

KRYLOV SUBSPACE TECHNIQUES FOR
MODEL REDUCTION AND THE SOLUTION OF
LINEAR MATRIX EQUATIONS

Mian Ilyas Ahmad

Thesis submitted for the degree of
Doctor of Philosophy
Imperial College London

Control and Power Group
Department of Electrical and Electronic Engineering
Imperial College London
June 2011

Abstract

This thesis focuses on the model reduction of linear systems and the solution of large scale linear matrix equations using computationally efficient Krylov subspace techniques. Most approaches for model reduction involve the computation and factorization of large matrices. However Krylov subspace techniques have the advantage that they involve only matrix-vector multiplications in the large dimension, which makes them a better choice for model reduction of large scale systems. The standard Arnoldi/Lanczos algorithms are well-used Krylov techniques that compute orthogonal bases to Krylov subspaces and, by using a projection process on to the Krylov subspace, produce a reduced order model that interpolates the actual system and its derivatives at infinity. An extension is the rational Arnoldi/Lanczos algorithm which computes orthogonal bases to the union of Krylov subspaces and results in a reduced order model that interpolates the actual system and its derivatives at a predefined set of interpolation points. This thesis concentrates on the rational Krylov method for model reduction.

In the rational Krylov method an important issue is the selection of interpolation points for which various techniques are available in the literature with different selection criteria. One of these techniques selects the interpolation points such that the approximation satisfies the necessary conditions for $\mathcal{H}_2$ optimal approximation. However it is possible to have more than one approximation for which the necessary optimality conditions are satisfied. In this thesis, some conditions on the interpolation points are derived, that enable us to compute all approximations that satisfy the necessary optimality conditions and hence identify the global minimizer to the $\mathcal{H}_2$ optimal model reduction problem. It is shown that for an $\mathcal{H}_2$ optimal approximation that interpolates at m interpolation points, the interpolation points are the simultaneous solution of m multivariate polynomial equations in m unknowns. This condition reduces to the computation of zeros of a linear system, for a first order approximation. In case of second order approximation the condition is to compute the simultaneous solution of two bivariate polynomial equations. These two cases are analyzed in detail and it is shown that a global minimizer to the $\mathcal{H}_2$ optimal model reduction problem can be identified. Furthermore, a computationally efficient iterative algorithm is also proposed for the $\mathcal{H}_2$ optimal model reduction problem that converges to a local minimizer.

In addition to the effect of interpolation points on the accuracy of the rational interpolating approximation, an ordinary choice of interpolation points may result in a reduced order model that loses the useful properties such as stability, passivity, minimum-phase

and bounded real character as well as structure of the actual system. Recently in the literature it is shown that the rational interpolating approximations can be parameterized in terms of a free low dimensional parameter in order to preserve the stability of the actual system in the reduced order approximation. This idea is extended in this thesis to preserve other properties and combinations of them. Also the concept of parameterization is applied to the minimal residual method, two-sided rational Arnoldi method and $\mathcal{H}_2$ optimal approximation in order to improve the accuracy of the interpolating approximation.

The rational Krylov method has also been used in the literature to compute low rank approximate solutions of the Sylvester and Lyapunov equations, which are useful for model reduction. The approach involves the computation of two set of basis vectors in which each vector is orthogonalized with all previous vectors. This orthogonalization becomes computationally expensive and requires high storage capacity as the number of basis vectors increases. In this thesis, a restart scheme is proposed which restarts without requiring that the new vectors are orthogonal to the previous vectors. Instead, a set of two new orthogonal basis vectors are computed. This reduces the computational burden of orthogonalization and the requirement of storage capacity. It is shown that in case of Lyapunov equations, the approximate solution obtained through the restart scheme approaches monotonically to the actual solution.

Acknowledgements

Several people have contributed to the completion of my PhD dissertation. However, the most prominent personality deserving due recognition is my worth supervisor, Dr. Imad M. Jaimoukha. Thank you Imad for your endless help, valuable guidance, constant encouragement, precious advice, sincere and affectionate attitude.

I acknowledge with thanks the cooperation and assistance of my colleague and co-worker Dr. Michalis Frangos. As a result of our mutual understanding, cooperation and with the help and able guidance of Dr. Imad M. Jaimoukha we have been able to do some interesting work.

I am thankful and highly indebted to Dr. Martin van Gijzen and Prof. Kees Vuik at the Department of Applied Mathematical Analysis, Delft University of Technology (TU Delft), Netherland for their generous help, cooperation and guidance during my stay at TU Delft.

Thanks to all the staff and students in Imperial College London and to all my friends especially Dr. Asim, Dr. Ansar, Dr. Amir, Dr. Jawad, Shakil, Ammar, Dr. Inam, Dr. Faraz, Israr, Masroor, Atif and Zhi.

Special thanks to my father Prof. Mushtaq A. Mian who provided a conducive environment, valuable guidance and crucial support at all levels of my educational career and my very beloved mother whose love, affection and prayers have been my most precious assets. Also I am thankful to my elder brothers Dr. Ashfaq, Dr. Aftab, Dr. Ishaq, Dr. Afaq and my sister Dr. Haleema who have been the source of inspiration for me to persue higher studies. My immediate elder brother Mr. Asbat and my younger sister Ayesha both studying in the UK, deserve recognition for their help, well wishes and moral support.

Finally I would like to express my profound thanks and appreciation to the National University of Science and Technology (NUST) Pakistan for sponsoring my PhD at the UK and the Idea league for the financial support during my stay in the Netherlands.

Contents

Abstract	2
Acknowledgements	4
List of Figures	9
List of Tables	10
List of Algorithms	11
Notations	12
Abbreviations	13
1 Introduction	14
1.1 Problem Statement	15
1.2 Overview of Model Reduction Techniques	16
1.2.1 Balanced Truncation	17
1.2.2 Hankel norm Approximation Method	19
1.2.3 $\mathcal{H}_2$ Optimal Model Reduction	20
1.2.4 Moment Matching Methods	22
1.3 Thesis Outline	24
2 Krylov Subspace Techniques for Model Reduction	26
2.1 Krylov Projection	27
2.1.1 Approximation of linear dynamical systems	30
2.1.2 Forward error expressions	32
2.2 Arnoldi Method	32

2.2.1	Single-sided Arnoldi algorithm	33
2.2.2	Two-sided Arnoldi Algorithm	34
2.2.3	Moment matching properties	36
2.2.4	Error Analysis	37
2.3	Lanczos Algorithm	38
2.4	Rational Krylov Method	41
2.4.1	Rational Arnoldi algorithm	41
2.4.2	Moment Matching Properties	42
2.4.3	Arnoldi-like equations	43
2.4.4	Residual Error in the Rational Case	46
2.5	Rational Krylov method and $\mathcal{H}_2$ Optimal model reduction	47
2.5.1	Lyapunov-based Necessary Conditions for Optimality	48
2.5.2	Interpolation based Necessary Optimality Conditions	49
2.5.3	Iterative Rational Krylov Algorithm	50
2.6	Parameterization of rational Arnoldi method	51
2.6.1	Preserving Stability	52
2.7	Rational Krylov methods for the Sylvester and Lyapunov equations	54
2.8	Conclusions	57
3	Iterative Krylov Methods for $\mathcal{H}_2$ Optimal Model Reduction	58
3.1	Shifted variants of the Bi-CG algorithm	59
3.2	Iterative shifted Bi-CG for $\mathcal{H}_2$ optimal model reduction	62
3.3	Partition of the iterative shifted Bi-CG method	64
3.4	Conclusions	66
4	All Optimal Local Minimizers to $\mathcal{H}_2$ Optimal Model Reduction	68
4.1	Error system in the optimal case	68
4.2	First order $\mathcal{H}_2$ optimal model reduction	72
4.2.1	Computation of Fixed points	72
4.2.2	Numerical results	75
4.3	Second order $\mathcal{H}_2$ optimal model reduction	78
4.3.1	Numerical results	83
4.4	Conclusions	85

5	Parameterization of Rational Krylov Approximations	86
5.1	Error Analysis	86
5.1.1	Residual and forward errors	87
5.1.2	Perturbation analysis	88
5.2	Choice of the Parameter L to Preserve Model Properties	89
5.2.1	Zero assignment/minimum-phase approximation	89
5.2.2	Preserving bounded realness	92
5.2.3	Preserving passivity	94
5.2.4	Preserving symmetry	96
5.2.5	Preserving multiple properties	96
5.3	Choice of L to enhance the accuracy of interpolating approximations . . .	97
5.3.1	Interpolating minimum residual solution	97
5.3.2	An $\mathcal{H}_2$ -norm optimal modal truncation-like approximation	99
5.3.3	Two-sided Arnoldi	100
5.4	Numerical examples	102
5.4.1	Example 1	102
5.4.2	Example 2	104
5.5	Conclusions	108
6	Krylov Based Restart Scheme for the Solution of Linear Matrix Equations	109
6.1	Restart Scheme	109
6.2	Low Rank Residual Error Analysis	113
6.2.1	Solution of Nonsymmetric Algebraic Riccati Equation	114
6.3	Solution of Lyapunov Equation	115
6.4	Numerical Results	116
6.4.1	Example 1	117
6.5	Conclusions	117
7	Conclusions and Future Research Directions	118
7.1	Summary	118
7.2	Thesis Contribution	121
7.3	Future Work	122
7.4	Publications Arising from this Work	123

List of Figures

2.1	Orthogonal and oblique projections	28
2.2	Singular value plot for single-sided and two-sided Arnoldi approximations	36
2.3	Singular value plot for two-sided rational Arnoldi approximations with two different choices of interpolation points	48
3.1	Singular value plot for the iterative rational Krylov approximation (IRKA) and the iterative shifted bi-conjugate gradient approximation (IBi-CGS) .	65
5.1	RLC circuit	102
5.2	Step and frequency responses of the RLC-system in Figure 5.1 and its approximations for $m = 12$	105
5.3	Frequency responses for <code>heat-cont</code> and its approximations for $m = 2$. .	107
5.4	Frequency responses for <code>heat-cont</code> and its approximations for $m = 8$. .	107

List of Tables

4.1	Four different expressions for $\mathcal{H}_2$ norm of approximation error in terms of controllability, observability and cross grammians	72
4.2	Fixed Points for $F_1(s)$, $F_2(s)$ $F_3(s)$	76
4.3	Relative Error for $F_1(s)$, $F_2(s)$ $F_3(s)$	76
4.4	Fixed Points for Benchmark Examples	76
4.5	Relative Error for Benchmark Examples	77
4.6	Fixed points for $\hat{F}_7$, $\hat{F}_8$, $\hat{F}_9$	84
4.7	Relative error for $\hat{F}_7$, $\hat{F}_8$, $\hat{F}_9$	84
4.8	Fixed Points for $\hat{F}_4$, $\hat{F}_5$, $\hat{F}_{10}$	84
4.9	Relative error for $\hat{F}_4$, $\hat{F}_5$, $\hat{F}_{10}$	85
5.1	Approximation accuracy for the RLC-system in Figure 5.1	103
6.1	The 2-norm of the forward error and the residual error for the restart scheme and the standard Krylov technique to solve the Sylvester equation	116

List of Algorithms

2.1	Single-sided Arnoldi Algorithm	33
2.2	Two-sided Arnoldi Algorithm	35
2.3	The Non-symmetric Lanczos Algorithm	39
2.4	The Rational Arnoldi Algorithm	42
2.5	The iterative rational Krylov algorithm	51
3.1	The bi-conjugate gradient algorithm	60
3.2	The shifted bi-conjugate gradient algorithm	63
3.3	Iterative shifted bi-conjugate gradient method	64
3.4	Krylov basis for Shifted Bi-CG method	65
3.5	IRKA using shifted bi-conjugate gradient method	66
4.1	$\mathcal{H}_2$ optimal first order approximation of linear systems	75
5.1	Algorithm to preserve model properties	96
6.1	Restart scheme for the solution of Sylvester equation	112

Notations

Symbol	Parameter
$\in$	Belongs to
$:=$	Equal by definition
$\stackrel{s}{=}$	Equal to the state space realization
$\mathcal{R}, \mathcal{C}$	Set of all real, complex numbers
$\mathcal{R}^n, \mathcal{C}^n$	Set of all real, complex vectors of size $n \times 1$
$\mathcal{R}^{n \times m}, \mathcal{C}^{n \times m}$	Set of all real, complex matrices of size $n \times m$
$\text{Ran}(M)$	Range of $M \in \mathcal{C}^{n \times m}$
$\text{colsp}(M)$	Column span of $M \in \mathcal{C}^{n \times m}$
σ_k	k -th singular value
λ_k	k -th eigenvalue
M^T	Transpose of $M \in \mathcal{R}^{n \times m}$
M^H	Conjugate transpose of $M \in \mathcal{C}^{n \times m}$
I_k	The identity matrix of size $k \times k$
e_k	The k -th column of the identity matrix
e_k^i	$e_k^i \in \mathcal{R}^k$ with each element equal to $i \in \mathcal{R}$
$G(s)$	Transfer function with respect to $s \in \mathcal{C}$
$G'(s)$	Derivative of transfer function $G(s)$ with respect to s
$\ \cdot\ _2$	The Euclidean norm (2-norm)
$\ \cdot\ _{\mathcal{H}_2}$	The $\mathcal{H}_2$ norm
$\ \cdot\ _{\mathcal{H}_\infty}$	The $\mathcal{H}_\infty$ norm
$\ \cdot\ _H$	The Hankel norm

Abbreviations

SISO	Single input single output system
MIMO	Multiple input multiple output system
VLSI	Very large scale integration
CT	Continuous time
LTI	Linear time invariant
OR	Orthogonal residual
MR	Minimal residual
SVD	Singular value decomposition
IRKA	Iterative rational Krylov algorithm
FOM	Full orthogonalization method
CG	Conjugate gradient method
Bi-CG	Bi-conjugate gradient method
QMR	Quasi-minimal residual method
GMRES	Generalized minimal residual method
MINRES	Minimal residual method
LMI	Linear matrix inequality

Chapter 1

Introduction

A mathematical representation of physical/artificial system is called a system model and the act of observing the behavior and/or controlling the response of a system through its model is called numerical simulation. Large scale complex models arise due to the desire of having more details and accurate measurements. Numerical simulations of such complex models are unmanageable in most cases due to computational and storage limitations. Although computational power is rising sharply, there are many applications for which detailed large scale models are still required. Furthermore, numerical errors also increase for complex computations.

The finite element and finite difference methods [40, 72] are two well known modeling techniques that can generate very large scale complex models. The problem of model reduction is to replace such large scale complex models by reduced order approximate models in such a way that the main features of the actual model are retained in the approximation. The reduced model is then used to predict the behavior of the actual system. This makes the task easier, especially if one wants to control the response of the model as well. In case of controlling the system response, the common practice is to interconnect the model with another dynamical system (controller) whose complexity is roughly the same as that of the model. Dealing with large scale models and designing large controllers is unmanageable and computationally expensive. This give further importance to the problem of model reduction. Some examples of dynamical systems that require model reduction to observe/control their behavior include simulation of passive electrical devices such as VLSI circuits, weather prediction systems, air quality and the International Space Station, see [8, 35] and the references therein.

A variety of model reduction techniques are available in the literature, with trade-off

among some useful parameters such as computational cost, accuracy and stability. Some of these techniques guarantee stable approximations and achieve a priori error bound such as Balanced Truncation [71] and Hankel Norm Approximation [41]. However they require the solution of Lyapunov equations which makes them computationally inefficient for large scale systems. Various iterative techniques based on Krylov subspace projection [11] have also been extensively used for model reduction. These techniques are computationally efficient as they involve only matrix-vector multiplications and utilize sparsity which is common in large scale systems. However Krylov based approximations can lose some properties of the actual system such as stability, passivity and minimum phase character.

This chapter formulates the problem of model reduction for linear systems and describes some well-known and well-used techniques for its solution. The outline of the thesis is given at the end which shows the contributions of the thesis related to the problem of model reduction.

1.1 Problem Statement

This thesis addresses the problem of model reduction for continuous time (CT) linear and time invariant (LTI) systems. To formulate the problem, we consider an LTI system that can be represented by the state space model:

$$E\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t), \quad (1.1)$$

where $E, A \in R^{n \times n}$, $B \in R^{n \times n_u}$, $C \in R^{n_y \times n}$ and $D \in R^{n_y \times n_u}$. The model represents a descriptor system if the matrix E is singular. It is assumed that E is the identity matrix (or nonsingular so that it can be transformed to the identity matrix). The length of the state vector $x(t) \in R^n$ is called the order or dimension of the state space model.

The transfer function representation of (1.1) with the assumption that E is the identity matrix is,

$$G(s) = C(sI - A)^{-1}B + D \stackrel{s}{=} \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]. \quad (1.2)$$

Also a shorthand is used, for the above linear system as $G(s) \stackrel{s}{=} (A, B, C, D)$. It is assumed that the system has single input and single output (SISO) i.e., $n_y = n_u = 1$. The problem of model reduction is to obtain an order $m \ll n$, approximate system

$$\dot{x}_m(t) = A_m x_m(t) + B_m u(t), \quad y_m(t) = C_m x_m(t) + Du(t), \quad (1.3)$$

such that for a given input $u(t)$, the output $y_m(t)$ is comparable to $y(t)$. In the frequency domain, the problem is to compute an m^{th} order transfer function,

$$G_m(s) = C_m(sI_m - A_m)B_m + D := \left[\begin{array}{c|c} A_m & B_m \\ \hline C_m & D \end{array} \right], \quad (1.4)$$

that approximate the original transfer function, $G(s)$ in some norm.

Definition 1.1. *The error associated with the approximation $G_m(s)$ is defined as the forward error and is given by*

$$E_m(s) := G(s) - G_m(s). \quad (1.5)$$

In some cases, it is possible to compute a priori forward error bounds for the approximation $G_m(s)$, which is very useful for comparing accuracy. Thus ideally we have the following requirements for the computation of m^{th} order $G_m(s)$:

- Minimization of the forward error in some norm.
- Existence of a priori error bound expression for $G_m(s)$.
- Retention of the properties of $G(s)$ such as stability, passivity, etc in $G_m(s)$.
- Efficient computational procedure for $G_m(s)$.

Computational efficiency varies with the order of $G(s)$ and for model reduction of very large scale systems we usually ignore one or two of the above requirements. Some well known approaches for model reduction are discussed in the following section with reference to the above properties.

1.2 Overview of Model Reduction Techniques

This section briefly reviews model reduction techniques for linear systems. A number of articles are available in the literature on model reduction of linear systems, introducing different approaches for this problem. Most of these methods utilize projections onto lower dimensional subspaces to obtain a reduced order model. Projection arises due to the following considerations:

- State transformation that results in a realization that can be classified according to certain criteria, which is followed by truncation with respect to the selected criteria. These include modal truncation [96] and balanced truncation [71] techniques.
- Interpolating and moment matching [6, 20] approximations using Krylov subspaces [91].
- Minimizing an optimization criterion, in particular, a norm of the forward error, such as the Hankel norm [41] and the $\mathcal{H}_2$ norm [51].

The details are in the books [8, 68, 104] and the review papers [7, 12, 35]. Some well known techniques are described below.

1.2.1 Balanced Truncation

In modeling a linear system, it is possible to have redundant states in the state vector. These redundant states can be removed by a state transformation followed by truncation of the unwanted states. The resulting reduced order model is called a minimal state space model, or a minimal realisation, and the length of the state vector associated with the minimal model is called the McMillan degree of the model. If the order of a model is greater than the McMillan degree, then the state vector contains one or more uncontrollable/unobservable states. Let the controllability and the observability gramians be defined as,

$$\begin{aligned}\mathcal{P} &= \int_0^\infty e^{At} B B^T e^{A^T t} dt, \\ \mathcal{Q} &= \int_0^\infty e^{A^T t} C^T C e^{At} dt.\end{aligned}$$

These gramians can be used to classify a minimal state space model according to the energy associated with each state of the system. The states which require a large amount of energy to control the system and those which produce small amount of energy by observing the output of the system without any excitation, can be removed to obtain a reduced order model for a minimal system [8]. However this requires a state transformation that gives a realization for which the controllability and observability gramians are equal and diagonal. Such a realization is called a balanced realization.

The importance of balanced realization in system theory was first observed by Moore [71]. He proposed that a natural first step of model reduction is to compute a balanced realization which is possible for any minimal and stable realization through a state transformation. The balanced realization has equal and diagonal gramians and the diagonal entries

are the Hankel singular values that are arranged in a sequence of decreasing magnitude. The Balanced truncation method [71] involves the partition of the balanced system into two subsystems followed by truncation of the subsystem with small Hankel singular values. This is equivalent to the removal of the states with high controllability energy and low observation energy.

The controllability and observability gramians are solutions of the Lyapunov equations,

$$A\mathcal{P} + \mathcal{P}A^T + BB^T = 0, \quad (1.6)$$

$$A^T\mathcal{Q} + \mathcal{Q}A + C^TC = 0. \quad (1.7)$$

In case of balanced realization, $\mathcal{P} = \mathcal{Q} = \Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$, where σ_i is the i^{th} Hankel singular value and for $i = 1, \dots, n-1$, we have $\sigma_i \geq \sigma_{i+1}$. Such a nonsingular transformation T , that results in a balanced realization, is possible for any minimal and stable state space model [71]. Let $G(s) = (A, B, C, 0)$ be a stable and minimal system and T be the balancing transformation, then,

$$G(s) \stackrel{s}{=} \left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right] = \left[\begin{array}{c|c} T^{-1}AT & T^{-1}B \\ \hline CT & 0 \end{array} \right] =: \left[\begin{array}{c|c} A_b & B_b \\ \hline C_b & 0 \end{array} \right] \quad (1.8)$$

is the representation of the balanced realization. To compute an m -th order approximation for $G(s)$, let the gramians are partitioned as $\Sigma = \text{diag}(\Sigma_m, \Sigma_{\hat{m}})$ where Σ_m and $\Sigma_{\hat{m}}$ are diagonal matrices with the diagonal elements of $\Sigma_m = \text{diag}(\sigma_1, \dots, \sigma_m)$ and has no singular values common to the diagonal elements of $\Sigma_{\hat{m}} = \text{diag}(\sigma_{m+1}, \dots, \sigma_n)$. The partition of the system compatibly with the partition of the gramians gives,

$$G(s) \stackrel{s}{=} \left[\begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline C_1 & C_2 & 0 \end{array} \right] \quad (1.9)$$

The reduced order model obtained by truncation is then,

$$G_m(s) \stackrel{s}{=} \left[\begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & 0 \end{array} \right]. \quad (1.10)$$

By construction, the reduced order system $G_m(s)$ is guaranteed to be minimal and stable. Also the $\mathcal{H}_\infty$ norm of the forward error $G(s) - G_m(s)$ is bounded by twice the sum of

the truncated Hankel singular values:

$$\|G(s) - G_m(s)\|_{\mathcal{H}_\infty} \leq 2(\sigma_{m+1} + \cdots + \sigma_n). \quad (1.11)$$

Although the balanced truncation technique guarantees stability and has a priori error bound, numerically it is inefficient, especially for large scale systems. The main reason for this is the computational complexity involved in the solution of Lyapunov equations, (1.6) and (1.7). There are various extensions of the Balanced truncation method in the literature such as Stochastic balancing [26], bounded real balancing [73] and frequency weighted balancing [28, 69, 98, 103]. These are reviewed in [50].

1.2.2 Hankel norm Approximation Method

The balanced truncation method gives good approximation but it does not minimize any norm. An alternative is to use the Hankel norm approximation method [1, 2, 41] which gives an optimal approximation in terms of the induced norm of Hankel operator [77]. The Hankel norm approximation method computes a reduced order stable model $G_m(s)$ for a given stable system $G(s)$ such that, if σ_i represents the i -th Hankel singular value of $G(s)$,

$$\|G - G_m\|_H \geq \sigma_{m+1}, \quad (1.12)$$

The approximation $G_m(s)$ is the optimal or suboptimal approximation of $G(s)$. In case of optimal approximation, $\|G - G_m\|_H = \sigma_{m+1}$ since σ_{m+1} is the best achievable lower bound of Hankel norm approximation error (1.12).

The solution of the problem of Hankel norm approximation requires the construction of an all pass system, which involves the solution of large Lyapunov equations. This makes the process of computing the optimal approximations very expensive, especially for large scale systems.

The two main advantages of this method are (i) the retention of stability and (ii) the existence of forward error bound in terms of the Hankel norm (1.12). The forward error bound in terms of the $\mathcal{H}_\infty$ norm is due to [41] and is given by,

$$\|G - G_m\|_{\mathcal{H}_\infty} = \sigma_{m+1} + \cdots + \sigma_n. \quad (1.13)$$

The advantages described above are similar to the balanced truncation method and it is possible that the approximation obtained through the balanced truncation method is better in terms of the $\mathcal{H}_\infty$ norm of the forward error as compared to the one obtained

from the Hankel norm approximation method. However in terms of Hankel norm, the approximation obtained through the Hankel norm approximation method is optimal.

1.2.3 $\mathcal{H}_2$ Optimal Model Reduction

Before introducing the problem of $\mathcal{H}_2$ norm optimal model reduction, we first define the $\mathcal{H}_2$ norm and give its expressions. The $\mathcal{H}_2$ norm is associated with the $\mathcal{H}_2$ (Hilbert) space that includes the set of all functions $f(s)$ that are analytic in the open right half plane and at each $\text{Re}(s) > 0$ the function is square integrable on the imaginary axis [51]. The transfer functions of all stable and strictly proper SISO finite dimensional systems belongs to the $\mathcal{H}_2$ space. If $G_1(s)$ and $G_2(s)$ are stable and strictly proper transfer functions, then the $\mathcal{H}_2$ inner product is defined as,

$$\langle G_1(s), G_2(s) \rangle := \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{trace}[G_1^*(-jw)G_2(jw)]dw, \quad (1.14)$$

and the $\mathcal{H}_2$ norm is defined as,

$$\|G_1(s)\|_{\mathcal{H}_2}^2 := \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{trace}[G_1^*(jw)G_1(jw)]dw, \quad (1.15)$$

where $G_1^*(jw)$ is a shorthand for $G_1^T(-jw)$. The $\mathcal{H}_2$ inner product (norm) can also be represented in terms of the solution of matrix equations as shown in the following lemma.

Lemma 1.2.1. [51] *Let $G_1(s) = C(sI - A)^{-1}B$ and $G_2(s) = \hat{C}(sI - \hat{A})^{-1}\hat{B}$ be stable and strictly proper transfer functions in which $A \in \mathcal{R}^{n \times n}, B, C^T \in \mathcal{R}^n$ and $\hat{A} \in \mathcal{R}^{m \times m}, \hat{B}, \hat{C}^T \in \mathcal{R}^m$. Then the inner product $\langle G_1, G_2 \rangle$ is related to the solution of the following Sylvester matrix equations.*

- If $\tilde{P}$ solves $A\tilde{P} + \tilde{P}\hat{A}^T + B\hat{B}^T = 0$, then $\langle G_1, G_2 \rangle = C\tilde{P}\hat{C}^T$.
- If $\tilde{Q}$ solves $A^T\tilde{Q} + \tilde{Q}\hat{A} + C^T\hat{C} = 0$, then $\langle G_1, G_2 \rangle = B^T\tilde{Q}\hat{B}$.
- If $\tilde{R}$ solves $A\tilde{R} + \tilde{R}\hat{A} + B\hat{C} = 0$, then $\langle G_1, G_2 \rangle = C\tilde{R}\hat{B}$.

