defines an optimal lower rank approximant of M. Optimal rank k approximants are by no means unique. Indeed, if δ_i, $i = 1, \ldots, k$, satisfy $|\delta_i| \leq \sigma_{k+1}$ then

$$M_k' := \sum_{i=1}^{k} (\sigma_i + \delta_i) y_i x_i^* \tag{3.4}$$

satisfies $\|M - M_k'\|_{\text{2-ind}} = \sigma_{k+1}^{(2)}$ and is therefore also an optimal rank k approximant of M.

3.4.1 A Lower Bound on the Error

If $p \neq 2$, the problem is more difficult. We first establish a lower bound on the mismatch between a matrix M and its lower rank approximations in the

p-induced norm. We then derive a sufficient condition for which this lower bound becomes sharp. Finally, we show that optimal rank $n-1$ approximants always attain this lower bound. Throughout this section, $\mathcal{X}$ and $\mathcal{Y}$ will be finite dimensional vector spaces of dimension n and m, respectively.

Proposition 3.2 *Let $M : \mathcal{X} \to \mathcal{Y}$ have rank r and let $M_k : \mathcal{X} \to \mathcal{Y}$ have rank at most k with $k < r$. Then*

$$\|M - M_k\|_{p\text{-}ind} \geq \sigma_{k+1}^{(p)}.$$

Proof. Let $\mathcal{K}_k = \ker M_k$. Then $\dim \mathcal{K}_k \geq n - k$ and we note that

$$\begin{aligned}
\|M - M_k\|_{p\text{-ind}} &= \sup_{0 \neq x \in \mathcal{X}} \frac{\|(M - M_k)x\|_p}{\|x\|_p} \\
&\geq \sup_{0 \neq x \in \mathcal{K}_k} \frac{\|(M - M_k)x\|_p}{\|x\|_p} \\
&= \sup_{0 \neq x \in \mathcal{K}_k} \frac{\|Mx\|_p}{\|x\|_p}
\end{aligned}$$

Since $\dim \mathcal{K}_k \geq n - k$, it follows that

$$\sup_{0 \neq x \in \mathcal{K}_k} \frac{\|Mx\|_p}{\|x\|_p} \geq \inf_{\substack{\mathcal{L} \in \mathcal{X} \\ \dim \mathcal{L} \geq n-k}} \sup_{0 \neq x \in \mathcal{L}} \frac{\|Mx\|_p}{\|x\|_p}$$

which shows that $\|M - M_k\|_{p\text{-ind}} \geq \sigma_{k+1}^{(p)}$. ∎

A natural question is whether the lower bound in Proposition 3.2 can actually be attained for a rank k matrix M_k. To answer this question, recall that two subspaces $\mathcal{L}'$ and $\mathcal{L}''$ of $\mathcal{X}$ are said to be *complementary* if $\mathcal{L}' \cap \mathcal{L}'' = \{0\}$ and $\mathcal{L}' + \mathcal{L}'' = \mathcal{X}$. If $(\mathcal{L}', \mathcal{L}'')$ is a complementary pair, every $x \in \mathcal{X}$ admits a unique decomposition $x = x' + x''$ with $x' \in \mathcal{L}'$ and $x'' \in \mathcal{L}''$. In that case, we write $x' = \Pi_{\mathcal{L}'|\mathcal{L}''} x$ and $x'' = \Pi_{\mathcal{L}''|\mathcal{L}'} x$ where $\Pi_{\mathcal{L}'|\mathcal{L}''} : \mathcal{X} \to \mathcal{L}'$ and $\Pi_{\mathcal{L}''|\mathcal{L}'} : \mathcal{X} \to \mathcal{L}''$ define the natural projections on $\mathcal{L}'$ along $\mathcal{L}''$ and on $\mathcal{L}''$ along $\mathcal{L}'$, respectively.

The following theorem provides a sufficient condition under which the lower bound in Proposition 3.2 will be sharp.

Theorem 3.1 *Given M, and define the sets $\mathbb{L}_k^{(p)}$ by* (3.3)*. If there exist $\mathcal{L}' \in \mathbb{L}_{k+1}^{(p)}$ and $\mathcal{L}'' \subseteq \mathcal{X}$ such that*

1. $(\mathcal{L}', \mathcal{L}'')$ *are complementary and*
2. $\|\Pi_{\mathcal{L}'|\mathcal{L}''}\|_{p\text{-}ind} \leq 1$

then there exists $M_k : \mathcal{X} \to \mathcal{Y}$ *of rank at most* k *such that*

$$\|M - M_k\|_{p\text{-}ind} = \sigma_{k+1}^{(p)}.$$

In particular, M_k *given by* $M_k|_{\mathcal{L}'} = 0$ *and* $M_k|_{\mathcal{L}''} = M|_{\mathcal{L}''}$ *is an optimal rank* k *approximant of* M*.*

Proof. In view of Proposition 3.2, it suffices to show that M_k, as specified, has rank $\leq k$ and satisfies $\|M - M_k\|_{p\text{-ind}} \leq \sigma_{k+1}^{(p)}$. To see this, first note that $\dim \mathcal{L}' \geq n - k$ which means that $\dim \mathcal{L}'' \leq k$ so that $\operatorname{rank} M_k \leq k$. Second, observe that

$$\begin{aligned}
\|M - M_k\|_{p\text{-ind}} &= \sup_{0 \neq x \in \mathcal{X}} \frac{\|Mx - M_k x\|_p}{\|x\|_p} = \sup_{\substack{x' \in \mathcal{L}', x'' \in \mathcal{L}'' \\ x' + x'' \neq 0}} \frac{\|Mx'\|_p}{\|x' + x''\|_p} \leq \sup_{0 \neq x' \in \mathcal{L}'} \frac{\|Mx'\|_p}{\|x'\|_p} \\
&= \sup_{0 \neq x' \in \mathcal{L}'} \frac{\|Mx'\|_p}{\|x'\|_p} \\
&= \sigma_{k+1}^{(p)}.
\end{aligned}$$

Here, we used in the third inequality that $\Pi_{\mathcal{L}'|\mathcal{L}''}$ is a contraction, i.e., $\|x'\|_p = \|\Pi_{\mathcal{L}'|\mathcal{L}''} x\|_p \leq \|x\|_p$. The last equality follows from the definition of $\mathbb{L}_{k+1}^{(p)}$. It follows that M_k is an optimal rank k approximant of M. ■

The main issue of the above result is the existence of a subspace $\mathcal{L}''$, complementary to $\mathcal{L}' \in \mathbb{L}_{k+1}^{(p)}$ such that the projection $\Pi_{\mathcal{L}'|\mathcal{L}''}$ defines a contraction on $\mathcal{X}$. We will investigate these conditions for a number of special cases.

3.4.2 Nonexistence of Contractive Projection

Theorem 3.1 provides sufficient conditions for which the lower bound in Proposition 3.2 will be attained. These conditions will not always be satisfied. In fact, to see how strict these conditions are, consider the case where $n = 3$, p is even, $p \neq 2$, and $\mathcal{L}'$ is a two dimensional subspace of $\mathcal{X} = \mathbb{R}^3$, spanned by the non-zero vectors x and y, i.e. $\mathcal{L}' = \operatorname{span}(x, y)$. A subspace $\mathcal{L}''$ of $\mathcal{X}$ will satisfy the conditions 1 and 2 of Theorem 3.1 if and only if $\mathcal{L}'' = \operatorname{span}(z)$ with $z \neq 0$ such that

1. $\det(x, y, z) \neq 0$ and

2. $\sum_{i=1}^{3}(\alpha x_i + \beta y_i + \gamma z_i)^p - (\alpha x_i + \beta y_i)^p \geq 0$ for all $\alpha, \beta, \gamma \in \mathbb{R}$.

In particular, the latter condition implies that

$$\sum_{i=1}^{3} x_i^{p-1} z_i = 0$$
$$\sum_{i=1}^{3} y_i^{p-1} z_i = 0$$
$$\sum_{i=1}^{3} (x_1 + y_i)^{p-1} z_i = 0$$

which yields (generically) that $z_1 = z_2 = z_3 = 0$; i.e. $z = 0$. Hence there does not exist a complementary subspace $\mathcal{L}''$ such that the projection $\Pi_{\mathcal{L}'|\mathcal{L}''}$ is contractive. Alternatively, from geometrical point of view, Figure 3.1 shows a two dimensional subspace $\mathcal{L}'$ in three dimensional space $\mathcal{X} = \mathbb{R}^3$, which does not satisfy the sufficient conditions in Theorem 3.1 when $p = \infty$. In this case it is not possible to project every point on infinity ball, along any one dimensional subspace, to the subspace $\mathcal{L}'$ in Figure 3.1 such that the projection point is inside or on the ball.

3.4.3 The Case $p = 2$ and Arbitrary k

If $p = 2$, $\mathcal{X}$ becomes a Hilbert space with the natural inner product $\langle \cdot, \cdot \rangle$. For every subspace $\mathcal{L} \subseteq \mathcal{X}$, its orthogonal complement $\mathcal{L}^\perp := \{x \in \mathcal{X} \mid \langle x, y \rangle = 0 \text{ for all } y \in \mathcal{L}\}$ is complementary to $\mathcal{L}$ and the orthogonal projection $\Pi_{\mathcal{L}|\mathcal{L}^\perp}$ is obviously a contraction. Hence, optimal rank k approximants always exist in this case and are given by the expression (3.4). This case is well understood and can be found in many text books (e.g. [12, 40]).

3.4.4 The Case $k = n - 1$ and Arbitrary p

Let p be arbitrary, suppose that $n = \operatorname{rank} M$ and consider the optimal rank approximation problem with $k = n - 1$. The set $\mathbb{L}_n^{(p)}$ then consists of subspaces of dimension ≥ 1 only. Let $\mathcal{L}' \in \mathbb{L}_n^{(p)}$ be a one dimensional subspace

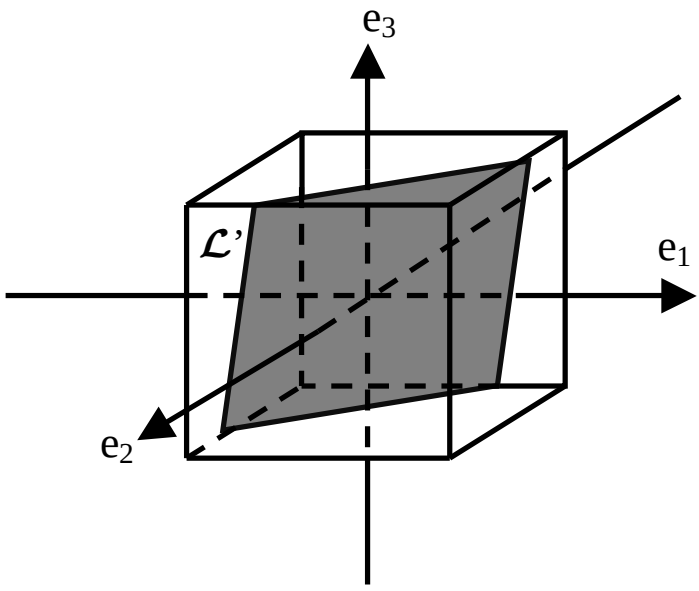

Figure 3.1: Nonexistence of contractive projection for the case $p = \infty$

and let $x' \in \mathcal{L}'$ be a nonzero element. Then $\mathcal{L}' = \text{span}(x')$. The following lemma is easily seen.

Lemma 3.1 *If $\mathcal{L}' = \text{span}(x')$ for a nonzero $x' \in \mathcal{X}$ then $\mathcal{L}'' \subseteq \mathcal{X}$ will be complimentary to $\mathcal{L}'$ if and only if*

$$\mathcal{L}'' = \{x'' \in \mathcal{X} \mid \langle w, x'' \rangle = 0\} \tag{3.5}$$

where $w \in \mathcal{X}$ is a nonzero vector such that $\langle w, x' \rangle \neq 0$.

Hence, Lemma 3.1 provides a parametrization of all complements of a given one-dimensional subspace spanned by a nonzero vector $x' \in \mathcal{X}$. In order to characterize complementary suspaces $(\mathcal{L}', \mathcal{L}'')$ for which the projection $\Pi_{\mathcal{L}'|\mathcal{L}''}$ is contractive, we resort to some terminology from convex analysis [35].

Definition 3.2 *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. A vector $w \in \mathcal{X}$ is said to be a* subgradient of f at $x \in \mathcal{X}$ *if*

$$f(z) \geq f(x) + \langle w, z - x \rangle \tag{3.6}$$

for all $z \in \mathcal{X}$. The set of all subgradients of f at $x \in \mathcal{X}$ is called the subdifferential of f at x *and denoted by $\nabla f(x)$, i.e.,*

$$\nabla f(x) := \{w \in \mathcal{X} \mid f(z) \geq f(x) + \langle w, z - x\rangle \\ \textit{for all } z \in \mathcal{X}\}.$$

Inequality (3.6) is usually referred to as the *subgradient inequality* and has the simple geometric interpretation that the graph of f lies on or above the affine function $g(z) := f(x) + \langle w, z - x\rangle$ which is the tangent hyperplane of f at x. We remark that the subdifferential of f at x is a closed convex set. If $\nabla f(x)$ is non-empty, f is said to be *subdifferentiable* at x.

The next proposition shows that subdifferentials of the mapping $f : x \mapsto \|x\|_p$ precisely parameterize the complements of $\mathcal{L}'$ for which $\Pi_{\mathcal{L}'|\mathcal{L}''}$ is contractive. The following lemma shall be used to prove the proposition.

Lemma 3.2 *Consider $f : x \mapsto \|x\|_p$. Then $w \in \nabla f(x)$ if and only if $\langle w, x\rangle = \|x\|_p$ and $\langle w, z\rangle \leq \|z\|_p$ for all $z \in \mathcal{X}$.*

Proof. *(if).* Obvious.

(only if). Substitute $z = \alpha x$ $(0 \leq \alpha \leq 1)$ into (3.6) then we have $\langle w, x\rangle \geq \|x\|_p$. Take $z = \alpha x$ where $\alpha > 1$ then we have $\langle w, x\rangle \leq \|x\|_p$. Thus we have $\langle w, x\rangle = \|x\|_p$ and $\langle w, z\rangle \leq \|z\|_p$ for all $z \in \mathcal{X}$. ■

Proposition 3.3 *Let $x' \in \mathcal{X}$ be nonzero and $\mathcal{L}' = \mathrm{span}(x')$. Then the pair $(\mathcal{L}', \mathcal{L}'')$ is complementary and $\|\Pi_{\mathcal{L}'|\mathcal{L}''}\|_{p\text{-}ind} \leq 1$ if and only if $\mathcal{L}''$ is given by* (3.5) *with $0 \neq w \in \nabla\|x'\|_p$.*

Proof. *(if).* From Lemma 3.2 we have $\langle w, x'\rangle = \|x'\|_p \neq 0$ which yields that the pair $(\mathcal{L}', \mathcal{L}'')$ is complimentary. Now the subgradient inequality (3.6) yields

$$\|z\|_p \geq \|x'\|_p + \langle w, z - x'\rangle$$

For any given nonzero $\lambda \in \mathbb{R}$ we have

$$\begin{aligned} & \|\lambda z\|_p \geq \|\lambda x'\|_p + \frac{|\lambda|}{\lambda}\langle w, \lambda z - \lambda x'\rangle \\ \Longrightarrow \quad & \|\bar{z}\|_p \geq \|\lambda x'\|_p + \langle v, \bar{z} - \lambda x'\rangle \end{aligned} \tag{3.7}$$

where we set $\bar{z} = \lambda z$ and $v = |\lambda| w/\lambda$. Since z is arbitrary and $\lambda \neq 0$, (3.7) yields the subgradient inequality for all $\bar{z} \in \mathcal{X}$ with $\bar{z} \notin \mathcal{L}''$. By the fact that

the pair $(\mathcal{L}', \mathcal{L}'')$ is complimentary any $\bar{z} \in \mathcal{X}$ has a unique decomposition $\bar{z} = \bar{z}' + \bar{z}''$ with $\bar{z}' = \Pi_{\mathcal{L}'|\mathcal{L}''}\bar{z}$ and $\bar{z}'' = \bar{z} - \bar{z}' \in \mathcal{L}''$. Since $\bar{z}' \in \mathcal{L}'$ it follows that there exists $\lambda \in \mathbb{R}$ such that $\bar{z}' = \lambda x'$. Now, from (3.7) we then have $\|\bar{z}\|_p \geq \|\bar{z}'\|_p$. Since $\bar{z}$ is arbitrary, it follows that $\|\Pi_{\mathcal{L}'|\mathcal{L}''}\|_{p\text{-ind}} \leq 1$.

(only if). By Lemma 3.1, there exists $w \in \mathcal{X}$, with $\langle w, x' \rangle \neq 0$ such that $\mathcal{L}''$ is given by (3.5). Since $\langle w, x' \rangle$ is nonzero, we may as well assume that

$$\|x'\|_p = \langle w, x' \rangle$$

By complimentary of $(\mathcal{L}', \mathcal{L}'')$ we can decompose uniquely any $z \in \mathcal{X}$ in terms of $z = z' + z''$ where $z' = \lambda x'$ for $\lambda \in \mathbb{R}$ and $z'' \in \mathcal{L}$. Since $\|\Pi_{\mathcal{L}'|\mathcal{L}''}\|_{p\text{-ind}} \leq 1$ we have $\|z\|_p \geq \|z'\|_p$. Then

$$\|z\|_p \geq \|\lambda x'\|_p \geq \lambda \|x'\|_p = \lambda \langle w, x' \rangle = \langle w, \lambda x' \rangle = \langle w, z \rangle$$

From this point we have obtained $\langle w, x' \rangle = \|x'\|_p$ and $\langle w, z \rangle \leq \|z\|_p$ for all $z \in \mathcal{X}$. Consequently, by Lemma 3.2, $w \in \nabla \|x'\|_p$ as desired. ■

The following theorem is an immediate consequence of Theorem 3.1 and Proposition 3.3.

Theorem 3.2 *Let $M : \mathcal{X} \to \mathcal{Y}$ have rank n. For every p there exists $M^* : \mathcal{X} \to \mathcal{Y}$ with $\operatorname{rank} M^* < n$ such that*

$$\|M - M^*\|_{p\text{-}ind} = \min_{\operatorname{rank} M' \leq n-1} \|M - M'\|_{p\text{-}ind} = \sigma_n^{(p)}.$$

Moreover, any M^* given by $M^*|_{\mathcal{L}'} = 0$ with $\mathcal{L}' = \operatorname{span}(x') \in \mathbb{L}_n$ and $M^*|_{\mathcal{L}''} = M|_{\mathcal{L}''}$ with $\mathcal{L}''$ given by (3.5) with $w \in \nabla \|x'\|_p$ is an optimal approximant of rank $< n$.

At this stage it is unclear whether for arbitrary p, the p-induced singular values $\sigma^{(p)}(M)$ precisely characterize the minimal achievable approximation errors in that

$$\min_{\operatorname{rank}(M') \leq k} \|M - M'\|_{p\text{-ind}} = \sigma_{k+1}^{(p)}$$

holds for all k. This question is currently under investigation.

3.5 Optimal System Identification

Consider the optimal system identification formulated in Section 3.3. Let $\tilde{w}(t)$, $t = 0, \ldots N$ be a real valued observed time series of dimension q, i.e.,

$\tilde{w}(t) \in \mathbb{R}^q$, and suppose that we wish to find an optimal autoregressive model

$$\sum_{i=0}^{n} x_i w(t+i) = 0$$

where the model coefficients x_i are row vectors of dimension q, and n is the model order. Let $x = (x_0, \ldots, x_n)^\top \in \mathbb{R}^q(n+1)$ denote the model coefficient vector and set

$$M = \begin{pmatrix} \tilde{w}^\top(0) & \cdots & \tilde{w}^\top(n) \\ \vdots & & \vdots \\ \tilde{w}^\top(N-n) & \cdots & \tilde{w}^\top(N) \end{pmatrix}.$$

It is immediate that the misfit

$$\mu(x, \tilde{w}) = \frac{\|Mx\|_p}{\|x\|_p}.$$

Consequently, $\mathcal{L} \subseteq \mathbb{R}^{q(n+1)}$ satisfies $\mu(\cdot, \tilde{w})|_{\mathcal{L}} \leq \varepsilon$ if and only if $\|M|_{\mathcal{L}}\|_{p\text{-ind}} \leq \varepsilon$. Hence, by definition, all subsets $\mathcal{L} \subseteq \mathbb{R}^{q(n+1)}$ with this property are characterized by $\mathcal{L} \in \mathbb{L}_j^{(p)}$, $j \geq k$ where k is such that

$$\sigma_{k-1}^{(p)}(M) > \varepsilon \geq \sigma_k^{(p)}(M). \tag{3.8}$$

This proves the following result:

Theorem 3.3 *If k satisfies* (3.8)*, then all $x \in \mathcal{L}$ with $\mathcal{L} \in \mathbb{L}_j^{(p)}$, $j \geq k$, solve the identification problem in that the misfit*

$$\mu(x, \tilde{w}) \leq \varepsilon.$$

The identification problem has no solution if no such k exists. Furthermore, every $x^* \in \mathcal{L}$ with $\mathcal{L} \in \mathbb{L}_{q(n+1)}^{(p)}$ defines an optimal *model of (minimal) misfit $\mu(x^*, \tilde{w}) = \sigma_{n+1}^{(p)}$.*

Note that this result provides a complete solution to the system identification problem for any p.