If $G_1(s)$ and $G_2(s)$ have the same realization, that is $G_1(s) = G_2(s) = C(sI - A)^{-1}B = G(s)$, then $\tilde{P} = P$, $\tilde{Q} = Q$ and $\tilde{R} = R$ are the controllability, observability and cross

grammians of $G(s)$, respectively and,

$$\|G\|_{\mathcal{H}_2}^2 = CPC^T = B^TQB = CRB. \quad (1.16)$$

□

Another expression for $\|G\|_{\mathcal{H}_2}$ has been developed [8], that is based on the poles and residues of the transfer function $G(s)$. That is if $G(s)$ has simple poles at $\lambda_1, \dots, \lambda_n$, then

$$\|G\|_{\mathcal{H}_2}^2 = \sum_{k=1}^n \text{res}[G(s), \lambda_k] G(-\lambda_k), \quad (1.17)$$

where $\text{res}[G(s), \lambda_k]$ represent the residue of $G(s)$ corresponding to the simple pole λ_k and is given as, $\text{res}[G(s), \lambda_k] = \lim_{s \rightarrow \lambda_k} (s - \lambda_k)G(s)$.

The optimal $\mathcal{H}_2$ model reduction problem is to identify a stable reduced order system $G_m(s)$, which is the best approximation of $G(s)$ in terms of the $\mathcal{H}_2$ -norm. I.e.,

$$G_m(s) = \arg \min_{\substack{\hat{G}_m(s) \text{ is stable} \\ \dim(\hat{G}_m(s))=m}} \|G - \hat{G}_m\|_{\mathcal{H}_2} \quad (1.18)$$

where $\dim(\cdot)$ denotes the order of a transfer function. This problem has received considerable attention in the literature [14, 37, 51, 56, 92, 101, 102]. The set of stable m -th order dynamical systems is not convex and therefore the problem in (1.18) has multiple minimizers. An iterative algorithm may not converge to the global minimizer that can solve (1.18) but may result in a local minimizer.

Definition 1.2. [51] *A stable reduced order approximation $G_m(s)$, is a local minimizer for (1.18), if, for all $\epsilon > 0$,*

$$\|G - G_m\|_{\mathcal{H}_2} \leq \|G - \tilde{G}_m^{(\epsilon)}\|_{\mathcal{H}_2} \quad (1.19)$$

for all stable $\tilde{G}_m^{(\epsilon)}(s)$ of order m , such that $\|G_m - \tilde{G}_m^{(\epsilon)}\|_{\mathcal{H}_2} \leq C\epsilon$ for some constant C .

Note that a global minimizer is also a local minimizer but the converse is not true. Most approaches for solving (1.18), derive first order necessary conditions for local optimality. Some of these conditions have been described in Chapter 2.

1.2.4 Moment Matching Methods

The model reduction of linear systems using moment matching involves the retention of some moments, coefficients of a series expansion, of the transfer function $G(s)$ in the reduced order system $G_m(s)$.

Definition 1.3. Let $G(s) = C(sI - A)^{-1}B$ be a transfer function of a system that is differentiable with respect to a given frequency s_0 such that $s_0 \notin \lambda(A)$. Then the i -th moment ($i \geq 1$) of $G(s)$ at s_0 is defined as,

$$\mu_i^{s_0} := (i!)^{-1}G^{(i)}(s_0), \quad (1.20)$$

where $G^{(i)}(s_0)$ is the i -th derivative of $G(s)$ at s_0 and $i!$ represent the i -th factorial. If $s_0 = \infty$, the i -th moment is given as

$$\mu_i^\infty = C(A)^{i-1}B. \quad (1.21)$$

In the following, a Taylor series expansion is used to describe moment matching in four different cases. These include the series expansion of $G(s)$ and $G_m(s)$ at (i) $s = \infty$, (ii) $s = 0$, (iii) $s = s_0$, $s_0 \in \mathcal{C}$ and (iv) $s = \{s_1, \dots, s_K\}$ in which $s_i \in \mathcal{C}$ for $i = 1, \dots, K$.

Let the transfer function $G(s) \stackrel{s}{=} (A, B, C, 0)$ be analytic around infinity, that is the Taylor series expansion,

$$G(s) = \mu_1^\infty s^{-1} + \mu_2^\infty s^{-2} + \dots \quad (1.22)$$

converges where μ_i^∞ represent the i -th moment evaluated at infinity that are called the Markov parameters. Let the reduced order transfer function $G_m(s)$ have a convergent series expansion,

$$G_m(s) = C_m(sI_m - A_m)^{-1}B_m = \sum_{i=1}^{\infty} \hat{\mu}_i^\infty s^{-i}. \quad (1.23)$$

and it matches the first $2m$ Markov parameters of $G(s)$, that is $\hat{\mu}_i^\infty = \mu_i^\infty$ for $i = 1, \dots, 2m$ and $m < n$. Then the reduced order model is called a partial realization [43] of $G(s)$. The partial realization gives better approximation at $t = 0$ ($s = \infty$) as compared to approximation at some later time.

An alternative is to use a series expansion at $s = 0$,

$$G(s) = \sum_{i=0}^{\infty} \mu_i^0 s^i, \quad (1.24)$$

where μ_i^0 is the i -th moment evaluated at $s = 0$. These moments are given by $\mu_i^0 = C(A)^{-(i+1)}B$ for $i \geq 0$. A reduced order model is called Padé approximation [13] if its series expansion at $s = 0$ has the first $2m$ moments that satisfy,

$$\hat{\mu}_i^0 := C_m(A_m)^{-(i+1)}B_m = \mu_i^0,$$

for $i = 1, \dots, 2m$.

The series expansion of $G(s)$ at a shifted frequency s_0 ,

$$G(s) = \sum_{i=0}^{\infty} \mu_i^{s_0} (s - s_0)^i, \quad (1.25)$$

involves shifted moments given by $\mu_i^{s_0} = C(s_0 I - A)^{-(i+1)}B$ for $i \geq 0$. The problem is to compute a reduced order model $G_m(s)$ that retains the first $2m$ shifted moments or in other words, interpolates $G(s)$ and its $2m-1$ derivatives at a user defined interpolating frequency s_0 .

An interesting extension to this problem is to compute a reduced order model that interpolates $G(s)$ and its derivatives at a predefined set of interpolating frequencies $S = \{s_1, \dots, s_k\}$. Let the multiplicity of the interpolating frequencies be respectively given by $\{m_{s_1}, \dots, m_{s_k}\}$. Then the problem is to compute a reduced order model that matches the first $2m_{s_1}$ moments at s_1 , the first $2m_{s_2}$ moments at s_2 and so on, where $m = m_{s_1} + \dots + m_{s_k}$. The resulting approximation is called multipoint Padé or rational interpolating approximation [82, 83].

The implementation techniques for the problems described above include (i) explicit moment matching techniques [79, 89] and (ii) implicit moment matching techniques [35, 45, 57, 59] based on Krylov subspaces [91].

Definition 1.4. For a given matrix $A \in \mathcal{R}^{n \times n}$ and vector $B \in \mathcal{R}^n$, an m -dimensional Krylov subspace is defined as

$$\mathcal{K}_m(A, B) = \text{colsp}([B \ AB, \ \dots, \ A^{m-1}B]), \ m < n. \quad (1.26)$$

Explicit moment matching involves the computation of the first $2m$ moments for the actual system as well as the approximate system and the moments associated with the approximate system are forced to be equal to the ones computed for the actual system. The computation of moments is ill-conditioned and computationally expensive. In contrast, Krylov based techniques achieve moment matching without explicit computation of moments. Krylov based projection techniques, in particular the Arnoldi [9] and the Lanczos [65, 66] algorithms, are reviewed in Chapter 2.

1.3 Thesis Outline

- *Krylov subspace techniques for model reduction:* Chapter 2 presents a literature review on Krylov subspace projection techniques for model reduction. The concept of projection is first described for the order reduction of linear dynamical systems and then iterative algorithms that compute orthogonal bases to the Krylov subspaces such as the Arnoldi and Lanczos algorithms are given. The main focus is the rational Arnoldi algorithm which compute orthogonal basis to unions of Krylov subspaces.
- *Iterative Krylov methods for $\mathcal{H}_2$ optimal model reduction:* Chapter 3 utilizes a link between rational interpolation and the $\mathcal{H}_2$ optimal model reduction problem and develops an iterative implementation based on the link to compute a local minimizer to the $\mathcal{H}_2$ optimal model reduction problem. The proposed algorithm is compared with the iterative rational Krylov algorithm (IRKA) [51] and it is observed that the proposed algorithm improves the computational cost of the IRKA method.
- *All optimal local minimizers to $\mathcal{H}_2$ optimal model reduction:* Chapter 4 develops the theory to compute all first and second order local minimizers to the $\mathcal{H}_2$ optimal model reduction problem for single input and single output systems. Useful expressions for the forward error in terms of the $\mathcal{H}_2$ norm are derived that can be used to identify the global minimizer to the $\mathcal{H}_2$ optimal model reduction problem.
- *Parameterization of rational Krylov approximations and its applications:* Chapter 5 extends the application of the recently introduced parameterization of rational interpolating approximation for the retention of stability, to preserve other properties such as passivity, minimum phase and bounded real character as well as structure of the actual system where required. The idea is also extended to the minimal residual method and the two-sided Arnoldi method in order to improve the accuracy of the approximation.

- *Krylov Based Restart Scheme for the Solution of Linear Matrix Equations:* Chapter 6 utilizes the parameterization of rational interpolating approximations to propose a restart scheme for the low rank Krylov based approximate solution of Sylvester and Lyapunov equations. It is proved that in case of Lyapunov equations the approximate solution through the restart scheme approaches monotonically to the actual solution of the Lyapunov equation.
- *Conclusions and future work:* Chapter 7 gives a summary of the work presented in this thesis. Future directions for the research work are also considered.

Chapter 2

Krylov Subspace Techniques for Model Reduction

The balanced truncation and optimal Hankel norm approximation methods discussed in Section 1.2 preserve stability of the actual system and have a priori error bound. However they are computationally expensive for large scale systems. Also the techniques do not utilize sparsity or any structure of large scale systems. This leads to the iterative Krylov based projection algorithms which involve only matrix-vector multiplications. This chapter gives a review of Krylov based algorithms for model reduction of linear systems. The idea of Krylov projection is first described in the context of model reduction and then two important Krylov based algorithms, the Arnoldi [9] and the Lanczos algorithms [65, 66] are presented, that compute orthogonal and bi-orthogonal bases, respectively, to the Krylov subspace defined in (1.26). These algorithms have extensions to the rational case [38, 82, 83], that compute bases to unions of Krylov subspaces and the resulting approximations achieve moment matching at a predefined set of frequencies known as interpolation points. The Arnoldi and Lanczos algorithms can be found in most textbooks on advanced linear algebra but are discussed here to get insight to rational Arnoldi/Lanczos algorithm, which is the focus of this thesis.

An important issue with the rational Krylov method is the selection of interpolation points. Various approaches have been developed in the literature [33, 47, 48] suggesting different choices of interpolation points. Recently an iterative rational Krylov algorithm (IRKA) has been proposed [51] that iteratively converges to a set of interpolating points for which the reduced order model is optimal in terms of $\mathcal{H}_2$ norm among all local models. This links Krylov projection to the $\mathcal{H}_2$ optimal model reduction problem. We described

this link and some first order necessary conditions for the problem of $\mathcal{H}_2$ optimal model reduction.

In addition to the effect of interpolation points on the accuracy of the approximation, the choice of interpolation points can also result in a reduced order approximation that loses useful properties of the actual system such as stability, minimum phase and bounded real character, passivity as well as structure of the system. In [30] a parameterization of rational Krylov approximations is proposed where the degrees of freedom are used to preserve stability in the approximate model. This idea is reviewed in this chapter and utilized later so that in addition to stability other properties can also be preserved.

Finally we describe the use of rational Krylov methods for the low rank approximate solution of linear matrix equations, in particular the Sylvester and the Lyapunov equations which are important for model reduction.

2.1 Krylov Projection

Projection techniques are often used to solve various matrix problems. This section in particular shows its use for model reduction. The idea of projection techniques is to project a problem onto a lower dimensional subspace and extract an approximate solution for the problem, from the lower dimensional subspace.

Definition 2.1. *A projection operator, P is a linear transformation mapping from $\mathcal{C}^n$ to itself, with P being idempotent, i.e., $P^2 = P$.*

Suppose that a vector $x \in \mathcal{C}^n$ is projected onto an m -dimensional subspace $\mathcal{V}_m = \text{Ran}(P)$ and let $\tilde{x} \in \mathcal{C}^n$ be the projected vector. Then

$$\tilde{x} = Px \quad \text{and} \quad e(x) = (I - P)x,$$

where $e(x) \in \text{Ran}(I - P)$ and is the error associated with the projection. It is useful to associate the error with an m -dimensional subspace $\mathcal{W}_m$ which is orthogonal to $\text{Ran}(I - P)$. That is $e(x) \perp \mathcal{W}_m$. A projection is called an orthogonal projection if $\mathcal{W}_m$ is equal to $\mathcal{V}_m$ otherwise it is called an oblique projection. Figure 2.1 illustrates the two types of projection.

To represent the projection process in matrix form, let the columns of $V_m \in \mathcal{R}^{n \times m}$ form a basis for $\mathcal{V}_m$. Then there exists a vector $x_m \in \mathcal{R}^m$ such that $\tilde{x} = V_m x_m$. Also let the

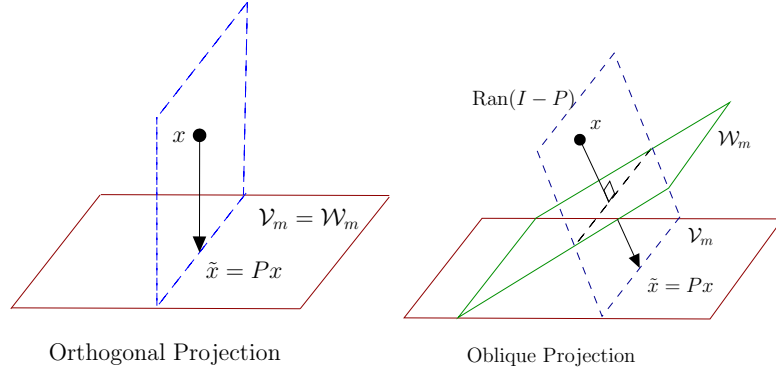


Figure 2.1: Orthogonal and oblique projections

columns of $W_m \in \mathcal{R}^{n \times m}$ form a basis for $\mathcal{W}_m$. Then $W_m^T e(x) = 0$. Thus we have

$$\tilde{x} = V_m x_m \in \mathcal{V}_m \quad (2.1)$$

$$W_m^T (x - V_m x_m) = 0. \quad (2.2)$$

Assuming that $W_m^T V_m$ is nonsingular the above condition implies that $x_m = (W_m^T V_m)^{-1} W_m^T x$ and hence $P = V_m (W_m^T V_m)^{-1} W_m^T$.

A projection technique for linear systems extracts an approximate solution to,

$$Ax = B, \quad (2.3)$$

where $A \in \mathcal{R}^{n \times n}$ and $x, B \in \mathcal{R}^n$. The problem is to compute an m -dimensional approximate solution $\tilde{x} \in \mathcal{V}_m$ ($m \ll n$) to (2.3) such that either $B - A\tilde{x}$ is minimized or $B - A\tilde{x} \perp \mathcal{W}_m$. In case of Krylov projection techniques the subspace $\mathcal{V}_m$ is a Krylov subspace or union of Krylov subspaces. The subspace $\mathcal{W}_m$ may or may not be a Krylov subspace.

Definition 2.2. Let $\tilde{x} \in \mathcal{V}_m$ be an approximation to x , then the residual error associated with $\tilde{x}$ is defined as,

$$r(\tilde{x}) = B - A\tilde{x}. \quad (2.4)$$

Thus we have two common approaches for the computation of $\tilde{x}$, an orthogonal residual approach and a minimal residual approach [91].

Orthogonal residual approach: In this approach the approximation $\tilde{x} = V_m x_m \in \mathcal{V}_m$

is such that it satisfies some orthogonality constraints on the residual error,

$$r(x) \perp W_m \Leftrightarrow W_m^T(B - A\tilde{x}) = 0 \quad (2.5)$$

The above condition is called the Petrov-Galerkin condition. Assuming that $W_m^T A V_m$ is nonsingular this condition implies that $x_m = (W_m^T A V_m)^{-1} W_m^T B$ and hence,

$$\tilde{x} = V_m (W_m^T A V_m)^{-1} W_m^T B. \quad (2.6)$$

When $W_m = V_m$, the Petrov-Galerkin condition reduces to the Galerkin condition, $V_m^T(B - A V_m x_m) = 0$.

The full orthogonalization (FOM) method [84], the conjugate gradient (CG) method [55], the quasi-minimal residual (QMR) method [34] and the bi-conjugate gradient (Bi-CG) method [29] are some well known techniques for solving linear systems using similar conditions on the residual error.

Minimal residual approach: In this approach the approximation $\tilde{x} = V_m x_m$ is selected such that the norm of the associated residual error is minimized, i.e.,

$$x_m = \arg \min_{\tilde{x}_m} \|B - A V_m \tilde{x}_m\|. \quad (2.7)$$

The minimal residual (MINRES) method [74] and the generalized minimal residual (GMRES) method [87] are some of the techniques based on the minimal residual approach.

For each of the above two approaches, we require V_m and W_m , whose columns form bases for the subspaces $\mathcal{V}_m$ and $\mathcal{W}_m$, respectively. These bases are computed by iterative Krylov projection algorithms. Among these algorithms, the Arnoldi and the Lanczos algorithms have received considerable attention in the literature [19, 44, 86, 95]. In standard form, they have the following properties.

1. The Arnoldi algorithm computes an orthogonal V_m for any A and has $A_m = W_m^T A V_m$ in upper Hessenberg form.
2. The Lanczos algorithm deals with symmetric matrices $A = A^H$, in which case A_m has a symmetric tridiagonal structure. An extension to the non-symmetric matrices is the bi-Lanczos algorithm [66] which require $A \in \mathcal{R}^{n \times n}$ and two vectors $B, C^T \in \mathcal{R}^n$ to compute a tridiagonal A_m .

In addition to the solution of linear systems, Krylov projection algorithms can also be

used to approximate some of the eigenvalues of A [85] and most importantly for the moment matching of linear dynamical systems in model reduction.

2.1.1 Approximation of linear dynamical systems

This subsection describe the use of Krylov subspace algorithms for model reduction of linear systems. Let

$$G(s) = C(sI - A)^{-1}B = CF(s), \quad (2.8)$$

where $F(s)$ is a solution to the linear system of equations $(sI - A)F(s) = B$. The Krylov projection technique gives an m -dimensional approximate solution to this linear system and thus an approximation to the transfer function $G(s)$. This approximation is called a single-sided projection. Note that (2.8) can also be written as,

$$G(s) = C(sI - A)^{-1}B = H(s)^T(sI - A)F(s), \quad (2.9)$$

where $H(s)$ is solution to the linear system, $(sI - A)^T H(s) = C^T$. Now we have two linear systems of equations and the system $G(s)$ can be approximated by using Krylov projection technique for each linear system. The resulting approximation is called a two-sided projection. The two-sided projection problem gives better approximation (high number of moments are matched) at the expense of increased computational cost. The approximations obtained through the two problems have the same structure but different properties as given below.

- **Single-sided projection:** Let $\tilde{F}(s)$ be an approximate solution to $F(s)$ obtained via Krylov projection onto the subspace $\mathcal{V}_m$ such that the residual error is orthogonal to $\mathcal{W}_m$ so that

$$\tilde{F}(s) = V_m F_m(s) \in \mathcal{V}_m, \quad r_m(s) \perp \mathcal{W}_m \Leftrightarrow W_m^T(B - (sI - A)V_m F_m(s)) = 0, \quad (2.10)$$

where

$$r_m(s) = B - (sI - A)V_m F_m(s) \quad (2.11)$$

is the residual error associated with $\tilde{F}(s)$. The Petrov-Galerkin conditions in (2.10) imply that, $F_m(s) = (sE_m - A_m)^{-1}B_m$, where $E_m = W_m^T V_m$, $A_m = W_m^T A V_m$ and $B_m = W_m^T B$. Replacing $F(s)$ by the m -dimensional approximation $\tilde{F}(s)$ in (2.8),

we get a single-sided reduced order transfer function $G_m(s)$ given by,

$$G_m(s) = C\tilde{F}(s) = CV_m(sE_m - A_m)^{-1}B_m \quad (2.12)$$

$$\stackrel{s}{=} \left[\begin{array}{c|c} E_m^{-1}A_m & E_m^{-1}B_m \\ \hline C_m & 0 \end{array} \right] \quad (2.13)$$

where $C_m = CV_m$ and E_m is assumed to be invertible.

- **Two-sided projection:** Let $\tilde{H}(s)$ be an approximate solution to $H(s)$ obtained via Krylov projection onto the subspace $\mathcal{W}_m$ such that the residual error is orthogonal to $\mathcal{V}_m$ so that

$$\tilde{F}(s) = V_m F_m(s) \in \mathcal{V}_m, \quad r_m(s) \perp \mathcal{W}_m \quad (2.14)$$

$$\tilde{H}(s) = W_m H_m(s) \in \mathcal{W}_m, \quad s_m(s) \perp \mathcal{V}_m \quad (2.15)$$

where $r_m(s)$ is as defined in (2.11) and

$$s_m(s) = C^T - (sI - A)^T W_m H_m(s) \quad (2.16)$$

is the residual errors associated with $\tilde{H}(s)$. The Petrov-Galerkin conditions in (2.14) and (2.15) imply that,

$$F_m(s) = (sE_m - A_m)^{-1}B_m, \quad (2.17)$$

$$H_m(s) = (sE_m^T - A_m^T)^{-1}C_m^T. \quad (2.18)$$

Using the m -dimensional approximations to $F(s)$ and $H(s)$ in (2.9), we get a reduced order transfer function $G_m(s)$ given by,

$$G_m(s) = \tilde{H}^T(sI - A)\tilde{F} = C_m(sE_m - A_m)^{-1}B_m \quad (2.19)$$

$$\stackrel{s}{=} \left[\begin{array}{c|c} E_m^{-1}A_m & E_m^{-1}B_m \\ \hline C_m & 0 \end{array} \right] \quad (2.20)$$

The above equations show that for a given transfer function $G(s)$, the model reduction problem requires the computation of V_m and W_m , with the properties discussed above. These matrices can be computed by Krylov based projection algorithms such as the Arnoldi and the Lanczos algorithms.

2.1.2 Forward error expressions

An important objective of model reduction techniques is to minimize the forward error in some norm. In this subsection we give some expressions of the forward error in terms of the residual errors. The forward error $E_m(s) := G(s) - G_m(s)$ can be expressed as,

$$\begin{aligned} E_m(s) &= C(sI - A)^{-1}(B - (sI - A)V_m F_m(s)), \\ &= C(sI - A)^{-1}r_m(s). \end{aligned} \tag{2.21}$$

Similarly we may write,

$$\begin{aligned} E_m(s) &= (C - H_m^T(s)W_m^T(sI - A))(sI - A)^{-1}B, \\ &= s_m^T(s)(sI - A)^{-1}B \end{aligned} \tag{2.22}$$

The forward error can also be written as,

$$\begin{aligned} E_m(s) &= C(sI - A)^{-1}B - CV_m F_m - H_m^T W_m^T B + H_m^T W_m^T (sI - A)V_m F_m, \\ &= C(sI - A)^{-1}r_m(s) - H_m^T W_m^T r_m(s), \\ &= s_m^T(s)(sI - A)^{-1}r_m(s), \end{aligned} \tag{2.23}$$

Later, we will derive various expressions for the residual errors, which give further representations of the forward error.

2.2 Arnoldi Method

Arnoldi's technique [9] for model reduction is based on orthogonal projection onto the Krylov subspace, $\mathcal{K}_m(A, B)$ defined in (1.26). The method involves the computation of an orthogonal basis to $\mathcal{K}_m(A, B)$ since, because of numerical ill-conditioning, the basis vectors in $\mathcal{K}_m(A, B)$ become dependant as m increases. This section describes the computation of single-sided and two-sided Arnoldi approximations and presents the moment matching properties of the approximations. The residual error and the forward error expressions are also given.

2.2.1 Single-sided Arnoldi algorithm

In standard form, the single-sided Arnoldi algorithm iteratively computes $V_m \in \mathcal{R}^{n \times m}$ for a given $A \in \mathcal{R}^{n \times n}$ and $B \in \mathcal{R}^n$, such that the following hold:

- Orthonormal matrix: $V_m^T V_m = I_m$
- Basis to Krylov subspace:

$$\text{colsp}(V_m) = \mathcal{K}_m(A, B) = \text{colsp}([B \ AB, \ \dots, \ A^{m-1}B]),$$

where $\text{colsp}(\cdot)$ represent the column span.

In the following, a variant of the standard Arnoldi algorithm is given (Algorithm 2.1) and it is shown that the algorithm computes $V_m \in \mathcal{R}^{n \times m}$ such that the above properties are satisfied. The main step of the Arnoldi algorithm is the Gram-Schmidt orthogonalization

Algorithm 2.1 Single-sided Arnoldi Algorithm

```

1: Inputs:  $A, B, m$  and  $tol$ 
2: Initialize:  $V_1 = [B]$ ;
3: set  $v_1 = B/\|B\|$ ;
4: for  $i = 1$  to  $m$  do
5:    $V_i = [V_i \ v_i]$ ;
6:    $\hat{v}_{i+1} = Av_i - V_i V_i^T Av_i$ ;
7:   if  $\|\hat{v}_{i+1}\| < tol$  then stop
8:    $v_{i+1} = \hat{v}_{i+1}/\|\hat{v}_{i+1}\|$ ;
9: end for
10: Outputs:  $V_m; v_{m+1}$ ;
```

process defined on line 6 of the Algorithm 2.1. The Gram-Schmidt process is often replaced by the modified Gram-Schmidt approach which is more stable [42]. For details on the orthogonalization techniques, see [42]. Note that V_m by construction is orthonormal and its columns form a basis to the Krylov subspace because each column vector of V_m can be written as a polynomial in A of degree $(m-1)$ times B [86].