Chapter 4

$\mathcal{L}_2$ Gain Approximation of Nonlinear Systems: a Heuristic Approach

4.1 Introduction

Modelling of physical systems may result in high dimension of states. In many cases high dimension of states in a model is not desirable as analysis can be much more difficult, not to mention synthesis for control. This has prompted people to reduce the number of states by removing some of the states, under certain mechanisms, while keeping the reduced model as close as possible to the original one. Many techniques of reducing high dimensional system have been developed for linear systems. Most of the works in linear systems are related to balanced truncation [25, 29], Hankel norm approximation [11] and $\mathcal{H}_\infty$-norm model approximation [13, 17]. On the other hand, there are not many results for nonlinear version despite the fact that many models in industry are nonlinear with large number of states.

While model reduction for nonlinear systems is essential for a wide range of applications, it is still far away from maturity as the subject is more difficult to understand and to solve than that of linear systems. As such, there are only a few results on scheme for constructing a reduced order model for nonlinear system, such as [20], [33] and [38].

In this chapter a computational method for obtaining a reduced order model for a class of nonlinear system with polynomial vector fields is presented.

The origin of the unforced original nonlinear system is globally asymptotically stable. First a necessary condition for the system to have a reduced model is derived. The result only holds under a strict condition in that the error model should have strong $\mathcal{H}_\infty$ performance with a certain property. The necessary condition is given by the existence of generalized reachability and observability functions obtained from the same type of Hamilton-Jacobi inequalities described in [33]. From this necessary condition an approach to construct a reduced model such that the error model is finite gain $\mathcal{L}_2$ stable is deduced. The method is heuristic in nature as one of the step of construction is based on that necessary condition. Thus if the condition is satisfied it does not mean that a reduced model exists and the method may sometimes fail to give a reduced model. The advantage of this method is that it benefits from the use of sum of squares programming as the reachability and observability functions to be computed should satisfy the same type of Hamilton-Jacobi inequalities like in [33]. Moreover under an additional condition the origin of the resulting reduced order model is guaranteed to be globally asymptotically stable whenever the input is zero.

4.2 $\mathcal{L}_2$ Gain Approximation

This chapter is concerned with a polynomial nonlinear system

$$\dot{x} = f(x) + B(x)\,u, \tag{4.1a}$$
$$y = h(x) + D(x)\,u, \tag{4.1b}$$

where $x = [x_1, \ldots, x_n]^T \in \mathbb{R}^n$ is the state vector of the system, $u \in \mathbb{R}^{n_u}$ is the input to the system and $y \in \mathbb{R}^{n_y}$ is the output of the system. We assume the functions f and h can be expressed in the form

$$f(x) = A(x)\,x,$$
$$h(x) = C(x)\,x,$$

with $A(x) \in \mathbb{R}^{n\times n}[x]$, $C(x) \in \mathbb{R}^{n_y\times n}[x]$ are polynomial matrices in x while $B(x) \in \mathbb{R}^{n\times n_u}[x]$ and $D(x) \in \mathbb{R}^{n_y\times n_u}[x]$ are also polynomial matrices in x, and therefore smooth. In this case ways to represent $A(x)$ and $C(x)$ are not unique. Furthermore, we assume that the origin is globally asymptotically stable (in the sense of Lyapunov) whenever $u = 0$.

We consider a reduced order model

$$\dot{x}_r = A_r(x_r)\,x_r + B_r(x_r)\,u, \tag{4.2a}$$
$$y_r = C_r(x_r)\,x_r + D_r(x_r)\,u, \tag{4.2b}$$

where $x_r = [x_{r1}, \ldots, x_{rn_r}]^T \in \mathbb{R}^{n_r}$ with $n_r < n$, $A_r(x_r) \in \mathbb{R}^{n_r \times n_r}[x_r]$, $B_r(x_r) \in \mathbb{R}^{n_r \times n_u}[x_r]$, $C_r(x_r) \in \mathbb{R}^{n_y \times n_r}[x_r]$ and $D_r(x_r) \in \mathbb{R}^{n_y \times n_u}[x_r]$. The error system is given by

$$\dot{\chi} = \mathcal{F}(\chi) + \mathcal{B}(\chi) u, \tag{4.3a}$$
$$e = \mathcal{H}(\chi) + \mathcal{D}(\chi) u, \tag{4.3b}$$

where

$$\chi = \begin{bmatrix} x \\ x_r \end{bmatrix},\ \mathcal{F}(\chi) = \mathcal{A}(\chi)\chi,\ \mathcal{H}(\chi) = \mathcal{C}(\chi)\chi,$$
$$\mathcal{A}(\chi) = \begin{bmatrix} A(x) & 0 \\ 0 & A_r(x_r) \end{bmatrix},\ \mathcal{B}(\chi) = \begin{bmatrix} B(x) \\ B_r(x_r) \end{bmatrix},$$
$$\mathcal{C}(\chi) = \begin{bmatrix} C(x) & -C_r(x_r) \end{bmatrix},\ \mathcal{D}(\chi) = D(x) - D_r(x_r).$$

This chapter aims at obtaining a reduced order model (4.2) such that $e \in \mathcal{L}_2[0,T]$ whenever $u \in \mathcal{L}_2[0,T]$ for $T \in [0,\infty)$. The quality of the approximant (4.2), in this case, is quantified by means of $\mathcal{L}_2$ gain of the error system (4.3). The $\mathcal{L}_2$ gain is defined as follows.

Definition 4.1 *[36] The error system (4.3) with $\chi(0) = 0$ is finite gain $\mathcal{L}_2$ stable with gain at most $\epsilon \geq 0$ if*

$$\int_0^T \|e(t)\|^2 dt \leq \epsilon^2 \int_0^T \|u(t)\|^2 dt$$

for all $u \in \mathcal{L}_2[0,T]$ and $T \in [0,\infty)$.

Throughout the chapter, we assume that $\epsilon^2 I_{n_u} - \mathcal{D}(\chi)^T \mathcal{D}(\chi) \succ 0$ for all χ is always satisfied for the given ϵ.

The following condition is sufficient for the error system to be finite gain $\mathcal{L}_2$ stable. Though the condition is only sufficient for finite gain $\mathcal{L}_2$ stable we will employ this condition throughout the chapter because it has a nice structure which can be exploited for computational purposes.

Proposition 4.1 *[22] System (4.3) is finite gain $\mathcal{L}_2$ stable with gain at most $\epsilon \geq 0$ if there exists a continuosly differentiable storage function $V(\chi)$ such that $\frac{\partial V(\chi)}{\partial \chi} = 2\chi^T M(\chi)$ where $M(\chi) = M(\chi)^T$ is positive definite and satisfies*

$$\begin{bmatrix} \mathcal{A}(\chi)^T M(\chi) + M(\chi)\mathcal{A}(\chi) & M(\chi)\mathcal{B}(\chi) & \mathcal{C}(\chi)^T \\ \mathcal{B}(\chi)^T M(\chi) & -\epsilon^2 I_{n_u} & \mathcal{D}(\chi)^T \\ \mathcal{C}(\chi) & \mathcal{D}(\chi) & -I_{n_y} \end{bmatrix} \preceq 0 \tag{4.4}$$

for all $\chi \in R^{n+n_r}$.

Proof. The short version of the proof is as follows. The sufficient condition implies

$$\frac{\partial V(\chi)}{\partial \chi}\mathcal{F}(\chi) + \mathcal{H}(\chi)^T \mathcal{H}(\chi) + \left(\frac{1}{2}\frac{\partial V(\chi)}{\partial \chi}\mathcal{B}(\chi) + \mathcal{H}(\chi)^T \mathcal{D}(\chi)\right) \times$$
$$\left(\epsilon^2 I_{n_u} - \mathcal{D}(\chi)' \mathcal{D}(\chi)\right)^{-1} \left(\frac{1}{2}\mathcal{B}(\chi)^T \frac{\partial V(\chi)}{\partial \chi}^T + \mathcal{D}(\chi)^T \mathcal{H}(\chi)\right) \leq 0 \quad (4.5)$$

for all χ. By completion of square, (4.5) implies that the system (4.3) is finite gain $\mathcal{L}_2$ stable with gain at most $\epsilon \geq 0$ ■

In [22], whenever $\epsilon = 1$, this sufficient condition for finite gain $\mathcal{L}_2$ stability is referred to as strong $\mathcal{H}_\infty$ performance. Throughout, the system (4.3) is said to have strong $\mathcal{H}_\infty$ performance of gain ϵ whenever the condition in Proposition 4.1 is satisfied for a prescribed ϵ which is not necessarily one.

We need to point out that any positive definite $M(\chi) = M(\chi)^T$ which satisfies (4.4) does not automatically guarantee that the system (4.3) is finite gain $\mathcal{L}_2$ stable. Additionally, we require that the vector field $\vartheta(x) = 2M(\chi)\chi$ should be conservative [19], that is there exists a scalar potential $V(\chi)$ such that

$$\vartheta(\chi) = \frac{\partial V(\chi)}{\partial \chi}^T.$$

The following condition is necessary and sufficient for a vector field to be conservative.

Proposition 4.2 *[22] Suppose* $\vartheta(\chi) = \left[\begin{array}{ccc} \vartheta_1(\chi) & \dots & \vartheta_N(\chi) \end{array}\right]^T$ *belongs to class* C^k *for some positive integer* k. *Then there exists a* C^{k+1} *function* $V(\chi)$ *such that*

$$\frac{\partial V(\chi)}{\partial \chi} = \vartheta(\chi)^T$$

if and only if

$$\frac{\partial \vartheta_i(\chi)}{\partial \chi_j} = \frac{\partial \vartheta_j(\chi)}{\partial \chi_i}$$

for all χ *and* $i, j = 1, \dots, N$. *In this case* V *is given by*

$$V(\chi) = \chi^T \int_0^1 \vartheta(\tau\chi)\, d\tau$$

where $V(0) = 0$. *In addition, if* $\vartheta(x) = 2M(\chi)\chi$ *where* $M(\chi)$ *is positive definite then* $V(\chi)$ *is also positive definite function.*

4.3 A Necessary Characterization

In linear systems a necessary and sufficient condition for $\mathcal{H}_\infty$ model reduction is given by the existence of the generalized gramians as described in Proposition 2.8. Motivated by the relation of the generalized gramians of linear systems with the $\mathcal{H}_\infty$ model reduction problem in linear system as stated in Proposition 2.8, the relation of generalized functions of nonlinear systems with the strong $\mathcal{H}_\infty$ model reduction problem in nonlinear systems will be shown. In this case a necessary condition for the nonlinear system (4.1) to have a reduced order model (4.2) with the error system (4.3) achieving strong $\mathcal{H}_\infty$ performance is presented.

Definition 4.2 *[33] A positive definite polynomial function $L_o(x)$ with $L_o(0) = 0$ is a generalized observability function to the system (4.1) if it satisfies*

$$\frac{\partial L_o(x)}{\partial x} f(x) + \frac{1}{2} h^T(x) h(x) \leq 0 \tag{4.6}$$

for all $x \in D_x$.

Definition 4.3 *[33] A positive definite polynomial function $L_c(x)$ with $L_c(0) = 0$ is a generalized reachability function to the system (4.1) if it satisfies*

$$\frac{\partial L_c(x)}{\partial x} f(x) + \frac{\partial L_c(x)}{\partial x} B(x) u - \frac{1}{2} u^T u \leq 0 \tag{4.7}$$

for all $x \in D_x$ and $u \in \mathbb{R}^{n_u}$.

Equivalently, (4.7) can be expressed by

$$\frac{\partial L_c(x)}{\partial x} f(x) + \frac{1}{2} \frac{\partial L_c(x)}{\partial x} B(x) B(x)^T \frac{\partial L_c(x)}{\partial x}^T \leq 0 \quad \forall x \in D_x. \tag{4.8}$$

Throughout this chapter, we set $D_x = \mathbb{R}^n$. The degree of the polynomial generalized functions $L_o(x)$ and $L_c(x)$ should not be less than two to guarantee positive definiteness and being vanished at the origin. The result of the following holds under strict condition on the structure of matrix M.

Theorem 4.1 *Suppose that the system (4.3) has strong $\mathcal{H}_\infty$ performance of gain ϵ with*

$$M(x) = M(x)^T = \begin{bmatrix} X(x) & X_2 \\ X_2^T & X_3 \end{bmatrix} \succ 0$$

where $X_2 \in \mathbb{R}^{n\times n_r}$, $X_3 = X_3^T \in \mathbb{R}^{n_r\times n_r}$, $X(x) = X(x)^T \in \mathbb{R}^{n\times n}[x]$ *with* $x^T X(x)$ *being a conservative vector field then there exist a generalized observability function* $L_o(x)$ *and a generalized reachability function* $L_c(x)$ *to the system (4.1).*

Proof. The proof is a generalization of the proof of necessary condition in Proposition 2.8. The proof can also be seen as a special case of the proof of solvability conditions for the strong $\mathcal{H}_\infty$ control problem in [22]. The inequality

$$\begin{bmatrix} \mathcal{A}(\chi)^T M(x) + M(x)\mathcal{A}(\chi) & M(x)\mathcal{B}(\chi) & \mathcal{C}(\chi)^T \\ \mathcal{B}(\chi)^T M(x) & -\epsilon^2 I_{n_u} & \mathcal{D}(\chi)^T \\ \mathcal{C}(\chi) & \mathcal{D}(\chi) & -I_{n_y} \end{bmatrix} \preceq 0$$

can be expressed as

$$\Phi(\chi) := H(x) + Q^T J_r(x_r) P + P^T J_r(x_r) Q \preceq 0$$

where

$$H(x) = \begin{bmatrix} X(x)A(x) + A(x)^T X(x) & A(x)^T X_2 & X(x)B(x) & C(x)^T \\ X_2^T A(x) & 0 & X_2^T B(x) & 0 \\ B(x)^T X(x) & B(x)^T X_2 & -\epsilon^2 I_{n_u} & D(x)^T \\ C(x) & 0 & D(x) & -I_{n_y} \end{bmatrix},$$

$$P = \begin{bmatrix} 0 & I_{n_r} & 0 & 0 \\ 0 & 0 & I_{n_u} & 0 \end{bmatrix},$$

$$Q = \begin{bmatrix} X_2^T & X_3 & 0 & 0 \\ 0 & 0 & 0 & -I_{n_y} \end{bmatrix},$$

$$J_r(x_r) = \begin{bmatrix} A_r(x_r) & B_r(x_r) \\ C_r(x_r) & D_r(x_r) \end{bmatrix}.$$

Let the inverse of $M(x)$ be

$$M(x)^{-1} = \begin{bmatrix} Y(x) & Y_2(x) \\ Y_2(x)^T & Y_3(x) \end{bmatrix}.$$

By

$$N_P = \begin{bmatrix} I_n & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & I_{n_y} \end{bmatrix},\ N_Q = \begin{bmatrix} I_n & 0 \\ Y_2(x)^T Y(x)^{-1} & 0 \\ 0 & I_{n_u} \\ 0 & 0 \end{bmatrix},$$

which are full rank and satisfy

$$PN_P = 0,$$
$$QN_Q = 0,$$

it is necessary that

$$N_P^T \Phi(\chi) N_P = N_P^T H(x) N_P \preceq 0,$$
$$N_Q{}^T \Phi(\chi) N_Q = N_Q{}^T H(x) N_Q \preceq 0.$$

From

$$N_P^T H(x) N_P = \begin{bmatrix} A(x)^T X(x) + X(x) A(x) & C(x)^T \\ C(x) & -I_{n_y} \end{bmatrix},$$
$$N_Q{}^T H(x) N_Q = \begin{bmatrix} A(x)^T Y(x)^{-1} + Y(x)^{-1} A(x) & Y(x)^{-1} B(x) \\ B(x)^T Y(x)^{-1} & -\epsilon^2 I_{n_u} \end{bmatrix},$$

and by Schur complement [6] we obtain

$$A(x)^T X(x) + X(x) A(x) + C(x)^T C(x) \preceq 0,$$

$$A(x)^T Y(x)^{-1} + Y(x)^{-1} A(x) + \frac{1}{\epsilon^2} Y(x)^{-1} B(x) B(x)^T Y(x)^{-1} \preceq 0,$$

which imply

$$x^T A(x)^T X(x) x + x^T X(x) A(x) x + x^T C(x)^T C(x) x \leq 0,$$

$$x^T A(x)^T Y(x)^{-1} x + x' Y(x)^{-1} A(x) x + \frac{1}{\epsilon^2} x^T Y(x)^{-1} B(x) B(x)^T Y(x)^{-1} x \leq 0.$$

Next we define the positive semidefinite matrix Λ which has rank at most n_r

$$\Lambda := X(x) - Y(x)^{-1} = X_2 X_3^{-1} X_2^T.$$

From the fact that $x^T X(x)$ is conservative and $X(x)$ is positive definite then there exists a positive definite function $L_1(x)$ such that $\frac{\partial L_1(x)}{\partial x} = x^T X(x)$. By defining

$$L_o(x) := L_1(x),$$
$$L_c(x) := \frac{1}{\epsilon^2} L_1(x) - \frac{1}{2\epsilon^2} x^T \Lambda x,$$

we have

$$\frac{\partial L_o(x)}{\partial x} = x^T X(x),$$
$$\frac{\partial L_c(x)}{\partial x} = \frac{1}{\epsilon^2} x^T (X(x) - \Lambda) = \frac{1}{\epsilon^2} x^T Y(x)^{-1},$$

and thus the result follows. ■

Theorem 4.1 basically says that if there exists a reduced order model (4.2) for the nonlinear system (4.1) with the error system (4.3) having strong $\mathcal{H}_\infty$ performance of gain ϵ with a particular structure of M then there always exist generalized reachability and observability functions $L_c(x)$ and $L_o(x)$, respectively satisfying the Hamilton-Jacobi inequalities (4.6-4.7). Here, the structure of $L_c(x)$ and $L_o(x)$ is related to the structure of $M(x)$ as shown in the proof.

It is important to point that Theorem 4.1 is quite conservative in that we do not consider a general structure of M. For simplification, the matrix M is dependent only on x but not on x_r. Moreover it is only the matrix X which is dependent on x while X_2 and X_3 are independent of x. Hence, it is sufficient to require $x^T X(x)$ to be a conservative vector field to guarantee the existence of a storage function V.

A truncation scheme based on generalized reachability function $L_c(x)$ and generalized observability function $L_o(x)$ satisfying (4.6-4.7) is introduced in [33]. In the scheme the search of the functions $L_c(x)$ and $L_o(x)$ are completely independent with each other. In our case in Theorem 4.1 necessarily the structure of $L_c(x)$ and $L_o(x)$ is identical modulo the quadratic form $x^T \Lambda x$ due to simplification of the matrix M. In this case the matrix Λ is dependent on X_2 and X_3 which are the elements of M while the order n_r of the reduced model will be the upper bound for the rank of the matrix Λ.

A natural question arises whether, for a prescribed ϵ, there exists any reduced model (4.2) where its error system (4.3) satisfies (4.5) if there exist positive definite functions $L_c(x)$ and $L_o(x)$ satisfying (4.6-4.7) without any additional requirement on the structure of $L_c(x)$ and $L_o(x)$. And if the answer is negative what kind of additional condition on the structure of $L_c(x)$ and $L_o(x)$ should be imposed so that a reduced model exists. This is a very difficult question and any answer on this will be paving the way towards a better construction mechanism of a reduced model for the nonlinear system (4.1). So far, we can only provide necessary characterization of a nonlinear system which has a reduced model when the error system has strong $\mathcal{H}_\infty$ performance with a certain structure of M, as described in Theorem 4.1.

4.4 A Heuristic Approach

Construction of a reduced model is difficult in general. In this section we propose a heuristic for constructing a reduced model (4.2) for the nonlinear system (4.1) of order n where the order of the reduced model is $n_r < n$ such that (4.5) is satisfied for a prescribed ϵ. The approach which is deduced from Theorem 4.1 is given as follows.

1. Find a positive definite function $L(x)$ and a positive semidefinite matrix $\Lambda \in R^{n\times n}$ whose rank is at most $n_r < n$ such that

$$L_c(x) = L(x),$$
$$L_o(x) = \epsilon^2 L(x) + \frac{1}{2}x^T \Lambda x,$$

   and (4.6)-(4.7) are satisfied.

2. The matrix $X_2 \in \mathbb{R}^{n\times n_r}$ can be obtained from

$$X_2 X_2^T = \Lambda.$$

   and we set $X_3 = I_{n_r}$.

3. Define $V(\chi) := 2L_o(x) + 2x^T X_2 x_r + x_r^T x_r$ and construct $A_r(x_r)$, $B_r(x_r)$, $C_r(x_r)$, $D_r(x_r)$ satisfying (4.5).

In step 1 we employ the condition in Theorem 4.1 where we put a structure on $L_c(x)$ and $L_o(x)$ in order to give the same structure of M as in Theorem 4.1. The constraint on the structure of $L_c(x)$ and $L_o(x)$ makes step 1 only work for a very restricted class of system. Moreover, as the nature of the condition in Theorem 4.1 is of necessity it is important to note that even if step 1 is satisfied it does not mean that the existence of a reduced model is guaranteed.

In step 3 we use another condition instead of the condition of strong $\mathcal{H}_\infty$ performance. Condition (4.5) is sufficient for the system (4.3) to be finite gain $\mathcal{L}_2$ stable with gain at most $\epsilon \geq 0$. But (4.5) is also a necessary condition for strong $\mathcal{H}_\infty$ performance of gain ϵ. Thus (4.5) is a better inequality to characterise $\mathcal{L}_2$ gain stability of (4.3).