It can be shown that the so-called Arnoldi equations [58, 60]

$$AV_m = V_m A_{Vm} + v_{m+1} C_{Vm}, \quad B = V_m B_{Vm}, \quad (2.24)$$

where $A_{Vm} = V_m^T A V_m$, $B_{Vm} = V_m^T B$, $C_{Vm} = v_{m+1}^T A V_m$ and $v_{m+1} \in \mathcal{R}^n$ is part of the orthonormal matrix $V_{m+1} = [V_m \ v_{m+1}]$, are satisfied. To demonstrate that the Arnoldi

equations are satisfied, note that lines 6 and 8 of Algorithm 2.1 imply,

$$Av_i = V_i V_i^T Av_i + \|\hat{v}_{i+1}\| v_{i+1} = [V_i \ v_{i+1}] \begin{bmatrix} V_i^T Av_i \\ \|\hat{v}_{i+1}\| \end{bmatrix}$$

where $V_i = [v_1 \ v_2 \cdots v_i]$. For $i = 1, \dots, m$, we may write,

$$A[v_1 \ v_2 \cdots v_m] = [v_1 \ v_2 \cdots v_m \ v_{m+1}] \begin{bmatrix} v_1^T Av_1 & v_1^T Av_2 & \cdots & v_1^T Av_m \\ \|\hat{v}_2\| & v_2^T Av_2 & \cdots & v_2^T Av_m \\ & \ddots & \ddots & \vdots \\ & & \|\hat{v}_m\| & v_m^T Av_m \\ & & & \|\hat{v}_{m+1}\| \end{bmatrix} \quad (2.25)$$

$$AV_m = V_m A_{V_m} + v_{m+1} C_{V_m} \quad (2.26)$$

where A_{V_m} is an upper Hessenberg matrix and $C_{V_m} = \|\hat{v}_{m+1}\| e_m^T$ in which e_m is the last column of the m -th order identity matrix. Line 3 of the above algorithm implies that $B = v_1 \|B\|$ which in terms of V_m can be written as $B = V_m \|B\| e_1 = V_m B_{V_m}$.

Note that the Arnoldi algorithm may breakdown if $\|\hat{v}_{j+1}\|$ defined on line 7 of the Algorithm 2.1 is less than a predefined tolerance or $\|\hat{v}_{j+1}\| \approx 0$. The algorithm stops because the vector v_{j+1} cannot be computed. If this is the case, then the subspace $\mathcal{V}_j$ is invariant under A and a projection method onto $\mathcal{V}_j$ is exact [86].

2.2.2 Two-sided Arnoldi Algorithm

The accuracy of the Arnoldi approximation can be improved by using the two-sided Arnoldi method [24, 59]. The two-sided Arnoldi method iteratively computes orthogonal basis V_m , $W_m \in \mathcal{R}^{n \times m}$ for $\mathcal{K}_m(A, B)$ and $\mathcal{K}_m(A^T, C^T)$, respectively:

$$\text{colsp}(V_m) = \text{colsp}([B \ AB, \ \cdots, \ A^{m-1}B]) \quad V_m^T V_m = I_m, \quad (2.27)$$

$$\text{colsp}(W_m) = \text{colsp}([C^T \ A^T C^T, \ \cdots, \ (A^T)^{m-1} C^T]) \quad W_m^T W_m = I_m. \quad (2.28)$$

Thus in the two-sided Arnoldi algorithm, W_m is not arbitrary. A version of the two-sided Arnoldi algorithm is given in Algorithm 2.2. By construction, it can be shown that the Arnoldi equations:

$$AV_m = V_m A_{V_m} + v_{m+1} C_{V_m}, \quad B = V_m B_{V_m}, \quad (2.29)$$

$$A^T W_m = W_m A_{W_m}^T + w_{m+1} B_{W_m}^T, \quad C^T = W_m C_{W_m}^T, \quad (2.30)$$

Algorithm 2.2 Two-sided Arnoldi Algorithm

```

1: Inputs:  $A, B, C^T, m$  and  $tol$ 
2: Initialize:  $V_1 = [], W_1 = []$ 
3: set  $v_1 = B/\|B\|, w_1 = C^T/\|C^T\|$ 
4: for  $i = 1$  to  $m$  do
5:    $V_i = [V_i \ v_i], W_i = [W_i \ w_i]$ 
6:    $\hat{v}_{i+1} = Av_i - V_i V_i^T Av_i$ 
7:    $\hat{w}_{i+1} = A^T w_i - W_i W_i^T A^T w_i$ 
8:   if  $\|\hat{v}_{i+1}\| < tol$  or  $\|\hat{w}_{i+1}\| < tol$  then stop
9:    $v_{i+1} = \hat{v}_{i+1}/\|\hat{v}_{i+1}\|$ 
10:   $w_{i+1} = \hat{w}_{i+1}/\|\hat{w}_{i+1}\|$ 
11: end for
12: Outputs:  $V_m; v_{m+1}; W_m; w_{m+1};$ 

```

where $AV_m = V_m^T AV_m, A_{Wm} = W_m^T AW_m, B_{Vm} = V_m^T B, C_{Vm} = v_{m+1}^T AV_m, C_{Wm} = C W_m, B_{Wm} = W_m^T A w_{m+1}$ in which $w_{m+1} \in R^n$ is part of the orthonormal matrix $W_{m+1} = [W_m \ w_{m+1}]$, are satisfied [59]. The above equations can be transformed to represent the oblique projection process. Multiplying W_m^T and V_m^T from the left of (2.29) and (2.30), respectively and expressing (A_{Vm}, B_{Vm}) and (A_{Wm}, C_{Wm}) in terms of (A_m, B_m) and (A_m, C_m) , respectively and using these in (2.29) and (2.30), we have:

$$AV_m = V_m E_m^{-1} A_m + (I_n - P_m) v_{m+1} C_{Vm}, \quad B = V_m E_m^{-1} B_m, \quad (2.31)$$

$$A^T W_m = W_m E_m^{-T} A_m^T + (I_n - P_m)^T w_{m+1} B_{Wm}^T, \quad C^T = W_m E_m^{-T} C_m^T, \quad (2.32)$$

where $P_m = V_m E_m^{-1} W_m^T$. To show that the approximation obtained through the two-sided Arnoldi method is better than the single-sided Arnoldi method, a numerical test example is considered in the following and later the moment matching properties of the two approximations are compared.

Example 2.2.1. *This example is taken from [78]. The system has order 1006 and has state matrices given by,*

$$A = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & A_3 & \\ & & & A_4 \end{bmatrix}, \quad A_1 = \begin{bmatrix} -1 & 100 \\ -100 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -1 & 200 \\ -200 & -1 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} -1 & 400 \\ -400 & -1 \end{bmatrix}, \quad A_4 = \text{diag}(-1, \dots, -1000), \quad B^T = C = \begin{bmatrix} 10(e_6^1)^T & (e_{1000}^1)^T \end{bmatrix}$$

where $e_k^1 \in R^k$ is a vector with each element equal to 1. We used single-sided and two-sided Arnoldi methods to compute two approximate models represented by $G_{m1}(s)$ and $G_{m2}(s)$, respectively each having order 10. The singular value plot of the actual system $G(s)$ and the two approximate models $G_{m1}(s)$ and $G_{m2}(s)$ is shown in figure 2.2. As shown in the plot the two methods provide an accurate approximation at infinity but crude at low frequencies. The two-sided approximation is more accurate at high frequencies as compared to the single-sided approximation.

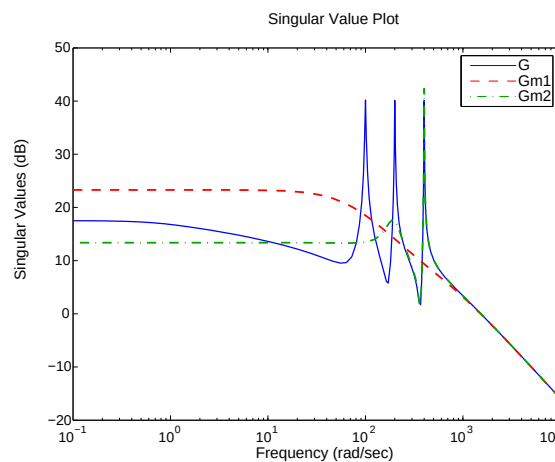


Figure 2.2: Singular value plot for single-sided and two-sided Arnoldi approximations

2.2.3 Moment matching properties

The Arnoldi algorithm in its standard form can solve the partial realization problem, defined in Subsection 1.2.4. The high frequency moments, called Markov parameters, are implicitly matched by the Arnoldi method, without explicit computation of moments. In this subsection, the moment matching properties of both single-sided and two-sided Arnoldi methods are described.

The next lemma shows the moment matching properties of the single-sided Arnoldi method

Lemma 2.2.1. *Let $G_m(s)$ be an m -th order approximation to $G(s)$ satisfying (2.13) and let $V_m \in R^{n \times m}$ be the orthogonal matrix obtained via the single-sided Arnoldi algorithm given in Algorithm 2.1. Then the first m Markov parameters of the two systems are*

matched. i.e.,

$$C_m(E_m^{-1}A_m)^{i-1}E_m^{-1}B_m = CA^{i-1}B, \quad \text{for } i = 1, 2, \dots, m \quad (2.33)$$

Or in other words,

$$G_m^{(i)}(\infty) = G^{(i)}(\infty), \quad i = 0, \dots, m-1, \quad (2.34)$$

where $G^{(i)}(\infty)$ represent the i -th derivative of $G(s)$ with respect to s , evaluated at infinity and $G^{(0)}(\infty) = G(\infty)$.

Proof. The proof can be found in [30, 97]. $\square$

The following lemma describe the moment matching properties of the two-sided Arnoldi method,

Lemma 2.2.2. *Let $V_m, W_m \in \mathbb{R}^{n \times m}$ be the orthogonal matrices obtained via the two-sided Arnoldi algorithm given in 2.2 such that $\mathcal{K}_m(A, B) \subseteq \text{colsp}(V_m)$ and $\mathcal{K}_m(A^T, C^T) \subseteq \text{colsp}(W_m)$. Suppose that $G_m(s)$ be an m -th order two-sided Arnoldi approximation for $G(s)$ satisfying (2.20). Then $G(s)$ and its $2m-1$ derivatives interpolate $G_m(s)$ and its $2m-1$ derivatives at infinity, i.e.,*

$$C_m(E_m^{-1}A_m)^{i-1}E_m^{-1}B_m = CA^{i-1}B, \quad \text{for } i = 1, 2, \dots, 2m. \quad (2.35)$$

Or in other words,

$$G_m^{(i)}(\infty) = G^{(i)}(\infty), \quad i = 0, 1, \dots, 2m-1, \quad (2.36)$$

Proof. The proof can be found in [30, 97]. $\square$

2.2.4 Error Analysis

In this subsection we used the Arnoldi equations (2.24) to express the residual errors in (2.11) and (2.16) in terms of the reduced-dimensional parameters only. This allows the use of the residual errors as stopping criteria for the Arnoldi algorithm.

The residual error associated with the single-sided Arnoldi projection in (2.11) satisfies the Galerkin condition $V_m^T r_m(s) = 0$ which gives $F_m(s) = (sI_m - A_m)^{-1} B_m$. Using the expression of $F_m(s)$ and the Arnoldi equations (2.24) we have,

$$r_m(s) = v_{m+1} C_{V_m} (sI_m - V_m^T A V_m)^{-1} V_m^T B. \quad (2.37)$$

In case of the two-sided Arnoldi approximation, the use of the Arnoldi equations (2.31) implies that the residual errors in (2.11) and (2.16) can be written as

$$r_m(s) = (I_n - P_m) v_{m+1} C_{V_m} (sI_m - E_m^{-1} A_m) E_m^{-1} B_m. \quad (2.38)$$

$$s_m^T(s) = C_m (sI_m - E_m^{-1} A_m)^{-1} B W_m w_{m+1}^T (I_n - P_m), \quad (2.39)$$

respectively

The above residual error expressions are all in terms of reduced-dimensional parameters only. The corresponding forward error expressions can be obtained by using the above equations in the expressions given in Subsection 2.1.2.

2.3 Lanczos Algorithm

The Lanczos method [65,66] is similar to the Arnoldi method in the sense that it involves the computation of orthonormal basis to the Krylov subspace $\mathcal{K}_m(A, B)$ and achieves moment matching without explicit computation of moments. However as mentioned before, the Lanczos algorithm deals with symmetric A and computes a symmetric tridiagonal A_m . The computational and storage cost for the Lanczos algorithm is relatively low as compared to the Arnoldi algorithm as it involves short term recurrences. That is, the Lanczos algorithm require a fixed, short number of vectors in computing the next iterate. The non-symmetric Lanczos algorithm [66] is an extension to an arbitrary A and is based on oblique projection onto Krylov subspaces. The useful property of short term recurrences is preserved by the non-symmetric Lanczos algorithm.

The non-symmetric or bi-Lanczos algorithm iteratively compute two matrices $V_m, W_m \in \mathcal{R}^{n \times m}$ for a given $A \in \mathcal{R}^{n \times n}$ and two vectors $B, C^T \in \mathcal{R}^n$, such that they have the following properties:

- Bi-orthogonality: $W_m^T V_m = I_m$

- Basis to Krylov subspaces:

$$\begin{aligned}\text{colsp}(V_m) &= \mathcal{K}_m(A, B) = \text{colsp}([B \ AB, \ \dots, \ A^{m-1}B]), \\ \text{colsp}(W_m) &= \mathcal{K}_m(A^T, C^T) = \text{colsp}([C^T \ A^T C^T, \ \dots, \ (A^T)^{m-1}C^T])\end{aligned}$$

where $\text{colsp}(\cdot)$ represent the column span.

A version of the non-symmetric Lanczos algorithm is given in Algorithm 2.3. Note that

Algorithm 2.3 The Non-symmetric Lanczos Algorithm

```

1: Inputs:  $A, B, C, m$  and  $tol$ 
2: Initialize:  $V_m = W_m = \emptyset, \beta_1 = \gamma_1 = 0, v_0 = w_0 = 0$ 
3: set  $v_1 = B/\|B\|, w_1 = C^T/\|C^T\|$ 
4: for  $i = 1$  to  $m$  do
5:    $V_m = [V_m \ v_i], W_m = [W_m \ w_i]$ 
6:    $\alpha_i = w_i^T v_i$ 
7:    $\hat{v}_{i+1} = Av_i - \alpha_i v_i - \beta_i v_{i-1}$ 
8:    $\hat{w}_{i+1} = A^T w_i - \alpha_i w_i - \gamma_i w_{i-1}$ 
9:    $\gamma_{i+1} = \|\hat{v}_{i+1}\|$ 
10:  if  $\gamma_{i+1} < tol$  then stop
11:   $v_{i+1} = \hat{v}_{i+1}/\gamma_{i+1}$ 
12:   $\beta_{i+1} = \hat{w}_{i+1}^T v_{i+1}$ 
13:  if  $\beta_{i+1} < tol$  then stop
14:   $w_{i+1} = \hat{w}_{i+1}/\beta_{i+1}$ 
15: end for
16: Outputs:  $V_m; v_{m+1}; W_m; w_{m+1};$ 

```

the parameters γ_{i+1} and β_{i+1} are any scalars satisfying $\gamma_{i+1}\beta_{i+1} = \hat{w}_{i+1}^T \hat{v}_{i+1}$. The algorithm breakdowns if either $\hat{w}_{i+1}$ or $\hat{v}_{i+1}$ are less than a tolerance value or they are greater than the tolerance but $\hat{w}_{i+1}^T \hat{v}_{i+1} = 0$. The first case is called lucky breakdown which gives orthogonal basis V_m (W_m) that are invariant under A (A^T) and the approximation to the linear system associated with A (A^T) is accurate. However in the second case, the orthogonal basis are not A/A^T -invariant and the Lanczos algorithm comes to a halt without any solution. This problem is considered in [76]. Since our main focus is not on breakdown, we assume no breakdown. For detailed analysis see [86].

In the Algorithm 2.3, the choice of the parameters γ_{i+1} and β_{i+1} result in a nonsymmetric tridiagonal A_m . The tridiagonal structure of A_m follows from line 7 and 11 of the above

algorithm. That is we have $Av_i = \alpha_i v_i + \beta_i v_{i-1} + \gamma_{i+1} v_{i+1}$ and for $i = 1, \dots, m$,

$$AV_m = V_m \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \gamma_2 & \alpha_2 & \beta_3 & & \\ & \cdot & \cdot & \cdot & \\ & & \gamma_{m-1} & \alpha_{m-1} & \beta_m \\ & & & \gamma_m & \alpha_m \end{bmatrix} + v_{m+1} \gamma_{m+1} e_m^T.$$

Thus the Algorithm 2.3 satisfies the Lanczos equations [75]:

$$AV_m = V_m A_m + v_{m+1} C_{V_m}, \quad B = V_m B_m, \quad (2.40)$$

where $C_{V_m} = w_{m+1}^T AV_m$ and $w_{m+1}, v_{m+1} \in R^n$ are the basis vectors computed at the m -th iteration in Algorithm 2.3. Similarly it can be shown that the algorithm satisfies,

$$A^T W_m = W_m A_m^T + w_{m+1} B_{W_m}^T, \quad C^T = W_m C_m^T, \quad (2.41)$$

where $B_{W_m} = W_m^T A v_{m+1}$. The following proposition gives the biorthogonality condition of the algorithm.

Proposition 2.1. *If the algorithm does not break down before m iterations, then the vectors v_i and w_j , $i, j = 1, \dots, m$ are biorthogonal, i.e.,*

$$w_j^T v_i = \delta_{ij}, \quad i, j = 1 \dots, m$$

See [86] for a proof. As $\delta_{ij} = 1$ only if $i = j$, in matrix form we have $W_m^T V_m = I_m$. Pre multiplication of W_m^T on both side of (2.40) implies $A_m = W_m^T AV_m$. The matrix A_m is obtained by the oblique projection of A onto $\mathcal{K}_m(A, B)$ and orthogonally to $\mathcal{K}_m(A^T, C^T)$. Similarly, A_m^T is obtained by the oblique projection of A^T onto $\mathcal{K}_m(A^T, C^T)$ and orthogonally to $\mathcal{K}_m(A, B)$. Thus the algorithm essentially involve two projection problems that can solve two linear systems associated with A and A^T . It can be shown that the approximation obtained through the Lanczos algorithm (Algorithm 2.3) interpolates the actual system and its derivatives at infinity in the same way as in Lemma 2.2.2.

2.4 Rational Krylov Method

The Arnoldi and Lanczos algorithm give good approximations to the actual system only at high frequencies. An extension is the rational Krylov algorithm [30, 38, 45] that is accurate at a predefined set of interpolating frequencies. Both the Arnoldi and Lanczos algorithms can be extended to their rational variant. In this section we will describe the more general rational Arnoldi algorithm for model reduction.

2.4.1 Rational Arnoldi algorithm

The Krylov subspace associated with A , B and a shift frequency σ can be written as,

$$\mathcal{K}_{m_\sigma}((\sigma I - A)^{-1}, (\sigma I - A)^{-1}B) = \text{colsp}[(\sigma I - A)^{-1}B \ \cdots \ (\sigma I - A)^{-m_\sigma}B], \quad (2.42)$$

where m_σ is called the multiplicity of the interpolating frequency σ . The Arnoldi method based on this shifted Krylov subspace gives accurate approximation to the original system at the given shift frequency. The following lemma consider the moment matching properties of the Arnoldi method at a shift frequency or interpolating frequency σ .

Lemma 2.4.1. (*[45].*) *Let $G(s) = C(sI - A)^{-1}B$ and let $V_m, W_m \in \mathcal{R}^{n \times m}$ assuming that $W_m^T V_m$ is non-singular. Define $G_m(s) = C_m(sI - A_m)^{-1}B_m$ where $A_m = (W_m^T V_m)^{-1}W_m^T A V_m$, $B_m = (W_m^T V_m)^{-1}W_m^T B$, and $C_m = C V_m$. Suppose $\sigma \in \mathbb{C}$ is not an eigenvalue of either A or A_m . If,*

- $(\sigma I - A)^{-1}B \in \text{span}(V_m)$ and
- $(\bar{\sigma} I - A_m^T)^{-1}C_m^T \in \text{span}(W_m)$, where $\bar{\sigma}$ denotes conjugation on σ ,

then

$$G_m(\sigma) = G(\sigma) \quad G'_m(\sigma) = G'(\sigma) \quad (2.43)$$

where $G'(\sigma)$ represent the derivative of $G(s)$ with respect to s , evaluated at $s = \sigma$. $\square$

A generalization to this approach is the rational Arnoldi method which involves union of shifted Krylov subspaces with different shift frequencies and multiplicities. The rational Arnoldi algorithm iteratively computes $V_m \in \mathcal{R}^{n \times m}$ for a given $A \in \mathcal{R}^{n \times n}$ and $B \in \mathcal{R}^n$

and a set of K distinct interpolating points, $\mathcal{S} = \{s_1, \dots, s_K\}$ that are closed under conjugation with multiplicities $m_{s_1}, \dots, m_{s_K}$, such that the following holds:

$$\mathcal{V}_U = \bigcup_{k=1}^K \mathcal{K}_{m_{s_k}}((s_k I - A)^{-1}, (s_k I - A)^{-1} B) \bigcup \mathcal{K}_{m_\infty}(A, B) \subseteq \text{colsp}(V_m)$$

where $m = \sum_{k=1}^K \alpha_k m_{s_k}$, and $\alpha_k = 1$ if s_k is real or infinite and $\alpha_k = 2$ otherwise.

In the following, a version of the rational Arnoldi algorithm is given (Algorithm 2.4).

Algorithm 2.4 The Rational Arnoldi Algorithm

```

1: Inputs:  $A \in \mathcal{R}^{n \times n}$ ,  $B, C^T \in \mathcal{R}^n$ ,  $S = \{s_1, \dots, s_K\} \subset \mathcal{C}$  closed under conjugation
   with multiplicities  $\mathcal{M} = \{m_{s_1}, \dots, m_{s_K}\}$  and  $tol > 0$ .
2: Initialize:  $V_m = [ ]$ ;  $j = 0$ ;
3: for  $k = 1$  to  $K$  do
4:   if  $\{s_k = \infty\}$ ;  $\tilde{v}_{j+1} = B$ ;  $\tilde{v}_{m+1} = \tilde{v}_{j+1}$ ; else  $\tilde{v}_{j+1} = (s_k I - A)^{-1} B$  end
5:   if  $\{k > 1\}$ ;  $\hat{v}_{j+1} = \tilde{v}_{j+1} - V_m V_m^T \tilde{v}_{j+1}$  end
6:   if  $\|\hat{v}_{j+1}\| < tol$ , Stop, end
7:    $v_{j+1} = \hat{v}_{j+1} / \|\hat{v}_{j+1}\|$ ;  $V_m = [V_m \ v_{j+1}]$ ;  $j = j + 1$ ;
8:   for  $i = 1$  to  $m_{s_k} - 1$  do
9:     if  $\{s_k = \infty\}$ ;  $\tilde{v}_{j+1} = A v_j$ ;  $\tilde{v}_{m+1} = \tilde{v}_{j+1}$ ; else  $\tilde{v}_{j+1} = (s_k I - A)^{-1} v_j$  end
10:     $\hat{v}_{j+1} = \tilde{v}_{j+1} - V_m V_m^T \tilde{v}_{j+1}$ 
11:    if  $\|\hat{v}_{j+1}\| < tol$ , Stop, end
12:     $v_{j+1} = \hat{v}_{j+1} / \|\hat{v}_{j+1}\|$ ;  $V_m = [V_m \ v_{j+1}]$ ;  $j = j + 1$ ;
13:   end for
14: end for
15:  $\hat{v}_{m+1} = A \tilde{v}_{m+1} - V_m V_m^T A \tilde{v}_{m+1}$ ;
16: if  $\|\hat{v}_{m+1}\| < tol$ , Stop, end
17:  $v_{m+1} = \hat{v}_{m+1} / \|\hat{v}_{m+1}\|$ 
18: Outputs:  $V_m$ ;  $v_{m+1}$ ;

```

2.4.2 Moment Matching Properties

This subsection describe the moment matching properties of the rational Arnoldi method. The next result describes these properties for multiple interpolating frequencies with different multiplicities.

Theorem 2.1. *Let all variables be as defined in Lemma 2.4.1 and let $V_m, W_m \in \mathcal{R}^{n \times m}$*

satisfy,

$$\bigcup_{\substack{k=1 \\ s_k \neq \infty}}^K \mathcal{K}_{m_{s_k}}((s_k I - A)^{-1}, (s_k I - A)^{-1} B) \bigcup \mathcal{K}_{m_\infty}(A, B) \subseteq \text{colsp}(V_m),$$

$$\bigcup_{\substack{\hat{k}=1 \\ \hat{s}_{\hat{k}} \neq \infty}}^{\hat{K}} \mathcal{K}_{\hat{m}_{\hat{s}_{\hat{k}}}}((\bar{\hat{s}}_{\hat{k}} I - A^T)^{-1}, (\bar{\hat{s}}_{\hat{k}} I - A^T)^{-1} C^T) \bigcup \mathcal{K}_{\hat{m}_\infty}(A^T, C^T) \subseteq \text{colsp}(W_m),$$

where $m = \sum_{k=1}^K m_{s_k} = \sum_{\hat{k}=1}^{\hat{K}} \hat{m}_{\hat{s}_{\hat{k}}}$ in which $S = \{s_1, \dots, s_K\} \subset \mathcal{C}$ and $\hat{S} = \{\hat{s}_1, \dots, \hat{s}_{\hat{K}}\} \subset \mathcal{C}$ are set of K and $\hat{K}$ distinct interpolating frequencies that are closed under conjugation with multiplicities, $\mathcal{M} = \{m_{s_1}, \dots, m_{s_K}\}$ and $\hat{\mathcal{M}} = \{\hat{m}_{\hat{s}_1}, \dots, \hat{m}_{\hat{s}_{\hat{K}}}\}$ respectively. Then assuming that $(sI - A)^{-1}$ and $(sW_m^T V_m - W_m^T A V_m)^{-1}$ exist, for all s in S and $\hat{S}$,

1. if $s_k = \hat{s}_{\hat{k}}$ for all k and $\hat{k}$,

$$G_m^{(i)}(s_k) = G^{(i)}(s_k), \quad i = 0, 1, \dots, m_{s_k} + \hat{m}_{\hat{s}_{\hat{k}}} - 1. \quad (2.44)$$

2. if $s_k \neq \hat{s}_{\hat{k}}$ for all k and $\hat{k}$,

$$G_m^{(i)}(s_k) = G^{(i)}(s_k), \quad i = 0, 1, \dots, m_{s_k} - 1, \quad (2.45)$$

$$G_m^{(i)}(\hat{s}_{\hat{k}}) = G^{(i)}(\hat{s}_{\hat{k}}), \quad i = 0, 1, \dots, \hat{m}_{\hat{s}_{\hat{k}}} - 1. \quad (2.46)$$

Proof. The proof can be found in [30, 45]. □

2.4.3 Arnoldi-like equations

The next lemma derives Arnoldi-like equations in the rational case. These equations will be used throughout this thesis especially in Chapter 4 to derive error expressions in the $\mathcal{H}_2$ optimal case and in Chapter 6 to justify the results of the proposed restart scheme.