It is important to note that feasibility test of (4.5), (4.6) and (4.7) is a hard problem for computation. Yet these inequalities have advantage from computational point of view as they can be relaxed by means of sum of squares

programming. It is very obvious that a polynomial expressed as a sum of squares of other polynomials is nonnegative everywhere. The works of [30] have showed that determining whether a polynomial can be expressed as a sums of squares is an LMI problem. Thus the problem of testing whether a polynomial is sum of squares becomes relatively easy while testing nonnegativity of a polynomial is a hard problem. (A more detailed discussion on sum of squares programming can be read from [30].) Within this direction the requirements of (4.5), (4.6) and (4.7) can be relaxed by means of certain polynomials being sum of squares (SOS). The relaxation is given as follows.

1. Find a positive definite $L(x)$ and a positive semidefinite matrix $\Lambda \in R^{n\times n}$ whose rank is at most $n_r < n$ such that

$$L_c(x) = L(x),$$
$$L_o(x) = \epsilon^2 L(x) + \frac{1}{2}x^T \Lambda x,$$

and

$$-\frac{\partial L_o(x)}{\partial x} f(x) - \frac{1}{2} h(x)^T h(x) \text{ is SOS}, \tag{4.9}$$
$$v^T \begin{bmatrix} -\frac{\partial L_c(x)}{\partial x} f(x) & \frac{\partial L_c(x)}{\partial x} B(x) \\ B(x)^T \frac{\partial L_c(x)}{\partial x}^T & 2I_{n_u} \end{bmatrix} v \text{ is SOS}, \tag{4.10}$$

where $v \in \mathbb{R}^{1+n_u}$.

2. The matrix $X_2 \in \mathbb{R}^{n\times n_r}$ can be obtained from

$$X_2 X_2^T = \Lambda.$$

and we set $X_3 = I_{n_r}$.

3. Define $V(\chi) := 2L_o(x) + 2x^T X_2 x_r + x_r^T x_r$ and construct $A_r(x_r)$, $B_r(x_r)$, $C_r(x_r)$, $D_r(x_r)$ such that

$$-w^T \begin{bmatrix} \frac{\partial V(\chi)}{\partial \chi}\mathcal{F}(\chi) & \frac{1}{2}\frac{\partial V(\chi)}{\partial \chi}\mathcal{B}(\chi) & \mathcal{H}(\chi)^T \\ \frac{1}{2}\mathcal{B}(\chi)^T \frac{\partial V(\chi)}{\partial \chi}^T & -\epsilon^2 I_{n_u} & \mathcal{D}(\chi)^T \\ \mathcal{H}(\chi) & \mathcal{D}(\chi) & -I_{n_y} \end{bmatrix} w$$
$$\text{is SOS for } w \in \mathbb{R}^{1+n_u+n_y}. \tag{4.11}$$

The requirements (4.10) of step 1 and (4.11) of step 3 are based on the following. If a polynomial $w^T \Psi(\chi) w$ is SOS, where $\Psi(\chi)$ is a matrix whose

entries are polynomials, then $\Psi(\chi) \succeq 0$ for all χ [32]. By Schur complement [6] the inequalities

$$\begin{bmatrix} \frac{\partial L_c(x)}{\partial x} f(x) & \frac{\partial L_c(x)}{\partial x} B(x) \\ B(x)^T \frac{\partial L_c(x)}{\partial x}^T & -2I_{n_u} \end{bmatrix} \preceq 0,$$

$$\begin{bmatrix} \frac{\partial V(\chi)}{\partial \chi} \mathcal{F}(\chi) & \frac{1}{2} \frac{\partial V(\chi)}{\partial \chi} \mathcal{B}(\chi) & \mathcal{H}(\chi)^T \\ \frac{1}{2} \mathcal{B}(\chi)^T \frac{\partial V(\chi)}{\partial \chi}^T & -\epsilon^2 I_{n_u} & \mathcal{D}(\chi)^T \\ \mathcal{H}(\chi) & \mathcal{D}(\chi) & -I_{n_y} \end{bmatrix} \preceq 0,$$

are equivalent to (4.8) and (4.5), respectively.

The rank condition of Λ in step 1 is hard to compute as it is nonconvex. General computational procedure for rank condition is still an open problem and will not be covered in this chapter. Though it will bring more conservativeness, to replace nonconvexity of rank condition we can set

$$\Lambda = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Lambda_1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where $\Lambda_1 \in \mathbb{R}^{n_r \times n_r}$ is positive semidefinite.

To determine stability properties of the reduced model it follows from (4.5) that

$$\frac{\partial V(\chi)}{\partial \chi} \mathcal{F}(\chi) \leq -\mathcal{H}(\chi)^T \mathcal{H}(\chi) \tag{4.12}$$

for all χ. For $x = 0$, (4.12) gives

$$\frac{\partial v(x_r)}{\partial x_r} f_r(x_r) \leq -h_r(x_r)^T h_r(x_r)$$

for all x_r, where $v(x_r) = x_r^T x_r$ and

$$f_r(x_r) = A_r(x_r) x_r, \; h_r(x_r) = C_r(x_r) x_r.$$

Invoking LaSalle's invariance principle [18] if we require that no solution of $\dot{x}_r = f_r(x_r)$ can stay identically in

$$S = \{x_r \in \mathbb{R}^{n_r} |\; h_r(x_r) = 0\}$$

other than the trivial solution $x_r(t) \equiv 0$ then the origin of $\dot{x}_r = f_r(x_r)$ is globally asymptotically stable.

4.5 Example

In this section a numerical example is given to illustrate the applicability of the proposed approach. Consider the system

$$f(x) = \begin{bmatrix} -x_1 - x_2 \\ 2x_1 - 3x_2 - x_2^3 \end{bmatrix}, \; g(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$
$$h(x) = 2x_2 - x_1,$$

with $L(x)$ and Λ in the following structure

$$L(x) = \alpha_1 x_1^2 + \alpha_2 x_2^2,$$
$$\Lambda = \begin{bmatrix} 0 & 0 \\ 0 & \gamma \end{bmatrix}.$$

The equilibrium point at the origin is globally asymptotically stable when $u = 0$. We want to construct a reduced order model of one dimension with the structure

$$A_r = a_1 + a_2 x_r^2, \quad B_r = b,$$
$$C_r = c, \quad D_r = d,$$

where $a_1, a_2, b, c, d \in \mathbb{R}$. Feasibility tests of (4.9), (4.10) and (4.11) are carried out using SOS programming tools [23, 31]. For $\epsilon = 0.2$ we obtain $A_r = -3.7141 - 0.7793x_r^2$, $B_r = -1.2387$, $C_r = -1.8926$, $D_r = 0.0022$ with

$$L = 8.4829x_1^2 + 1.3206x_2^2,$$
$$\gamma = 1.5203,$$

The origin of the reduced model is globally asymptotically stable whenever $u = 0$. The response of the system and the reduced model to inputs $u = e^{-1.5t}sin(1.5t)$ can be seen in Figure 4.1.

For comparison we consider linearization of the system around the origin. The linearized system is controllable and observable. Now we resort to $\mathcal{H}_\infty$ model reduction where we are to compute a reduced model of one dimension for the linearized system by means of Proposition 2.8. All the conditions in Proposition 2.8 can be casted in terms of LMIs but the rank condition. In this case a numerical scheme based on alternating projection [13] is used to approach the rank condition. As the algorithm will not give accurate results for $\hat{\epsilon} < 0.06$, the reduced model is computed for $\hat{\epsilon} = 0.06$. The resulting reduced model gives a good approximation for the original nonlinear system when it is excited around the origin. But for the region which include nonlinearity the reduced model shows poor performance as shown in Figure 4.2.

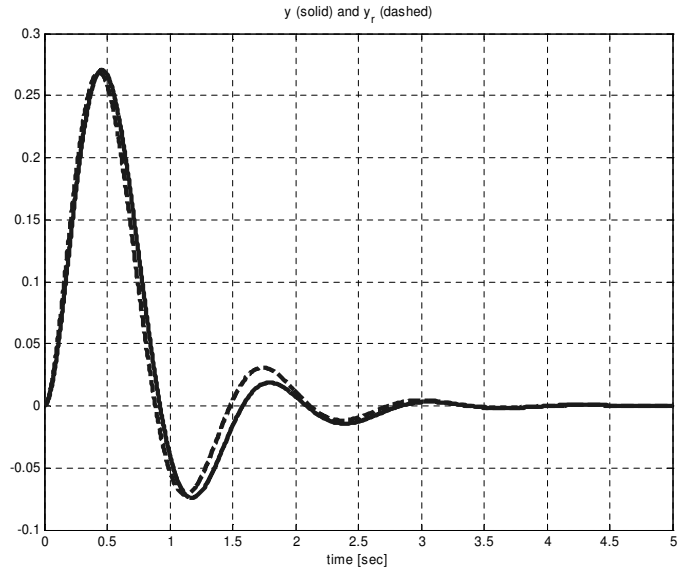

Figure 4.1: Response of the output to the input $u = e^{-1.5t}sin(1.5t)$

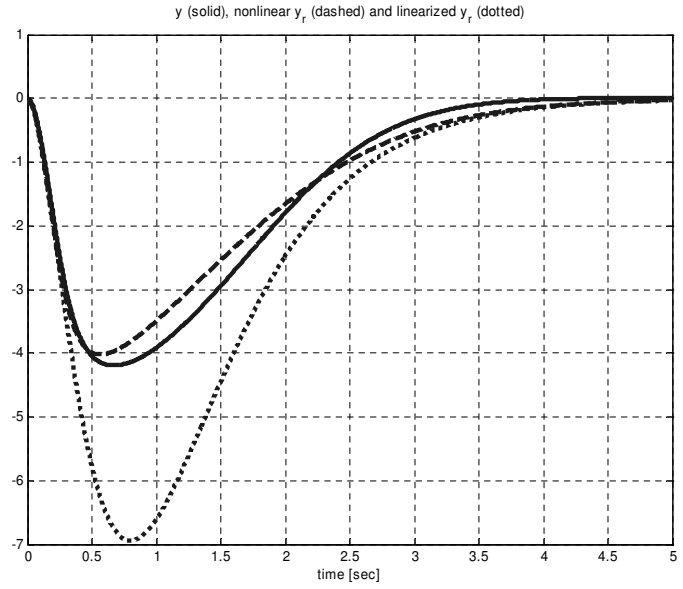

Figure 4.2: Response of the output to the input $u = 50e^{-3t} - 50e^{-1.5t}$

Chapter 5

Reachability-Based Approach for $\mathcal{L}_2$ Gain Approximation of Polynomial Systems

5.1 Introduction

In attempting to utilize the power of sum of squares programming a heuristic approach is introduced in the previous chapter. The approach computes a reduced model for the polynomial nonlinear system such that the error model satisfies a finite gain $\mathcal{L}_2$ stability condition. Though verification of this condition can be done through sum of squares programming, the method suffers from the coupling of the unknown storage function with the unknown structure of the reduced model which makes the computation untractable. To avoid this coupling of the unknown variables, the generalized reachability and observability functions which satisfy the same type of Hamilton-Jacobi inequalities like in [33] are computed through sum of squares programming. Based on the generalized functions, a storage function is constructed. The reduced model can then be computed such that the error model satisfies the finite gain $\mathcal{L}_2$ stability condition.

In this chapter we try a different approach to decouple the unknown variables for verifying the finite gain $\mathcal{L}_2$ stability condition. Instead of constructing the storage function as the first step to avoid the coupling of the unknown

variables, we construct partially the structure of the reduced model which is coupled with the storage function. The construction is based on an estimate of the reachability set of the system when the initial condition is set to the origin. In this case we seek the part of the system which is strongly reachable. This part of the system will be the state space of the reduced model while the output of the reduced model is obtained through sum of squares programming of relaxation of the finite gain $\mathcal{L}_2$ stability condition.

The method in this chapter is restricted to a certain class of polynomial nonlinear system (4.1) where $B(x) = B \in \mathbb{R}^{n\times n_u}$, $D(x) = D \in \mathbb{R}^{n_y\times n_u}$ are constant matrices and $h(x) = Cx$ with $C \in \mathbb{R}^{n_y\times n}$. In this case we can write the system

$$\dot{x} = f(x) + Bu, \tag{5.1a}$$
$$y = Cx + Du. \tag{5.1b}$$

We assume the following about the system.

Assumption 5.1 *There exists $Q = Q^T \succ 0$ such that*

$$x^T f(x) \leq -x^T Q x$$

for all $x \in \mathbb{R}^n$.

By the assumption, the origin of the unforced system is globally asymptotically stable. For the linear system $f(x) = Ax$ the assumption means that $A + A^T$ should be negative definite. Assumption 5.1 is needed to guarantee that the model reduction method in this chapter preserves global asymptotic stability.

The reduced order model to be computed is in the form (4.2) where $B_r(x_r) = B_r \in \mathbb{R}^{n_r\times n_u}$, $D_r(x_r) = D_r \in \mathbb{R}^{n_y\times n_u}$ are constant matrices. In this case we can write the reduced order model

$$\dot{x}_r = f_r(x_r) + B_r u, \tag{5.2a}$$
$$y_r = C_r x_r + D_r u. \tag{5.2b}$$

As in the previous chapter we also use the following assumption.

Assumption 5.2 $\epsilon^2 I_{n_u} - \mathcal{D}^T\mathcal{D} \succ 0$.

5.1.1 Estimate of the Reachable Set

Consider the inequality (4.7) or, equivalently

$$\frac{\partial L_c(x)}{\partial x} f(x) + \frac{\partial L_c(x)}{\partial x} Bu \leq \frac{1}{2} u^T u \quad \forall (x,u)$$

where the function $L_c(x)$ is a positive definite polynomial in x with $L_c(0) = 0$. By setting $x(0) = 0$ we obtain

$$L_c(x(t)) \leq \frac{1}{2} \int_0^t \|u(\tau)\|^2 \, d\tau.$$

If we denote

$$\mathcal{R}_c(\delta) = \left\{ x \in \mathbb{R}^n | \; L_c(x) \leq \frac{1}{2}\delta \right\}$$

then $\mathcal{R}_c(\delta)$ is a set which contains all reachable states from the origin when the input u to the system

$$\dot{x} = f(x) + Bu, \quad x(0) = 0,$$

satisfies

$$\int_0^\infty \|u(\tau)\|^2 \, d\tau \leq \delta.$$

In this case we can use $\mathcal{R}_c(\delta)$ as an estimate of the reachable set from the origin. It is important to point out that there are many choices for $\mathcal{R}_c(\delta)$ as the function $L_c(x)$ is nonunique. Since the estimates are not unique we may consider the smallest set $\mathcal{R}_c(\delta)$ which contains the reachable set.

5.1.2 Overview of the Approach

For a linear system G with a realization $\{A, B, C, D\}$ where $A + A^T$ is negative definite the error system is given by the realization $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}\}$ where

$$\mathcal{A} = \begin{bmatrix} A & 0 \\ 0 & A_r \end{bmatrix}, \; \mathcal{B} = \begin{bmatrix} B \\ B_r \end{bmatrix},$$
$$\mathcal{C} = \begin{bmatrix} C & -C_r \end{bmatrix}, \; \mathcal{D} = D - D_r,$$

and $\{A_r, B_r, C_r, D_r\}$ is a realization of the reduced model G_r of order $n_r < n$. In this case the $\mathcal{L}_2$ gain approximation problem becomes $\mathcal{H}_\infty$ model reduction problem. Necessary and sufficient condition for the error system

to have the $\mathcal{H}_\infty$-norm at most $\epsilon \geq 0$ is given by the existence of a positive definite matrix $M = M^T \in \mathbb{R}^{(n+n_r)\times(n+n_r)}$ such that [9]

$$\begin{bmatrix} \mathcal{A}^T M + M\mathcal{A} & M\mathcal{B} & \mathcal{C}^T \\ \mathcal{B}^T M & -\epsilon^2 I & \mathcal{D}^T \\ \mathcal{C} & \mathcal{D} & -I \end{bmatrix} \prec 0. \tag{5.3}$$

Indeed, the $\mathcal{H}_\infty$ model reduction problem is to find $\{A_r, B_r, C_r, D_r\}$ and positive definite M such that (5.3) is satisfied for a minimum value of ϵ. But this problem is not easy to solve in terms of computation as the inequality (5.3) is not convex in the unknown variables M, A_r, B_r because of the coupling terms $M\mathcal{A}$ and $M\mathcal{B}$.

Bearing the nonconvexity of our condition in mind we introduce an approach to avoid the problem of the coupling of the unknown variables M, A_r, B_r in (5.3). Our approach is divided into two steps:

1. Compute A_r and B_r based on an estimate of the reachability set.
2. For the given A_r and B_r, compute M, C_r and D_r which satisfy (5.3).

It is important to note that our approach will introduce conservatism as the computation of the unknowns A_r and B_r is based on an estimate instead of the exact reachability set. Moreover A_r, B_r, C_r, D_r and M are not simultaneously computed while minimizing ϵ in (5.3). So in this case the minimum value of ϵ obtained through this scheme is not guaranteed to be optimum.

The rest of the chapter is devoted to discussing this approach. First, we will elaborate this approach for linear systems. By using the same way of reasoning we will extend the use of this approach to polynomial systems in the form (5.1). Indeed, by Schur complement [6] and Assumption 5.2, the inequality

$$\begin{bmatrix} \frac{\partial V(\chi)}{\partial \chi}\mathcal{F}(\chi) & \frac{1}{2}\frac{\partial V(\chi)}{\partial \chi}\mathcal{B} & \mathcal{H}(\chi)^T \\ \frac{1}{2}\mathcal{B}^T\frac{\partial V(\chi)}{\partial \chi}^T & -\epsilon^2 I_{n_u} & \mathcal{D}^T \\ \mathcal{H}(\chi) & \mathcal{D} & -I_{n_y} \end{bmatrix} \preceq 0, \tag{5.4}$$

is equivalent to (4.5). A relaxation of (5.4) in terms of sum of squares is given in the previous chapter, that is

$$-w^T \begin{bmatrix} \frac{\partial V(\chi)}{\partial \chi}\mathcal{F}(\chi) & \frac{1}{2}\frac{\partial V(\chi)}{\partial \chi}\mathcal{B} & \mathcal{H}(\chi)' \\ \frac{1}{2}\mathcal{B}^T\frac{\partial V(\chi)}{\partial \chi}^T & -\epsilon^2 I_{n_u} & \mathcal{D}^T \\ \mathcal{H}(\chi) & \mathcal{D} & -I_{n_y} \end{bmatrix} w \tag{5.5}$$

is SOS for $w \in \mathbb{R}^{1+n_u+n_y}$ and $\chi \in \mathbb{R}^{n+n_r}$.

Yet this relaxation is not possible to verify by means of tractable computation because of the coupling of the unknowns $V(\chi)$, $f_r(x_r)$ and B_r in the terms $\frac{\partial V(\chi)}{\partial \chi}\mathcal{F}(\chi)$ and $\frac{\partial V(\chi)}{\partial \chi}\mathcal{B}$. Hence we face the same type of problem like in linear systems where the coupling of the unknowns renders the computation intractable. Like in linear part our approach to avoid the coupling terms for nonlinear systems is to compute $f_r(x_r)$ and B_r independently from the computational scheme of $V(\chi)$, $h_r(x_r)$ and D_r. To be more precise the approach for the class of nonlinear systems we are considering is given as follow.

1. Compute $f_r(x_r)$ and B_r based on an estimate of the reachability set.
2. For the given $f_r(x_r)$ and B_r, compute $V(\chi), C_r, D_r$ and minimizing ϵ which satisfy (5.5).

5.2 Reachability Based Approach

5.2.1 Linear System

We consider again

$$\dot{x} = Ax + Bu. \tag{5.6}$$

For an estimate of the reachable set from the origin we can select the quadratic $L_c(x) = \frac{1}{2}x^T\hat{Y}^{-1}x$ where $\hat{Y}$ is a symmetric positive definite matrix of size n by n. We can write (4.8) in the form

$$A^T\hat{Y}^{-1} + \hat{Y}^{-1}A + \hat{Y}^{-1}BB^T\hat{Y}^{-1} \preceq 0$$

or equivalently

$$\hat{Y}A^T + A\hat{Y} + BB^T \preceq 0.$$

The estimate of the reachable set from the origin when $\int_0^\infty \|u(\tau)\|^2 d\tau \leq \delta$ is given by

$$\mathcal{R}_c(\delta) = \left\{x \in \mathbb{R}^n |\ x^T\hat{Y}^{-1}x \leq \delta\right\}.$$

Without loosing generality we can set $\delta = 1$ and we denote

$$\mathcal{R}_c = \left\{x \in \mathbb{R}^n |\ x^T\hat{Y}^{-1}x \leq 1\right\}.$$

Since $\hat{Y}^{-1}$ is a symmetric positive definite matrix the set $\mathcal{R}_c$ is a hyper-ellipsoid [5] where the directions and the lengths of its principal axes are

defined by the eigenvectors and the inverse of the square root of the eigenvalues, respectively, of the matrix $\hat{Y}^{-1}$. If we denote T as an orthogonal matrix ($T^TT = I$) whose columns are the normalized eigenvectors of the matrix $\hat{Y}^{-1}$ then we can define a new coordinate system $z = T^{-1}x$ where its main axis coincide with the principal axis of the ellipsoid. By denoting $S = diag\,(\lambda_1, \ldots, \lambda_n)$ where λ_i is the eigenvalue of the matrix $\hat{Y}^{-1}$ and the $i-$th column of the matrix T is the eigenvector with respect to λ_i, we have $\hat{Y}^{-1}T = TS$. Indeed, the length of the axis with respect to the $i-$th eigenvector is equal to $1/\sqrt{\lambda_i}$.

With respect to the new coordinate system we can rewrite the linear system (5.6) in the form

$$\dot{z} = \hat{A}z + \hat{B}u$$

where $\hat{A} = T^{-1}AT$ and $\hat{B} = T^{-1}B$. The estimate of the reachable set in the new coordinate system is given by

$$\mathcal{R}_{c,z} = \left\{ z \in \mathbb{R}^n |\ z^T T^T \hat{Y}^{-1} T z \leq 1 \right\} = \left\{ z \in \mathbb{R}^n |\ z^T S z \leq 1 \right\}.$$

Suppose we order the eigenvalues in such a way that $\lambda_i \leq \lambda_j$ whenever $i \leq j \leq n$. Then from the set $\mathcal{R}_{c,z}$ we may claim that the trajectories of the system are more accumulated around the z_i-axis rather than z_j-axis for $i \leq j$. This forms the foundation of our approach where we remove the axes which are weakly reachable.

Next we partition the part of size n into two parts of size n_r and $n - n_r$ based on the following

$$z = \begin{bmatrix} z_{[1]}^T & z_{[2]}^T \end{bmatrix}^T,$$
$$\hat{A} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix},\ \hat{B} = \begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix},$$
$$S = \begin{bmatrix} \lambda_{[1]} & 0 \\ 0 & \lambda_{[2]} \end{bmatrix},$$

and the system can be expressed as

$$\dot{z}_{[1]} = \hat{A}_{11} z_{[1]} + \hat{A}_{12} z_{[2]} + \hat{B}_1 u,$$
$$\dot{z}_{[2]} = \hat{A}_{21} z_{[1]} + \hat{A}_{22} z_{[2]} + \hat{B}_2 u.$$

Removing the least reachable part $z_{[2]}$ we have the dynamics of our new reduced model $x_r = z_{[1]}$ represented by

$$\dot{x}_r = \hat{A}_{11} x_r + \hat{B}_1 u, \tag{5.7}$$

where

$$\hat{A}_{11} = T_1^T A T_1,\ \hat{B}_1 = T_1^T B.$$

with T being partitioned by

$$T = \begin{bmatrix} T_1 & T_2 \end{bmatrix}. \tag{5.8}$$

To sum up, we have approximated (5.6) with another system of lower dimension given by (5.7) with the argument that the least reachable part in (5.6) is removed from (5.7) while the most influential part in (5.6) is preserved in (5.7).

5.2.2 Extension to Nonlinear Systems

To extend the ideas in linear systems to nonlinear systems we consider again (5.1) with all the assumptions. We assume the existence of a positive definite polynomial function $L_c(x)$ which satisfies (4.8). Associated with $L_c(x)$ we denote the set

$$\mathcal{R}_c = \{x \in \mathbb{R}^n |\ 2L_c(x) \leq 1\}. \tag{5.9}$$

As the function $L_c(x)$ can be nonquadratic for a nonlinear system and if we express such a function in a way like in the linear case, that is

$$L_c(x) = \frac{1}{2} x^T \hat{Y}(x)^{-1} x,$$

we will have difficulties in computing the eigenvalues and eigenvectors of the polynomial matrix $\hat{Y}(x)^{-1}$ as it is not a constant matrix anymore. Instead of dealing with nonquadratic $L_c(x)$ to determine 'the most important' axis we introduce another quadratic function $\hat{L}_c(x) = \frac{1}{2} x^T \Psi^{-1} x$ where

$$\mathcal{R}_c \subseteq \left\{x \in \mathbb{R}^n |\ 2\hat{L}_c(x) \leq 1\right\} = \hat{\mathcal{R}}_c.$$

Hence the set $\hat{\mathcal{R}}_c$ is also an estimate of the reachable set as $\hat{\mathcal{R}}_c$ contains $\mathcal{R}_c$. Though $\hat{\mathcal{R}}_c$ is more conservative than $\mathcal{R}_c$, it has a nice shape in a way that the set

$$\hat{\mathcal{R}}_c = \left\{x \in \mathbb{R}^n |\ x^T \Psi^{-1} x \leq 1\right\}$$

is a hyperellipsoid where the directions and the lengths of its principal axes are defined by the eigenvectors and the inverse of square root of the eigenvalues, respectively, of the matrix Ψ^{-1}. The rest will follow in the same way with those in linear systems where we denote T as an orthogonal matrix ($T^T T = I$) whose columns are the normalized eigenvectors of matrix Ψ^{-1}

and we define a new coordinate system $z = T^{-1}x$ where its main axis coincide with the principal axis of the ellipsoid. Indeed, we have $\Psi^{-1}T = TS$ where $S = diag\,(\lambda_1, \ldots, \lambda_n)$. In line with that in linear systems the transformed nonlinear system is given by

$$\dot{z} = T^T f\,(Tz) + T^T B u, \tag{5.10}$$

with

$$z^T T^T f\,(Tz) = x^T f\,(x) \leq -x^T Q x = -z T^T Q T z. \tag{5.11}$$

From the fact that $T^{-1} = T^T$ we have

$$\hat{Q} = T^T Q T \succ 0.$$

The set $\hat{\mathcal{R}}_c$ can be written in terms of the new coordinate

$$\hat{\mathcal{R}}_{c,z} = \left\{ z \in \mathbb{R}^n |\ z^T T' \Psi^{-1} T z \leq 1 \right\} = \left\{ z \in \mathbb{R}^n |\ z' S z \leq 1 \right\}.$$

By ordering $\lambda_i \leq \lambda_j$ whenever $i \leq j \leq n$ then the trajectories of the system are more accummulated around the z_i-axis rather than z_j-axis. By removing the weakly reachable parts of (5.10) we can truncate (5.10) to obtain a reduced order model of dimension $n_r < n$ in the form

$$\dot{x}_r = f_r\,(x_r) + B_r u \tag{5.12}$$

where

$$f_r\,(x_r) = T_1^T f\,(T_1 x_r)\,,\ B_r = T_1^T B$$

and T comes in the form (5.8). Removing the least reachable part from (5.11) and partitioning

$$\hat{Q} = \begin{bmatrix} \hat{Q}_1 & \hat{Q}_2 \\ \hat{Q}_2 & \hat{Q}_3 \end{bmatrix}$$

it follows that

$$x_r^T f_r\,(x_r) \leq -x_r^T \hat{Q}_1 x_r$$

where $\hat{Q}_1 \succ 0$ and thus the origin of the unforced truncated system (5.12) is globally asymptotically stable.

To reduce conservatism of the set $\hat{\mathcal{R}}_c$ we require that the set $\hat{\mathcal{R}}_c$ should be contained in as small ball B_γ as possible. Therefore we need to minimize $\gamma > 0$ such that $\mathcal{R}_c \subseteq \hat{\mathcal{R}}_c \subseteq B_\gamma$. A sufficient condition for the required containment is given as follows.

Proposition 5.1 *If*

$$\frac{1}{\gamma}\left\|x\right\|_2^2 \leq 2\hat{L}_c\left(x\right) \leq 2L_c\left(x\right) \tag{5.13}$$

for all x then $\mathcal{R}_c \subseteq \hat{\mathcal{R}}_c \subseteq B_\gamma$.

Proof. *All x in $\mathcal{R}_c$ satisfies $2L_c\left(x\right) \leq 1$. From $2\hat{L}_c\left(x\right) \leq 2L_c\left(x\right)$ it follows that $2\hat{L}_c\left(x\right) \leq 1$. Hence x is in $\hat{\mathcal{R}}_c$. Furthermore, from $\frac{1}{\gamma}\left\|x\right\|_2^2 \leq 2\hat{L}_c\left(x\right)$ we have $\frac{1}{\gamma}\left\|x\right\|_2^2 \leq 1$. Thus x is also in the ball B_γ.* ■

It has been already indicated that verifying nonnegativity is a hard problem. Instead, the inequalities (4.8) and (5.13) can be relaxed by means of sum of squares (SOS). We summarize our approach as follows.

1. a) Maximize $\theta > 0$ such that

$$\begin{gathered} L_c\left(x\right) - \hat{L}_c\left(x\right) \text{ is SOS for all } x \in \mathbb{R}^n \\ 2\hat{L}_c\left(x\right) - \theta\left\|x\right\|_2^2 \text{ is SOS for all } x \in \mathbb{R}^n, \\ v^T \begin{bmatrix} -\frac{\partial L_c(x)}{\partial x} f\left(x\right) & \frac{\partial L_c(x)}{\partial x} B \\ B^T \frac{\partial L_c(x)}{\partial x}^T & 2I_{n_u} \end{bmatrix} v \text{ is SOS for all } x \in \mathbb{R}^n \text{ and } v \in \mathbb{R}^{1+n_u}, \end{gathered}$$

   where $L_c\left(x\right)$ is a polynomial in x and $\hat{L}_c\left(x\right)$ is a quadratic polynomial in x.

   b) Compute the transformation T from

$$\hat{\mathcal{R}}_c = \left\{x \in \mathbb{R}^n | 2\hat{L}_c\left(x\right) = \ x^T \Psi^{-1} x \leq 1\right\},$$

   and truncate the transformed system $z = T^T x$ at $n_r < n$. In this case we obtain $f_r\left(x_r\right)$ and B_r.

2. Compute C_r, D_r and positive semidefinite $V\left(\chi\right)$, and minimize ϵ such that

$$-w^T \begin{bmatrix} \frac{\partial V(\chi)}{\partial \chi}\mathcal{F}\left(\chi\right) & \frac{1}{2}\frac{\partial V(\chi)}{\partial \chi}\mathcal{B} & \mathcal{H}\left(\chi\right)' \\ \frac{1}{2}\mathcal{B}^T \frac{\partial V(\chi)}{\partial \chi}^T & -\epsilon^2 I_{n_u} & \mathcal{D}^T \\ \mathcal{H}\left(\chi\right) & \mathcal{D} & -I_{n_y} \end{bmatrix} w$$

   is SOS for all $x \in \mathbb{R}^n$ and $w \in \mathbb{R}^{1+n_u+n_y}$.

It is important to point out that for the given $f_r(x_r)$ and B_r from step 1 it is not clear so far whether step 2 will work. In this case it is not immediate that we can guarantee the existence of $C_r \in \mathbb{R}^{n_y \times n_r}$ and $D_r \in \mathbb{R}^{n_y \times n_u}$ such that the error system (4.3) is finite gain $\mathcal{L}_2$ stable. The following result resolves this issue.

Proposition 5.2 *Consider the system (5.1) and (5.12). For any $C_r \in \mathbb{R}^{n_y \times n_r}$ and $D_r \in \mathbb{R}^{n_y \times n_u}$ such that*

$$\begin{aligned} \left\| \begin{bmatrix} C & -C_r \end{bmatrix} \right\|_{2-ind} &< \infty, \\ \|D - D_r\|_{2-ind} &< \infty, \end{aligned}$$

the error system (4.3) is finite gain $\mathcal{L}_2$ stable.

Proof. *We have*

$$\chi^T \mathcal{F}(\chi) \leq -x^T Q x - x_r' \hat{Q}_1 x_r \leq -c \|\chi\|^2$$

where c is the minimum eigenvalue of the positive definite symmetric matrices Q and $\hat{Q}_1$. Denoting

$$\begin{aligned} L &= \left\| \begin{bmatrix} B \\ B_r \end{bmatrix} \right\|_{2-ind}, \\ \eta_1 &= \left\| \begin{bmatrix} C & -C_r \end{bmatrix} \right\|_{2-ind}, \\ \eta_2 &= \|D - D_r\|_{2-ind}, \end{aligned}$$

it follows from Theorem 5.1 of [18] that (4.3) is finite gain $\mathcal{L}_2$ stable with gain bounded above by

$$\hat{\epsilon} = \eta_2 + \frac{\eta_1 L}{c}.$$

■

Proposition 5.2 shows that any bounded C_r and D_r will guarantee finiteness of the $\mathcal{L}_2$ gain of the error system (4.3).

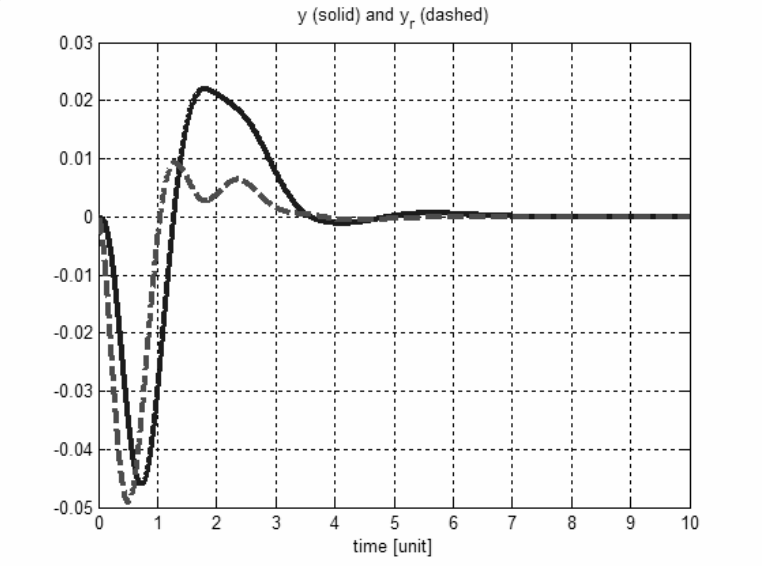

Figure 5.1: Response of the system in Example 1 to the input $u = e^{-1.5t}sin(5t)$

5.3 Numerical Example

5.3.1 Example 1

Consider the system

$$\begin{aligned}
\dot{x}_1 &= -x_2 - x_3 - x_1(x_1^2 + x_2^2 + x_3^2 + 1),\\
\dot{x}_2 &= x_1 - x_3 - x_2(x_1^2 + x_2^2 + x_3^2 + 1),\\
\dot{x}_3 &= x_1 + x_2 - x_3(x_1^2 + x_2^2 + x_3^2 + 1) + u,\\
y &= x_1.
\end{aligned}$$

We want to compute a reduced model of order two. By feasibility test we obtain

$$L_c(x) = \hat{L}_c(x) = 27x_1^2 + 5x_2^2 + 3x_3^2 - 6x_1x_2 + 10x_1x_3 + 2x_2x_3$$

and $\theta = 2.5666$. The transformation T is given by

$$T = \begin{bmatrix} -0.2201 & 0.0256 & -0.9752 \\ -0.4151 & 0.9022 & 0.1174 \\ 0.8827 & 0.4306 & -0.1879 \end{bmatrix}$$

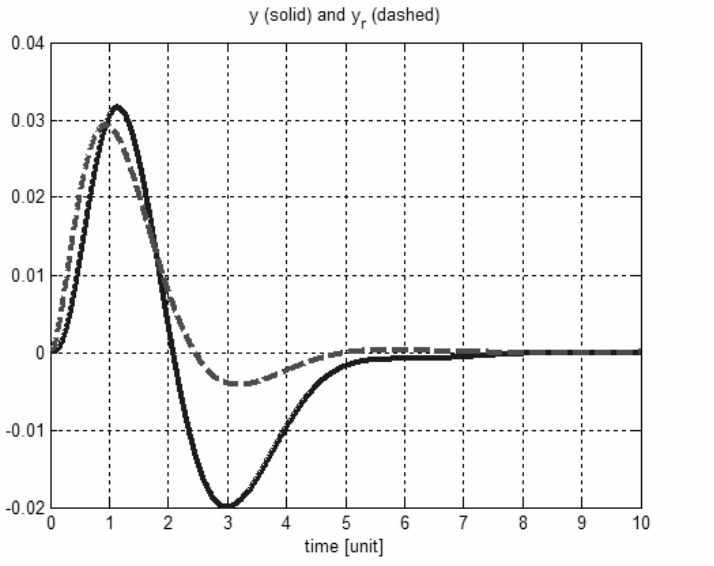

Figure 5.2: Response of the system in Example 1 to the input $u = e^{-3t} - e^{-1.5t}$

and truncation of the transformed system gives

$$\dot{x}_{r1} = 1.2804x_{r2} - x_{r1}\left(x_{r1}^2 + x_{r2}^2 + 1\right) + 0.8827u,$$
$$\dot{x}_{r2} = -1.2804x_{r1} - x_{r2}\left(x_{r1}^2 + x_{r2}^2 + 1\right) + 0.4306u.$$

with

$$\begin{aligned} V(\chi) = {} & 0.68152x_1^2 + 1.5916x_2^2 + 0.4388x_3^2 - 0.76919x_1x_2 - 0.73432x_1x_3 \\ & + 0.65377x_2x_3 + 0.53614x_{r1}^2 - 0.20907x_{r1}x_{r2} + 1.816x_{r2}^2 + 0.64768x_1x_{r1} \\ & + 1.011x_1x_{r2} + 0.56639x_2x_{r1} - 3.2278x_2x_{r2} - 0.69682x_3x_{r1} - 1.0061x_3x_{r2} \end{aligned}$$

and

$$y_r = -0.2498x_{r1} - 0.1297x_{r2}$$

with $\epsilon = 0.1014$. The response of the system and the reduced model to inputs $u = e^{-1.5t}sin(5t)$ and $u = e^{-3t} - e^{-1.5t}$ can be seen in Figure 5.1 and Figure 5.2, respectively. Though the responses are not too much in agreement, our scheme still outperforms the one in the previous chapter as the scheme in the previous chapter fails to compute a reduced model of order two for the original system.

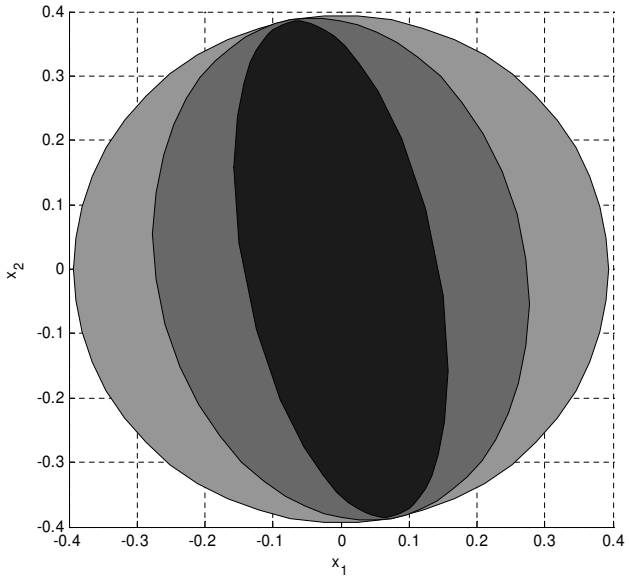

Figure 5.3: Inclusion $\mathcal{R}_c \subseteq \hat{\mathcal{R}}_c \subseteq B_r$ in Example 2

5.3.2 Example 2

Consider again the system from the previous chapter

$$\begin{aligned} \dot{x}_1 &= -x_1 - x_2, \\ \dot{x}_2 &= 2x_1 - 3x_2 - x_2^3 + u, \\ y &= x_1. \end{aligned}$$

We want to compute a reduced model of order one. For $L_c(x)$ in quadratic form (thus $L_c(x) = \hat{L}_c(x)$) we obtain $\theta = 6.000$. For $L_c(x)$ with maximum degree of four we obtain $\theta = 6.4593$. Hence $L_c(x)$ with maximum degree of four gives a better estimate of the reachable set than that of quadratic $L_c(x)$. Increasing the maximum degree of $L_c(x)$ higher than four will give the same value of θ as in $L_c(x)$ of maximum degree four. In this case we will use $L_c(x)$ with maximum degree of four where we obtain

$$\begin{aligned} L_c(x) &= 24x_1^2 + 8x_1x_2 + 4x_2^2 + 6.3537x_1^4 + 0.29051x_2^4, \\ \hat{L}_c(x) &= 6.6696x_1^2 + 1.3249x_1x_2 + 3.3572x_2^2. \end{aligned}$$

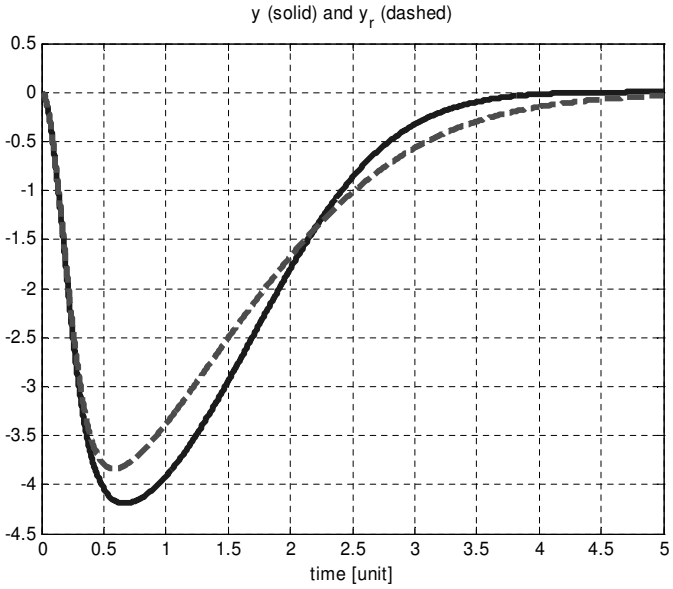

Figure 5.4: Response of the system in Example 2 to the input $u = 50e^{-3t} - 50e^{-1.5t}$

The plot of the inclusion $\mathcal{R}_c \subseteq \hat{\mathcal{R}}_c \subseteq B_\gamma$ where $\gamma = \frac{1}{\theta}$ can be seen in Figure 5.3. The reduced model is given by

$$\begin{aligned} \dot{x}_r &= -3.1142x_r - 0.9298x_r^3 - 0.9820u, \\ y_r &= -2.0512x_r, \end{aligned}$$

with $\epsilon = 0.1298$. The response of the system and the reduced model to input $u = 50e^{-3t} - 50e^{-1.5t}$ can be seen in Figure 5.4 which, qualitatively, is almost similar with that in the previous chapter.