Lemma 2.4.2. [33] Let $V_m \in \mathcal{R}^{n \times n}$ be as defined in Theorem 2.1 and define v_{m+1} such that

$$\text{colsp}([V_m \quad A^{m_\infty} B]) = \text{colsp}(V_{m+1})$$

where

$$V_{m+1} = \begin{bmatrix} V_m & v_{m+1} \end{bmatrix}$$

Then,

$$AV_m = V_m A_{V_m} + v_{m+1} C_{V_m}, \quad B = V_m B_{V_m} + v_{m+1} D_{V_m}. \quad (2.47)$$

where

$$\left[\begin{array}{c|c} A_{V_m} & B_{V_m} \\ \hline C_{V_m} & D_{V_m} \end{array} \right] = \left[\begin{array}{c|c} V_m^T A_{V_m} & V_m^T B \\ \hline v_{m+1}^T A_{V_m} & v_{m+1}^T B \end{array} \right]. \quad (2.48)$$

Proof. The proof can be found in [30, 33] however to clarify our notations and for completeness, it is given below. From the algorithm, we have

$$\begin{bmatrix} \tilde{v}_1 & \tilde{v}_2 & \cdots & \tilde{v}_{m+1} \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & \cdots & v_{m+1} \end{bmatrix} \begin{bmatrix} \|\hat{v}_1\| & v_1^T \tilde{v}_2 & \cdots & \tilde{v}_1^T \tilde{v}_{m+1} \\ 0 & \|\hat{v}_2\| & \cdots & \tilde{v}_2^T \tilde{v}_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\hat{v}_{m+1}\| \end{bmatrix} \quad (2.49)$$

where we assume no breakdown, that is $\|\hat{v}_i\| \neq 0$ for $i = 1, \dots, m$. Also note that $\tilde{v}_1 = \hat{v}_1$.

Let us represent the i -th vector of the above equation as,

$$\tilde{v}_i = \begin{bmatrix} V_{i-1} & | & v_i \end{bmatrix} \begin{bmatrix} r_{i-1} \\ \rho_i \end{bmatrix} \quad (2.50)$$

We first consider the $i = 1$ case, for which we have two possibilities,

$$\tilde{v}_1 = \begin{cases} B & \text{if } s_1 = \infty \\ (s_1 I - A)^{-1} B & \text{if } s_1 \neq \infty \end{cases} \quad (2.51)$$

Using the expression of $\tilde{v}_1$ from (2.50) into the above equation, we have

$$Av_1 = \frac{1}{\rho_1} AB \quad (2.52)$$

$$Av_1 = s_1 v_1 + \frac{1}{\rho_1} B \quad (2.53)$$

Now $\tilde{v}_{m+1}$ (described on line 22 of the algorithm) also have two possibilities, for $m = 1$. i.e.,

$$\tilde{v}_2 = \begin{cases} AB & \text{if } s_1 = \infty \\ B & \text{if } s_1 \neq \infty \end{cases} \quad (2.54)$$

The above two possibilities along with the expression of $\tilde{v}_2$ from (2.50) can be used to express (2.52) and (2.53) as,

$$Av_1 = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} \frac{r_1}{\rho_1} \\ \frac{\rho_2}{\rho_1} \end{bmatrix} \quad (2.55)$$

$$Av_1 = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} s_1 + \frac{r_1}{\rho_1} \\ \frac{\rho_2}{\rho_1} \end{bmatrix} \quad (2.56)$$

Thus in case of $i = 1$, $Av_i \in \text{colsp}(V_{m+1})$. Now we assume that for $i = 2, \dots, m-1$, Av_i also belongs to the $\text{colsp}(V_{m+1})$. Note that in this case we have four possibilities,

$$\tilde{v}_i = \begin{cases} B & \text{if } s_i = \infty \neq s_{i-1} \\ (s_i I - A)^{-1} B & \text{if } s_i \neq \infty \text{ and } s_i \neq s_{i-1} \\ A\tilde{v}_{i-1} & \text{if } s_i = s_{i-1} = \infty \\ (s_i I - A)^{-1} \tilde{v}_{i-1} & \text{if } s_i = s_{i-1} \neq \infty \end{cases} \quad (2.57)$$

We need to show that our assumption is true for $i = k \leq m$. Using (2.50) into the above equation, we have

$$\begin{aligned} \tilde{v}_k = B &\Rightarrow Av_k = \frac{1}{\rho_k} AB - \frac{1}{\rho_k} AV_{k-1} r_{k-1} \\ \tilde{v}_k = (s_k I - A)^{-1} B &\Rightarrow Av_k = v_k s_k + \frac{1}{\rho_k} B + \frac{s_k}{\rho_k} V_{k-1} r_{k-1} + \frac{1}{\rho_k} AV_{k-1} r_{k-1} \\ \tilde{v}_k = A\tilde{v}_{k-1} &\Rightarrow Av_k = \frac{1}{\rho_k} A^2 \tilde{v}_{k-1} - \frac{1}{\rho_k} AV_{k-1} r_{k-1} \\ \tilde{v}_k = (s_k I - A)^{-1} \tilde{v}_{k-1} &\Rightarrow Av_k = v_k s_k + \frac{1}{\rho_k} \tilde{v}_{k-1} + \frac{s_k}{\rho_k} V_{k-1} r_{k-1} + \frac{1}{\rho_k} AV_{k-1} r_{k-1}. \end{aligned}$$

It follows from our assumption at $k-1$ and the construction of v_{m+1} in the lemma that $Av_k \in \text{colsp}(V_{m+1})$ for $k \leq m$. Thus for $k = 1, \dots, m$, we may write

$$AV_m = V_m AV_m + v_{m+1} C_{V_m} \quad (2.58)$$

for some A_{V_m} and C_{V_m} . Since by construction V_{m+1} is orthogonal, $A_{V_m} = V_m^T A V_m$ and $C_{V_m} = v_{m+1}^T A V_m$. This proves the first equation in (2.47). To show that $B \in \text{colsp}(V_{m+1})$ we have the following two possibilities,

$$B = \begin{cases} \tilde{v}_i & \text{if } \infty \in S \text{ and } i \leq m \\ \tilde{v}_{m+1} & \text{if } \infty \notin S \end{cases} \quad (2.59)$$

In each case, (2.50) shows that $B \in \text{colsp}(V_{m+1})$ and there exists B_{V_m} and D_{V_m} such that $B = V_m B_{V_m} + v_{m+1} D_{V_m}$. $\square$

2.4.4 Residual Error in the Rational Case

In this section, we describe the use of the Arnoldi-like equations to obtain the residual error expression in the rational case, involving only the reduced parameters. The next lemma describe these expressions,

Lemma 2.4.3. [33] *Let all variables be as defined in Lemma 2.4.2 and let $G_m(s) = C V_m F_m(s)$ where $F_m(s) = (sI - \tilde{A})^{-1} \tilde{B}$. Define the residual error associated with $F_m(s)$ as $r_m(s) := B - (sI - A) V_m F_m(s)$. Then*

1. *The residual error of the approximation is given by*

$$r_m(s) = V_{m+1} S(F_m(s)) \quad (2.60)$$

where

$$S(F_m(s)) = \begin{bmatrix} B_m - (sI - A_m) F_m(s) \\ D_{V_m} + C_{V_m} F_m(s) \end{bmatrix} \stackrel{s}{=} \left[\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline A_m - \tilde{A} & B_m - \tilde{B} \\ \hline C_{V_m} & D_{V_m} \end{array} \right]. \quad (2.61)$$

2. *The Galerkin condition $W_m^T r_m(s) = 0$ is satisfied if and only if $F_m(s) = (sI_m - W_m^T A V_m)^{-1} W_m^T B$, provided that $W_m^T V_m = I_m$, in which case*

$$r_m(s) = v_{m+1} (D_{V_m} + C_{V_m} (sI - W_m^T A V_m)^{-1} W_m^T B).$$

Proof. The proof for (1) can be obtained from the Arnoldi equations (2.47) by replacing B and AV_m in the residual error $r_m(s) := B - (sI - A)V_mF_m(s)$. The structure of $S(F_m(s))$ in the second equality of (2.61) follows by using $F_m(s) = (sI - \tilde{A})^{-1}\tilde{B}$ in,

$$\begin{bmatrix} B_m - (sI - A_m)F_m(s) \\ D_{V_m} + C_{V_m}F_m(s) \end{bmatrix} = \begin{bmatrix} B_m - ((sI - \tilde{A}) + \tilde{A} - A_m)F_m(s) \\ D_{V_m} + C_{V_m}F_m(s) \end{bmatrix}. \quad (2.62)$$

(2) follows by using the residual error expression (2.60) in the Galerkin condition and solving for $F_m(s)$. The implication of (2) in other direction follows by using $F_m(s) = (sI - A_m)^{-1}B_m$ in (2.61) and the resulting residual error in the Galerkin condition. $\square$

Example 2.4.1. *To see the effect of interpolation points on the rational interpolating approximation, consider the use of the two-sided rational Arnoldi method to compute two approximate models for the linear system $G(s)$ with state matrices given in Example 2.2.1. Let the reduced order approximations be $G_{m1}(s)$ and $G_{m2}(s)$ each having order 10 with interpolation points $S_1 = \hat{S}_1$ and $S_2 = \hat{S}_2$ respectively. The two sets of interpolation points are randomly selected such that $\text{real}(S_1)$ lies between 1 and 10, while $\text{real}(S_2)$ lies between 1 and 100. The singular value plot of the actual system $G(s)$ and the approximate models is shown in figure 2.3. The plot shows that the quality of the approximation is highly dependent on the choice of interpolation points.*

2.5 Rational Krylov method and $\mathcal{H}_2$ Optimal model reduction

Recently it is shown in [51] that an iterative implementation of the rational Krylov method can be used to compute local minimizers (in the sense of (1.19)) to the problem of $\mathcal{H}_2$ optimal model reduction defined in (1.18). Several researchers have worked on the optimal $\mathcal{H}_2$ model reduction problem. Most approaches for the computation of $\mathcal{H}_2$ optimal reduced order models are based on some first order necessary conditions. The resulting reduced order systems are not necessarily the global minimizers but often they are effective reduced order approximations. Some well known first order necessary conditions for computing the reduced order systems, classified as Lyapunov-based conditions and interpolation-based conditions, are described below.

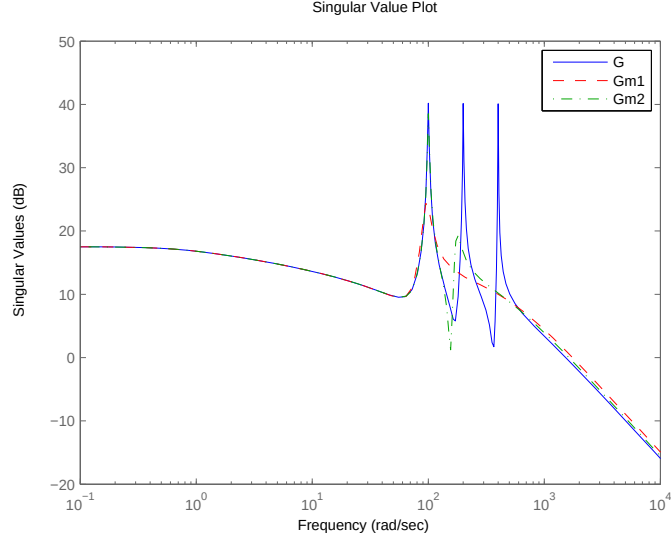


Figure 2.3: Singular value plot for two-sided rational Arnoldi approximations with two different choices of interpolation points

2.5.1 Lyapunov-based Necessary Conditions for Optimality

Let the error system associated with the reduced order model $G_m(s)$ be defined as,

$$E_m(s) := G(s) - G_m(s) \stackrel{s}{=} \left[\begin{array}{c|c} A_E & B_E \\ \hline C_E & D_E \end{array} \right] := \left[\begin{array}{cc|c} A & 0 & B \\ 0 & A_m & B_m \\ \hline C & -C_m & 0 \end{array} \right]. \quad (2.63)$$

Also let

$$P_E = \left[\begin{array}{cc} P & X \\ X^T & P_m \end{array} \right] \quad \text{and} \quad Q_E = \left[\begin{array}{cc} Q & Y \\ Y^T & Q_m \end{array} \right], \quad (2.64)$$

be the controllability and observability grammians, respectively for $E_m(s)$ satisfying,

$$\begin{aligned} A_E P_E + P_E A_E^T + B_E B_E^T &= 0, \\ A_E^T Q_E + Q_E A_E + C_E^T C_E &= 0. \end{aligned}$$

The Lyapunov-based Wilson conditions for $\mathcal{H}_2$ optimality are given by,

$$X^T Y + P_m Q_m = 0, \quad (2.65)$$

$$Y^T B + Q_m B_m = 0, \quad (2.66)$$

$$C_m^T P_m - C^T X = 0. \quad (2.67)$$

Then according to Wilson [101], the reduced order model,

$$G_m(s) \stackrel{s}{=} \left[\begin{array}{c|c} W_m^T A V_m & W_m^T B \\ \hline C V_m & 0 \end{array} \right], \quad (2.68)$$

with

$$V_m = X P_m^{-1}, \quad W_m = -Y Q_m^{-1}, \quad (2.69)$$

satisfies the Lyapunov-based necessary Wilson conditions for $\mathcal{H}_2$ optimality and furthermore $W_m^T V_m = I_m$.

Hyland and Bernstein [56] gave similar necessary conditions for $\mathcal{H}_2$ optimality. They used the oblique projection $\Pi = V_m W_m^T$ to compute a reduced order model as in (2.68). The model satisfies the Lyapunov-based Hyland-Bernstein necessary conditions for $\mathcal{H}_2$ optimality given below, if there exist two nonnegative definite matrices $P, Q \in \mathbb{R}^{n \times n}$ such that $PQ = V_m R W_m^T$, where $R \in \mathbb{R}^{m \times m}$ is similar to a positive definite matrix and V_m, W_m are two $n \times m$ matrices such that $W_m^T V_m = I_m$. The Hyland-Bernstein necessary conditions for $\mathcal{H}_2$ optimality are given by,

$$\text{rank}(P) = \text{rank}(Q) = \text{rank}(PQ), \quad (2.70)$$

$$\Pi[AP + PA^T + BB^T] = 0, \quad (2.71)$$

$$[A^T Q + QA + CC^T]\Pi = 0. \quad (2.72)$$

In both cases [56, 101], the necessary conditions are given in the form of large-scale Lyapunov equations which make their algorithms expensive for computation of the reduced order model [105].

2.5.2 Interpolation based Necessary Optimality Conditions

It is shown in [51] that the $\mathcal{H}_2$ -norm of the error is small if the reduced order model $G_m(s)$ interpolates $G(s)$ at $-\lambda_i(A)$ and $-\lambda_i(A_m)$. Since $-\lambda_i(A_m)$ is not known a priori, several approaches [47, 48] are developed to minimize the $\mathcal{H}_2$ error by choosing interpolation points to be $-\lambda_i(A)$. However, Meier and Luenberger [70] showed that interpolation at $-\lambda_i(A_m)$ is more useful and is a necessary condition for $\mathcal{H}_2$ optimality. The following theorem gives this condition.

Lemma 2.5.1. (*[70].*) *Given a stable SISO system $G(s) = C(sI - A)^{-1}B$, let $G_m(s) = C_m(sI_m - A_m)^{-1}B_m$ be a local minimizer (in the sense defined in (1.19)) of dimension*

m for the optimal $\mathcal{H}_2$ model reduction problem (1.18) and suppose that $G_m(s)$ has simple poles at $-s_i$, $i=1, \dots, m$. Then $G_m(s)$ interpolates both $G(s)$ and its first derivative at s_i , $i=1, \dots, m$:

$$G_m(s_i) = G(s_i), \quad G'_m(s_i) = G'(s_i), \quad \forall s_i \in S_m := \{s_1, \dots, s_m\}. \quad (2.73)$$

□

Recently it has been shown in [51] that the necessary optimality conditions of Hyland-Bernstein [56] and Wilson [101] are equivalent to the Meier-Luenberger conditions [70] in the case of continuous time SISO systems having simple poles. Also some new necessary conditions based on the $\mathcal{H}_2$ inner product have been stated. The equivalence of necessary optimality conditions for continuous time MIMO systems with multiple poles is derived in [99]. Similar results about the equivalence of necessary optimality conditions are formulated [21] for discrete time MIMO systems. Also an iterative algorithm based on rational Krylov method has been proposed [51], which on convergence achieves moment matching and satisfies the interpolation based optimality conditions. The algorithm utilizes rational Krylov steps in constructing $G_m(s)$ and does not require Lyapunov solvers or dense matrix decompositions, which shows its effectiveness, especially for large-scale systems.

2.5.3 Iterative Rational Krylov Algorithm

The necessary conditions given in (2.73) are equivalent to the root finding problem of the following expression.

$$\lambda(A_m(S_m)) = -S_m, \quad (2.74)$$

where $\lambda(\cdot)$ represent the eigenvalues and S_m are the required roots and where we have written $A_m(S_m)$ to emphasize the dependence of A_m on the interpolation points. A Newton framework is used in [51] to compute the interpolation points (roots) that successively update the roots using $S_m^{i+1} = \lambda(-A_m(S_m^i))$. Using Theorem 2.1 and Lemma 2.5.1, the successive updates $S_m^{i+1} = \lambda(-A_m(S_m^i))$ are possible via the rational Krylov method. This leads to the iterative rational Krylov algorithm (IRKA) [51] that is given in Algorithm 2.5.

In Chapter 3 an iterative algorithm is proposed that converges to a local minimizer of the $\mathcal{H}_2$ optimal model reduction problem with improvement in the computational cost. In

Algorithm 2.5 The iterative rational Krylov algorithm

- 1: **Inputs:** A, B, C, m and tol
- 2: **Initialize:** Select $S_m^0 = \{s_1^0, \dots, s_m^0\} \in \mathcal{C}^m$ that are closed under conjugation and set $j = 0$ and $error > tol$.
- 3: **while** $error < tol$ **do**
- 4: Use the rational Arnoldi algorithm (Algorithm 2.4) to obtain V_m and W_m such that

$$\begin{aligned} \text{Ran}(V_m) &= \text{colsp} \left[(s_1^j I - A)^{-1} B, \dots, (s_m^j I - A)^{-1} B \right], \\ \text{Ran}(W_m) &= \text{colsp} \left[(s_1^j I - A^T)^{-1} C^T, \dots, (s_m^j I - A^T)^{-1} C^T \right]. \end{aligned}$$

- 5: Compute $A_m = (W_m^T V_m)^{-1} W_m^T A V_m$
 - 6: $error = \|\text{sort}(\text{eig}(-A_m)) - \text{sort}(S_m^j)\|$
 - 7: Update the interpolation points, $S_m^{j+1} = \text{eig}(-A_m)$; $j = j + 1$;
 - 8: **end while**
 - 9: Compute $A_m = (W_m^T V_m)^{-1} W_m^T A V_m$; $B_m = (W_m^T V_m)^{-1} W_m^T B$; $C_m = C V_m$;
 - 10: **Outputs:** A_m, B_m, C_m .
-

solving a sequence of shifted linear systems, the proposed algorithm involves the matrix-vector multiplications required for solving only a single system.

2.6 Parameterization of rational Arnoldi method

This section describes the idea of introducing a free low dimensional parameter [30] in the reduced order model obtained via the rational Arnoldi method. It is shown that the stability of the actual system can be retained in the approximation by using a proper choice of the free parameter. The following theorem defines such a parameterization of the reduced order model that does not affect the interpolating properties of the rational Krylov method. A special case of this parameterization has also appeared in [10], where it is assumed that the interpolation is at one finite frequency only. The next result gives a parameterization without these assumptions.

Theorem 2.2. [30] *Let $A \in \mathcal{R}^{n \times n}$ and $B, C^T \in \mathcal{R}^n$ be given and define $G(s) := CF(s)$, where $F(s) = (sI - A)^{-1}B$. Let $S = \{s_1, \dots, s_K\} \subset \mathcal{C}$ be a set of distinct K interpolating points with multiplicities $\mathcal{M} = \{m_{s_1}, \dots, m_{s_K}\}$ such that S is closed under conjugation and $S \cap \lambda(A) = \emptyset$. Let*

$$F_m(s) = (sI - W_m^T A V_m)^{-1} W_m^T B, \tag{2.75}$$

be the Galerkin/Arnoldi approximation of $F(s)$ with $W_m^T V_m = I$. Assume that $S \cap$

$\lambda(A_m) = \emptyset$ so that $V_m F_m(s)$ matches the first m_{s_k} moments of $F(s)$ at s_k . Then $G_m^L(s) := CV_m F_m^L(s)$ in which,

$$F_m^L(s) = (sI - A_m^L)^{-1} B_m^L, \quad (2.76)$$

where

$$A_m^L = A_m + LC_{V_m}, \quad B_m^L = B_m + LD_{V_m}, \quad (2.77)$$

and $\{A_m, B_m, C_{V_m}, D_{V_m}\}$ are as defined in (2.48), interpolates $G(s)$, for all $L \in \mathcal{R}^{m \times 1}$ with $S \cap \lambda(A_m^L) = \emptyset$. That is,

$$G^{(i)}(s_k) = G_m^{L(i)}(s_k), \quad i = 0, \dots, m_{s_k} - 1, \quad k = 1, \dots, K \quad (2.78)$$

Proof. These results are proved in [30]. □

The fact that L is free and is of low dimension will prove useful in the following when we consider the choice of L to preserve model properties.

2.6.1 Preserving Stability

It is shown in [30] that the parameterization of interpolating approximation described above can preserve the stability of the actual system and this work is reviewed in this subsection. The next result derives necessary and sufficient conditions for the existence of L such that $F_m^L(s)$ is stable if $F(s)$ is stable and gives the extent to which we can assign the poles of $F_m^L(s)$ through choice of L .

Proposition 2.2. *Let all variables be as defined in Theorem 2.2. Then*

1. *The following are equivalent:*

- (a) *There exists L such that $F_m^L(s)$ is stable.*
- (b) *The pair (A_m, C_{V_m}) is detectable.*

(c) There exist $0 \prec X_m = X_m^T \in \mathcal{R}^{m \times m}$ and $Y_m \in \mathcal{R}^{m \times 1}$ such that the following linear matrix inequality (LMI)

$$A_m^T X_m + X_m A_m + Y_m C_{V_m} + C_{V_m}^T Y_m^T \prec 0 \quad (2.79)$$

is satisfied, in which case $L = X_m^{-1} Y_m$.

2. If $F(s)$ is stable then there exists L such that $F_m^L(s)$ is stable.
3. All eigenvalues of A_m that are not eigenvalues of A are observable through C_{V_m} . In particular, if A is stable then the pair (A_m, C_{V_m}) is detectable.

Proof. Although the proposition was proved in [30], the following is a more concise proof.

1.

(a) $\leftrightarrow$ (b) This is immediate since $A_m^L = A_m + LC_{V_m}$ and, by definition, the pair (A_m, C_{V_m}) is detectable if and only if there exists L such that $A_m + LC_{V_m}$ is stable.

(a) $\leftrightarrow$ (c) $F_m^L(s)$ is stable if and only if there exists $0 \prec X_m = X_m^T \in \mathcal{R}^{m \times m}$ such that

$$(A_m^L)^T X_m + X_m (A_m^L) = (A_m^T + C_{V_m}^T L^T) X_m + X_m (A_m + LC_{V_m}) \prec 0. \quad (2.80)$$

Setting $Y_m = X_m L$ proves the result.

2. If $F(s)$ is stable, there exists $0 \prec X = X^T \in \mathcal{R}^{n \times n}$ such that $A^T X + X A \prec 0$. It follows that

$$V_m^T (A^T X + X A) V_m \prec 0 \quad (2.81)$$

since V_m has full column rank. Substituting the first equation of Arnoldi-like equations (2.47) in (2.81) and rearranging gives (2.80) with $X_m := V_m^T X V_m \in \mathcal{R}^{m \times m} = X_m^T \succ 0$ and $L := X_m^{-1} V_m^T X v_{m+1}$. It follows that $F_m^L(s)$ is stable.

3. Let $\lambda \in \mathcal{C}$ be an unobservable mode of the pair (A_m, C_{V_m}) so that $A_m x_m = \lambda x_m$ and $C_{V_m} x_m = 0$ for some $x_m \neq 0$. Post-multiplying the first equation in the

Arnoldi-like equations (2.47) by x_m gives $Ax = \lambda x$ where $x = V_m x_m \neq 0$. Thus λ is an eigenvalue of A . If A is stable, then all eigenvalues of A_m in the closed right half of the complex plane are observable through C_{V_m} so the pair (A_m, C_{V_m}) is detectable.

□

It follows that Part (3) of Proposition 2.2 gives us the freedom to use L to assign all the eigenvalues of A_m^L which are not eigenvalues of A anywhere in the complex plane, e.g., using the method described in [63]. This is acceptable since any eigenvalues of A_m which are eigenvalues of A do not, in general, need to be changed.

Note however, that for $F_m^L(s)$ to be interpolating, we should ensure that the eigenvalues of A_m^L are not placed at the interpolating points in S .

In Chapter 5, the idea of parameterization is extended to derive conditions on the free parameter such that some other useful properties such as minimum phase and bounded real character, passivity and system structure are preserved in the interpolating approximation. Also it is shown that the parameterization of the interpolating approximations can be used to improve the accuracy of the approximations.

2.7 Rational Krylov methods for the Sylvester and Lyapunov equations

In this section, the use of rational Krylov methods for the low rank approximate solution of the Sylvester and the Lyapunov equation is described. The Sylvester equation can be written as,

$$AP + P\hat{A}^T + B\hat{B}^T = 0, \quad (2.82)$$

where $A \in \mathcal{R}^{n \times n}$, $\hat{A} \in \mathcal{R}^{\hat{n} \times \hat{n}}$, $B \in \mathcal{R}^n$, $\hat{B} \in \mathcal{R}^{\hat{n}}$ and $P \in \mathcal{R}^{n \times \hat{n}}$. It is assumed that A and $\hat{A}$ are large and sparse. Note that the Sylvester equation reduces to the Lyapunov equation if $\hat{A} = A$ and $\hat{B} = B$.

The Sylvester and Lyapunov equations play an important role in control theory, signal processing, model reduction, filtering and image restoration; see, for example, [17, 22, 64] and the references therein. Exact solutions of the Sylvester equation can be computed by direct methods such as the Bartels Stewart method [15] and the Hammerling Method [54].

However these techniques are not suitable for sufficiently large dimensional problems due to their computational complexity. To overcome the computational complexity associated with the direct methods, iterative algorithms, based on Krylov subspace techniques, have been proposed in the literature, e.g., [46, 58, 61, 90].

Krylov subspace projection methods are based on obtaining low-rank approximate solutions to the large-scale Sylvester equation, such that the approximate solution belongs to a subspace of dimension that is much smaller than the dimension of the space associated with the exact solution. This technique, as will be described below, requires the solution of a reduced order Sylvester equation. The reduced order Sylvester equation is obtained by projection using bases of suitable Krylov subspaces [8].