Chapter 6

Approximate Balanced Truncation of Polynomial Nonlinear Systems

6.1 Introduction

Model reduction by balanced truncation for linear systems was introduced in [25]. This method is systematic on its construction of reduced model and very popular among other reduction schemes for linear systems due to its simplicity. On the other hand, there is no similar systematic procedure of balanced truncation like in linear system which can be implemented for nonlinear system as the problem in nonlinear systems becomes more difficult to understand and solve and it is still subject to further research. In particular, a mechanism for balancing nonlinear system is given by [38] and an empirical approach for truncating nonlinear system is given by [20].

One of the main advantage of the empirical approach introduced in [20] is that its computational scheme to construct a reduced order model is not expensive as it only requires linear matrix computations. Its limitation is on the resulting reduced order model. The reduced model is expected to work well within a working region of state space as the method relies heavily on the snapshots of data. In this case the quality of its reduced model depends on the collection of data obtained through generating trajectories of the original system.

On the other hand, the approach in [38] does not depend on the snapshots of

data. The method first computes a reachability function and an observability function from Hamilton-Jacobi equations. In general those functions are not entirely balanced. By seeking a coordinate transformation the original system is made balance to some extent of definition. A reduced order model is obtained by truncating the balanced system. The drawback of this method is on the computation of the controllability and observability functions which in most cases are very difficult.

One way to avoid this problem is introduced in [33] where the authors consider generalized controllability and observability functions which are obtained through Hamilton-Jacobi inequalities instead of Hamilton-Jacobi equalities. Despite the fact that the truncation scheme based on these generalized functions will not guarantee to give a stable reduced order model for a stable original system, the advantage of this approach is that it exploits the use of sum of squares programming [30] to compute the generalized functions and thus is amenable to computer solution in case the original system to be reduced has polynomial vector fields. This approach shows promising direction in developing computational scheme for model reduction of polynomial systems. However this approach still leaves an open problem where there is no constructive procedure yet on how to balance the generalized functions which are nonquadratic.

This chapter introduces an approach of balancing the nonquadratic generalized functions which are obtained through the same procedure in [33]. In this case we use another quadratic functions which can be viewed as conservative version of the nonquadratic generalized functions. Instead of balancing the generalized functions we focus on the conservative quadratic functions as the basis of performing truncation for polynomial nonlinear systems.

6.2 Balanced Truncation Based on Approximate Generalized Functions

We consider polynomial nonlinear system in the form

$$\dot{x} = f(x) + B(x)\,u \tag{6.1a}$$
$$y = h(x) \tag{6.1b}$$

where $x = [x_1, \ldots, x_n]^T \in D_x \subset \mathbb{R}^n$ is the state vector of the system, $u \in \mathbb{R}^{n_u}$ is the input to the system and $y \in \mathbb{R}^{n_y}$ is the output of the system. The functions $f : D_x \to \mathbb{R}^n$, $B : D_x \to \mathbb{R}^{n \times n_u}$ and $h : D_x \to \mathbb{R}^{n_y}$ are polynomials

in x and therefore smooth. We assume that the origin of the unforced system $\dot{x} = f(x)$ is asymptotically stable on D_x.

We define two functions which characterize the minimum energy of the input to reach particular state and the energy of the output generated by a particular initial condition.

Definition 6.1 *A positive definite function $W_o(x)$ with $W_o(0) = 0$ is an observability function to the system (6.1) if it satisfies*

$$\frac{\partial W_o(x)}{\partial x} f(x) + \frac{1}{2} h(x)^T h(x) = 0 \tag{6.2}$$

for all $x \in D_x$.

Definition 6.2 *A positive definite function $W_c(x)$ with $W_c(0) = 0$ is a reachability function to the system (6.1) if it satisfies*

$$\frac{\partial W_c(x)}{\partial x} f(x) + \frac{1}{2} \frac{\partial W_c(x)}{\partial x} B(x) B(x)^T \frac{\partial W_c(x)}{\partial x} = 0. \tag{6.3}$$

for all $x \in D_x$.

The reachability function satisfies

$$W_c(x_0) = \frac{1}{2} \int_{-\infty}^{0} \|u_{\min}(t)\|^2 dt \tag{6.4}$$

where $u_{\min} \in L_2(-\infty, 0]$, is the input with minimum energy required for the system (6.1) with $x(-\infty) = 0$ such that $x(0) = x_0$. The observability function satisfies

$$W_o(x_0) = \frac{1}{2} \int_{0}^{\infty} \|y(t)\|^2 dt$$

where y is the output of the system (6.1) with $x(0) = x_0$ and $u(t) = 0$ for $t \in [0, \infty)$. Indeed, for the linear system $f(x) = Ax$, $B(x) = B$, $h(x) = Cx$ the reachability function $W_c(x) = \frac{1}{2} x^T Y^{-1} x$ and observability function $W_o(x) = \frac{1}{2} x^T X x$ satisfy (2.2) and (2.3), respectively.

For more details on the existence of the solutions $W_o(x)$ and $W_c(x)$ the reader can consult [38]. As we move to practicality we will encounter difficulty in computing the observability and reachability functions as there is no tractable computational scheme serving the purpose yet. Instead of computing both functions for the purpose of model reduction through balanced

truncation, we consider an approach where we use generalized observability and rechability functions as defined in Definition 4.2-4.3.

Some properties pertaining to the generalized functions are summarized as follows.

- A generalized observability function $L_o(x)$ to the system (6.1) satisfies

$$L_o(x) \geq W_o(x)$$

  for all $x \in D_x$.

- A generalized reachability function $L_c(x)$ to the system (6.1) satisfies

$$L_c(x) \leq W_c(x)$$

  for all $x \in D_x$.

It is important to note that there are many choices of generalized functions which satisfy (4.6-4.7). A means to classify a closer representation of the functions $L_o(x)$ and $L_c(x)$ with the functions $W_o(x)$ and $W_c(x)$ is by introducing

$$\gamma_{opt} = \sup_{0 \neq x \in D_x} \frac{W_o(x)}{W_c(x)}.$$

Then the generalized functions $L_o(x)$ and $L_c(x)$ which satisfy (4.6-4.7) are computed such that

$$L_o(x) \leq \gamma L_c(x) \tag{6.5}$$

for all $x \in D_x$ where $\gamma > 0$. Hence γ is an upper bound for the gain γ_{opt}. In this case we minimize the constant γ so that the upper bound is as tight as possible.

The generalized functions characterize the states, in the domain of interest, which are weakly observable and reachable in the following ways. The states which have small value of $L_o(x)$ are considered to be less observable. The states which have larger value of $L_c(x)$ are considered to be less reachable.

Now let us consider the class of input with

$$\int_0^T \|u(t)\|^2 dt \leq K_u,$$
$$u(t) = 0 \ \forall t > T,$$

for all $T \in [0, \infty)$. Suppose that we generate the trajectories from the origin with this class of input then for $t \in [0, T]$ all the trajectories of the system will be inside of the set

$$\mathcal{R}_c = \left\{ x \in \mathbb{R}^n \mid L_c(x) \leq \frac{1}{2} K_u \right\}.$$

It is easy to see that the trajectories will also remain in $\mathcal{R}_o$ for $t \geq T$ because $\mathcal{R}_o$ is a positively invariant set for zero input. Hence for the given class of input with initial condition at the origin all the trajectories of the system will be inside of $\mathcal{R}_o$. Then for any $x \in \mathcal{R}_o$ we have $L_o(x) \leq \frac{1}{2}\gamma K_u$.

For the reachability set

$$\mathcal{R}_c = \left\{ x \in \mathbb{R}^n \mid L_c(x) \leq \frac{1}{2} K_u \right\}$$

which is compact a smaller value of $L_c(x)$ means that the state x is easier to reach for the given admissible input. For the observability set

$$\mathcal{R}_o = \left\{ x \in \mathbb{R}^n \mid L_o(x) \leq \frac{1}{2} \gamma K_u \right\}$$

which is compact a larger value of $L_o(x)$ means that the state x is easier to observe from the output. In this paper instead of dealing with the sets $\mathcal{R}_c$ and $\mathcal{R}_o$ to analyze the most/least important part of the system which is reachable and observable we will consider another sets Ω_c and Ω_o as the estimate of the reachability set $\mathcal{R}_c$ and observability set $\mathcal{R}_o$, respectively, in a way that $\mathcal{R}_c \subseteq \Omega_c$ and $\mathcal{R}_o \subseteq \Omega_o$ where the sets Ω_c and Ω_o are to be as close as possible to the sets $\mathcal{R}_c$ and $\mathcal{R}_o$, respectively. The sets Ω_c and Ω_o will be characterized by quadratic functions $\hat{L}_o(x)$ and $\hat{L}_c(x)$ such that

$$\Omega_c = \left\{ x \in \mathbb{R}^n \mid \hat{L}_c(x) = \frac{1}{2} x^T Y^{-1} x \leq 1 \right\},$$
$$\Omega_o = \left\{ x \in \mathbb{R}^n \mid \hat{L}_o(x) = \frac{1}{2} x^T X x \leq 1 \right\}$$

where $X, Y \succ 0$ are symmetric matrices.

Though the sets Ω_c and Ω_o are more conservative they are easier to analyze because the sets are in the form of hyperellipsoid. It is easy to see from the principal axes of the hyperellipsoid Ω_c that a longer axis represent a more reachable part of the state. On the other hand, from the principal axes of the hyperellipsoid Ω_o, we can see that a shorter axis represent a more observable

part of the state. Equivalently we can analyze the length of the axes from the eigenvalues of the matrices Y^{-1} and X. The associated eigenvector of a smaller eigenvalues of the matrix Y^{-1} represents the direction of a longer axis of hyperellipsoidal set Ω_c. (Or equivalently the associated eigenvector of a larger eigenvalues of the matrix Y represents the direction of a longer axis of hyperellipsoidal set Ω_c.) Similarly the associated eigenvector of a larger eigenvalues of the matrix X represents the direction of a shorter axis of hyperellipsoidal set Ω_o. Thus the associated eigenvector of a larger eigenvalues of the matrices Y and X represent the direction of the states which are more strongly reachable and observable, respectively.

To satisfy the containment $\mathcal{R}_c \subseteq \Omega_c$ and $\mathcal{R}_o \subseteq \Omega_o$ we use the following sufficient condition which is easy to implement numerically.

Proposition 6.1 *Let $L(x)$ be a positive definite polynomial function with $L(0) = 0$. Let*

$$\mathcal{R} = \{x \in D_x |\ L(x) \leq \epsilon\},$$
$$\Omega = \left\{x \in D_x |\ \frac{1}{2}x^T \Phi x \leq 1\right\},$$

for positive constant $\epsilon \in \mathbb{R}$ and positive definite matrix $\Phi = \Phi^T \in \mathbb{R}^{n\times n}$. If there exists a positive semidefinite polynomial $s(x)$ such that

$$1 - \frac{1}{2}x^T \Phi x + s(x)(L(x) - \epsilon) \geq 0 \tag{6.6}$$

for all $x \in D_x$ then $\mathcal{R} \subseteq \Omega$.

Proof. *For any $x \in \mathcal{R}$ we have $L(x) - \epsilon \leq 0$. It follows that $0 \leq 1 - \frac{1}{2}x^T\Phi x + s(x)(L(x) - \epsilon) \leq 1 - \frac{1}{2}x^T\Phi x$ or $x \in \Omega$.* ■

To reduce the conservatism of the set Ω we require that the set $\mathcal{R}$ should be contained in as small Ω as possible.

We now consider a change of basis $x = \Gamma z$ where Γ is given by

$$\Gamma = Y^{\frac{1}{2}} U S^{-\frac{1}{2}}$$

where U and S are obtained from the singular value decomposition of

$$Y^{\frac{1}{2}} X Y^{\frac{1}{2}} = U S^2 U^T.$$

It is easy to see that

$$\begin{aligned}\Gamma^T X \Gamma &= \left(S^{-\frac{1}{2}} U^T Y^{\frac{1}{2}}\right) X \left(Y^{\frac{1}{2}} U S^{-\frac{1}{2}}\right) \\ &= S^{-\frac{1}{2}} U^T U S^2 U^T U S^{-\frac{1}{2}} = S\end{aligned}$$

and

$$\begin{aligned}\Gamma^{-1} Y \left(\Gamma^{-1}\right)^T &= \left(S^{\frac{1}{2}} U^T Y^{-\frac{1}{2}}\right) Y \left(Y^{-\frac{1}{2}} U S^{\frac{1}{2}}\right) \\ &= S.\end{aligned}$$

With respect to the change of basis, (6.1) can be transformed into

$$\dot{z} = \hat{f}(z) + \hat{B}(z)\, u, \tag{6.7a}$$
$$y = \hat{h}(z), \tag{6.7b}$$

where

$$\hat{f}(z) = \Gamma^{-1} f(\Gamma z),\ \hat{B}(z) = \Gamma^{-1} B(\Gamma z),\ \hat{h}(z) = h(\Gamma z).$$

The generalized functions for the system (6.7) are given as follows.

Proposition 6.2 *For a coordinate transformation $x = \Gamma z$, which brings the system (6.1) into (6.7) we define $\hat{L}_o(z) = L_o(\Gamma z)$ and $\hat{L}_c(z) = L_c(\Gamma z)$. Then $\hat{L}_o(z)$ and $\hat{L}_c(z)$ are the generalized observability and reachability function, respectively for (6.7).*

Proof. *For reachability*

$$\begin{aligned}\frac{\partial \hat{L}_c(z)}{\partial z}\left(\hat{f}(z) + \hat{B}(z)\, u\right) - \frac{1}{2} u^T u &= \left(\frac{\partial L_c(\nu)}{\partial \nu}\frac{\partial \nu}{\partial z}\left(\Gamma^{-1} f(\Gamma z) + \Gamma^{-1} B(\Gamma z)\, u\right)\right)_{\nu=\Gamma z} \\ &\quad - \frac{1}{2} u^T u \\ &= \frac{\partial L_c(\nu)}{\partial \nu} f(\nu) + \frac{\partial L_c(\nu)}{\partial \nu} B(\nu)\, u - \frac{1}{2} u^T u \leq 0.\end{aligned}$$

For observability

$$\begin{aligned}\frac{\partial \hat{L}_o(z)}{\partial z}\hat{f}(z) + \frac{1}{2}\hat{h}(z)^T \hat{h}(z) &= \left(\frac{\partial L_o(\nu)}{\partial \nu}\frac{\partial \nu}{\partial z}\Gamma^{-1} f(\Gamma z)\right)_{\nu=\Gamma z} + \frac{1}{2} h(\Gamma z)^T h(\Gamma z) \\ &= \frac{\partial L_o(\nu)}{\partial \nu} f(\nu) + \frac{1}{2} h(\nu)^T h(\nu) \leq 0.\end{aligned}$$

■

Furthermore, it is easy to see that

$$\begin{aligned}0 &\leq 1-\frac{1}{2}x^T Y^{-1}x+s_c\left(x\right)\left(L_c\left(x\right)-\frac{1}{2}K_u\right)\\ &= 1-\frac{1}{2}z\Gamma^T Y^{-1}\Gamma z+\hat{s}_c\left(z\right)\left(\hat{L}_c\left(z\right)-\frac{1}{2}K_u\right)\\ &= 1-\frac{1}{2}zS^{-1}z+\hat{s}_c\left(z\right)\left(\hat{L}_c\left(z\right)-\frac{1}{2}K_u\right)\end{aligned}$$

where $\hat{s}_c\left(z\right)=s_c\left(\Gamma z\right)$ is positive semidefinite in z and

$$\begin{aligned}0 &\leq 1-\frac{1}{2}x^T Xx+s_o\left(x\right)\left(L_o\left(x\right)-\frac{1}{2}\gamma K_u\right)\\ &= 1-\frac{1}{2}z^T\Gamma^T X\Gamma z+\hat{s}_o\left(z\right)\left(\hat{L}_o\left(z\right)-\frac{1}{2}\gamma K_u\right)\\ &= 1-\frac{1}{2}z^T Sz+\hat{s}_o\left(z\right)\left(\hat{L}_o\left(z\right)-\frac{1}{2}\gamma K_u\right)\end{aligned}$$

where $\hat{s}_o\left(z\right)=s_o\left(\Gamma z\right)$ is positive semidefinite in z. Using the same argument like in Proposition 3 it follows that

$$\begin{aligned}\left\{z\in\mathbb{R}^n|\hat{L}_c\left(z\right)\leq\frac{1}{2}K_u\right\}&\subseteq\left\{z\in\mathbb{R}^n|\frac{1}{2}z^T S^{-1}z\leq 1\right\},\\ \left\{z\in\mathbb{R}^n|\hat{L}_o\left(z\right)\leq\frac{1}{2}\gamma K_u\right\}&\subseteq\left\{z\in\mathbb{R}^n|\frac{1}{2}z^T Sz\leq 1\right\},\end{aligned}$$

which indicate that the transformed system (6.7) have a balanced representation in that the states which are more strongly reachable and observable are more or less in the same direction. In this case the associated eigenvector of a larger eigenvalue of S represents the direction of the states which is both more strongly reachable and observable.

Next we partition the part of size n into two parts of size n_r and $n-n_r$ with $n_r<n$ based on the following

$$\begin{aligned}z&=\left[\begin{array}{cc} z_{[1]}^T & z_{[2]}^T\end{array}\right]^T,\\ \Gamma&=\left[\begin{array}{cc}\Gamma_1 & \Gamma_2\end{array}\right],\\ \Gamma^{-1}&=\left[\begin{array}{cc}\Upsilon_1 & \Upsilon_2\end{array}\right].\end{aligned}$$

By removing the weakly reachable and observable part $z_{[2]}$ we have the dynamic of our new reduced model $x_r=z_{[1]}$ of dimension n_r given by

$$\dot{x}_r = f_r\left(x_r\right)+B_r\left(x_r\right)u, \tag{6.8a}$$

$$y_r = h_r\left(x_r\right), \tag{6.8b}$$

where

$$f_r(x_r) = \Upsilon_1 f(\Gamma_1 x_r),\ B_r(x_r) = \Upsilon_1 B(\Gamma_1 x_r),$$
$$h_r(x_r) = h(\Gamma_1 x_r).$$

To sum up we have obtained a reduced order model (6.8) for the system (6.1) where the least reachable and observable parts in (6.1) are removed while the most influential parts are preserved in (6.8).

6.3 Sum of Squares Formulation

Since we are concerned with polynomial systems, computation of the generalized functions $L_o(x)$ and $L_c(x)$ can be done efficiently by relaxing the left hand side of (4.6-4.7) being sum of squares. We will consider first the global case when $D_x = \mathbb{R}^n$. Then we will formulate sum of squares relaxation for the local case.

6.3.1 Global Case

For the case $D_x = \mathbb{R}^n$ computational scheme for the generalized functions can be summarized as follows.

- Minimize $\gamma \geq 0$ and find positive definite polynomials $L_o(x)$ and $L_c(x)$ with $L_o(0) = 0$ and $L_c(0) = 0$ such that

$$-\frac{\partial L_o(x)}{\partial x} f(x) - \frac{1}{2} h(x)^T h(x) \quad \text{is SOS } \forall x \in \mathbb{R}^n, \tag{6.9}$$

$$-\frac{\partial L_c(x)}{\partial x} f(x) - \frac{1}{2} \frac{\partial L_c(x)}{\partial x} B(x) B(x)^T \frac{\partial L_c(x)}{\partial x} \quad \text{is SOS } \forall x \in \mathbb{R}^n, \tag{6.10}$$

$$\gamma L_c(x) - L_o(x) \text{ is SOS } \forall x \in \mathbb{R}^n. \tag{6.11}$$

Unfortunately, this scheme may sometimes fail to give any solutions. In the next proposition we will give necessary conditions for the scheme to work. These conditions come in the form of degree constraint of the generalized functions. With the conditions in mind we can avoid certain classes of polynomial systems which will not give any solution. The following lemma shall be used to prove the proposition.

Lemma 6.1 *Consider polynomials $q(x)$ and $r(x)$ such that*

$$q(x) - r(x)^T r(x) \text{ is sum of squares.}$$

Then $\mu_{\min}(q(x)) \leq 2\mu_{\min}(r(x))$ and $\mu_{\max}(q(x)) \geq 2\mu_{\max}(r(x))$.

Proof. *Write*

$$\begin{aligned} q(x) &= \hat{q}(x) + \sum_k a_k x_1^{2m_{1k}} x_2^{2m_{2k}} \dots x_n^{2m_{nk}} \\ r(x)^T r(x) &= \hat{\varphi}_r(x) + \sum_l \left(c_l x_1^{q_{1l}} x_2^{q_{2l}} \dots x_n^{q_{nl}}\right)^2 \end{aligned}$$

where $a_k, c_l \in \mathbb{R}$ and m_{ik}, q_{il} are nonnegative integer with $\sum_{i=1}^{n} 2m_{ik} = \mu_{\min}(q(x))$ for all k and $\sum_{i=1}^{n} q_{il} = \mu_{\min}(r(x))$ for all l, while $\hat{q}(x)$ and $\hat{\varphi}_r(x)$ contain other different terms of monomial. Let $\mu_{\min}(q(x)) > 2\mu_{\min}(r(x))$. Then

$$\begin{aligned} q(x) - r(x)^T r(x) =& \hat{q}(x) + \sum_k a_i x_1^{2m_{1k}} x_2^{2m_{2k}} \dots x_n^{2m_{nk}} - \hat{\varphi}_r(x) \\ & - \sum_l \left(c_l x_1^{q_{1l}} x_2^{q_{2l}} \dots x_n^{q_{nl}}\right)^2 \\ =& \hat{q}(x) - \hat{\varphi}_r(x) + \sum_k a_i x_1^{2m_{1k}} x_2^{2m_{2k}} \dots x_n^{2m_{nk}} \\ & - \sum_l c_l^2 x_1^{2q_{1l}} x_2^{2q_{2l}} \dots x_n^{2q_{nl}} \end{aligned}$$

is not a sum of squares because each negative term $-c_l^2 x_1^{2q_{1l}} x_2^{2q_{2l}} \dots x_n^{2q_{nl}}$ can not be cancelled out by any term $a_i x_1^{2m_{1k}} x_2^{2m_{2k}} \dots x_n^{2m_{nk}}$ as $\sum_{i=1}^{n} m_{ik} > \sum_{i=1}^{n} q_{il}$ for all k and l. The maximum degree can be proved in the same way. ■

The lemma only provides necessary condition for the polynomial $q(x) - r(x)^T r(x)$ to be sum of squares as the converse, in general, is not true. However we may still use it to indicate the minimum and maximum degree of $q(x)$ and $r(x)$ for possibility of having $q(x) - r(x)^T r(x)$ being sum of squares. Furthermore, the lemma can be applied to characterize the degree of the generalized functions as we have in the following.