The Sylvester equation (2.82) has a unique solution P if and only if the eigenvalues of A and $\hat{A}$ satisfy $\lambda_i(A) + \lambda_j(\hat{A}) \neq 0 \forall i, j$. The following integral formula has been derived in [81] for the solution of the Sylvester equation (2.82),

$$P = \frac{1}{2\pi i} \int_{\Gamma} F(s) \hat{F}^{\sim}(s) ds, \quad (2.83)$$

where Γ represent a rectifiable contour which includes all $\lambda(A)$ but none of $\lambda(\hat{A})$ and $F(s) = (sI - A)^{-1}B$ while $\hat{F}^{\sim}(s) = \hat{F}(-s)^T$ in which $\hat{F}(-s) = (-sI - \hat{A})^{-1}\hat{B}$ (see Theorem 5.2.2 in [64]). In [58], it is proposed that an approximate solution of the Sylvester equation (2.82) can be computed by using the Galerkin approximation to the functions $F(s)$ and $\hat{F}(s)$, obtained via the standard Arnoldi method (see Section (2.2)). That is,

$$P \approx \mathcal{P} = V_m Y \hat{V}_{\hat{m}}^T, \quad (2.84)$$

where $Y \in \mathcal{R}^{m \times \hat{m}}$ and $V_m \in \mathcal{R}^{n \times m}$ (and $\hat{V}_{\hat{m}} \in \mathcal{R}^{\hat{n} \times \hat{m}}$) form an orthogonal basis to the Krylov subspace $\mathcal{K}_m(A, B) = \text{span}\{B, AB, \dots, A^{m-1}B\}$ ($\mathcal{K}_{\hat{m}}(\hat{A}, \hat{B})$) and is the output of the standard Arnoldi algorithm (see Algorithm 2.1) with inputs $A, B, m (\hat{A}, \hat{B}, \hat{m})$. The residual error associated with the above approximation can be written as,

$$\mathcal{R} := AP + \mathcal{P}\hat{A}^T + B\hat{B}^T = AV_m Y \hat{V}_{\hat{m}}^T + V_m Y \hat{V}_{\hat{m}}^T \hat{A}^T + B\hat{B}^T. \quad (2.85)$$

It is found in [58] that the residual error associated with $\mathcal{P}$ satisfies the Galerkin-like condition,

$$V_m^T \mathcal{R} \hat{V}_{\hat{m}} = 0, \quad (2.86)$$

if Y satisfies the following reduced order Sylvester equation,

$$A_m Y + Y \hat{A}_{\hat{m}}^T + B_m \hat{B}_{\hat{m}}^T = 0 \quad (2.87)$$

where A_m, B_m (and similarly $\hat{A}_{\hat{m}}, \hat{B}_{\hat{m}}$) are as defined by the Arnoldi equations (2.47).

In [39], the rational Arnoldi method is utilized to obtain an approximate solution of the Sylvester equation, however the computational and storage issues are not considered. Using the rational Arnoldi method an approximate solution of the Sylvester equation of the form (2.84) can be written as,

$$P \approx \tilde{P} = V_m X \hat{V}_m^T, \quad (2.88)$$

where $X \in \mathcal{R}^{m \times \hat{m}}$, $V_m \in \mathcal{R}^{n \times m}$ and $\hat{V}_{\hat{m}} \in \mathcal{R}^{\hat{n} \times \hat{m}}$. The matrices V_m and $\hat{V}_{\hat{m}}$ are the orthogonal matrices computed by the rational Arnoldi algorithm (see Algorithm 2.4) with inputs $A, B, m, S = \{s_1, \dots, s_m\}$ and $\hat{A}, \hat{B}, \hat{m}, \hat{S} = \{\hat{s}_1, \dots, \hat{s}_{\hat{m}}\}$ respectively. The interpolation points S and $\hat{S}$ are selected such that they are subsets of $-\lambda(A)$ and $-\lambda(\hat{A})$, respectively. Note that the computation of a subset of eigenvalues of A can be efficiently carried out by using Krylov subspace techniques [67, 85]. The residual error associated with $\tilde{P}$ can be expressed in the following form, by using the Arnoldi-like equations (2.47),

$$\tilde{R} = V_{m+1} \tilde{S} \hat{V}_{\hat{m}+1}^T, \quad (2.89)$$

where $V_{m+1} = [V_m \ v_{m+1}]$, $\hat{V}_{\hat{m}+1} = [\hat{V}_{\hat{m}} \ \hat{v}_{\hat{m}+1}]$ in which v_{m+1} and $\hat{v}_{\hat{m}+1}$ are the outputs of the rational Arnoldi algorithm (Algorithm 2.4) and

$$\tilde{S} = \begin{bmatrix} A_m X + X(\hat{A}_{\hat{m}})^T + B_m(\hat{B}_{\hat{m}})^T & X \hat{C}_{\hat{V}_{\hat{m}}}^T + B_m \hat{D}_{\hat{V}_{\hat{m}}}^T \\ C_{V_m} X + D_{V_m}(\hat{B}_{\hat{m}})^T & D_{V_m} \hat{D}_{\hat{V}_{\hat{m}}}^T \end{bmatrix}, \quad (2.90)$$

where $A_m, B_m, C_{V_m}, D_{V_m}$ (and similarly $\hat{A}_{\hat{m}}, \hat{B}_{\hat{m}}, \hat{C}_{\hat{V}_{\hat{m}}}, \hat{D}_{\hat{V}_{\hat{m}}}$) are defined by the Arnoldi-like equations (2.47) for given A, B, V_m and v_{m+1} ($\hat{A}, \hat{B}, \hat{V}_{\hat{m}}$ and $\hat{v}_{\hat{m}+1}$). It is easy to see that the residual error (2.89) satisfies the Galerkin-like condition (2.86) if the following holds,

$$A_m X + X(\hat{A}_{\hat{m}})^T + B_m(\hat{B}_{\hat{m}})^T = 0.$$

The size of the residual error, measured using a suitable norm, is usually compared with a predefined tolerance to obtain the desired size of the projected reduced Sylvester equation. If the size of the residual error is greater than the tolerance level, the values of m and $\hat{m}$ are increased. The increase in the size of the projected Sylvester equation increases the computational burden associated with ensuring that the new columns of the matrices V_m and $\hat{V}_{\hat{m}}$ are orthogonal to all the previous columns. Also the storage requirements for storing the columns of the matrices V_m and $\hat{V}_{\hat{m}}$ become excessive. These issues are addressed in Chapter 6, where an algorithm is proposed which restarts without requiring that the new columns are orthogonal to the previous columns. Instead, a set

of two new orthogonal matrices is constructed.

2.8 Conclusions

In this chapter, we have described Krylov projection techniques for the model reduction of large-scale linear dynamical systems. In particular, the Arnoldi/Lanczos algorithm and its extension, the rational Arnoldi/Lanczos algorithm have been discussed. In each case, it is shown that the approximation achieves moment matching without explicit computation of moments. Various expressions are presented for the residual error and the forward error that will later be used for further analysis. It is shown that an iterative implementation of the rational Arnoldi method can compute local minimizers to the $\mathcal{H}_2$ optimal model reduction. In the next chapter we use a similar iterative method that computes local minimizers to the $\mathcal{H}_2$ optimal model reduction problem with improved computational cost. Later we compute all local minimizers to the $\mathcal{H}_2$ optimal model reduction problem in special cases. Also in this chapter we described the idea of introducing a free low dimensional parameter that can preserve stability of the actual system in the approximate system. This idea is extended in Chapter 5 for the retention of some other properties such as passivity, minimum phase and bounded real characters as well as structure of the system. Finally this chapter presents the use of rational Arnoldi method to compute low rank approximate solution to the Sylvester and Lyapunov equations. In Chapter 6 a restart algorithm is proposed that updates the approximation solution of the Sylvester and Lyapunov equation with improved computational and storage requirements.

Chapter 3

Iterative Krylov Methods for $\mathcal{H}_2$ Optimal Model Reduction

The iterative rational Krylov algorithm (IRKA) described in Section 2.5 links Krylov methods with the $\mathcal{H}_2$ optimal model reduction problem. This chapter suggests an alternative implementation to the IRKA method by using a variant of bi-conjugate gradient (Bi-CG) method [29]. In IRKA (Algorithm 2.5), in every iteration j , the two systems $(s_i^j I - A)x = B$ and $(\bar{s}_i^j I - A^T)\tilde{x} = C^T$ ($\bar{s}$ represents conjugate of s) have to be solved for every interpolating frequency $s_i^j \in S_m^j = \{s_1^j, \dots, s_m^j\}$. The Bi-CG method can also solve such shifted linear systems simultaneously and with the nice property that it is based on short recurrences, meaning that the storage requirements are modest and fixed, and also that the number of arithmetic operations is constant in every iteration. This is in contrast to the methods based on Arnoldi algorithms such as FOM or GMRES, for which the storage requirements and computational complexity increase with the number of iterations.

Both IRKA and the proposed iterative algorithm converge to the same local minimizer of the $\mathcal{H}_2$ optimal model reduction problem. However, the proposed algorithm is based on the observation [36, 62] that Krylov subspaces are shift invariant, meaning that

$$\mathcal{K}_m(A, B) = \mathcal{K}_m((sI - A), B).$$

The proposed algorithm utilizes a variant of the Bi-CG algorithm for shifted linear systems that can exploit this property by constructing approximate solutions for the sequence of shifted systems by utilizing only the basis vectors for $\mathcal{K}_m(A, B)$ and $\mathcal{K}_m(A^T, C^T)$.

This means that algorithms which require the solution of many sequences of shifted linear systems (such as IRKA) can be efficiently solved by using the matrix-vector multiplications associated with only one system. Thus the proposed algorithm improves the computational cost of the IRKA method.

The construction of a new basis vector requires a matrix-vector multiplication with A for $\mathcal{K}_m(A, B)$ and with A^T for $\mathcal{K}_m(A^T, C^T)$, therefore by using this vector for m shifted systems we save $m - 1$ matrix-vector product with A and A^T in every iteration, compared to solving these systems one-by-one with the standard Bi-CG algorithm (or IRKA). It is shown that all the basis vectors computed in each iteration that are utilized by the m shifted systems can be computed in advance. This is useful because if we have basis vectors for solving $Ax = B$ and $A^T\tilde{x} = C^T$, we do not need to compute them again for solving a sequence of m shifted systems.

3.1 Shifted variants of the Bi-CG algorithm

This section describes a variant of the standard Bi-CG algorithm that allows us to simultaneously solve a sequence of shifted systems [36, 62] using only the basis vectors for $\mathcal{K}_m(A, B)$ and $\mathcal{K}_m(A^T, C^T)$. We begin with a brief description of the standard Bi-CG method.

The Bi-CG method is based on the non-symmetric Lanczos algorithm that can simultaneously solve a linear system $Ax = B$ and a transposed system $A^T\tilde{x} = C^T$. The technique computes two approximate solutions, x_m and $\tilde{x}_m$ for a given $A \in \mathcal{R}^{n \times n}$ and two vectors $B, C^T \in \mathcal{R}^n$ such that the following holds:

- A -orthogonal search directions:

Let x_0 and $\tilde{x}_0$ are given, then for $i = 1, 2, \dots$, until convergence

$$x_i = x_{i-1} + \alpha_i d_i \quad \text{and} \quad \tilde{x}_i = \tilde{x}_{i-1} + \bar{\alpha}_i \tilde{d}_i \quad (3.1)$$

where $\alpha_i \in \mathcal{C}$, $\bar{\alpha}_i$ is the complex conjugate and where $d_i, \tilde{d}_j \in \mathcal{C}^n$ are A -orthogonal, that is,

$$\tilde{d}_i^T A d_j = 0 \quad \text{for } i \neq j \quad (3.2)$$

- Orthogonal residual errors:

Let $\tilde{r}_i = C^T - A^T \tilde{x}_i$ and $r_j = B - Ax_j$ are the residual errors for $i, j = 0, 1, 2, \dots$, until convergence, then

$$\tilde{r}_i^T r_j = 0 \quad \text{for } i \neq j \quad (3.3)$$

A version of the Bi-CG algorithm is given in Algorithm 3.1.

Algorithm 3.1 The bi-conjugate gradient algorithm

- 1: **Inputs:** $A, B, C, x_0, \tilde{x}_0, \text{max_it}$ and tol
 - 2: **Initialize:** $r_0 = B - Ax_0 = d_0$, $\tilde{r}_0 = C^T - A^T \tilde{x}_0 = \tilde{d}_0$, and $\text{error} = \max\{\|r_0\|/\|B\|, \|\tilde{r}_0\|/\|C\|\}$
 - 3: **for** $i = 0$ to max_it **do**
 - 4: $\alpha_i = (\tilde{r}_i^T r_i) / (\tilde{d}_i^T A d_i)$
 - 5: $x_{i+1} = x_i + \alpha_i d_i$, $\tilde{x}_{i+1} = \tilde{x}_i + \bar{\alpha}_i \tilde{d}_i$
 - 6: $r_{i+1} = r_i - \alpha_i A d_i$, $\tilde{r}_{i+1} = \tilde{r}_i - \bar{\alpha}_i A^T \tilde{d}_i$.
 - 7: $\beta_{i+1} = (\tilde{r}_{i+1}^T r_{i+1}) / (\tilde{r}_i^T r_i)$.
 - 8: $d_{i+1} = r_{i+1} - \beta_{i+1} d_i$, $\tilde{d}_{i+1} = \tilde{r}_{i+1} - \bar{\beta}_{i+1} \tilde{d}_i$.
 - 9: $\text{error} = \max\{\|r_{i+1}\|/\|B\|, \|\tilde{r}_{i+1}\|/\|C\|\}$
 - 10: **if** $\text{error} < \text{tol}$ **then** break **end**
 - 11: **end for**
 - 12: **Outputs:** $x_{i+1}, \tilde{x}_{i+1}$.
-

In order to solve a sequence of shifted linear systems by using basis vectors associated with A and A^T , we consider a linear system and a shifted linear system that are given by,

$$Ax = B \quad (sI - A)x^s = B, \quad (3.4)$$

where $A \in \mathcal{R}^{n \times n}$, and $x, x^s, B \in \mathcal{R}^n$ while $s \in \mathcal{C}$. Let $x_i \in \mathcal{K}_i(A, B)$ and $\tilde{x}_i^s \in \mathcal{K}_i((sI - A), B)$ be the approximations of the above linear systems obtained by using i iterations of the Bi-CG algorithm (Algorithm 3.1). Then the associated residual errors $r_i = B - Ax_i$ and $r_i^s = B - (sI - A)\tilde{x}_i^s$ belong to $\mathcal{K}_{i+1}(A, B)$ and $\mathcal{K}_{i+1}((sI - A), B)$, respectively. That is, if we express,

$$x_i = p_{i-1}(A)B \quad \text{and} \quad \tilde{x}_i^s = p_{i-1}^s(sI - A)B, \quad (3.5)$$

where p_{i-1} (and p_{i-1}^s) are polynomials of degree less than or equal to $i-1$, then the residual errors are,

$$r_i = q_i(A)B \quad \text{and} \quad r_i^s = q_i^s(sI - A)B, \quad (3.6)$$

where $q_i(t) = 1 - tp_{i-1}(t)$ (and q_i^s) are polynomials of degree i . Note that $q_i(0) = q_i^s(0) = 1$. Since Krylov subspaces are shift invariant, $\mathcal{K}_i(A, B) = \mathcal{K}_i((sI - A), B)$ we must have collinear residuals [62], $r_i = \zeta_i^s r_i^s$ in which $\zeta_i^s \in \mathcal{C}$. This relation is due to the residual errors satisfying $r_i^s \perp \mathcal{K}_i((\bar{s}I - A^T), C^T)$ and $r_i^s \perp \mathcal{K}_i(A^T, C^T)$ together with the shift invariant property of Krylov subspaces, $\mathcal{K}_i(A^T, C^T) = \mathcal{K}_i((\bar{s}I - A^T), C^T)$. Thus

$$r_i^s = (1/\zeta_i^s)r_i, \quad (3.7)$$

which implies,

$$r_i^s = q_i^s(sI - A)B = (1/\zeta_i^s)q_i(A)B.$$

Since $q_i^s(s - t) = (1/\zeta_i^s)q_i(t)$ and $q_i^s(0) = 1$,

$$\zeta_i^s = q_i(s). \quad (3.8)$$

The above expression and (3.7) show that the residual error associated with the shifted linear system in (3.4) can be expressed in terms of the residual error associated with the linear system $Ax = B$. In the following, we give some expressions of the parameters associated with the shifted Bi-CG method.

From the Bi-CG algorithm (Algorithm 3.1), we have for $i = 1, 2, \dots$

$$d_{i-1} = \frac{1}{\beta_i}(r_i - d_i), \quad Ad_i = \frac{1}{\alpha_i}(r_i - r_{i+1}),$$

and therefore

$$\begin{aligned} r_i = r_{i-1} - \alpha_{i-1}Ad_{i-1} &= r_{i-1} - \alpha_{i-1}A\left(\frac{1}{\beta_i}(r_i - d_i)\right), \\ &= r_{i-1} - \frac{\alpha_{i-1}}{\beta_i}Ar_i + \frac{\alpha_{i-1}}{\beta_i\alpha_i}(r_i - r_{i+1}), \end{aligned}$$

which results in the following three-term recurrence expression for the residual error,

$$r_{i+1} = -\alpha_i Ar_i + \frac{\beta_i \alpha_i}{\alpha_{i-1}} r_{i-1} + \left(1 - \frac{\beta_i \alpha_i}{\alpha_{i-1}}\right) r_i. \quad (3.9)$$

In terms of the polynomial representation used in (3.6), the above expression can be written as,

$$q_{i+1}(t) = -\alpha_i t q_i(t) + \frac{\beta_i \alpha_i}{\alpha_{i-1}} q_{i-1}(t) + \left(1 - \frac{\beta_i \alpha_i}{\alpha_{i-1}}\right) q_i(t).$$

Using $t = s$ and (3.8), we have

$$\zeta_{i+1}^s = \left(1 - \alpha_i s - \frac{\beta_i \alpha_i}{\alpha_{i-1}}\right) \zeta_i^s + \frac{\beta_i \alpha_i}{\alpha_{i-1}} \zeta_{i-1}^s. \quad (3.10)$$

Since we have collinear residuals (3.7), the residual error for the shifted system is,

$$\begin{aligned} r_{i+1}^s &= \frac{1}{\zeta_{i+1}^s} \left(-\alpha_i \zeta_i^s A r_i^s + \frac{\beta_i \alpha_i \zeta_{i-1}^s}{\alpha_{i-1}} r_{i-1}^s + \left(\zeta_i^s - \frac{\zeta_i^s \beta_i \alpha_i}{\alpha_{i-1}} \right) r_i^s \right), \\ &= \frac{\alpha_i \zeta_i^s}{\zeta_{i+1}^s} (sI - A) r_i^s + \frac{\beta_i \alpha_i \zeta_{i-1}^s}{\alpha_{i-1} \zeta_{i+1}^s} r_{i-1}^s + \left(\frac{\zeta_i^s}{\zeta_{i+1}^s} - \frac{\zeta_i^s \beta_i \alpha_i}{\alpha_{i-1} \zeta_{i+1}^s} - s \frac{\alpha_i \zeta_i^s}{\zeta_{i+1}^s} \right) r_i^s. \end{aligned}$$

Now comparing the above equation with the three term recurrence relation (similar to (3.9)) for the shifted system given by,

$$r_{i+1}^s = -\alpha_i^s (sI - A) r_i^s + \frac{\beta_i^s \alpha_i^s}{\alpha_{i-1}^s} r_{i-1}^s + \left(1 - \frac{\beta_i^s \alpha_i^s}{\alpha_{i-1}^s} \right) r_i^s, \quad (3.11)$$

we have,

$$\alpha_i^s = -\alpha_i \left(\frac{\zeta_i^s}{\zeta_{i+1}^s} \right), \quad (3.12)$$

$$\beta_i^s = \left(\frac{\alpha_i}{\alpha_i^s} \right) \left(\frac{\alpha_{i-1}^s}{\alpha_{i-1}} \right) \frac{\zeta_{i-1}^s}{\zeta_{i+1}^s} \beta_i = \left(\frac{\zeta_{i-1}^s}{\zeta_i^s} \right)^2 \beta_i, \quad (3.13)$$

$$\zeta_{i+1}^s = \left(1 - \alpha_i s - \frac{\beta_i \alpha_i}{\alpha_{i-1}} \right) \zeta_i^s + \frac{\beta_i \alpha_i}{\alpha_{i-1}} \zeta_{i-1}^s. \quad (3.14)$$

For more details on the derivation of the above expressions, see [36]. It is easy to see that the above expressions are also true for $m = 0$, if we initialize $\zeta_{-1}^s = 1$. A shifted variant of the Bi-CG method is given in Algorithm 3.2.

Note that in finding an update for the approximate solution of shifted systems, Algorithm 3.2 uses $x_{i+1}^s = x_i^s + \alpha_i^s d_i^s$ where $d_i^s = r_i^s - \beta_i^s d_{i-1}^s$. Since r_i^s can be expressed in terms of r_i , no matrix-vector multiplications are required for the shifted systems, if the residual error r_i is known. In the following we use this useful result to propose iterative algorithms for the $\mathcal{H}_2$ optimal model reduction problem.

3.2 Iterative shifted Bi-CG for $\mathcal{H}_2$ optimal model reduction

In this section, an iterative implementation of the shifted bi-conjugate gradient (Bi-CGS) method has been developed to compute local minimizers to the $\mathcal{H}_2$ optimal model reduction problem. The proposed method is compared with the IRKA method described in Section 2.5 and it is shown that the algorithm has better computational cost as compared to IRKA.

Algorithm 3.2 The shifted bi-conjugate gradient algorithm

```
1: Inputs:  $A, B, C, S_m = \{s_1, \dots, s_m\} \subset \mathcal{C}^m$ ,  $max\_it$  and  $tol$ 
2: Initialize:  $x_0 = \tilde{x}_0 = \text{zeros}(n, 1)$ ,  $r_0 = B = d_0$ ,  $\tilde{r}_0 = C^T = \tilde{d}_0$ ,  $\zeta_o^s = \zeta^s = \text{ones}(m, 1)$ ,  $\zeta_n^s = \alpha^s = \beta^s = \text{zeros}(m, 1)$ ,  $x^s = \tilde{x}^s = d^s = \tilde{d}^s = \text{zeros}(n, m)$ 
3:  $error = \max\{\|r_0\|/\|B\|, \|\tilde{r}_0\|/\|C\|\}$ 
4: for  $i = 0$  to  $max\_it$  do
5:   {Linear system with no shift}
6:    $\alpha_i = (\tilde{r}_i^T r_i) / (\tilde{d}_i^T A d_i)$ 
7:    $x_{i+1} = x_i + \alpha_i d_i$ ,  $\tilde{x}_{i+1} = \tilde{x}_i + \bar{\alpha}_i \tilde{d}_i$ 
8:    $r_{i+1} = r_i - \alpha_i A d_i$ ,  $\tilde{r}_{i+1} = \tilde{r}_i - \bar{\alpha}_i A^T \tilde{d}_i$ 
9:    $\beta_{i+1} = (\tilde{r}_{i+1}^T r_{i+1}) / (\tilde{r}_i^T r_i)$ 
10:   $d_{i+1} = r_{i+1} - \beta_{i+1} d_i$ ,  $\tilde{d}_{i+1} = \tilde{r}_{i+1} - \bar{\beta}_{i+1} \tilde{d}_i$ 
11:   $error = \max\{\|r_{i+1}\|/\|B\|, \|\tilde{r}_{i+1}\|/\|C\|\}$ 
12:  if  $error < tol$  then break end
13:  {Shifted system}
14:  for  $k = 1, \dots, m$  do
15:     $\zeta_n^s(k) = (1 - \alpha_i s_k) \zeta^s(k) + \left(\frac{\alpha_i \beta_i}{\alpha_{i-1}}\right) (\zeta_o^s(k) - \zeta^s(k))$ 
16:     $\alpha^s(k) = -\alpha_i \left(\frac{\zeta^s(k)}{\zeta_n^s(k)}\right)$ 
17:     $\beta^s(k) = \left(\frac{\zeta_o^s(k)}{\zeta_n^s(k)}\right)^2 \beta_{i+1}$ 
18:    for  $j = 1, \dots, n$  do
19:       $d^s(j, k) = (1/\zeta^s(k)) r_{i+1} - \beta^s(k) d^s(j, k)$ ,
20:       $\tilde{d}^s(j, k) = (1/\tilde{\zeta}^s(k)) \tilde{r}_{i+1} - \bar{\beta}^s(k) \tilde{d}^s(j, k)$ 
21:       $x^s(j, k) = x^s(j, k) + \alpha^s(k) d^s(j, k)$ ,
22:       $\tilde{x}^s(j, k) = \tilde{x}^s(j, k) + \bar{\alpha}^s(k) \tilde{d}^s(j, k)$ 
23:    end for
24:  end for
25: end for
26: Outputs:  $x_m^s, \tilde{x}_m^s$ .
```

A version of the proposed algorithm is given in Algorithm 3.3. Since the algorithm utilizes a shifted variant of the Bi-CG method it can deal with a sequence of shifted systems and transposed shifted systems using only matrix-vector multiplications associated with the solution of linear systems with no shifts. This saves the burden of computing matrix-vector multiplications corresponding to each shifted system.

Observe that the proposed algorithm involves Krylov iterations to compute inverses of the shifted matrices $(s_i I - A)$ $i = 1, \dots, m$. This allow us to tune the approximation of inverses by setting max_it and tol values to different values.

Algorithm 3.3 Iterative shifted bi-conjugate gradient method

- 1: **Inputs:** A, B, C, m, max_it and tol
 - 2: **Initialize:** $S_m^0 = \{s_1^0, \dots, s_m^0\} \subset \mathcal{C}^m$, $V_m = [\]$, $W_m = [\]$, set $j = 0$ and $error > tol$.
 - 3: **while** $error < tol$ **do**
 - 4: Call the shifted Bi-CG algorithm (Algorithm 3.2) to compute x_m^s and $\tilde{x}_m^s$ with A, B, C, S_m^j, max_it and tol as inputs.
 - 5: $V_m := [x_m^s]$, $W_m := [\tilde{x}_m^s]$
 - 6: Compute $A_m = (W_m^T V_m)^{-1} W_m^T A V_m$; $B_m = (W_m^T V_m)^{-1} W_m^T B$; $C_m = C V_m$;
 - 7: $error = \|\text{sort}(eig(-A_m)) - \text{sort}(S_m^j)\|$
 - 8: Update the interpolation points, $S_m^{j+1} = eig(-A_m)$; $j = j + 1$;
 - 9: **end while**
 - 10: **Outputs:** A_m, B_m, C_m .
-

3.3 Partition of the iterative shifted Bi-CG method

The iterative shifted Bi-CG algorithm (Algorithm 3.2) has the useful property of utilizing the basis vectors associated with a linear system $Ax = B$ and a transposed system $A^T \tilde{x} = C^T$ in order to solve a sequence of shifted linear systems and its transposed systems. In the previous section we showed how this can be exploited in one iteration of the IRKA method. This section shows that the required basis vectors can be computed in advance before using it for the solution of different shifted linear systems.

Suppose that the shifted BiCG algorithm (Algorithm 3.2) is divided into two parts. The first part involves the computation of all the residuals $\mathcal{R} = [r_1, r_2, \dots, r_m]$ and $\tilde{\mathcal{R}} = [\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_m]$ and all the scalar parameters, $\mathcal{A} = [\alpha_1, \dots, \alpha_m]$ and $\mathcal{B} = [\beta_1, \dots, \beta_m]$. This part is given by Algorithm 3.4. Note that the algorithm does not compute the approximations x_m and $\tilde{x}_m$, because only the residual errors and the constant parameters are required for the solution of shifted linear systems. The residual errors represent Bi-CG polynomials of degree m that are linked to the set of interpolation points $S_m = [s_1, \dots, s_m]$ according to (3.7), but are computed independant of S_m . This means that we can compute all the Krylov subspaces associated with S_m in advance.