Proposition 6.3 *Consider positive definite polynomials $L_o(x)$ and $L_c(x)$ with $L_o(0) = 0$ and $L_c(0) = 0$ satisfying (6.9-6.10). Then*

$$\mu_{\min}(L_o(x)) \leq 2\mu_{\min}(h(x)) + 1 - \mu_{\min}(f(x)), \tag{6.12}$$

$$\mu_{\max}(L_c(x)) \leq \mu_{\max}(f(x)) + 1 - 2\mu_{\max}(B(x)). \tag{6.13}$$

Proof. *From the previous lemma we have*

$$\begin{aligned}
&\mu_{\min}\left(\frac{\partial L_o(x)}{\partial x} f(x)\right) \leq \mu_{\min}\left(h(x)^T h(x)\right), \\
\Longleftrightarrow\ &\mu_{\min}(L_o(x)) - 1 + \mu_{\min}(f(x)) \leq 2\mu_{\min}(h(x)), \\
\Longleftrightarrow\ &\mu_{\min}(L_o(x)) \leq 2\mu_{\min}(h(x)) + 1 - \mu_{\min}(f(x)),
\end{aligned}$$

and

$$\begin{aligned}
&\mu_{\max}\left(\frac{\partial L_c(x)}{\partial x} f(x)\right) \geq \mu_{\max}\left(\frac{\partial L_c(x)}{\partial x} B(x) B(x)^T \frac{\partial L_c(x)}{\partial x}\right), \\
\Longleftrightarrow\ &\mu_{\max}(L_c(x)) - 1 + \mu_{\max}(f(x)) \geq 2\left(\mu_{\max}(L_c(x)) - 1 + \mu_{\max}(B(x))\right), \\
\Longleftrightarrow\ &1 + \mu_{\max}(f(x)) - 2\mu_{\max}(B(x)) \geq \mu_{\max}(L_c(x)).
\end{aligned}$$

■

Recall from Section 4.3 that the degree of the generalized functions should not be less than two. In this case we can classify whether a certain class of polynomial systems fits to our scheme. The following example shows a class of systems where our scheme will not work.

Example 6.1 *Consider a homogenous system in the form*

$$\begin{aligned}
f_i(x) &= \sum_{j=1}^{n} a_{ij} x_j^m,\ a_{ij} \in \mathbb{R},\ m \in \{3, 5, 7, \ldots\}, \\
B(x) &= B \in \mathbb{R}^{n \times n_u}, \\
h(x) &= Cx,\ C \in \mathbb{R}^{n_y \times n},
\end{aligned}$$

where the origin of its unforced system is globally asymptotically stable. In this case we have $\mu_{\min}(f(x)) = m$ *and* $\mu_{\min}(h(x)) = 1$. *It follows that*

$$\mu_{\min}(L_o(x)) \leq 2\mu_{\min}(h(x)) + 1 - \mu_{\min}(f(x)) = 3 - m \leq 0$$

which implies that there is no positive definite $L_o(x)$ *with* $L_o(0) = 0$ *satisfying (6.9). Hence this system does not belong to the class that we consider.*

For containment in Proposition 6.1 where the set $\mathcal{R}$ should be contained in as small Ω as possible we can maximize the trace of Φ so that the volume of Ω is minimized. Hence we can summarize our computational approach to get the sets Ω_c and Ω_o as follows.

1. Maximize trace Y^{-1} such that
$$1 - \frac{1}{2}x^T Y^{-1} x + s_c(x)\left(L_c(x) - \frac{1}{2}K_u\right) \text{ is SOS } \forall x \in \mathbb{R}^n$$
where $s_c(x)$ is a sum of squares.

2. Maximize trace X such that
$$1 - \frac{1}{2}x^T X x + s_o(x)\left(L_o(x) - \frac{1}{2}\gamma K_u\right) \text{ is SOS } \forall x \in \mathbb{R}^n$$
where $s_o(x)$ is a sum of squares.

Like in the computation of the generalized functions, this scheme may sometimes fail to give any solution. The following is a necessary condition for the scheme to work.

Lemma 6.2 *If* $1 - \frac{1}{2}x^T \Phi x + s(x)(L(x) - \epsilon)$ *is a sum of squares of polynomials then*
$$\mu_{\min}(L(x)) = 2.$$

Proof. To cancel out or to dominate the negative quadratic term $-\frac{1}{2}x^T \Phi x$ it is necessary that the positive definite polynomial function $L(x)$ has quadratic term as well. ■

Recall from the previous discussion that (6.12-6.13) are necessary. Then the following is immediate.

Proposition 6.4 *If*
$$1 - \frac{1}{2}x^T Y^{-1} x + s_c(x)\left(L_c(x) - \frac{1}{2}K_u\right),$$
$$1 - \frac{1}{2}x^T X x + s_o(x)\left(L_o(x) - \frac{1}{2}\gamma K_u\right),$$
are sum of squares of polynomials then
$$\mu_{\min}(f(x)) \leq 2\mu_{\min}(h(x)) - 1,$$
$$\mu_{\max}(f(x)) \geq 2\mu_{\max}(B(x)) + 1.$$

Thus our scheme will not work for the class of polynomial systems with
$$\mu_{\min}(f(x)) > 2\mu_{\min}(h(x)) - 1,$$
$$\mu_{\max}(f(x)) < 2\mu_{\max}(B(x)) + 1.$$

6.3.2 Local Case

For the case where D_x is a semialgebraic set given by

$$D_x = \{x \in \mathbb{R}^n |\ p_i(x) \geq 0; i = 1, \ldots, m\}$$

where $p_i \in \mathbb{R}[x]$ we can use the following relaxation.

Proposition 6.5 *If there exists sum of squares $s_i(x)$ for $i = 1, \ldots, m$ such that*

$$q(x) - p_1(x) s_1(x) - \ldots - p_m(x) s_m(x) \ \textit{ is sum of squares}$$

then $q(x) \geq 0\ \forall x \in D_x$.

Proof. *Use Positivestellensatz [39].* ■

This proposition can be easily applied for feasibility test of (4.6-4.7), (6.5) and (6.6) using sum of squares programming.

6.4 Structure Preservation

It is not yet clear what kind of property, in general, the reduced order model preserves from the original system. This is still under investigation. For a special case where there exists a symmetric $Q \succ 0$ such that

$$x^T f(x) \leq -x^T Q x$$

for all $x \in \mathbb{R}^n$ we can guarantee that the reduced model preserves asymptotic stability. This is discussed in detail in Subsection 5.2.2.

6.5 Example

In this section, two numerical examples are given to illustrate the applicability of the proposed approach.

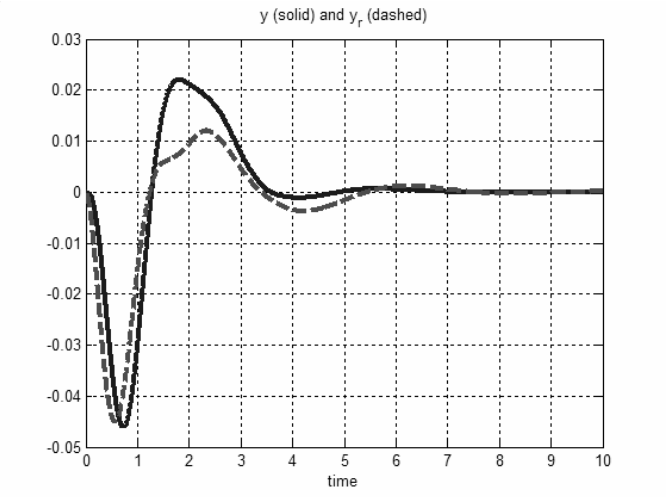

Figure 6.1: Response of the system in Example 1 to the input $u(t) = e^{-1.5t}sin(5t)$

6.5.1 Example 1

Consider again the system from the previous chapter

$$\begin{aligned}
\dot{x}_1 &= -x_2 - x_3 - x_1(x_1^2 + x_2^2 + x_3^2 + 1),\\
\dot{x}_2 &= x_1 - x_3 - x_2(x_1^2 + x_2^2 + x_3^2 + 1),\\
\dot{x}_3 &= x_1 + x_2 - x_3(x_1^2 + x_2^2 + x_3^2 + 1) + u,\\
y &= x_1.
\end{aligned}$$

We want to compute a reduced model of order two for the global case $D_x = \mathbb{R}^3$. The method gives the transformation

$$\Gamma = \begin{bmatrix} 0.3128 & -0.2998 & 0.4152 \\ 0.8310 & 0.3536 & -0.0721 \\ -0.9193 & 1.1060 & 0.7230 \end{bmatrix}$$

and truncation of the transformed system gives

$$\begin{aligned}
\dot{x}_{r1} &= -1.8505x_{r2} - x_{r1}\left(1.6335x_{r1}^2 - 1.6334x_{r1}x_{r2} + 1.4382x_{r2}^2 + 0.0799\right) - 0.1602u,\\
\dot{x}_{r2} &= 1.4356x_{r1} - x_{r2}\left(1.6335x_{r1}^2 - 1.6334x_{r1}x_{r2} + 1.4382x_{r2}^2 + 1.0699\right) + 0.4703u,\\
y_r &= 0.3128x_{r1} - 0.2998x_{r2}.
\end{aligned}$$

The response of the system and the reduced model to the input $u = e^{-1.5t}sin(5t)$ can be seen in Figure 6.1. Qualitatively, our scheme outperforms the one in the previous chapter.

6.5.2 Example 2

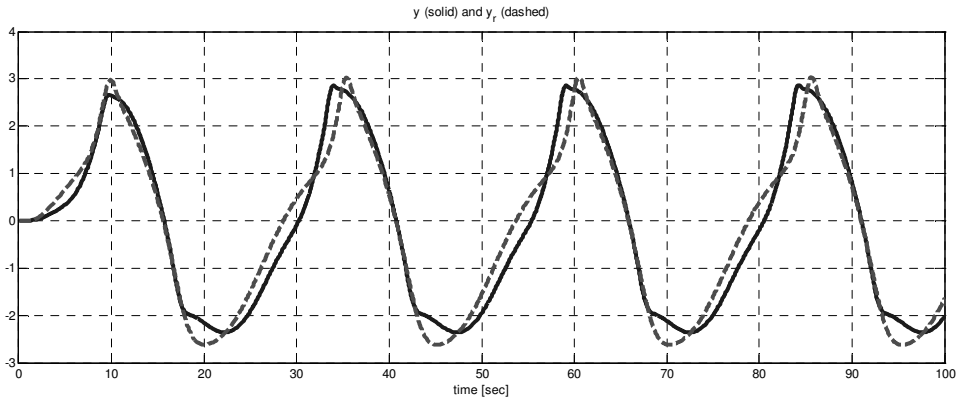

Figure 6.2: Response of the system in Example 2 to the sinusoidal input $u(t) = 2.5sin(0.25t)$

The following system is taken from [33]

$$\begin{aligned}
\dot{x}_1 &= x_2 - x_1x_2 - 3x_2x_3 - x_1x_4,\\
\dot{x}_2 &= x_3 + 0.5x_1x_2 + 0.5x_2x_3 + x_1x_4,\\
\dot{x}_3 &= x_4 + 0.5x_1x_2 + 0.5x_2x_3 - 0.25x_1x_4,\\
\dot{x}_4 &= -x_1 - 3x_2 - 5x_3 - 7x_4 - 3x_1x_2 + 0.1x_2x_3 + 0.3x_1x_4 + u,\\
y &= x_1,
\end{aligned}$$

where $D_x = \{x \in \mathbb{R}^4 |\ 12 - \|x\|^2 \geq 0\}$. For this system we can compute quadratic generalized functions but for higher order of generalized functions we can get a lower value of γ which is an upper bound for the Hankel norm. We compute generalized functions of order six with bound $\gamma = 1.2$. Choosing higher order than six will not give significant improvement of the bound γ. By applying the truncation scheme we obtain a reduced model of order three whose response to the sinusoidal input $u(t) = 2.5sin(0.25t)$ can be seen in Figure 6.2. Qualitatively, our scheme also outperforms the one in [33].

6.6 A Class of Polynomial Nonlinear Systems

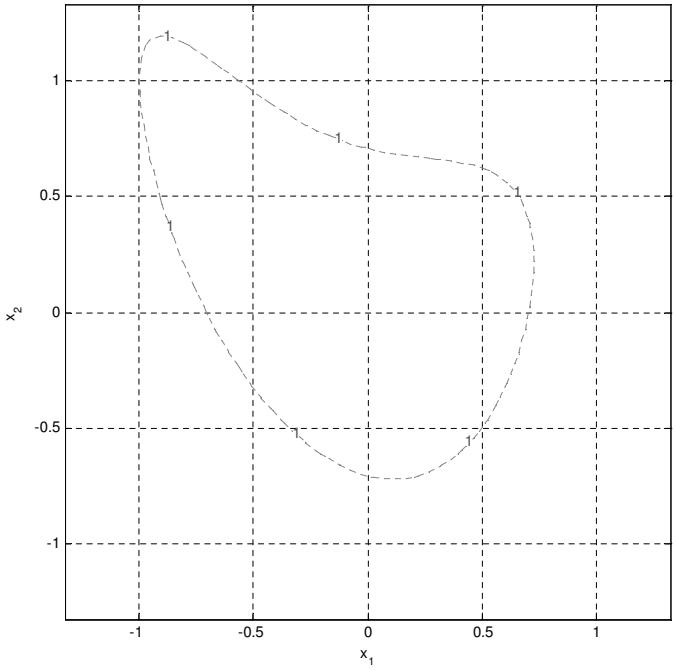

Figure 6.3: The set $\mathcal{R}$ whose shape is not symmetric

The method in this chapter has a drawback when the generalized function has a particular form. Consider, for example, the function

$$L(x) = 2x_1^4 - 3x_1^2x_2 + x_1^2 + x_1x_2 + 2x_2^2$$

which is positive definite. Now let

$$\mathcal{R} = \left\{x \in \mathbb{R}^2 | L(x) \leq 1\right\}.$$

The plot of the set $\mathcal{R}$ can be seen in Figure 6.3 where the shape of $\mathcal{R}$ is not symmetric while the shape of any set in the form

$$\Omega = \left\{x \in \mathbb{R}^2 | \frac{1}{2}x^T\Phi x \leq 1\right\}$$

characterized by the quadratic function $x^T\Phi x$ is always symmetric. If we want to identify the most/least important part of the set $\mathcal{R}$ by considering

the estimate Ω which is as close as $\mathcal{R}$, we might have less accurate result due to the asymmetric shape of the set $\mathcal{R}$. However, there is a certain class of polynomial systems where we can guarantee that the set $\mathcal{R}$ has symmetric shape, that is $x \in \mathcal{R} \Longleftrightarrow -x \in \mathcal{R}$.

Suppose, furthermore, we assume the following about the system (6.1).

Assumption 6.1 *The function f satisfies $f(-x) = -f(x)$. The functions g and h satisfy $g(-x)^T g(-x) = g(x)^T g(x)$ and $h(-x)^T h(-x) = h(x)^T h(x)$.*

Assumption 6.2 *The domain of interest D_x is symmetric where $x \in D_x \Leftrightarrow -x \in D_x$.*

The controllability and observability function for the class of systems considered are symmetric as shown by the following.

Proposition 6.6 *The reachability function satisfies $W_c(x) = W_c(-x)$ for all $x \in D_x$.*

Proof. *Consider the positive definite solution $W_c(x)$ which satisfies (6.3) for all $x \in D_x$ (hence, it satisfies (6.4) as well). Then (6.3) also holds for all $-x \in D_x$, that is*

$$\frac{\partial W_c}{\partial x}(-x) f(-x) + \frac{1}{2}\frac{\partial W_c}{\partial x}(-x) g(-x) g(-x)^T \frac{\partial W_c}{\partial x}(-x) = 0.$$

By the relation $\frac{\partial W_c}{\partial x}(-x) = -\frac{\partial [W_c(-x)]}{\partial x}$ then

$$\frac{\partial [W_c(-x)]}{\partial x} f(x) + \frac{1}{2}\frac{\partial [W_c(-x)]}{\partial x} g(x) g(x)^T \frac{\partial [W_c(-x)]}{\partial x} = 0$$

which implies that $W_c(-x)$ is also a solution to (6.3). It is easy to show that $W_c(-x)$ is positive definite since $W_c(x)$ is positive definite. Then $W_c(-x)$ also satisfies (6.4). Hence $W_c(-x) = W_c(x)$. ■

Proposition 6.7 *The observability function satisfies $W_o(x) = W_o(-x)$ for all $x \in D_x$.*

Proof. *The proof can be shown in the same way as in the proof of reachability proposition. Alternatively we can use the following. Let $x_+(t)$ and $x_-(t)$ be the solutions for*

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

whenever the initial conditions at time 0 *are* x_0 *and* $-x_0$*, respectively. Let* $y_+(t)$ *and* $y_-(t)$ *be the outputs associated with initial conditions* x_0 *and* $-x_0$*, respectively. From* $f(x) = -f(-x)$ *it follows that* $x_-(t) = -x_+(t)$ *and from* $h(-x)^T h(-x) = h(x)^T h(x)$ *we have*

$$\begin{aligned} y_-(t)^T y_-(t) &= h(x_-(t))^T h(x_-(t)) = h(-x_+(t))^T h(-x_+(t)) \\ &= h(x_+(t))^T h(x_+(t)) = y_+(t)^T y_+(t). \end{aligned}$$

Then

$$\begin{aligned} W_o(-x_0) &= \frac{1}{2}\int_0^\infty |y_-(t)|^2 dt = \frac{1}{2}\int_0^\infty |y_+(t)|^2 dt \\ &= W_o(x_0). \end{aligned}$$

■

As the reachability function and observability function for the class of system satisfy $W_c(x) = W_c(-x)$ and $W_o(x) = W_o(-x)$ for all $x \in D_x$ we can also impose the same condition for the generalized functions where we require that $L_c(x) = L_c(-x)$ and $L_o(x) = L_o(-x)$ for all $x \in D_x$. In this case we can pick any polynomial whose monomials have even degree as the candidate for the generalized function.

Chapter 7

Suppressing Riser-Based Slugging in Multiphase Flow by State Feedback

7.1 Introduction

The theoretical development of stabilization of multiphase flow in oil-production pipelines is still in its infancy. The stabilization is related to the purpose of suppressing an oscillation phenomenon, called severe slugging, that occurs in pipelines carrying multiphase flow. Severe slugging in pipelines is caused by inclined or vertical pipe sections, and is potentially damaging to downstream processing equipment such as separators. Moreover, large oscillations may cause lower oil production. While the traditional remedy is to manually choke the flow at the expense of lower production, automatic control has the potential of removing oscillations without production loss (see [15] for the potential benefit of suppressing slug). It is therefore essential to develop control strategies that guarantee attenuation of severe slugging.

An important step in the development of a stabilization scheme in this direction can be traced back to [16], where it was shown that active choking could remove oscillations in a vertical riser. In [7, 14, 15], it was shown that by stabilizing the riser base pressure by active choking, large oscillations are effectively removed. Despite the fact that active control manages to suppress slugging, none of the previous works, to the best of our knowledge, has proved from a mathematical point of view why the control scheme works.

In this paper, we design a state feedback control law which is able to suppress severe slugging occurring in the model developed in [41]. Theoretically, the feedback can achieves regulation of the output to its set-point. The feedback is based on the input-output linearization approach, where the output is chosen such that it satisfies certain conditions.

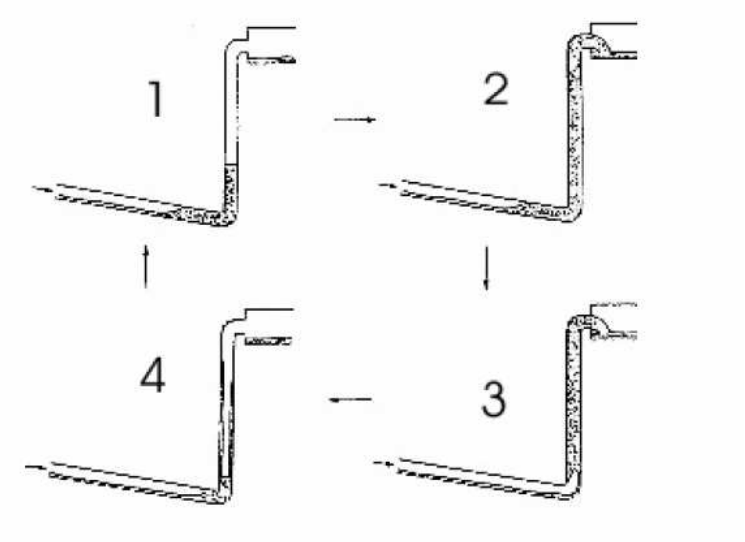

Figure 7.1: Severe slugging in the pipeline-riser system

7.2 Model

Mathematical models of multiphase flow can be found in for instance [1, 4, 41, 42], and are usually of different type depending on the application and the assumptions made. In this paper, we consider a mathematical model of multiphase flow [41] which captures gravity-induced slugging in a pipeline-riser system where the inclination of the pipe may vary from case to case, while the riser is vertical. Generally, severe slugging in the riser can be described as follows (see Figure 7.1). When multiphase flow (gas and liquid) enter the riser at relatively low rate, the liquid stays in the riser base. The liquid will block the gas from entering the riser until the pressure of the gas upstream the riser base can overcome the hydrostatic pressure of the liquid in the riser. When the pressure of the gas is high enough, the gas penetrates into the riser, violently pushing the accumulated liquid out of the riser. This behavior causes high fluctuations in the separator, and may damage it.

The model (see Figure 7.2) can be written as

$$\dot{x}_1 = w_{gc} - w_g(x), \tag{7.1}$$
$$\dot{x}_2 = w_g(x) - w_{gp}(x,u), \tag{7.2}$$
$$\dot{x}_3 = w_{oc} - w_{op}(x,u), \tag{7.3}$$

where $x = [x_1, x_2, x_3]^T$ is the state of the system, x_1 is the total mass of gas in the volume upstream of the riser base (volume one), x_2 is the total mass of gas in the riser (volume two), x_3 is the total mass of liquid, u is the opening position of the production orifice (control input to the system), w_{gc} is the constant mass flow rate of gas into volume one, w_g is the mass flow rate of gas from volume one into volume two, w_{gp} is the mass flow rate of gas through the production orifice, w_{oc} is the constant mass flow rate of liquid entering the riser, and w_{op} is the mass flow rate of produced liquid through the production orifice. The non-constant flows in equation (7.1)–(7.3) are expressed as

$$w_g(x) = v_{G1}(x)\,\rho_{G1}(x_1)\,\widehat{A}(x), \tag{7.4}$$
$$w_{gp}(x,u) = (1-\alpha_L^m(x))w_p(x)\,u, \tag{7.5}$$
$$w_{op}(x,u) = \alpha_L^m(x)\,w_p(x)\,u, \tag{7.6}$$

where v_{G1} is the gas velocity at the riser base, ρ_{G1} is the density of gas in volume one, $\widehat{A}$ is the gas flow area at the riser base, α_L^m is the oil fraction (mass basis) through the valve, and w_p is the total mass flow rate through the valve when it is fully open. They are given by

$$v_{G1}(x) = \begin{cases} K_2 \frac{H_1 - h_1(x)}{H_1} \sqrt{\frac{P_1(x_1) - P_2(x) - \rho_L g H_2 \alpha_L(x)}{\rho_{G1}(x_1)}}, \\ \qquad\qquad \text{if } h_1(x) < H_1 \\ 0, \qquad \text{otherwise} \end{cases},$$
$$\rho_{G1}(x_1) = \frac{x_1}{V_{G1}},$$
$$\alpha_L^m(x) = \alpha_{LT}(x)\frac{\rho_L}{\rho_T(x)},$$
$$\widehat{A}(x) = r^2\left[\pi - \varphi(x) - \cos(\pi - \varphi(x))\sin(\pi - \varphi(x))\right],$$
$$w_p(x) = K_1\sqrt{\rho_T(x)\left[P_2(x) - P_0\right]},$$

where K_2 is the internal gas flow constant, H_1 is the critical oil level, H_2 is the height of the riser, ρ_L is the density of oil, g is the specific gravity, V_{G1} is the size of volume one, r is the radius of the pipe, K_1 is the valve constant, and P_0 is the constant pressure after the valve. The liquid level at the riser base (h_1), the pressure in volume one (P_1), the pressure in volume

two (P_2), the average liquid fraction (volume basis) in the riser (α_L), the angle φ, the liquid fraction (volume basis) through the valve (α_{LT}) and the density through the valve (ρ_T) are given by

$$\begin{aligned}
h_1(x) &= \frac{V_L(x_3) - V_{LR}(x)}{A_1}, \\
P_1(x_1) &= \frac{x_1 RT}{M_G V_{G1}}, \\
P_2(x) &= \frac{x_2 RT}{M_G V_{G2}(x)}, \\
\alpha_L(x) &= \frac{V_{LR}(x)}{V_T}, \\
\varphi(x) &= \cos^{-1}\left(\frac{(H_1 - h_1)\cos\theta}{r} - 1\right), \\
\alpha_{LT}(x) &= \begin{cases} \frac{V_{LR}(x) - A_2 H_2}{A_3 H_3 (1+w(x))} + \frac{w(x)}{1+w(x)}\alpha_L, \\ \qquad\qquad \text{if } V_{LR}(x) > A_2 H_2 \\ \frac{w(x)}{1+w(x)}\alpha_L(x), \qquad \text{otherwise} \end{cases}, \\
\rho_T(x) &= \alpha_{LT}(x)\rho_L + (1 - \alpha_{LT}(x))\rho_{G2}(x),
\end{aligned}$$

where A_1 is the cross section area in the horizontal plane upstream the riser base, R is the gas constant, T is the constant system temperature, M_G is the molecular weight of gas, V_T is the total volume of the riser, θ is the inclination of the feed pipe, A_2 is the cross sectional area in the horizontal plane of the riser, A_3 is the cross sectional area of the horizontal top section and H_3 is the length of the horizontal top section. The volume occupied by liquid (V_L), the volume of liquid in the riser (V_{LR}), the size of volume two (V_{G2}), the friction function (w) and the gas density in volume two (ρ_{G2}) are given by

$$\begin{aligned}
V_L(x_3) &= \frac{x_3}{\rho_L}, \\
V_{G2}(x) &= V_T - V_{LR}(x), \\
V_{LR}(x) &= \frac{\rho_{mix}(x) V_T - x_2}{\rho_L}, \\
w(x) &= \frac{K_3 \rho_{G1}(x_1) v_{G1}^2(x)}{(\rho_L - \rho_{G1}(x_1))^n}, \\
\rho_{G2}(x) &= \frac{x_2}{V_{G2}(x)},
\end{aligned}$$

where ρ_L is the liquid density, n is the tuning parameter in the friction expression, and K_3 is the friction parameter. The average density in the

riser (ρ_{mix}) satisfies the relation

$$\rho_{mix}(x)\, g\,(H_2 + H_3) - \rho_L g h_1(x) = P_1(x_1) - P_2(x)\,.$$

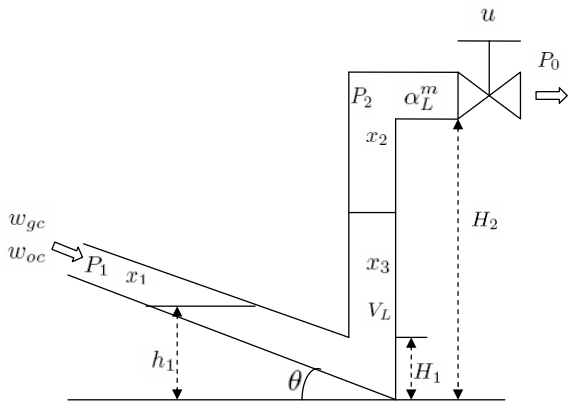

Figure 7.2: The pipeline-riser system with parameters

7.3 State Feedback Design Based on Input Output Linearization

In this section we design a state feedback control law which is based on the input-output linearization technique [18]. The output in this case is a variable ψ which is selected by the designer and it has to satisfy certain conditions. Possible candidates for the output are all variables which are dependent on the state x (see Section 7.2). The set-point of the selected variable is denoted by ψ^*.

7.3.1 Feedback Design

Suppose we select $\psi(x)$ and u as the variable to be controlled (the output) and the manipulated variable (the input), respectively. The selected variable $\psi(x)$ should be continuous and bounded on a domain D. Throughout the chapter, we assume that the set $\Omega_D := \{\psi(x) \in \mathbb{R}_+ \mid x \in D\}$ contains all admissible ψ.

We can rewrite the system as

$$\dot{x}_1 = w_{gc} - w_g(x,\psi), \tag{7.7}$$
$$\dot{x}_2 = w_g(x,\psi) - [1-\alpha_L^m(x,\psi)]\, w_p(x,\psi)\, u, \tag{7.8}$$
$$\dot{x}_3 = w_{oc} - \alpha_L^m(x,\psi)\, w_p(x,\psi)\, u, \tag{7.9}$$

and the dynamics of the variable to be controlled can be expressed as

$$\dot{\psi} = f_\psi(x) + g_\psi(x)\, u \tag{7.10}$$

where

$$f_\psi(x) = \frac{\partial\psi}{\partial x_1}(w_{gc} - w_g(x,\psi)) + \frac{\partial\psi}{\partial x_2} w_g(x,\psi) + \frac{\partial\psi}{\partial x_3} w_{oc},$$
$$g_\psi(x) = -\left[\frac{\partial\psi}{\partial x_2}(1-\alpha_L^m(x,\psi)) + \frac{\partial\psi}{\partial x_3}\alpha_L^m(x,\psi)\right] w_p(x,\psi).$$

We assume that the selection of $\psi(x)$ guarantees that the following assumptions are satisfied.

Assumption 7.1 *$g_\psi(x) \neq 0$ for $x \in D$.*

Assumption 7.2 *The sets*

$$\{x \in D \mid g_\psi(x) < 0, f_\psi(x) < 0, \psi(x) - \psi^* < 0\} \tag{7.11}$$
$$\{x \in D \mid g_\psi(x) > 0, f_\psi(x) > 0, \psi(x) - \psi^* > 0\} \tag{7.12}$$

are empty.

Assumption 7.3 *$|g_\psi(x)| \geq |f_\psi(x)|$ for $x \in D$.*

Proposition 7.1 *Under Assumption 7.1 and the feedback*

$$u = \frac{f_\psi(x) + \lambda(\psi(x) - \psi^*)}{-g_\psi(x)} \tag{7.13}$$

where $\lambda > 0$, the equilibrium point $\psi = \psi^*$ of (7.10) is asymptotically stable.

Proof. By Assumption 7.1 the feedback (7.13) does not have any singularity in the domain D. Applying the feedback scheme (7.13) in (7.10) yields

$$\dot{\psi} = -\lambda(\psi - \psi^*).$$

Consider the Lyapunov function candidate $V = (\psi - \psi^*)^2$. Then its time derivative

$$\dot{V} = -2\lambda (\psi - \psi^*)^2$$

is negative definite. ■

In applications, the feedback in the form (7.13) has to be saturated since u is the valve opening which is in the range between zero and one. The following theorem presents the result on saturated feedback.

Theorem 7.1 *Consider*

$$\dot{\psi} = f_\psi (x) + g_\psi (x) \tilde{u} \tag{7.14}$$

$$\tilde{u} := \begin{cases} 0, & \text{if } u < 0 \\ u, & \text{if } 0 \leq u \leq 1 \\ 1, & \text{if } u > 1 \end{cases} \tag{7.15}$$

Under Assumption 7.1, 7.2 and 7.3 and the feedback u in the form (7.13), the equilibrium point $\psi = \psi^*$ is asymptotically stable.

Proof. Consider the Lyapunov function candidate $V = (\psi - \psi^*)^2$. We have $\dot{V} = 2(\psi - \psi^*)\dot{\psi}$.

1. Case $0 \leq u \leq 1$: See the proof of Proposition 7.1.

2. Case $u < 0$: If $g_\psi (x) < 0$, then (7.13) gives $f_\psi (x) < \lambda (\psi^* - \psi (x))$. Then by (7.11) of Assumption 7.2 we have

   $$\dot{V} = 2 (\psi - \psi^*) f_\psi (x) < 0.$$

   The result in the case of $g_\psi (x) > 0$ is achieved similarly using (7.12).

3. Case $u > 1$: For the case $g_\psi (x) < 0$, (7.13) implies $f_\psi (x) + g_\psi (x) > \lambda (\psi^* - \psi (x))$. By Assumption 7.3 we then have $0 > f_\psi (x) + g_\psi (x) > \lambda (\psi^* - \psi (x))$ which implies $\psi (x) - \psi^* > 0$. Then

   $$\dot{V} = 2 (\psi - \psi^*) (f_\psi (x) + g_\psi (x)) < 0.$$

   The result in the case of $g_\psi (x) > 0$ is achieved similarly.

■

Remark 7.1 *Assumption 7.1 is imposed to avoid singularity in the feedback (7.13).*

Remark 7.2 *Assumption 7.2 is imposed for the case of saturation whenever the feedback (7.13) has negative value.*

Remark 7.3 *In Subsection 7.3.2 we use another condition to replace Assumption 7.1 and 7.2. It is sufficient to have*

$$f_\psi(x)g_\psi(x) < 0 \text{ for } x \in D \tag{7.16}$$

to guarantee Assumption 7.1 and 7.2 to hold. The main reason for this is that, based on exhaustive simulation runs, the new condition is also necessary for Assumption 1 and 2.

Remark 7.4 *Assumption 7.3 is imposed when $u > 1$ since we need to show that $[f_\psi(x) + g_\psi(x)]g_\psi(x) > 0$ in the proof of Theorem 7.1. However, the saturated feedback scheme may still guarantee convergence when $|g_\psi(x)| < |f_\psi(x)|$. For example, in the case $g_\psi(x) < 0$ where $f_\psi(x) + g_\psi(x) > 0$ at some x it follows that $f_\psi(x) + g_\psi(x) > \max[0, \lambda(\psi^* - \psi(x))]$ which gives*

$$\dot{V} = 2(\psi - \psi^*)\dot{\psi} = 2(\psi - \psi^*)(f_\psi(x) + g_\psi(x))$$

and thus

$$\begin{aligned} \dot{V} &> 0, && \text{if } \psi - \psi^* > 0 \\ \dot{V} &< -2(\psi - \psi^*)^2, && \text{if } \psi - \psi^* < 0. \end{aligned}$$

Consequently, convergence is always guaranteed whenever $\psi - \psi^* < 0$. In the case of $\psi - \psi^* > 0$, no conclusion can be made.

In practical applications, the model of multiphase flow in Section 7.2 may be modified to meet certain objectives. For instance, the equation for the production valve may vary depending on the type of valve being used. In the case of modification of the terms w_g, w_{gp}, w_{op} in the model, the feedback (7.13) can easily be modified. Thus, our feedback is quite flexible to modification of the model.

The approach discussed here still leaves an open problem where the stability of the zero dynamics is not covered due to the difficulties of the problem. However, for the case of $\psi(x) = x_3$, we can disregard this issue as discussed in the last section of this chapter.

7.3.2 Selection of Variable-To-Be-Controlled

This subsection provides some approaches on selecting the controlled variable ψ which fits the proposed feedback scheme. In short, the approaches should meet the conditions in Assumption 7.1, 7.2 and 7.3, at least in some region of a given set-point. These approaches are not exact, but to some extent they can be used as guidelines for selecting the controlled variables.

Empirical Approach

The first approach is to plot the functions f_ψ and g_ψ for a given data set and then select the variables, which satisfy Assumption 7.1 and 7.2, from the plots. For this purpose we run a simulation of the system with a given set of parameters. In this case the parameters are set so that

$V_{G1} = 12.158$ m^3	$\rho_L = 750$ kg/m^3
$\theta = 0.0274$ rad	$H_1 = 0.12$ m
$H_2 = 300$ m	$H_3 = 0.12$ m
$L_3 = 100$ m	$A_1 = 0.4128$ m^2
$A_2 = 0.0113$ m^2	$A_3 = 12$ m^2
$R = 8314$ J/(K*Kmol)	$T = 308$ K
$M_G = 35$ kg/Kmol	$g = 9.81$ m/s^2
$P_0 = 50 \times 10^5$ N/m^2	$V_T = 4.8329$ m^3
$n = 2.55$	$K_1 = 0.0051$
$K_2 = 4.3983$	$K_3 = 0.2030$

The opening of the production orifice is set to 50% ($u = 0.5$). The simulation is run for 30 minutes and the states oscillate as shown in Figure 7.3. This means that the constant input of 50% opening induces severe slugging. Note that our purpose in the end is to stabilize a variable at a certain set-point and then to see whether it will suppress the slugging or not. A set-point here means a point which is associated with the equilibrium condition $\dot{x} = 0$, when a certain constant input u is applied. In this case the input u could be a constant value between zero and one.

The candidates for the controlled variable are ρ_{G1}, P_1, V_L, ρ_{mix}, V_{LR}, h_1, V_{G2}, P_2, α_L, ϕ, A, v_{G1}, ρ_{G2}, w, α_{LT}, ρ_T, α_L^m and w_g. All these variables are dependent on the state x (see Section 7.2). Based on the available data set for 30 minutes, the only variables which satisfy (7.16) are V_L, h_1, P_2 and ρ_{G2} (see Figure 7.4). Note that we skip the plots associated with ρ_{G2} as they are equivalent to those with P_2 (see also from the equations of ρ_{G2}

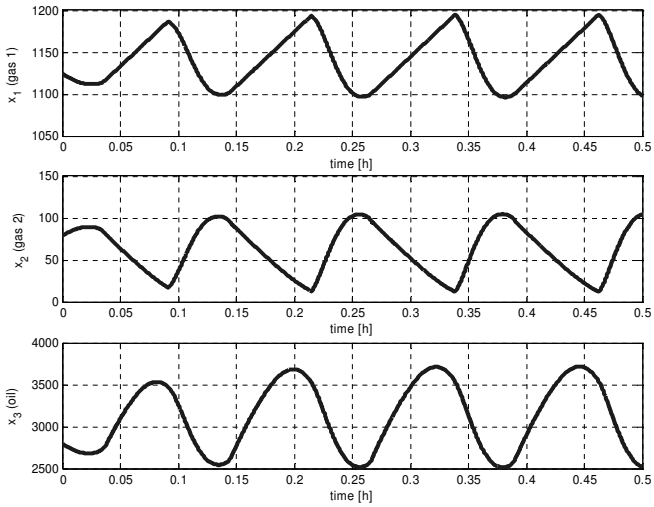

Figure 7.3: The state x for $u = 0.5$

and P_2 in Section 7.2). The next step is to check whether Assumption 7.3 is also satisfied for V_L, h_1 and P_2. Figure 7.4 shows that none of the chosen variables satisfy the required condition of Assumption 7.3.

We should keep in mind that the current selection process is based on the data set of severe slugging where large magnitude of oscillatory behavior of the state x is expected. This is the reason why at some times Assumption 7.3 is not satisfied for V_L, h_1 and P_2. As a comparison we perform another simulation for 10% of opening of the production orifice which is in the stable region. The response of the state to the input $u = 0.1$, when the initial condition is associated with an oscillatory behavior, can be seen from Figure 7.5 where the state finally converges to a set-point after oscillating. Figure 7.6 shows the corresponding plots of $g_\psi(x)$, $f_\psi(x)$ and $|f_\psi(x)|/|g_\psi(x)|$ for V_L, h_1 and P_2. It indicates that during the small magnitude of oscillation in the state x, the magnitude of f_ψ is always smaller than that of g_ψ. Thus we can say that whenever the magnitude of oscillation is small we can select V_L, h_1 and P_2 as the controlled variable in our scheme.

Analytical and Empirical Approach

The next approach is a mixture of analytical and empirical nature. The candidates to be assessed by this approach is the total mass of the liquid x_3

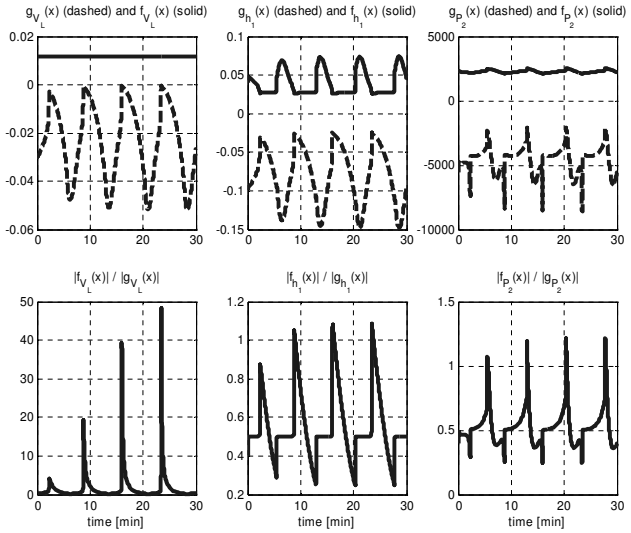

Figure 7.4: $g_\psi(x)$, $f_\psi(x)$ and $|f_\psi(x)|/|g_\psi(x)|$ for V_L, h_1 and P_2 in the case $u = 0.5$

and the total mass $M = x_1 + x_2 + x_3$. We then obtain

$$\begin{aligned} f_{x_3}(x) = w_{oc} > 0, \quad g_{x_3}(x) = -\alpha_L^m(x) w_p(x) < 0 \\ f_M(x) = w_{gc} + w_{oc} > 0, \quad g_M(x) = -w_p(x) < 0 \end{aligned}$$

where $f_\psi(x)$ is a constant for both cases. By (7.16), Assumption 7.1 and 7.2 are satisfied. In the event of severe slugging, as a result of blocking, the total mass flow rate through the production valve ($w_p(x)$) and the oil fraction (mass basis) through the production valve ($\alpha_L^m(x)$) become very small. As a result the absolute value of $g_\psi(x)$ is very small compared to that of $f_\psi(x)$ for x_3 and M. Consequently Assumption 7.3 is not satisfied. On the other hand, when there is no slugging, using the same reasoning we can guarantee that the absolute value of $g_\psi(x)$ is big enough to satisfy Assumption 7.3 for x_3 and M.