The second part of Algorithm 3.2 involves the computation of all search directions d_i^s and $\tilde{d}_i^s$ and all the scalar parameters $\mathcal{Z}_s = [\zeta_1^s, \dots, \zeta_m^s]$, $\mathcal{A}_s = [\alpha_1^s, \dots, \alpha_m^s]$ and $\mathcal{B}_s = [\beta_1^s, \dots, \beta_m^s]$ to get approximations to shifted and transposed shifted systems. This part also requires the output of Algorithm 3.4. This kind of partition can be useful because the outputs of Algorithm 3.4 can be used by different sequence of shifted linear systems without computing them for each sequence. However the basis vectors need to be stored, which increases its storage requirement.

The iterative shifted Bi-CG algorithm with the partition is given in Algorithm 3.5.

Algorithm 3.4 Krylov basis for Shifted Bi-CG method

- 1: **Inputs:** $A, B, C, max_it, error = \max\{\|r_0\|/\|B\|, \|\tilde{r}_0\|/\|C\|\}$ and tol
 - 2: $n := \text{size}(A, 1)$
 - 3: **Initialize:** $\mathcal{R} = \tilde{\mathcal{R}} = \text{zeros}(n, m + 1), \mathcal{A} = \mathcal{B} = \text{zeros}(m + 1, 1)$
 - 4: $\mathcal{R}(:, 1) = r_1 := B, \tilde{\mathcal{R}}(:, 1) = \tilde{r}_1 := C^T, \mathcal{A}(1) = 1, \mathcal{B}(1) = 1, d_1 = r_1$ and $\tilde{d}_1 = \tilde{r}_1$
 - 5: **for** $i = 1$ to max_it **do**
 - 6: $\alpha_i = (\tilde{r}_i^T r_i) / (\tilde{d}_i^T A d_i)$
 - 7: $r_{i+1} = r_i - \alpha_i A d_i, \tilde{r}_{i+1} = \tilde{r}_i - \bar{\alpha}_i A^T \tilde{d}_i$
 - 8: $\beta_{i+1} = (\tilde{r}_{i+1}^T r_{i+1}) / (\tilde{r}_i^T r_i)$
 - 9: $d_{i+1} = r_{i+1} - \beta_{i+1} d_i, \tilde{d}_{i+1} = \tilde{r}_{i+1} - \bar{\beta}_{i+1} \tilde{d}_i$
 - 10: $error = \max\{\|r_{i+1}\|/\|B\|, \|\tilde{r}_{i+1}\|/\|C\|\}$
 - 11: **if** $error < tol$ **then** break $max_it = i$ **end**
 - 12: $\mathcal{R}(:, i + 1) = r_{i+1}, \tilde{\mathcal{R}}(:, i + 1) = \tilde{r}_{i+1}, \mathcal{A}(i + 1) = \alpha_i, \mathcal{B}(i + 1) = \beta_i$
 - 13: **end for**
 - 14: **Outputs:** $\mathcal{R}, \tilde{\mathcal{R}}, \mathcal{A}, \mathcal{B}$
-

Example 3.3.1. Let us consider a benchmark model [23] of atmospheric storm track. The system is SISO and has 598 states. We use both IRKA and the iterative shifted biCG method to compute an order 10 approximate model that satisfies the necessary $\mathcal{H}_2$ optimality conditions. The elapsed time for IRKA method was 39.476791 seconds while the iterative shifted biCG method took 7.296713 seconds to compute the reduced order model. Both approaches take 19 iterations for convergence with a tolerance of 1×10^{-8} . The $\mathcal{H}_2$ error is the same in each case and the singular value plot for the actual system and the reduced systems in the two cases are shown in figure 3.1.

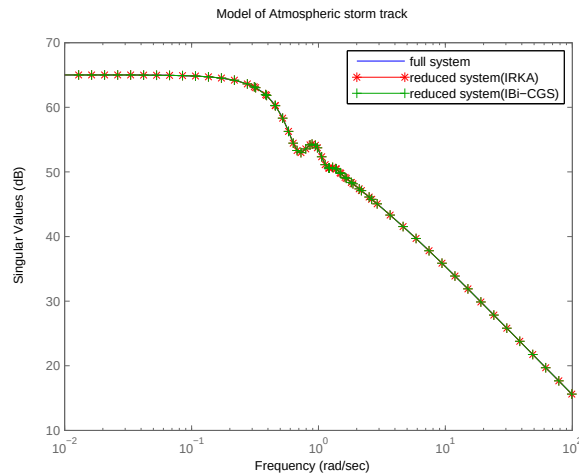


Figure 3.1: Singular value plot for the iterative rational Krylov approximation (IRKA) and the iterative shifted bi-conjugate gradient approximation (IBi-CGS)

Algorithm 3.5 IRKA using shifted bi-conjugate gradient method

```

1: Inputs:  $A, B, C, m, max\_it$  and  $tol$ 
2:  $n := \text{size}(A, 1)$ 
3: Call Algorithm 3.4 with inputs  $A, B, C, max\_it$  and  $tol$  to have  $\mathcal{R}, \tilde{\mathcal{R}}, \mathcal{A}, \mathcal{B}$ .
4: while  $error < tol$  do
5:   Initialize:  $S_m^0 = [s_1^0, \dots, s_m^0] \in \mathcal{C}^m$ ,  $\mathcal{Z}_s = \hat{\mathcal{Z}}_s = \tilde{\mathcal{Z}}_s = \text{ones}(m, 1)$ ,  $\mathcal{A}_s = \mathcal{B}_s =$ 
     zeros( $m, 1$ ),  $x_s = \tilde{x}_s = d_s = \tilde{d}_s = \text{zeros}(n, m)$ ,  $V_m = [ ]$ ,  $W_m = [ ]$  and  $j = 0$ 
6:   for  $i = 0$  to  $m$  do
7:      $\alpha = \mathcal{A}(i + 1)$ ,  $\alpha_{-1} = \mathcal{A}(i)$ ,  $\beta = \mathcal{B}(i + 1)$ 
8:      $r = \mathcal{R}(:, i)$ ,  $\tilde{r} = \tilde{\mathcal{R}}(:, i)$ 
9:      $\tilde{\mathcal{Z}}_s = (1 - \alpha * S_m(i)) * \mathcal{Z}_s + (\alpha * \beta / \alpha_{-1}) * (\hat{\mathcal{Z}}_s - \mathcal{Z}_s)$ ;
10:     $\mathcal{B}_s = (\hat{\mathcal{Z}}_s ./ \mathcal{Z}_s).^2 * \beta$ 
11:     $\mathcal{A}_s = -(\mathcal{Z}_s ./ \tilde{\mathcal{Z}}_s) * \alpha$ ,  $k = 1$ 
12:    if  $k < m + 1$  then
13:       $d_s(:, k) = (1 / \mathcal{Z}_s(k)) * r - \mathcal{B}_s(k) * d_s(:, k)$ 
14:       $\tilde{d}_s(:, k) = (1 / \mathcal{Z}_s(k)') * \tilde{r} - \mathcal{B}_s(k)' * \tilde{d}_s(:, k)$ 
15:       $x_s(:, k) = x_s(:, k) + \mathcal{A}_s(k) * d_s(:, k)$ 
16:       $\tilde{x}_s(:, k) = \tilde{x}_s(:, k) + \mathcal{A}_s(k)' * \tilde{d}_s(:, k)$ ,
17:       $k = k + 1$ 
18:    end
19:     $\hat{\mathcal{Z}}_s = \mathcal{Z}_s$ ,  $\mathcal{Z}_s = \tilde{\mathcal{Z}}_s$ .
20:  end for
21:   $V_m = x_s$  and  $W_m = \tilde{x}_s$ 
22:  Compute  $A_m = (W_m^T V_m)^{-1} W_m^T A V_m$ ;  $B_m = (W_m^T V_m)^{-1} W_m^T B$ ;  $C_m = C V_m$ ;
23:   $error = \|\text{sort}(\text{eig}(-A_m)) - \text{sort}(S_m^j)\|$ 
24:  Update the interpolation points,  $S_m^{j+1} = \text{eig}(-A_m)$ ;  $j = j + 1$ ;
25: end while
26: Outputs:  $A_m, B_m, C_m$ .

```

3.4 Conclusions

This chapter presented an iterative algorithm for the $\mathcal{H}_2$ optimal model reduction problem. The proposed algorithm updates the shift frequencies in the same way as in IRKA method and converges to the same local minimizers starting from a set of interpolation points. However it is shown that the proposed algorithm requires basis vectors associated with a linear system having no shift and its transposed system to solve a sequence of shifted linear systems. The approach utilizes a collinear relationship between the residual error of the linear system $Ax = B$ and the shifted linear system $(s_0 I - A)x = B$ which follows by noting that Krylov subspaces are shift invariant. Also it is shown that the basis vectors required in solving a sequence of shifted linear systems can be separately computed and stored for other sequences. The IRKA method and the iterative shifted BiCG method may not converge to the global minimizer of the $\mathcal{H}_2$ optimal model re-

duction problem (1.18). In the next chapter, some extra conditions are derived that guarantee a global minimizer to the problem in special cases.

Chapter 4

All Optimal Local Minimizers to $\mathcal{H}_2$ Optimal Model Reduction

The IRKA method (Section 2.5) and the iterative bi-conjugate gradient method (Chapter 3) can find a local solution to the $\mathcal{H}_2$ optimal model reduction problem. However the approximations obtained through these methods are not necessarily the global minimizers. In this chapter, some new expressions of the $\mathcal{H}_2$ norm of the error system are derived that can be used to compute a first order global minimizer to the $\mathcal{H}_2$ optimal model reduction problem. The approach involves the computation of all fixed points that satisfy the necessary conditions for local minimizers. To illustrate the procedure a few numerical examples are given. Also, using a similar framework, a solution of the second order optimal $\mathcal{H}_2$ approximation problem is derived that involves the computation of all simultaneous solutions to two bivariate polynomials. It is shown that this is equivalent to the simultaneous solution of two eigenvalue problems. Finally, a few examples are given to illustrate the procedure of solving second order optimal $\mathcal{H}_2$ approximation problem.

4.1 Error system in the optimal case

In this section, various expressions are derived for the $\mathcal{H}_2$ norm of the approximation error associated with m^{th} order approximations satisfying (2.74). An interesting result concerning the $\mathcal{H}_2$ norm of the approximation error is given in the following theorem.

Theorem 4.1. *Let $S_m = \{s_1, \dots, s_m\} \subset \mathbb{C}^+$ be a fixed point of (2.74) and assume,*

without loss of generality, that $W_m^T V_m = I_m$. Let $G_m(s) = C_m(sI_m - A_m)^{-1} B_m$, where $A_m = W_m^T A V_m$, $B_m = W_m^T B$ and $C_m = C V_m$, be the corresponding interpolating function given by Lemma 2.4.1, so that $G(s_i) = G_m(s_i)$, $G'(s_i) = G'_m(s_i)$, $i = 1, \dots, m$ and $\lambda(A_m) = -S_m$. Then

$$\|G - G_m\|_{\mathcal{H}_2}^2 = \|G\|_{\mathcal{H}_2}^2 - \|G_m\|_{\mathcal{H}_2}^2 \quad (4.1)$$

Proof. It follows from Lemma 2.4.2 that the Arnoldi-like equations

$$A V_m = V_m A_m + v_{m+1} C_{V_m}, \quad B = V_m B_m + v_{m+1} D_{V_m} \quad (4.2)$$

$$A^T W_m = W_m A_m^T + w_{m+1} B_{W_m}^T, \quad C^T = W_m C_m^T + w_{m+1} D_{W_m}^T \quad (4.3)$$

are satisfied for the rational Arnoldi algorithm with the biorthogonality condition $W_m^T V_m = I_m$. Using these equations, the derivation given in Section 2.1.2 shows that the error system $E_m(s) = G(s) - G_m(s)$ can be written as

$$\begin{aligned} E_m(s) &:= C(sI - A)^{-1} B - C_m(sI - A_m)^{-1} B_m \\ &= C(sI - A)^{-1} v_{m+1} \tilde{r}_m(s) \end{aligned}$$

where

$$\tilde{r}_m(s) = D_{V_m} + C_{V_m}(sI - A_m)^{-1} B_m. \quad (4.4)$$

Note that the interpolation points are the zeros of $\tilde{r}_m(s)$ [33] and, since $S_m = -\lambda(A_m)$ and the poles of $\tilde{r}_m(s)$ are the eigenvalues of A_m it follows that $D_{V_m}^{-1} \tilde{r}_m(s)$ is all pass. Thus if P_m is the controllability Grammian of the realisation of $\tilde{r}_m(s)$ in (4.4), then

$$A_m P_m + P_m A_m^T + B_m B_m^T = 0, \quad C_{V_m} P_m + D_{V_m} B_m^T = 0 \quad (4.5)$$

are the allpass equations for $\tilde{r}_m(s)$ [104]. Let

$$P_E = \begin{bmatrix} P & X \\ X^T & P_m \end{bmatrix} \quad (4.6)$$

be the controllability Grammian for the error system $E_m(s)$ given by (2.63), so that

$$\begin{bmatrix} AP + PA^T + BB^T & AX + XA_m^T + BB_m^T \\ \star & A_m P_m + P_m A_m^T + B_m B_m^T \end{bmatrix} = 0. \quad (4.7)$$

Then a manipulation using (4.2), (4.5) and (4.7) gives

$$A(XP_m^{-1} - V_m) = (XP_m^{-1} - V_m)(A_m - B_m D_{V_m}^{-1} C_{V_m}) \quad (4.8)$$

Since the interpolation points are the eigenvalues of $A_m - B_m D_{V_m}^{-1} C_{V_m}$ (i.e. the zeros of $\tilde{r}_m(s)$) and since it is assumed that the interpolation points are not eigenvalues of A , it follows from (4.8) that $V_m = XP_m^{-1}$. Thus

$$CX = CV_m P_m = C_m P_m. \quad (4.9)$$

Now using the expression for the $\mathcal{H}_2$ norm from Lemma 1.2.1 results in

$$\begin{aligned} \|G - G_m\|_{\mathcal{H}_2}^2 &= C_E P_E C_E^T \\ &= C P C^T - C_m X^T C^T - C X C_m^T + C_m P_m C_m^T \\ &= C P C^T - C_m P_m C_m^T \\ &= \|G\|_{\mathcal{H}_2}^2 - \|G_m\|_{\mathcal{H}_2}^2 \end{aligned}$$

where the third equation follows from (4.9) and this proves the theorem. $\square$

The next theorem develop a link between the problem of $\mathcal{H}_2$ optimal approximation of $G(s)$ and low rank approximate solution of linear matrix equations.

Theorem 4.2. *Let all variables be as defined in Theorem 4.1. Then the $\mathcal{H}_2$ norm of the approximation error (4.1) is equal to the $\mathcal{H}_2$ norm of the following systems,*

$$G_P(s) = C(sI - A)^{-1} \tilde{B}, \quad G_Q(s) = \tilde{C}(sI - A)^{-1} B, \quad G_X(s) = \tilde{C}(sI - A)^{-1} \tilde{B}, \quad (4.10)$$

where $\tilde{B} = v_{m+1} D_{V_m}$ and $\tilde{C} = w_{m+1}^T D_{W_m}$. That is,

$$\|G - G_m\|_{\mathcal{H}_2} = \|G_P(s)\|_{\mathcal{H}_2} = \|G_Q(s)\|_{\mathcal{H}_2} = \|G_X(s)\|_{\mathcal{H}_2} \quad (4.11)$$

Proof. Theorem 4.1 suggests that,

$$\begin{aligned}\|G - G_m\|_{\mathcal{H}_2}^2 &= C(P - V_m P_m V_m^T)C^T, \\ &= C(P - \tilde{P})C^T.\end{aligned}\tag{4.12}$$

where $\tilde{P} = V_m P_m V_m^T$. Note that P is the controllability grammian satisfying,

$$AP + PA + BB^T = 0.\tag{4.13}$$

Let

$$R_{\tilde{P}} := A\tilde{P} + \tilde{P}A + BB^T.\tag{4.14}$$

Using the rational Arnoldi-like equations in (4.2) to replace AV_m and B in the above equation, we have,

$$R_{\tilde{P}} = \begin{bmatrix} V_m & v_{m+1} \end{bmatrix} \begin{bmatrix} A_m P_m + P_m A_m^T + B_m B_m^T & P_m C_{V_m}^T + B_m D_{V_m}^T \\ C_{V_m} P_m + D_{V_m} B_m^T & D_{V_m} D_{V_m}^T \end{bmatrix} \begin{bmatrix} V_m^T \\ v_{m+1}^T \end{bmatrix}.\tag{4.15}$$

Using (4.5) now gives

$$R_{\tilde{P}} = v_{m+1} D_{V_m} D_{V_m}^T v_{m+1}^T = \tilde{B}\tilde{B}^T\tag{4.16}$$

Subtracting (4.14) from (4.13), we have

$$A(P - \tilde{P}) + (P - \tilde{P})A^T + \tilde{B}\tilde{B}^T = 0,\tag{4.17}$$

From (4.12) and (4.17), we have $\|G - G_m\|_{\mathcal{H}_2} = \|G_P\|_{\mathcal{H}_2}$. A similar procedure can be used to prove the other two equalities. $\square$

The above theorems suggest various expressions for the $\mathcal{H}_2$ norm of the error system and these are summarized in Table 4.1. The table assume that P_v and X_v are the controllability and cross grammians of $G_P(s)$, Q_w and X_w are the observability and cross grammians of $G_Q(s)$ and X_{vw} is the cross grammian associated with $G_X(s)$.

Table 4.1: Four different expressions for $\mathcal{H}_2$ norm of approximation error in terms of controllability, observability and cross grammians

No.	$\ G - G_m\ _{\mathcal{H}_2}^2$ in terms of Controllability grammian	$\ G - G_m\ _{\mathcal{H}_2}^2$ in terms of Observability grammian	$\ G - G_m\ _{\mathcal{H}_2}^2$ in terms of Cross grammian
1	$CP C^T - C_m P_m C_m^T$	$B^T Q B - B_m^T Q_m B_m$	$C X B - C_m X_m B_m$
2	$CP_v C^T$	$ D_{V_m} ^2 v_{m+1}^T Q v_{m+1}$	$ D_{V_m} C X_v v_{m+1}$
3	$ D_{W_m} ^2 w_{m+1}^T P w_{m+1}$	$B^T Q_w B$	$ D_{W_m} w_{m+1}^T X_w B$
4	$ D_{W_m} ^2 w_{m+1}^T P_v w_{m+1}$	$ D_{V_m} ^2 v_{m+1}^T Q_w v_{m+1}$	$ D_{V_m} D_{W_m} w_{m+1}^T X_{vw} v_{m+1}$

4.2 First order $\mathcal{H}_2$ optimal model reduction

The key challenge addressed in this section is the identification of first order global $\mathcal{H}_2$ optimal interpolating approximations. The problem of identification of global $\mathcal{H}_2$ optimal approximations, is still an open problem and has received considerable attention in the literature, e.g., [14, 37, 56, 92, 101, 102] and more recently [51]. We address this issue when the original system is single input single output (SISO) and the approximation has first order.

Lemma 2.5.1 suggests that if we can find all possible sets of interpolating points $S_m = \{s_1, \dots, s_m\} \subset \mathbb{C}^+$ such that (2.73) is satisfied, then the corresponding $G_m(s)$'s will include all the local minimizers for $G(s)$. Thus the problem reduces to the computation of all the fixed points satisfying (2.73).

4.2.1 Computation of Fixed points

The following theorem gives the solution to the global $\mathcal{H}_2$ model reduction problem for SISO $G(s)$ when $m = 1$.

Theorem 4.3. *Let $G(s) = C(sI_n - A)^{-1}B$ be a stable SISO and minimal system and let*

$m = 1$. Then all fixed points of (2.74) are given by the zeros of the transfer function

$$\begin{aligned} \hat{H}(s) = \hat{C}(sI_{2n} - \hat{A})^{-1}\hat{B} &\stackrel{s}{=} \left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & 0 \end{array} \right] \\ &:= \left[\begin{array}{cc|c} A & A & B \\ 0 & A & 2B \\ \hline C & 0 & 0 \end{array} \right] \end{aligned} \quad (4.18)$$

or, equivalently, by the eigenvalues of the matrix pencil

$$\left[\begin{array}{cc} \hat{A} - \lambda I_{2n} & \hat{B} \\ \hat{C} & 0 \end{array} \right].$$

Let $s_1, \dots, s_f$ be the positive zeros of $\hat{H}$ and let $G_1(s), \dots, G_f(s)$ be the corresponding interpolating transfer functions given by Lemma 2.4.1, so that

$$G_i(s_i) = G(s_i), \quad G'_i(s_i) = G'(s_i), \quad i = 1, \dots, f. \quad (4.19)$$

Then the global solution to the $\mathcal{H}_2$ model reduction problem (1.18) is one of the $G_i(s)$'s.

Proof. For $m = 1$, (2.73) implies

$$S_m = - \overbrace{(W_m^T V_m)^{-1} W_m^T A V_m}^{A_m}, \quad (4.20)$$

where $S_m = \sigma \in \mathcal{R}^+$. Rearranging (4.20) gives

$$W_m^T (\sigma I + A) V_m = 0.$$

Using Lemma 2.4.1 to define $V_m = (\sigma I - A)^{-1} B$ and $W_m = (\sigma I - A^T)^{-1} C^T$ we have,

$$\begin{aligned} C(\sigma I - A)^{-1}(\sigma I + A)(\sigma I - A)^{-1}B &= 0, \\ \Rightarrow C(\sigma I - A)^{-1}(\sigma I - A + 2A)(\sigma I - A)^{-1}B &= 0, \\ \Rightarrow C(\sigma I - A)^{-1}(B + 2A(\sigma I - A)^{-1}B) &= 0, \end{aligned}$$

or $\hat{H}(\sigma) = 0$ so that σ is a zero of $\hat{H}(s)$.

To prove that one of the $G_i(s)$'s must be the global minimizer, note first that the global minimizer is also a local minimizer. It follows that if the $G_i(s)$'s include all local minimizers then we are done.

Now, since $m = 1$, $G_i(s)$ has one (simple) pole and it follows from Lemma 2.5.1 that all local minimizers satisfy (4.19). It remains to prove that all the $G_i(s)$'s are projections in the sense of Lemma 2.4.1 since the implication in the lemma is only in one direction. Let $G_i(s)$ satisfy (4.19) and write $G_i(s) = C_i(sI_1 - A_i)^{-1}B_i$ where A_i , B_i and C_i are scalar since $G_i(s)$ is SISO and of order one. Then to satisfy (4.19) we have

$$\frac{C_i B_i}{s_i - A_i} = C(s_i I_n - A)^{-1} B, \quad \frac{C_i B_i}{(s_i - A_i)^2} = C(s_i I_n - A)^{-2} B. \quad (4.21)$$

First, assume that $C(s_i I_n - A)^{-2} B \neq 0$. Then a computation using (4.21) shows that

$$\begin{aligned} A_i &= (W_m^T V_m)^{-1} W_m^T A V_m, & B_i &= \alpha (W_m^T V_m)^{-1} W_m^T B \\ C_i &= \alpha^{-1} C V_m \end{aligned}$$

where $V_m = (s_i I_n - A)^{-1} B$ and $W_m = (s_i I_n - A^T)^{-1} C^T$ for some nonzero α . Thus $G_i(s)$ is interpolating in the sense of Lemma 2.4.1. Finally, suppose that $G'(s_i) = C(s_i I_n - A)^{-2} B = 0$. Then a computation using (4.21) verifies that there exists no strictly proper $G_i(s)$ satisfying (4.21), that is, interpolating $G(s)$ and $G'(s)$ at s_i . $\square$

The following algorithm summarizes our proposed method for the solution of the first order $\mathcal{H}_2$ optimal approximation problem.

Algorithm 4.1 $\mathcal{H}_2$ optimal first order approximation of linear systems

Data: A , B and C .

- 1: Compute all the zeroes of the linear dynamical system $\hat{H}(s)$ given by (4.18).
 - 2: Choose the set Z_f of positive and real zeroes and let f be the number of elements in Z_f .
 - 3: For A , B , C and each element of Z_f use the rational Arnoldi or Lanczos algorithms to obtain f different first order interpolating approximations $G_i(s) = C_i(sI_1 - A_i)^{-1}B_i$, $i = 1, \dots, f$.
 - 4: Compute the product $C_i P_i C_i^T$ for each approximation, where P_i is the controllability grammian satisfying $A_i P_i + P_i A_i^T + B_i B_i^T = 0$.
 - 5: Using (4.1), the optimum $\mathcal{H}_2$ approximation is the $G_i(s)$ corresponding to the maximum $C_i P_i C_i^T$.
-

4.2.2 Numerical results

The following (randomly generated) examples illustrate our proposed technique. Consider the following transfer functions,

$$\begin{aligned}
 F_1(s) &= \frac{-1.986s^2 + 19.17s - 0.1606}{s^3 + 4.857s^2 + 14.08s + 23.02} \\
 F_2(s) &= \frac{-0.3556s^3 + 9.402s^2 + 5.149s - 6.527}{s^4 + 9.486s^3 + 31.24s^2 + 32.85s + 5.857} \\
 F_3(s) &= \frac{0.6355s^2 - 5.769s + 1.119}{s^3 + 3.696s^2 + 5.037s + 1.557}
 \end{aligned}$$

For each of the above three systems, the fixed points and the optimal choice of fixed point are shown in Table 4.2. Let the order one optimal approximations be represented by $\hat{F}_1(s)$, $\hat{F}_2(s)$ and $\hat{F}_3(s)$. The relative $\mathcal{H}_2$ norm of the approximation error is then,

$$E_i := \|\hat{F}_i - F_i\|_{\mathcal{H}_2} / \|F_i\|_{\mathcal{H}_2},$$

and is given in Table 4.3 evaluated at the fixed points.

Table 4.2: Fixed Points for $F_1(s)$, $F_2(s)$ $F_3(s)$

Model	Fixed Points			Optimal Point
$F_i(s)$	s_1	s_2	s_3	
$F_1(s)$	36.2279	2.1364	0.0028	2.1364
$F_2(s)$	93.2228	2.8113	0.0928	2.8113
$F_3(s)$	32.1125	1.2538	0.0507	1.2538

Table 4.3: Relative Error for $F_1(s)$, $F_2(s)$ $F_3(s)$

Model	Local Minimizers			Global Minimizer
$F_i(s)$	$E_i^{s_1}$	$E_i^{s_2}$	$E_i^{s_3}$	
$F_1(s)$	0.9904	0.9339	1.0000	0.9339
$F_2(s)$	0.9986	0.8922	0.8962	0.8922
$F_3(s)$	0.9952	0.8673	0.9902	0.8673

Table 4.4: Fixed Points for Benchmark Examples

Model	Fixed Points			Optimal Point
$F_j(s)$	s_4	s_5	s_6	
$F_4(s)$	0.4952	-	-	0.4952
$F_5(s)$	0.5762	-	-	0.5762
$F_6(s)$	4.9981×10^3	0.4659	0.0050	4.9981×10^3

Table 4.5: Relative Error for Benchmark Examples

Model	Local Minimizers			Global Minimizer
$F_j(s)$	$E_j^{s_4}$	$E_j^{s_5}$	$E_j^{s_6}$	
$F_4(s)$	0.4268	-	-	0.4268
$F_5(s)$	0.4818	-	-	0.4818
$F_6(s)$	0.0975	0.9992	0.9950	0.0975

Consider the following three models from the literature.

- The state space representation of the linear system $F_4(s)$ is given by [56],

$$A = \begin{bmatrix} 0 & 0 & 0 & -150 \\ 1 & 0 & 0 & -245 \\ 0 & 1 & 0 & -113 \\ 0 & 0 & 1 & -19 \end{bmatrix}, \quad B = \begin{bmatrix} 4 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad C^T = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (4.22)$$

- The transfer function of the linear system $F_5(s)$ is given by [92],

$$F_5(s) = \frac{s^2 + 15s + 50}{s^4 + 5s^3 + 33s^2 + 79s + 50} \quad (4.23)$$

- The state space representation of the linear system $F_6(s)$ is given by [56],

$$A = \begin{bmatrix} -0.005 & -0.99 \\ -0.99 & -5000 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 100 \end{bmatrix}, \quad C = B^T. \quad (4.24)$$

For each of the above three systems, the fixed points and the optimal choice of fixed point are shown in Table 4.4. The relative $\mathcal{H}_2$ norms of the errors for each of the fixed points are also given in Table 4.5. Note that for Models 1 and 2 there is only one fixed point, while for Model 3 there are three. The first two local minima were reported in [56] and the table gives a third.

4.3 Second order $\mathcal{H}_2$ optimal model reduction

Although the solution of the $\mathcal{H}_2$ norm model reduction problem in the case $m = 1$ was given explicitly in the previous section, the solution in the general case is much more difficult. The following result derives a solution to the $m = 2$ problem that involves the evaluation of all simultaneous solutions of two equations in two unknowns.

Theorem 4.4. *Let all variables be as defined before and let $m = 2$. For $s_1, s_2 \in \mathbb{C}^+$ such that $s_1 \neq s_2$ and $S_m := \{s_1, s_2\}$ is closed under conjugation define*

$$\begin{aligned} V_m(s_1, s_2) &= \begin{bmatrix} (s_1 I - A)^{-1} b & (s_1 I - A)^{-1} (s_2 I - A)^{-1} b \end{bmatrix} \\ W_m(s_1, s_2) &= \begin{bmatrix} (s_1 I - A)^{-T} C^T & (s_1 I - A)^{-T} (s_2 I - A)^{-T} C^T \end{bmatrix} \\ A_m(s_1, s_2) &= (W_m^T(s_1, s_2) V_m(s_1, s_2))^{-1} W_m^T(s_1, s_2) A V_m(s_1, s_2) \\ f(s_1, s_2) &= \det(W_m^T(s_1, s_2)(s_1 I + A) V_m(s_1, s_2)). \end{aligned}$$

Then all $S_m = \{s_1, s_2\} \subset \mathbb{C}^+$ such that $S_m = -\lambda(A_m(s_1, s_2))$ are given by the simultaneous solutions of the two equations

$$f(s_1, s_2) = 0, \quad f(s_2, s_1) = 0. \quad (4.25)$$

Let $S_m^1, \dots, S_m^f$ be all simultaneous solutions to (4.25) and let $G_1(s), \dots, G_f(s)$ be the corresponding interpolating transfer functions given by Lemma 2.4.1. Then the global solution to the $\mathcal{H}_2$ model reduction problem (1.18) is one of the $G_i(s)$'s.

Proof: s_1 is an eigenvalue of $-A_m(s_1, s_2)$ if there exists a nonzero $x_m \in \mathbb{C}^m$ such that $(A_m(s_1, s_2) + s_1 I) x_m = 0$. A simple manipulation shows that this is equivalent to $f(s_1, s_2) = 0$. Note that since $\lambda(A_m(s_1, s_2)) = \lambda(A_m(s_2, s_1))$ (moment matching is independent of the order it is carried out), it follows that s_2 is an eigenvalue of $-A_m(s_1, s_2)$ if there exists a nonzero $x_m \in \mathbb{C}^m$ such that $(A_m(s_2, s_1) + s_2 I) x_m = 0$. A simple manipulation shows that this is equivalent to $f(s_2, s_1) = 0$. $\square$

The following result shows that the solution to the $m = 2$ problem reduces to the evaluation of all simultaneous solutions of two bivariate polynomial equations.

Theorem 4.5. *Let all variables be as defined before and let $m = 2$. Define*

$$\begin{aligned} \mathcal{A}_{11} &= \begin{bmatrix} A & 2A \\ 0 & A \end{bmatrix} & \mathcal{A}_{12} &= \begin{bmatrix} I & 0 \\ I & 0 \end{bmatrix} & \mathcal{B}_1 &= \begin{bmatrix} b & 0 \\ b & 0 \end{bmatrix} \\ \mathcal{A}_{21} &= \begin{bmatrix} 0 & 0 \\ I & 0 \end{bmatrix} & \mathcal{A}_{22} &= \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} & \mathcal{B}_2 &= \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} \\ \mathcal{C}_1 &= \begin{bmatrix} c & 0 \\ 0 & 0 \end{bmatrix} & \mathcal{C}_2 &= \begin{bmatrix} 0 & 0 \\ 0 & c \end{bmatrix} & \Delta &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

and

$$p(s_1, s_2) = \det \left(\begin{bmatrix} \mathcal{A}_{11} - s_1 I & \mathcal{A}_{12} & \mathcal{B}_1 \\ \mathcal{A}_{21} & \mathcal{A}_{22} - s_2 I & \mathcal{B}_2 \\ \mathcal{C}_1 & \mathcal{C}_2 & \Delta \end{bmatrix} \right). \quad (4.26)$$

Then all $s_1, s_2 \in \mathbb{C}^+$ such that $\{s_1, s_2\} = -\lambda(A_m(s_1, s_2))$ are given by the simultaneous solutions of the two bivariate polynomial equations

$$p(s_1, s_2) = 0, \quad p(s_2, s_1) = 0 \quad (4.27)$$

Proof. Let $A_{s_1} := (s_1 I - A)$ and $A_{s_2} := (s_2 I - A)$, then Theorem 4.4 implies,

$$\det \left(\begin{bmatrix} C \\ CA_{s_2}^{-1} \end{bmatrix} A_{s_1}^{-1} (s_1 I_n + A) A_{s_1}^{-1} \begin{bmatrix} B & A_{s_2}^{-1} B \end{bmatrix} \right) = 0. \quad (4.28)$$

Write

$$\begin{aligned} \begin{bmatrix} C \\ CA_{s_2}^{-1} \end{bmatrix} &= \underbrace{\begin{bmatrix} C \\ 0 \end{bmatrix}}_{C_1} + \underbrace{\begin{bmatrix} 0 \\ C \end{bmatrix}}_{C_2} A_{s_2}^{-1} \\ &= C_1 + C_2 A_{s_2}^{-1}, \\ \begin{bmatrix} B & A_{s_2}^{-1} B \end{bmatrix} &= \underbrace{\begin{bmatrix} B & 0 \end{bmatrix}}_{B_1} + A_{s_2}^{-1} \underbrace{\begin{bmatrix} 0 & B \end{bmatrix}}_{B_2} \\ &= B_1 + A_{s_2}^{-1} B_2 \end{aligned}$$

then (4.28) is satisfied if and only if there exists $0 \neq u \in R^2$ such that

$$(C_1 + C_2 A_{s_2}^{-1}) A_{s_1}^{-1} (s_1 I_n + A) A_{s_1}^{-1} (B_1 + A_{s_2}^{-1} B_2) u = 0. \quad (4.29)$$

Suppose we define vectors x_1, x_2, x_3 and $x_4 \in R^n$ such that,

$$x_1 := A_{s_2}^{-1} B_2 u \quad \Leftrightarrow \quad -A_{s_2} x_1 + B_2 u = 0, \quad (4.30)$$

$$x_2 := A_{s_1}^{-1} (B_1 u + x_1) \quad \Leftrightarrow \quad -A_{s_1} x_2 + B_1 u + x_1 = 0, \quad (4.31)$$

$$x_3 := A_{s_1}^{-1} (s_1 I_n + A) x_2 \quad \Leftrightarrow \quad -A_{s_2} x_3 + (s_1 I_n + A) x_2 = 0, \quad (4.32)$$

$$x_4 := A_{s_2}^{-1} x_3 \quad \Leftrightarrow \quad -A_{s_2} x_4 + x_3 = 0, \quad (4.33)$$

then (4.29) can be written as,

$$C_1 x_3 + C_2 x_4 = 0. \quad (4.34)$$

Using (4.30), (4.31), (4.32), (4.33) and (4.34) it follows that (4.28) is satisfied if and only if there exists $0 \neq u \in R^2$, $x_1, x_2, x_3, x_4 \in R^n$ such that

$$\begin{bmatrix} -A_{s_2} & 0 & 0 & 0 & B_2 \\ I_n & -A_{s_1} & 0 & 0 & B_1 \\ 0 & A + s_1 I_n & -A_{s_1} & 0 & 0 \\ 0 & 0 & I_n & -A_{s_2} & 0 \\ 0 & 0 & C_1 & C_2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ u \end{bmatrix} = 0 \quad (4.35)$$

Using row and column operations, we have

$$\begin{bmatrix} -A_{s_1} & 2A & I_n & 0 & B_1 \\ 0 & -A_{s_1} & I_n & 0 & B_1 \\ 0 & 0 & -A_{s_2} & 0 & B_2 \\ I_n & 0 & 0 & -A_{s_2} & 0 \\ C_1 & 0 & 0 & C_2 & 0 \end{bmatrix} \begin{bmatrix} x_3 + x_2 \\ x_2 \\ x_1 \\ x_4 \\ u \end{bmatrix} = 0$$

The above equation can be written as,

$$\begin{bmatrix} \mathcal{A}_{11} - s_1 I & \mathcal{A}_{12} & \mathcal{B}_1 \\ \mathcal{A}_{21} & \mathcal{A}_{22} - s_2 I & \mathcal{B}_2 \\ \mathcal{C}_1 & \mathcal{C}_2 & \Delta \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ u \end{bmatrix} = 0,$$

which proves that $p(s_1, s_2) = 0$. Similarly it can be shown that $p(s_2, s_1) = 0$. $\square$

Note that the solution is given in terms of the simultaneous solutions of two 2-dimensional matrix pencil problems. Under two mild assumptions, the next result gives the solution in terms of the simultaneous solutions of two 2-dimensional matrix eigenvalue problems.

Theorem 4.6. *Let all variables be as defined before and let $m = 2$. Assume that $p(0, 0) \neq 0$ so that the pair $s_1 = s_2 = 0$ is not a solution to (4.27). Let*

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

Assume that

$$\mathcal{C}\mathcal{A}^{-1}\mathcal{B} = \begin{bmatrix} -\mathcal{C}\mathcal{A}^{-1}\mathcal{B} & \mathcal{C}\mathcal{A}^{-2}\mathcal{B} \\ \mathcal{C}\mathcal{A}^{-2}\mathcal{B} & -\mathcal{C}\mathcal{A}^{-3}\mathcal{B} \end{bmatrix}$$

is nonsingular (which is equivalent to $f(0, 0) \neq 0$) and that there are no simultaneous solutions to the two bivariate polynomial equations

$$\det \begin{bmatrix} \mathcal{A}_{11} - s_1 I & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} - s_2 I \end{bmatrix} = 0, \quad \det \begin{bmatrix} \mathcal{A}_{11} - s_2 I & \mathcal{A}_{12} \\ \mathcal{A}_{21} & \mathcal{A}_{22} - s_1 I \end{bmatrix} = 0 \quad (4.36)$$

Let

$$\hat{\mathcal{A}} = \mathcal{A}^{-1} - \mathcal{A}^{-1}\mathcal{B}(\mathcal{C}\mathcal{A}^{-1}\mathcal{B})^{-1}\mathcal{C}\mathcal{A}^{-1}$$

and partition $\hat{\mathcal{A}}$ compatibly with $\mathcal{A}$ as

$$\hat{\mathcal{A}} = \begin{bmatrix} \hat{\mathcal{A}}_{11} & \hat{\mathcal{A}}_{12} \\ \hat{\mathcal{A}}_{21} & \hat{\mathcal{A}}_{22} \end{bmatrix}. \quad (4.37)$$

Let $\hat{s}_1$ and $\hat{s}_2$ denote s_1^{-1} and s_2^{-1} , respectively, and define

$$\hat{p}(\hat{s}_1, \hat{s}_2) = \det \left(\begin{bmatrix} \hat{\mathcal{A}}_{11} - \hat{s}_1 I & \hat{\mathcal{A}}_{12} \\ \hat{\mathcal{A}}_{21} & \hat{\mathcal{A}}_{22} - \hat{s}_2 I \end{bmatrix} \right).$$

Then all $s_1, s_2 \in \mathbb{C}^+$ such that $\{s_1, s_2\} = -\lambda(A_m(s_1, s_2))$ are given by $s_1 = \hat{s}_1^{-1}$ and $s_2 = \hat{s}_2^{-1}$ where $\{\hat{s}_1, \hat{s}_2\}$ are simultaneous solutions of the two bivariate polynomial equations

$$\hat{p}(\hat{s}_1, \hat{s}_2) = 0, \quad \hat{p}(\hat{s}_2, \hat{s}_1) = 0 \quad (4.38)$$

Proof. The polynomial $p(s_1, s_2)$ defined in Theorem 4.5 can be written as,

$$p(s_1, s_2) = \det \left(\begin{bmatrix} \mathcal{A} - s_1 F_1 - s_2 F_2 & \mathcal{B} \\ \mathcal{C} & 0 \end{bmatrix} \right), \quad (4.39)$$

where

$$F_1 = \begin{bmatrix} I_{2n} & 0 \\ 0 & 0 \end{bmatrix}, \quad F_2 = \begin{bmatrix} 0 & 0 \\ 0 & I_{2n} \end{bmatrix}$$

Let $\Lambda := s_1 F_1 + s_2 F_2$, then $\hat{\Lambda} = s_1^{-1} F_1 + s_2^{-1} F_2$ is the inverse of Λ because Λ is a diagonal matrix. Since we are assuming that there are no simultaneous solutions to the two bivariate polynomial equations in (4.36), we may assume that $\det(\mathcal{A} - \Lambda) \neq 0$. Thus we can write,

$$\begin{aligned} p(s_1, s_2) = 0 &\Leftrightarrow \det(\mathcal{A} - \Lambda) \det(-\mathcal{C}(\mathcal{A} - \Lambda)^{-1} \mathcal{B}) = 0 \\ &\Leftrightarrow \det(-\mathcal{C}(\mathcal{A} - \hat{\Lambda}^{-1})^{-1} \mathcal{B}) = 0, \\ &\Leftrightarrow \det(-\mathcal{C} \mathcal{A}^{-1} \mathcal{B} + \mathcal{C} \mathcal{A}^{-1} (\mathcal{A}^{-1} - \hat{\Lambda})^{-1} \mathcal{A}^{-1} \mathcal{B}) = 0 \\ &\Leftrightarrow \det \left(\begin{bmatrix} \mathcal{A}^{-1} - \hat{\Lambda} & \mathcal{A}^{-1} \mathcal{B} \\ -\mathcal{C} \mathcal{A}^{-1} & -\mathcal{C} \mathcal{A}^{-1} \mathcal{B} \end{bmatrix} \right) = 0. \\ &\Leftrightarrow \det(\mathcal{A}^{-1} - \hat{\Lambda} - \mathcal{A}^{-1} \mathcal{B} (\mathcal{C} \mathcal{A}^{-1} \mathcal{B})^{-1} \mathcal{C} \mathcal{A}^{-1}) = 0 \end{aligned} \quad (4.40)$$

where the third equivalence follows from the use of the matrix inversion lemma [52].

Using the partition in (4.37) and the definition of $\hat{\Lambda}$ we have,

$$p(s_1, s_2) = 0 \Leftrightarrow \det \begin{pmatrix} \hat{\mathcal{A}}_{11} - \hat{s}_1 I_{2n} & \hat{\mathcal{A}}_{12} \\ \hat{\mathcal{A}}_{21} & \hat{\mathcal{A}}_{22} - \hat{s}_2 I_{2n} \end{pmatrix} = 0. \quad (4.41)$$

which proves that $p(s_1, s_2) = 0$ if and only if $\hat{p}(\hat{s}_1, \hat{s}_2) = 0$. A similar proof shows that $p(s_2, s_1) = 0$ if and only if $\hat{p}(\hat{s}_2, \hat{s}_1) = 0$. $\square$

Remark 4.3.1. *Regarding the assumption on the non-existence of simultaneous solutions for (4.36), note that if we take $s_1 = s_2$ and writing $p(s_1)$ for $p(s_1, s_2)$ in (4.26), then finding the roots of $p(s) = 0$ corresponds to finding the system zeros for the state-space system $\mathcal{G} \stackrel{s}{=} (\mathcal{A}, \mathcal{B}, \mathcal{C}, 0)$. Thus the assumption that there are no simultaneous solutions to the two bivariate polynomial equations in (4.36) corresponds to the assumption that $\det(\mathcal{A} - s_1 I) \neq 0$, that is, that s_1 is not simultaneously both a pole and a zero of the system $\mathcal{G}$. Relaxing this assumption requires a generalisation of the concepts of controllability and observability to two-dimensional systems and falls outside the scope of this thesis. However, it should be mentioned that generically the assumption is satisfied.*

4.3.1 Numerical results

To illustrate the procedure described in this section, we give some numerical examples. Consider the following (randomly generated) transfer functions,

$$\begin{aligned} F_7(s) &= \frac{-2.9239s^3 - 39.5525s^2 - 97.5270s - 147.1508}{s^4 + 11.9584s^3 + 43.9119s^2 + 73.6759s + 44.3821}, \\ F_8(s) &= \frac{-1.2805s^3 - 6.2266s^2 - 12.8095s - 9.3373}{s^4 + 3.1855s^3 + 8.9263s^2 + 12.2936s + 3.1987}, \\ F_9(s) &= \frac{-1.3369s^3 - 4.8341s^2 - 47.5819s - 42.7285}{s^4 + 17.0728s^3 + 84.9908s^2 + 122.4400s + 59.9309}. \end{aligned}$$

For the three systems above, second order optimal approximations are constructed, $\hat{F}_7, \hat{F}_8$ and $\hat{F}_9$, following the approach presented in this work. The choice of fixed points are shown in Table 4.6. The relative $\mathcal{H}_2$ norms of the errors for each of the fixed points are also given in Table 4.7

Table 4.6: Fixed points for $\hat{F}_7, \hat{F}_8, \hat{F}_9$

Fixed Points		
	S_1	S_2
$\hat{F}_7$	$\{2.4437, 0.8883\}$	$\{42.8733, 0.9891\}$
$\hat{F}_8$	$\{1.2052, 0.2030\}$	$\{6.3626, 1.1693\}$
$\hat{F}_9$	$\{0.8261 \pm 0.6577i\}$	$\{39.2800, 0.7051\}$

Table 4.7: Relative error for $\hat{F}_7, \hat{F}_8, \hat{F}_9$

$\hat{F}_i$	Local Minimizers		Global
	$E_i^{S_1}$	$E_i^{S_2}$	Minimizer
$\hat{F}_7$	0.0546	0.0563	0.0546
$\hat{F}_8$	0.3271	0.3370	0.3271
$\hat{F}_9$	0.2998	0.2676	0.2676

Next, we repeat our approach to obtain second order optimal approximations for some linear systems used in the literature, given below,

- F_4 given in state space representation $F_4(s) \stackrel{s}{=} (A, b, c, 0)$, with state matrices given in (4.22)
- F_5 given by (4.23).
- F_{10} in [53], given by,

$$F_{10}(s) = \frac{41s^2 + 50s + 140}{(s^2 + s + 1)(s^2 + 10s + 100)}.$$

Table 4.8: Fixed Points for $\hat{F}_4, \hat{F}_5, \hat{F}_{10}$

Fixed Points					
	S_1	S_2	S_3	S_4	S_5
$\hat{F}_4$	$\{2.511, 1.099\}$	-	-	-	-
$\hat{F}_5$	$\{4.193, 1.153\}$	-	-	-	-
$\hat{F}_{10}$	$\{0.611, 0.323\}$	$\{4.952, 0.583\}$	$\{73.664, 0.897\}$	$\{27.206, 3.841\}$	$\{8.868 \pm 5.258i\}$

Table 4.9: Relative error for $\hat{F}_4, \hat{F}_5, \hat{F}_{10}$

Model	Relative Error for each Local Minimizer					Global
$\hat{F}_j$	$E_j^{S_1}$	$E_j^{S_2}$	$E_j^{S_3}$	$E_j^{S_4}$	$E_j^{S_5}$	Minimizer
$\hat{F}_4$	0.0393	-	-	-	-	0.0393
$\hat{F}_5$	0.2443	-	-	-	-	0.2443
$\hat{F}_{10}$	0.5434	0.5078	0.5336	0.5883	0.5840	0.5078

Similarly, we follow the approach presented in this section to construct second order optimal approximations $\hat{F}_4, \hat{F}_5$ and $\hat{F}_{10}$. The fixed points and their corresponding relative errors are shown in Table 4.8 and Table 4.9 respectively. Note that, there is only one fixed point for models $\hat{F}_4$ and $\hat{F}_5$. For comparison to the approach presented in this work, we mention that the iterative rational Krylov (IRKA) method $S^{(i+1)} = \lambda(-A_m(S^{(i)}))$ converges to the fixed points for any initial choice of the starting interpolation point set. However, in the case of $\hat{F}_{10}$, where there are five fixed points, the convergence of the iterative method to the optimal points depends on the choice of the initial interpolation points set.

4.4 Conclusions

This chapter has presented algorithms to compute all first and second order approximations that satisfy the necessary conditions for minimization to the $\mathcal{H}_2$ optimal model reduction problem given by Lemma 2.5.1. Some useful expressions of the $\mathcal{H}_2$ norm of the error system has been derived that give useful information about $\mathcal{H}_2$ optimal solution and its connection with the approximate solution of Lyapunov and Sylvester equations. Later we will use this information to propose a restart scheme that updates the approximate solution of Lyapunov and Sylvester equations.

Chapter 5

Parameterization of Rational Krylov Approximations

The parameterization of rational interpolating approximation, as described in Section 2.6, can ensure stability of the reduced order model when the actual system is stable. The free parameter introduced in a rational interpolating approximation is such that it can take any value without altering the interpolation properties of the reduced order model. In this chapter this procedure is extended to derive conditions in order to preserve some other useful properties such as the minimum phase and bounded real characters, passivity and symmetry of the actual system in the rational interpolating approximation. A related approach to preserving model properties was reported in [10], but under the assumption that the interpolation is at one finite frequency. In the following, these assumptions are removed. Also the concept of parameterization is extended to the minimal residual approximations (which, in general, are not interpolating approximations) and the two-sided rational Arnoldi approximations in order to further improve the accuracy of the approximations.

5.1 Error Analysis

This section extends the applicability of the parameterization derived in [30] by providing error expressions for the interpolating approximations as well as carrying out a perturbation analysis. This is useful as a stopping condition and for subsequent robustness analysis.

5.1.1 Residual and forward errors

In this subsection, the residual error and the forward error expressions are derived for the parameterized interpolating function $F_m^L(s)$ defined in Theorem 2.2.

The residual error associated with the parameterized interpolating function $F_m^L(s)$ is given by,

$$\begin{aligned} R(F_m^L(s)) &= B - (sI_n - A)V_m F_m^L(s), \\ &= V_{m+1} S(F_m^L(s)), \end{aligned} \quad (5.1)$$

where

$$S(F_m^L(s)) = \left[\begin{array}{c} B_m - (sI - A_m)F_m^L(s) \\ D_{Vm} + C_{Vm}F_m^L(s) \end{array} \right] \stackrel{s}{=} \left[\begin{array}{c|c} \frac{A_m + LC_{Vm}}{-LC_{Vm}} & \frac{B_m + LD_{Vm}}{-LD_{Vm}} \\ \hline C_{Vm} & D_{Vm} \end{array} \right]. \quad (5.2)$$

A derivation for the expression in (5.1) can be obtained by using the Arnoldi-like equations (4.2) and the definitions of A_m^L , B_m^L and $F_m^L(s)$ in Theorem 2.2.

An alternative expression for the residual error can be obtained from (5.2) as

$$\begin{aligned} R(F_m^L(s)) &= V_{m+1} S(F_m^L(s)) \\ &= \left[\begin{array}{cc} V_m & v_{m+1} \end{array} \right] \left(\left[\begin{array}{c} -LD_{Vm} \\ D_{Vm} \end{array} \right] + \left[\begin{array}{c} -LC_{Vm} \\ C_{Vm} \end{array} \right] (sI - A_m^L)^{-1} B_m^L \right) \\ &= v_{m+1}^L (D_{Vm} + C_{Vm}(sI - A_m^L)^{-1} B_m^L) \end{aligned}$$

where

$$v_{m+1}^L = (v_{m+1} - V_m L).$$

This shows that the residual belongs to the 1-dimensional subspace spanned by v_{m+1}^L .

A simple manipulation shows that the forward approximation error is given as

$$\begin{aligned} G(s) - G_m^L(s) &:= C(sI - A)^{-1} B - C V_m (sI - A_m^L)^{-1} B_m^L \\ &= C(sI - A)^{-1} v_{m+1}^L (D_{Vm} + C_{Vm}(sI - A_m^L)^{-1} B_m^L). \end{aligned}$$

Since the interpolating points are not eigenvalues of A or A_m^L , it follows that the in-

interpolating points are the zeros of the transfer function $D_{V_m} + C_{V_m}(sI - A_m^L)^{-1}B_m^L$, which is the frequency dependent part of the residual error. Note that these zeros are the eigenvalues of the matrix pencil

$$\begin{bmatrix} A_m + LC_{V_m} - \lambda I & B_m + LD_{V_m} \\ C_{V_m} & D_{V_m} \end{bmatrix}.$$

Expressing the matrix pencil as

$$\begin{bmatrix} I & L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A_m - \lambda I & B_m \\ C_{V_m} & D_{V_m} \end{bmatrix}$$

shows that the eigenvalues, and therefore the interpolating points, are independent of L .