Closed Loop Investigation

In real applications, especially for safety reasons, large oscillatory behavior is avoided. The large magnitude of oscillation can be avoided during stabilization of an unstable set-point by selecting an initial set-point which is associated with the stable region (the opening of the production orifice which does not induce slugging). By slowly changing the set-point from the stable

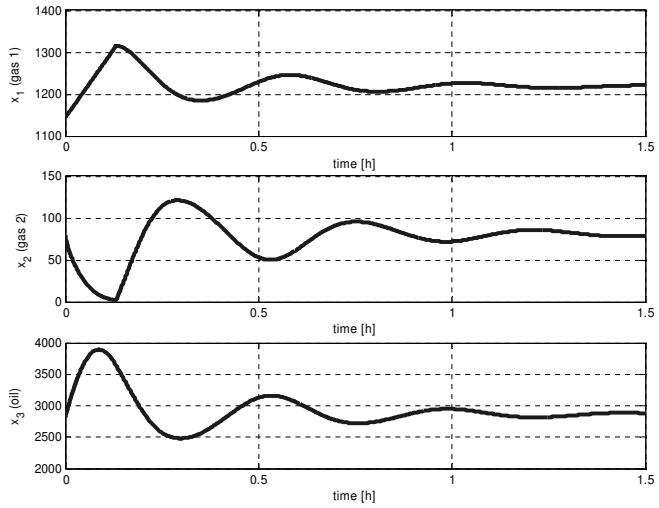

Figure 7.5: The state x for $u = 0.1$

one to the slugging region while the controller is working, stabilization is also achieved.

For closed loop simulation with the feedback when the controlled variable is h_1, we set $\lambda = 2 \times 10^8$. The initial condition is set to the point associated with the stable region of constant opening $u = 0.1$. The purpose of the feedback is to stabilize h_1 at the point 0.1099 which is associated with the equilibrium condition of $u = 0.2$ (slugging case). In Figure 7.7 we can see that the states converge to a point and the state feedback does not experience saturation (Figure 7.8). Unfortunately the state feedback u does not converge to the desired set point of 0.2. Instead, the state feedback u converges to the point 0.1877. To understand what happen we can observe the plot of h_1 in Figure 7.8. The variable h_1 converges to the point 0.1099 which is the point associated with the equilibrium condition of $u = 0.2$. But apparently the point $h_1 = 0.1099$ is also associated with the equilibrium condition of $u = 0.1877$. In this case the controller actually works well that the controlled variable h_1 converges to 0.1099 even though u converges to another point. In this case the internal dynamic of the system does not evolve as what is expected. Thus the selection of h_1 as the variable to be controlled has a drawback in that the state x is 'unobservable' from the information given by h_1.

With the same initial condition we try another feedback where the variable to be controlled is P_2. The constant λ is set to 3000. The feedback is designed

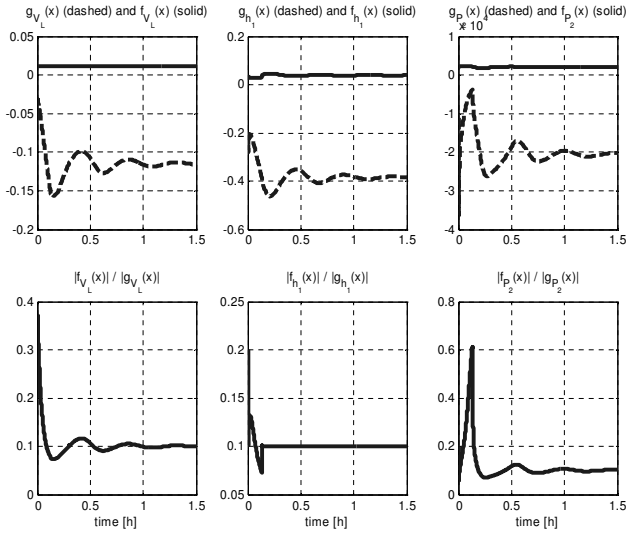

Figure 7.6: $g_\psi(x)$, $f_\psi(x)$ and $|f_\psi(x)|/|g_\psi(x)|$ for V_L, h_1 and P_2 in the case $u = 0.1$

in such a way that the system moves from the initial condition associated with the equilibrium condition of $u = 0.1$, gradually with step 0.1, to the final one of $u = 0.5$. The results of the simulation using the feedback can be seen in Figure 7.9 and Figure 7.10 where all the states are converging and the feedback also converges to the desired points. Thus the feedback works well when the variable to be controlled is P_2. The same type of results are also obtained when we select M and x_3 (equivalently V_L; see Section 7.2) as the controlled variable.

Clearly, the foregoing stabilization scenario of P_2, x_3 and M indicates that severe slugging can be suppressed as satisfactorily demonstrated by the results of simulation of the states. Validating the results analytically needs a further investigation which is not easy since the model is quite complicated. However, validation can be done easily in the case of selecting x_3 as the controlled variable. Stabilizing the controlled variable x_3 at a set-point guarantees that the total mass of liquid (oil) does not fluctuate and thus no blockage occurs. In this case severe slugging can be avoided and we can ignore whether x_1 and x_2 oscillate or not. Hence the stability of the zero dynamics can be disregarded here.

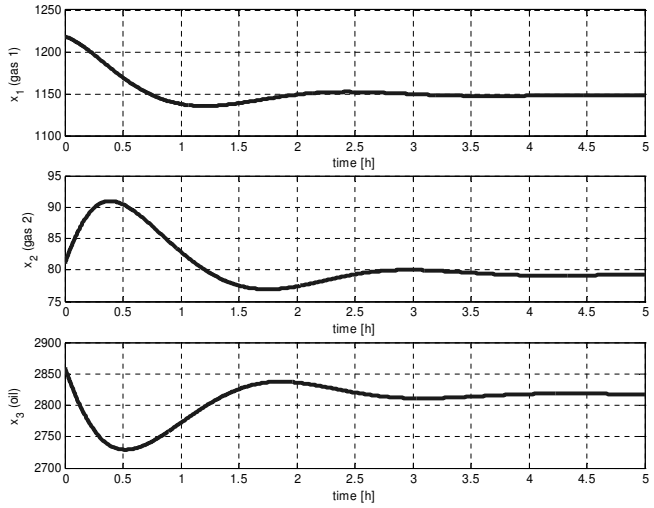

Figure 7.7: The state x (when h_1 is the controlled variable)

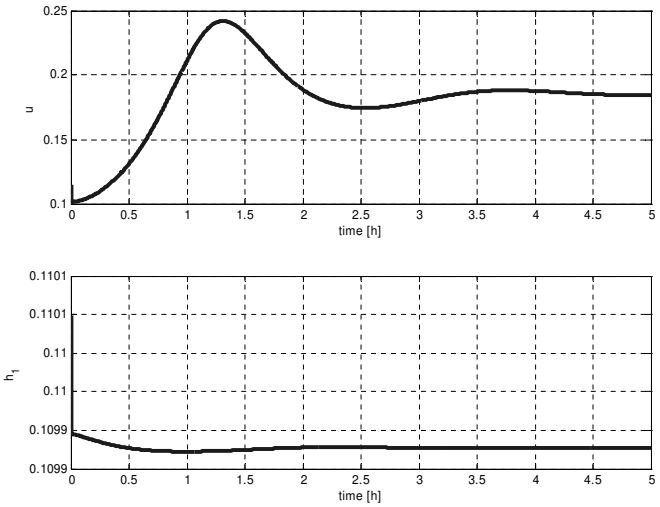

Figure 7.8: The feedback u and the liquid level h_1 as the controlled variable

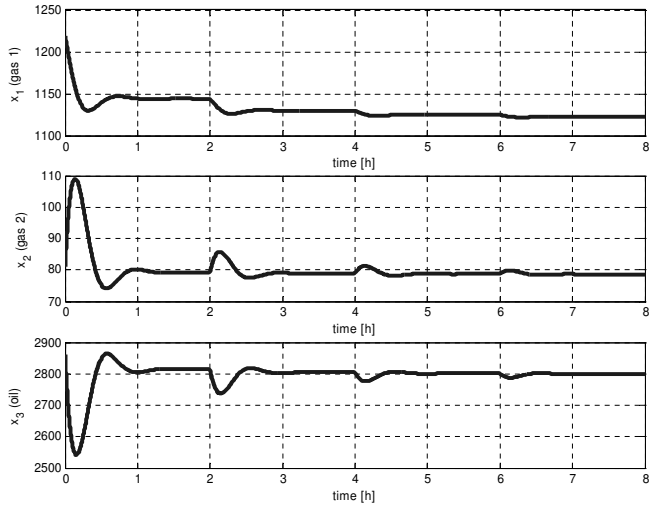

Figure 7.9: The state x (when P_2 is the controlled variable)

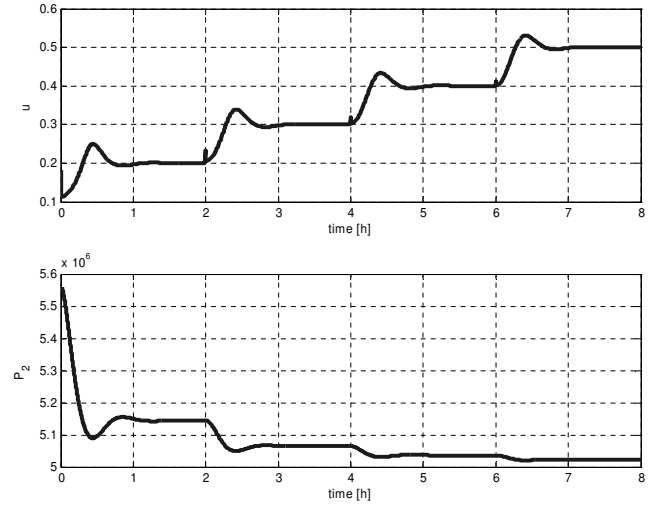

Figure 7.10: The feedback u and the pressure P_2 as the controlled variable

Chapter 8

Conclusions

8.1 Summary

This thesis consists of two contributions, the main contribution in model approximation, and the additional contribution in stabilization of multiphase flow in the riser. The contribution in model approximation is divided into two parts, the optimal rank approximation of linear operators, and the approximation of polynomial nonlinear systems.

In Chapter 3 we introduce the notion of induced p-norm singular values and show their relevance for a number of problems. In particular, we address the optimal rank approximation problem and derive sufficient conditions for the existence of optimal approximants which minimize the induced p-norm of the error.

In Chapter 4 an approach to construct a reduced order model for a class of nonlinear system is developed. The approach is heuristic in nature due to the coupling of the unknown variables in verifying a finite gain $\mathcal{L}_2$ stability condition. Despite the fact that the approach is heuristic it offers a systematic means of computation using sum of squares programming which amounts to LMI feasibility problem.

In Chapter 5 we propose an approach to decouple the unknown variables in verifying the finite gain $\mathcal{L}_2$ stability condition for model reduction of a class of polynomial systems. First we seek a transformation based on an estimate of the reachability set of the system. The estimate is computed by means of sum of squares programming. The transformed system is then truncated and the truncated system will be the state space system of the

reduced model. For the second step, through sum of squares programming, the output of the reduced model is determined such that the error model satisfies the relaxation of the finite gain $\mathcal{L}_2$ stability condition.

A direction to improve the first step of the approach in Chapter 5 would be to seek a transformation which separates the weakly reachable and observable part simultaneously from the strongly reachable and observable part. This direction is discussed in Chapter 6 which introduces a novel approach to balancing a polynomial nonlinear system and truncates the balanced representation.

And for the contribution on stabilization of multiphase flow in the riser pipeline in Chapter 7, an early phase in the design method of state feedback for the purpose of suppressing riser-induced slugging occurring in a vertical pipeline has been presented. The method utilizes the input-output linearization techniques. Under certain conditions the feedback scheme is guaranteed to work in the case of saturation. These conditions can be applied as the bases for selecting the variable to be controlled. By carefully selecting the variable to be controlled the feedback scheme can avoid severe slugging. Although the selection processes are not exact, they can be used as guidelines for suppressing riser-based slugging.

8.2 Future Directions

The methods of model reduction for polynomial nonlinear systems in this thesis rely on the information from the generalized observability and reachability functions as they can be efficiently constructed by means of sum of squares programming. Specifically, in Chapter 4, we have shown that the existence of the generalized functions is necessary for the strong $\mathcal{H}_\infty$ performance model reduction. However, the question whether the reverse also holds remains open. In this case the nonuniqueness of representation $A(x)$ and $C(x)$ may be investigated to show whether the existence of the generalized functions with certain structure will guarantee the $\mathcal{H}_\infty$ performance model reduction (and hence $\mathcal{L}_2$-gain model reduction). Furthermore, this issue of nonuniqueness still open a question whether there always exist a certain representation of $A(x)$ and $C(x)$ which guarantee the strong $\mathcal{H}_\infty$ performance.

In Chapter 5 the method of model reduction consists of two steps. The first step of the approach removes part of the original system which is weakly reachable. On the other hand, this part of the system can be exactly the

most observable part. It is unclear how well the second step can cope with this problem. Further investigation is needed to resolve this issue for future work.

The balanced truncation approach in Chapter 6 still needs further investigation especially on the issue of accuracy. It is not yet clear how far the model reduction scheme can preserve stability. Another important issue is on the error analysis between the reduced model and the original model which is not covered in the thesis.

In general, the power of sum of squares programming has not been exploited extensively for model reduction while it certainly has potential benefits for computational purposes. For this reason it becomes more apparent that other type of model reduction scheme can benefit from it. For example, the potential use of sum of squares programming for balancing certain energy functions [37] should be investigated. Furthermore, there are many type of energy functions related to the dissipation inequality [44] which can be exploited and might give a new computational scheme for model reduction.

Bibliography

[1] O.M. Aamo, G.O. Eikrem, H.B. Siahaan and B.A. Foss, "Observer Design for Multiphase Flow in Vertical Pipes with Gas-Lift: Theory and Experiments", *Journal of Process Control*, 5, Issue 3, pp. 247-257, 2005.

[2] V.M. Adamjan, D.Z. Arov and M.G. Krein, Infinite Block Hankel Matrices and Related Extension Problems, *American Math. Society Transactions,* 111, pp. 133-156, 1978.

[3] A.C. Antoulas, D.C. Sorensen and S. Gugercin, A Survey of Model Reduction Methods for Large-Scale Systems. In: V. Olshevsky (ed), *Structured Matrices in Mathematics, Computer Science and Engineering, Vol. I. Contemporary Mathematics Series*, pp. 193-219, AMS, 2001.

[4] K. Bendiksen, D. Malnes and O. Nydal, "On the Modelling of Slug Flow", *Chemical Engineering Science*, 40, pp. 59-75, 1985.

[5] D.S. Bernstein, *Matrix Mathematics*, Princeton Univ. Press, Princeton, 2005.

[6] S. Boyd, L. El Ghaoui, E. Feron and V. Balakrishnan, *Linear Matrix Inequalities in Systems and Control Theory*, SIAM, Philadelphia, 1994.

[7] M. Dalsmo, E. Halvorsen and O. Slupphaug, "Active Feedback Control of Unstable Wells at the Brage Field", *SPE Production and Facilities,* SPE 77650, 2002.

[8] C.A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*, Academic Press, London, 1975.

[9] G.E. Dullerud and F. Paganini, *A Course in Robust Control Theory: a Convex Approach*, Springer-Verlag, New York, 2000.

[10] B.A. Francis, *A Course in H_∞ Control Theory*, Springer-Verlag Berlin, Heidelberg, 1987.

[11] K. Glover, All Optimal Hankel-norm Approximations of Linear Multivariable Systems and Their L_∞Error Bounds, *International Journal of Control*, vol. 39, pp 1115-1193, 1984.

[12] G.H. Golub and C.F. van Loan, *Matrix Computations*, The John Hopkins University Press, Baltimore, 1989.

[13] K.M. Grigoriadis, Optimal H_∞Model Reduction via Linear Matrix Inequalities: Continuous and Discrete-Time Cases, *Systems and Control Letters*, 26, pp. 321-333, 1995.

[14] K. Havre and M. Dalsmo, "Active Feedback Control as a Solution to Severe Slugging", *SPE Production and Facilities,* SPE 79252, pp. 138-148, 2002.

[15] K. Havre, K.O. Stornes and H. Stray, "Taming Slug Flow in Pipelines", *ABB Review,* 4, Dec. 2000.

[16] P. Hedne and H. Linga, "Supression on Terrain Slugging with Automation and Manual Riser Choking", *Advances in Gas-Liquid Flows*, pp. 453-460, 1990.

[17] D. Kavranoglu and M. Bettayeb, Characterization of the Solution to the Optimal H_∞ Model Reduction Problem, *Systems and Control Letters,* 20, pp. 99-107, 1993.

[18] H.K. Khalil, *Nonlinear Systems*, Prentice Hall, New Jersey, 2002.

[19] E. Kreyszig, *Advanced Engineering Mathematics*, John Wiley & Sons, New York, 1999.

[20] S. Lall, J. Marsden and S. Glavaski, A Subspace Approach to Balanced Truncation for Model Reduction of Nonlinear Control Systems, *International Journal of Robust and Nonlinear Control*, vol. 12, no. 6, pp. 519-535, 2002.

[21] Y. Liu and B.D.O. Anderson, Singular Perturbation Approximation of Balanced Systems, *International Journal of Control*, vol. 50, pp. 1379–1405, 1989.

[22] W.-M. Lu and J.C. Doyle, H_∞ Control of Nonlinear Systems: a Convex Characterization, *IEEE Transaction on Automatic Control,* 40:9, pp. 1668-1675, 1995.

[23] J. Löfberg, YALMIP: A Toolbox for Modeling and Optimization in MATLAB, *Proceedings 2004 CACSD Conference*, Taipei, Taiwan, 2004. Available at `http://control.ee.ethz.ch/~joloef/yalmip.php`.

[24] D.G. Meyer, Fractional Balanced Reduction - Model Reduction via a Fractional Representation, *IEEE Transaction on Automatic Control*, vol. 35, pp. 1341-1345, 1990.

[25] B.C. Moore, Principal Component Analysis in Linear Systems: Controllability, Observability and Model Reduction, *IEEE Transaction on Automatic Control*, vol. 26, pp. 17-32, 1981.

[26] D. Mustafa and K. Glover, Controller Reduction by H_∞-Balanced Truncation, *IEEE Transaction on Automatic Control*, vol. 36, pp. 668-682, 1991.

[27] R. Ober and D. McFarlane, Balanced Canonical Forms for Minimal Systems: a Normalized Coprime Factor Approach, *Linear Algebra and its Applications*, 122-124, pp. 23-64, 1989.

[28] P. Opdenacker and E.A. Jonckheere, LQG Balancing and Reduced LQG Compensation of Symmetric Passive Systems, *International Journal of Control*, vol. 41, pp. 73–109, 1985.

[29] L. Parnebo and L.M. Silverman, Model Reduction by Balanced State Space Representations, *IEEE Transaction on Automatic Control*, vol. 27, pp. 382-387, 1982.

[30] P.A. Parrilo, Semidefinite Programming Relaxations for Semialgebraic Problems, *Mathematical Programming Ser. B*, 96:2, pp. 293-320, 2003.

[31] S. Prajna, A. Papachristodoulou and P.A. Parrilo, Introducing SOSTOOLS: a General Purpose Sum of Squares Programming Solver, *Proceedings 41st IEEE Conference on Decision and Control*, 2002. Available at `http://www.cds.caltech.edu/sostools`.

[32] S. Prajna, A. Papachristodoulou and F. Wu, Nonlinear Control Synthesis by Sum of Squares Optimization: a Lyapunov-based Approach, *Proceedings 2004 Asian Control Conference*, Melbourne, Australia, 2004.

[33] S. Prajna and H. Sandberg, On Model Reduction of Polynomial Dynamical Systems, *Proceedings 44th IEEE Conference on Decision and Control and European Control Conference 2005*, Seville, Spain, pp. 1666-1671, 2005.

[34] M. Rathinam and L.R. Petzold, A New Look at Proper Orthogonal Decomposition, *SIAM Journal on Numerical Analysis*, 41:5, pp. 1893-1925, 2003.

[35] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[36] A. van der Schaft, *L_2-Gain and Passivity Techniques in Nonlinear Control*, SpringerVerlag, London, 2000.

[37] A. van der Schaft, On Balancing of Passive Systems, *Proceedings European Control Conference 2007*, Kos, Greece, 2007.

[38] J. Scherpen, Balancing for Nonlinear Systems, *Systems and Control Letters*, 21, pp. 143-153, 1993.

[39] G. Stengle, A Nullstellensatz and a Positivestellensatz in Semialgebraic Geometry, *Math. Ann.*, 207, pp. 87-97, 1974.

[40] G.W. Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, London, 1990.

[41] E. Storkaas, S. Skogestad and J. Godhavn, A Low Dimentional Model of Severe Slugging for Controller Design and Analysis, *Proceedings Multi Phase'03*, San Remo, Italy, 11-13 June 2003.

[42] Y. Taitel and D. Barnea, Two Phase Slug Flow, *Advances in Heat Transfer*, 20, pp. 71-103, 1990.

[43] E.I. Verriest and T. Kailath, On Generalized Balanced Realizations, *IEEE Transaction on Automatic Control*, vol. 28, pp. 833-844, 1983.

[44] J.C. Willems, Dissipative Dynamical Systems - Part I: General Theory, *Archive for Rational Mechanics and Analysis*, vol. 45, pp. 321-351, 1972.

[45] K. Zhou, G. Salomon and E. Wu, Balanced Realization and Model Reduction for Unstable Systems, *International Journal of Robust and Nonlinear Control*, vol. 9, pp. 183–198, 1999.

-9432147212-