5.1.2 Perturbation analysis

The Arnoldi-like equations in (4.2) can be written as,

$$AV_m = [V_m \ v_{m+1}]I_{m+1} \begin{bmatrix} A_m \\ C_{V_m} \end{bmatrix} \quad B = [V_m \ v_{m+1}]I_{m+1} \begin{bmatrix} B_m \\ D_{V_m} \end{bmatrix} \quad (5.3)$$

where I_{m+1} is the $(m+1) \times (m+1)$ identity matrix and can be expressed as

$$I_{m+1} = \begin{bmatrix} I_m & -L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I_m & L \\ 0 & 1 \end{bmatrix} \quad (5.4)$$

Using this in (5.3) we have the Arnoldi-like equations,

$$AV_m = V_m A_m^L + v_{m+1}^L C_{V_m}, \quad B = V_m B_m^L + v_{m+1}^L D_{V_m}, \quad (5.5)$$

in terms of the parameterized interpolating approximations. A simple manipulation now gives

$$(A - \Delta_A^L)V_m = V_m A_m^L, \quad B - \Delta_B^L = V_m B_m^L$$

where

$$\begin{bmatrix} \Delta_A^L & \Delta_B^L \end{bmatrix} = v_{m+1}^L \begin{bmatrix} C_{V_m} & D_{V_m} \end{bmatrix} \begin{bmatrix} V_m^T & 0 \\ 0 & 1 \end{bmatrix}$$

is a rank 1 perturbation with

$$\left\| \begin{bmatrix} \Delta_A^L & \Delta_B^L \end{bmatrix} \right\| = \sqrt{1 + \|L\|^2} \left\| \begin{bmatrix} C_{V_m} & D_{V_m} \end{bmatrix} \right\|.$$

It follows that the reduced order interpolating approximation $V_m F_m^L(s)$ is equal to the large scale perturbed system $F_\Delta^L(s) = (sI - (A - \Delta_A^L))^{-1} (B - \Delta_B^L)$, which generalizes the results in [59, 60], derived for the standard Arnoldi algorithm, to the rational case. Note that $\left\| \begin{bmatrix} \Delta_A^L & \Delta_B^L \end{bmatrix} \right\|$ is minimized by $L = 0$.

5.2 Choice of the Parameter L to Preserve Model Properties

It is shown in Section 2.6 that the parameterization developed in Theorem 2.2 can preserve stability of the actual system in the approximate system [30]. This work extends the applicability of the parameterization derived in [30] by considering other choices of L that preserve some other desirable properties such as minimum-phase character, passivity and bounded realness as well as symmetry where appropriate. At the end of the section an algorithm is given to describe the computation procedure of L .

5.2.1 Zero assignment/minimum-phase approximation

If $G(s) = D + CF(s)$ is minimum-phase, that is, if all the (finite) zeros of $G(s)$ are in the open left half of the complex plane, then an interesting question is whether we can find an L such that $G_m^L(s) := D + CV_m F_m^L(s)$ is also minimum-phase. The next result affirms this, derives necessary and sufficient conditions for the existence of L such that $G_m^L(s)$ is minimum-phase in the case that $D \neq 0$ and gives the extent to which we can assign the zeros of $G_m^L(s)$ through choice for L .

Proposition 5.1. *Let all variables be as defined in Proposition 2.2 and let $G(s) = D + C(sI - A)^{-1}B$ and $G_m^L(s) = D + C_m(sI - A_m^L)^{-1}B_m^L$ where $C_m = CV_m$. Assume that $D \neq 0$. Then:*

1. *The following are equivalent:*

(a) *There exists L such that $G_m^L(s)$ is minimum-phase.*

(b) The pair $(A_m - B_m D^{-1} C_m, C_{Vm} - D_{Vm} D^{-1} C_m)$ is detectable.

(c) There exist $0 \prec X_m = X_m^T \in \mathcal{R}^{m \times m}$ and $Y_m \in \mathcal{R}^{m \times 1}$ such that the LMI

$$\begin{aligned} (A_m - B_m D^{-1} C_m)^T X_m + X_m (A_m - B_m D^{-1} C_m) + Y_m (C_{Vm} - D_{Vm} D^{-1} C_m) \\ + (C_{Vm} - D_{Vm} D^{-1} C_m)^T Y_m^T \prec 0 \end{aligned} \quad (5.6)$$

is satisfied, in which case $L = X_m^{-1} Y_m$.

2. If $G(s)$ is minimum-phase then there exists L such that $G_m^L(s)$ is minimum-phase.
3. All eigenvalues of $A_m - B_m D^{-1} C_m$ (equivalently, the zeros of $G_m^{L=0}(s)$) that are not eigenvalues of $A - B D^{-1} C$ (equivalently, zeros of $G(s)$) are observable through $C_{Vm} - D_{Vm} D^{-1} C_m$. In particular, if $A - B D^{-1} C$ is stable (equivalently, if $G(s)$ is minimum-phase) then the pair $(A_m - B_m D^{-1} C_m, C_{Vm} - D_{Vm} D^{-1} C_m)$ is detectable.

Proof. 1. Let $\hat{A}_m^L := A_m^L - B_m^L D^{-1} C_m = (A_m - B_m D^{-1} C_m) + L(C_{Vm} - D_{Vm} D^{-1} C_m)$.

(a) $\leftrightarrow$ (b) This follows since the zeros of $G_m^L(s)$ are the eigenvalues of $\hat{A}_m^L$.

(a) $\leftrightarrow$ (c) $G_m^L(s)$ is minimum-phase, or equivalently, $\hat{A}_m^L$ is stable, if and only if there exists $0 \prec X_m = X_m^T \in \mathcal{R}^{m \times m}$ such that $(\hat{A}_m^L)^T X_m + X_m (\hat{A}_m^L) \prec 0$, or equivalently,

$$\begin{aligned} (A_m - B_m D^{-1} C_m)^T X_m + X_m (A_m - B_m D^{-1} C_m) + X_m L (C_{Vm} - D_{Vm} D^{-1} C_m) \\ + (C_{Vm} - D_{Vm} D^{-1} C_m)^T (X_m L)^T \prec 0. \end{aligned} \quad (5.7)$$

Setting $Y_m = X_m L$ proves the result.

2. If $G(s)$ is minimum-phase so that $A - B D^{-1} C$ is stable, then there exists $X = X^T \succ 0$ such that

$$(A - B D^{-1} C)^T X + X (A - B D^{-1} C) \prec 0.$$

Then

$$V_m^T ((A - B D^{-1} C)^T X + X (A - B D^{-1} C)) V_m \prec 0$$

since V_m has full column rank. Substituting the Arnoldi equations (4.2) and rearranging gives (5.7) with $X_m := V_m^T X V_m \in \mathcal{R}^{m \times m} = X_m^T \succ 0$ and $L := X_m^{-1} V_m^T X v_{m+1}$. It follows from Part 1 that $G_m^L(s)$ is minimum-phase.

3. Let $\lambda \in \mathcal{C}$ be an unobservable mode of the pair $(A_m - B_m D^{-1} C_m, C_{V_m} - D_{V_m} D^{-1} C_m)$ so that $(A_m - B_m D^{-1} C_m)x_m = \lambda x_m$ and $(C_{V_m} - D_{V_m} D^{-1} C_m)x_m = 0$ for some $x_m \neq 0$. Post-multiplying the first expression of the Arnoldi equations (4.2) by x_m and carrying out a manipulation gives $(A - B D^{-1} C)x = \lambda x$ where $x = V_m x_m \neq 0$. Thus λ is an eigenvalue of $A - B D^{-1} C$. If $A - B D^{-1} C$ is stable, then all eigenvalues of $A_m - B_m D^{-1} C_m$ in the closed right half of the complex plane are observable through $C_{V_m} - D_{V_m} D^{-1} C_m$ so the pair $(A_m - B_m D^{-1} C_m, C_{V_m} - D_{V_m} D^{-1} C_m)$ is detectable.

□

It follows that Part (3) of Proposition 5.1 gives us the freedom to use L to assign all the zeros of $G_m^L(s)$ which are not zeros of $G(s)$ anywhere in the complex plane. Similar to the eigenvalue assignment case in the previous section, this is acceptable since any zeros of $G_m(s)$ which are zeros of $G(s)$ do not need to be changed.

When $D = 0$, then $G_m^L(s)$ is strictly proper and so has at least one zero at ∞ and the finite zeros of $G_m^L(s)$ are the finite eigenvalues of the matrix pencil

$$\begin{bmatrix} A_m + L C_{V_m} - \lambda I & B_m + L D_{V_m} \\ C_m & 0 \end{bmatrix}.$$

In this case, the zero assignment problem is complicated by the (possibly multiple) zeros at infinity and the analysis becomes more intricate and falls outside the scope of this work. However, the following result shows that, under a mild assumption, L can be used to assign up to m zeros of $G_m^L(s)$ at ∞ .

Proposition 5.2. *Let all variables be as defined in Proposition 5.1 and assume that $D = 0$. Let $1 < \zeta_\infty \leq m$ be given. Then:*

1. *The following are equivalent:*

(a) *There exists L such that $G_m^L(s)$ has ζ_∞ zeros at ∞ .*

(b) There exists L such that

$$\begin{bmatrix} C_m \\ \vdots \\ C_m (A_m^L)^{\zeta_\infty - 2} \end{bmatrix} B_m^L = 0. \quad (5.8)$$

2. Assume that $D_{V_m} \neq 0$ (that is, $\infty \notin S$; see Lemma 2.4.2) and define $\hat{A}_m = A_m - B_m D_{V_m}^{-1} C_{V_m}$. Then there exists L such that $G_m^L(s)$ has ζ_∞ zeros at ∞ . Furthermore, all such L are the solutions of the following linear systems of equations:

$$\begin{bmatrix} C_m \\ \vdots \\ C_m \hat{A}_m^{\zeta_\infty - 2} \end{bmatrix} (B_m + L D_{V_m}) = 0. \quad (5.9)$$

Proof.

1. This is immediate
2. A manipulation verifies that

$$C_m (A_m^L)^i B_m^L = 0, i = 0, \dots, j-1 \implies C_m (A_m^L)^j B_m^L = C_m \hat{A}_m^j B_m^L$$

and the result follows from Part 1.

□

Note that when $D_{V_m} = 0$, then the solution of the nonlinear system of equation for L in (5.8) is, in general, intractable. When $D_{V_m} \neq 0$, then the system of equations in (5.9) is linear in L and can be easily obtained.

5.2.2 Preserving bounded realness

The next result derives necessary and sufficient LMI conditions for the existence of L such that $G_m^L(s)$ is bounded real and shows that if $G(s)$ is bounded real then there exists L such that $G_m^L(s)$ is bounded real.

Proposition 5.3. *Let all variables be as defined in Proposition 5.1 and let $\gamma > 0$ be given. There exists L such that $G_m^L(s)$ is stable and $\|G_m^L\|_\infty < \gamma$ if and only if there exist $0 \prec X_m = X_m^T \in \mathcal{R}^{m \times m}$ and $Y_m \in \mathcal{R}^{m \times 1}$ such that*

$$\begin{bmatrix} A_m^T X_m + X_m A_m + C_{V_m}^T Y_m^T + Y_m C_{V_m} & X_m B_m + Y_m D_{V_m} & C_m^T \\ B_m^T X_m + D_{V_m}^T Y_m^T & -\gamma I & D^T \\ C_m & D & -\gamma \end{bmatrix} \prec 0, \quad (5.10)$$

in which case $L = X_m^{-1} Y_m$. Furthermore, if $G(s)$ is stable and $\|G\|_\infty < \gamma$ then there exists L such that $G_m^L(s)$ is stable and $\|G_m^L\|_\infty < \gamma$.

Proof. $G_m^L(s)$ is stable and $\|G_m^L\|_\infty < \gamma$ if and only if there exists an $X_m = X_m^T \succ 0$ such that [18],

$$\begin{bmatrix} (A_m^T + C_{V_m}^T L^T) X_m + X_m (A_m + L C_{V_m}) & X_m (B_m + L D_{V_m}) & C_m^T \\ (B_m^T + D_{V_m}^T L^T) X_m & -\gamma I & D^T \\ C_m & D & -\gamma \end{bmatrix} \prec 0.$$

Setting $Y_m = X_m L$ proves the first part. If $G(s)$ is stable and $\|G\|_\infty < \gamma$ then there exists an $X = X^T \succ 0$ such that

$$\begin{bmatrix} A^T X + X A & X B & C^T \\ B^T X & -\gamma I & D^T \\ C & D & -\gamma I \end{bmatrix} \prec 0.$$

It follows that

$$\begin{bmatrix} V_m^T & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} A^T X + X A & X B & C^T \\ B^T X & -\gamma I & D^T \\ C & D & -\gamma I \end{bmatrix} \begin{bmatrix} V_m & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \prec 0 \quad (5.11)$$

since V_m has full column rank. Substituting the Arnoldi equations in (5.11) and rear-

ranging,

$$\begin{bmatrix} A_m^T V_m^T X V_m + V_m^T X V_m A_m + C_{V_m}^T v_{m+1}^T X V_m + V_m^T X v_{m+1} C_{V_m} & \star & \star \\ B_m^T V_m^T X V_m + D_{V_m}^T v_{m+1}^T X V_m & -\gamma I & \star \\ C_m & D & -\gamma I \end{bmatrix} \prec 0$$

where $\star$ denotes terms readily inferred from symmetry. Define $X_m = X_m^T = V_m^T X V_m \succ 0$ and $L = X_m^{-1} V_m^T X v_{m+1}$. Then

$$\begin{bmatrix} (A_m^T + C_{V_m}^T L^T) X_m + X_m (A_m + L C_{V_m}) & X_m (B_m + L D_{V_m}) & C_m^T \\ (B_m^T + D_{V_m}^T L^T) X_m & -\gamma I & D^T \\ C_m & D & -\gamma I \end{bmatrix} \prec 0$$

and so $G_m^L(s)$ is stable and $\|G_m^L\|_\infty < \gamma$. $\square$

Note that preserving bounded realness is of interest when $G(s)$ represents a high order closed-loop transfer function and it is required to find a low order $\mathcal{H}_\infty$ controller.

5.2.3 Preserving passivity

Suppose that $G(s)$ is passive. An interesting question is whether we can find an interpolating function $G_m^L(s)$ that preserves the passivity property of $G(s)$. Recall that $G(s)$ is passive if and only if it is positive real, equivalently, if and only if there exists $P \succ 0$ such that [100],

$$\begin{bmatrix} A^T P + P A & P B - C^T \\ B^T P - C & -(D + D^T) \end{bmatrix} \preceq 0.$$

The next result derives necessary and sufficient LMI conditions for the existence of L such that $G_m^L(s)$ is passive and shows that if $G(s)$ is passive there always exists a passive interpolating approximation.

Proposition 5.4. *Let all variables be as defined in Proposition 5.3. There exists L such that $G_m^L(s)$ is passive if and only if there exist $0 \prec X_m = X_m^T \in \mathcal{R}^{m \times m}$ and $Y_m \in \mathcal{R}^{m \times 1}$*

such that

$$\begin{bmatrix} A_m^T X_m + X_m A_m + Y_m C_{V_m} + C_{V_m}^T Y_m^T & X_m B_m + Y_m D_{V_m} - C_m^T \\ (X_m B_m + Y_m D_{V_m} - C_m^T)^T & -(D + D^T) \end{bmatrix} \preceq 0, \quad (5.12)$$

in which case $L = X_m^{-1} Y_m$. Furthermore, if $G(s)$ is passive then there exists L such that $G_m^L(s)$ is passive.

Proof. $G_m^L(s)$ is passive if and only if there exists an $X_m = X_m^T \succ 0$ such that [100],

$$\begin{bmatrix} (A_m^T + C_{V_m}^T L^T) X_m + X_m (A_m + L C_{V_m}) & X_m (B_m + L D_{V_m}) - C_m^T \\ (B_m^T + D_{V_m}^T L^T) X_m^T - C_m & -(D + D^T) \end{bmatrix} \preceq 0.$$

Setting $Y_m = X_m L$ proves the first part. If $G(s)$ is passive, there exists a $X = X^T \succ 0$ such that

$$\begin{bmatrix} A^T X + X A & X B - C^T \\ B^T X - C & -(D + D^T) \end{bmatrix} \preceq 0.$$

It follows that

$$\begin{bmatrix} V_m^T & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A^T X + X A & X B - C^T \\ B^T X - C & -(D + D^T) \end{bmatrix} \begin{bmatrix} V_m & 0 \\ 0 & I \end{bmatrix} \preceq 0. \quad (5.13)$$

Substituting the Arnoldi equations in (5.13) and rearranging we have that

$$\begin{bmatrix} A_m^T V_m^T X V_m + V_m^T X V_m A_m + C_{V_m}^T v_{m+1}^T X V_m + V_m^T X v_{m+1} C_{V_m} & \star \\ B_m^T V_m^T X V_m + D_{V_m}^T v_{m+1}^T X V_m - C_m & -(D + D^T) \end{bmatrix} \preceq 0.$$

Define $X_m = X_m^T = V_m^T X V_m \succ 0$ and $L = X_m^{-1} V_m^T X v_{m+1}$. Then

$$\begin{bmatrix} (A_m^T + C_{V_m}^T L^T) X_m + X_m (A_m + L C_{V_m}) & X_m (B_m + L D_{V_m}) - C_m^T \\ (B_m^T + D_{V_m}^T L^T) X_m^T - C_m & -(D + D^T) \end{bmatrix} \preceq 0,$$

and so $G_m^L(s)$ is passive. □

5.2.4 Preserving symmetry

Suppose that A is symmetric. The next result gives a parameterization of interpolating approximations with a symmetric A -matrix.

Proposition 5.5. *Let all variables be as defined in Proposition 5.4. Let $G(s) = D + C(sI - A)^{-1}B$ and $G_m^L(s) = D + C_m(sI - A_m^L)^{-1}B_m^L$ and assume that A is symmetric. Then A_m^L is symmetric if and only if $L = \ell C_{V_m}^T$ for some $\ell \in \mathcal{R}$.*

Proof. Since $A = A^T$ and $A_m = V_m^T A V_m$ then $A_m = A_m^T$. So $A_m^L = (A_m^L)^T$ if and only if $LC_{V_m} = C_{V_m}^T L^T$. Since L and $C_{V_m}^T$ are column vectors the result follows. $\square$

5.2.5 Preserving multiple properties

In Section 2.6 and Sections 5.2.1-5.2.3 above, we derived necessary and sufficient conditions on the free parameter L to preserve stability, minimum-phase character, bounded realness and passivity, respectively. Each of these conditions requires the existence of $0 \prec X_m = X_m^T \in \mathcal{R}^{m \times m}$ and $Y_m \in \mathcal{R}^{m \times 1}$ such that the appropriate LMI (2.79), (5.6), (5.10) or (5.12) is satisfied, in which case $L = X_m^{-1}Y_m$. These LMIs involve reduced order matrices only and the computational cost for their solution is independent of the large dimensions of the original model; thus, they can be solved efficiently. We briefly describe the computation procedure in Algorithm 5.1.

Algorithm 5.1 Algorithm to preserve model properties

- 1: Inputs: interpolating points $\mathcal{S} = \{s_1, \dots, s_K\} \subset \mathcal{C}$ with multiplicities $\mathcal{M} = \{m_{s_1}, \dots, m_{s_K}\}$, system matrices $A \in \mathcal{R}^{n \times n}$, $B, C^T \in \mathcal{R}^n$, and $d \in \mathcal{R}, \gamma > 0$ and tolerance level $\epsilon > 0$.
 - 2: Compute V_m , A_m , B_m and C_m by using the rational Arnoldi algorithm 2.4 for given A , B , S , and ϵ
 - 3: Using Lemma 2.4.2 compute V_{m+1} , C_{V_m} and D_{V_m}
 - 4: For preserving stability, minimum-phase character, bounded realness or passivity, solve the LMIs in (2.79), (5.6), (5.10) or (5.12), respectively, to obtain X_m and Y_m and solve $Y_m X_m = L$ to obtain L .
 - 5: Compute the reduced order system $G_m^L = C_m(sI - A_m^L)^{-1}B_m^L + D$, where $C_m = C V_m$, $A_m^L = A_m + L C_{V_m}$ and $B_m^L = B_m + L D_{V_m}$
-

Suppose that it is required to preserve more than one property, say $G(s)$ is stable and minimum-phase, with $D \neq 0$, and it is required that $G_m^L(s)$ be also stable and minimum-phase. Then it follows from Propositions 2.2 and 5.1 that a necessary and sufficient

condition is the existence of $0 \prec X_S = X_S^T \in \mathcal{R}^{m \times m}$, $0 \prec X_M = X_M^T \in \mathcal{R}^{m \times m}$, $Y_S \in \mathcal{R}^{m \times 1}$ and $Y_M \in \mathcal{R}^{m \times 1}$ such that

$$\begin{aligned} A_m^T X_S + X_S A_m + Y_S C_{V_m} + C_{V_m}^T Y_S^T &\prec 0 \\ \hat{A}_m^T X_M + X_M \hat{A}_m + Y_M \hat{C}_{V_m} + \hat{C}_{V_m}^T Y_M^T &\prec 0 \\ X_S^{-1} Y_S = X_M^{-1} Y_M &= L \end{aligned}$$

where we have defined $\hat{A}_m := A_m - B_m D^{-1} C_m$ and $\hat{C}_{V_m} := C_{V_m} - D_{V_m} D^{-1} C_m$. This is a nonlinear feasibility problem because of the requirement $X_S^{-1} Y_S = X_M^{-1} Y_M$. To obtain a linear feasibility problem, we impose that $X_m = X_S$ and $Y_M = Y_S$, in which case we settle for sufficient, rather than necessary and sufficient, conditions for the existence of L .

5.3 Choice of L to enhance the accuracy of interpolating approximations

In the previous section we were concerned with the choice of L to enhance the properties of the interpolating approximation. In this section, we suggest techniques for choosing L so that we enhance the accuracy of the interpolating approximation. In particular, we consider the choice of L to minimize the $\mathcal{H}_\infty$ -norm of the residual error, an $\mathcal{H}_2$ -norm optimal modal truncation-like approximation, and a partial two-sided Arnoldi approximation.

5.3.1 Interpolating minimum residual solution

In this section we consider the choice of L to reduce the $\mathcal{H}_\infty$ -norm of the residual error. First, we define and obtain the minimum residual solution (not necessarily interpolating) which minimizes the $\mathcal{H}_\infty$ -norm of the residual error. This is useful as a comparison with our subsequent (suboptimal) solution for the interpolating minimum residual approximation.

The following proposition defines and obtains the minimum residual solution and generalizes the result in [57], derived for the standard Arnoldi algorithm, to the rational case.

Proposition 5.6. *Let all variables be as defined in Lemma 2.4.3 and let $\gamma > 0$ be given.*

Then $\tilde{A}$ is stable and $\|R(F_m(s))\|_{\mathcal{H}_\infty} < \gamma$ if and only if there exist feasible solutions $\hat{A} \in \mathcal{R}^{m \times m}$, $\tilde{B} \in \mathcal{R}^{m \times 1}$ and $\mathcal{R}^{m \times m} \ni Y = Y^T \succ 0$ such that the LMI

$$\begin{bmatrix} \hat{A} + \hat{A}^T & Y A_m^T - \hat{A}^T & Y C_{V_m}^T & \tilde{B} \\ A_m Y - \hat{A} & -\gamma I & 0 & B_m - \tilde{B} \\ C_{V_m} Y & 0 & -\gamma & D_{V_m} \\ \tilde{B}^T & B_m^T - \tilde{B}^T & D_{V_m}^T & -\gamma \end{bmatrix} \prec 0$$

is satisfied, in which case $\tilde{A} = \hat{A}Y^{-1}$.

Proof. It follows from the expression for the residual error in Lemma 2.4.3, the fact that V_{m+1} is orthogonal and the bounded real lemma [18], that $\tilde{A}$ is stable and $\|R(F_m(s))\|_{\mathcal{H}_\infty} < \gamma$ if and only if there exists $Y = Y^T \succ 0$ such that

$$\begin{bmatrix} \tilde{A}Y + Y\tilde{A}^T & Y(A_m - \tilde{A})^T & Y C_{V_m}^T & \tilde{B} \\ (A_m - \tilde{A})Y & -\gamma I & 0 & B_m - \tilde{B} \\ C_{V_m} Y & 0 & -\gamma & D_{V_m} \\ \tilde{B}^T & B_m^T - \tilde{B}^T & D_{V_m}^T & -\gamma \end{bmatrix} \prec 0.$$

The result immediately follows by the substitution $\hat{A} = \tilde{A}Y$. $\square$

The minimum residual approximation developed above is attractive in the sense that it is stable and minimizes the $\mathcal{H}_\infty$ -norm of the residual error. However, it suffers from the fact that it does not in general interpolate $G(s)$. The next result gives a procedure for finding the interpolating function with the least residual $\mathcal{H}_\infty$ -norm.

Proposition 5.7. *Let all variables be as defined in Theorem 2.2 and let $\gamma > 0$ be given. Then there exists $L \in \mathcal{R}^{m \times 1}$ such that $G_m^L(s) := C_m F_m^L(s)$, where $F_m^L(s) = (sI - A_m^L)^{-1} B_m^L$, is stable and $\|R(F_m^L(s))\|_{\mathcal{H}_\infty} < \gamma$ if and only if there exists $X_m = X_m^T \succ 0$*

such that the bilinear matrix inequality

$$\begin{bmatrix} (A_m + LC_{V_m})^T X_m + X_m (A_m + LC_{V_m}) & X_m (B_m + LD_{V_m}) & -C_{V_m}^T L^T & C_{V_m}^T \\ (B_m + LD_{V_m})^T X_m & -\gamma & -D_{V_m}^T L^T & D_{V_m}^T \\ -LC_{V_m} & -LD_{V_m} & -\gamma I & 0 \\ C_{V_m} & D_{V_m} & 0 & -\gamma \end{bmatrix} \prec 0 \quad (5.14)$$

is satisfied.

Proof. Substituting $\tilde{A} = A_m + LC_{V_m}$ and $\tilde{b} = B_m + LD_{V_m}$ into the expression for the residual error in (2.60) and (2.61) gives $R(F_m^L(s)) = V_{m+1} S(F_m^L(s))$ where

$$S(F_m^L(s)) \stackrel{s}{=} \left[\begin{array}{c|c} A_m + LC_{V_m} & B_m + LD_{V_m} \\ \hline -LC_{V_m} & -LD_{V_m} \\ C_{V_m} & D_{V_m} \end{array} \right].$$

The result now follows from the bounded real lemma. $\square$

Note that while calculating the optimal γ is in general intractable, the fact that (5.14) is bilinear (linear in L for fixed X_m and linear in X_m for fixed L) results in many excellent algorithms for obtaining suboptimal solutions [94].

5.3.2 An $\mathcal{H}_2$ -norm optimal modal truncation-like approximation

In this section we present a modal truncation-like approximation via the rational Arnoldi method. Using the parameterization method developed in Section 2.6 and by a proper selection of the free parameter L , the reduced order approximation is optimal in the $\mathcal{H}_2$ -norm, in a sense defined below, and preserves stability.

Recently the authors in [16, 48, 49, 51] have derived reduced order interpolating approximations which are optimal in the $\mathcal{H}_2$ -norm by satisfying the Meir-Luenberger conditions [70]. The Meir-Luenberger conditions state that for a given set of poles of the reduced order system, denoted by $\lambda(A_m)$, then $\|G - G_m\|_2$ is minimized if and only if $G(s_i) = G_m(s_i)$ for all $s_i \in \lambda(-A_m)$. However the poles of G_m are not known a priori and the authors propose in their work to run the Krylov methods successively until

convergence of the interpolation points to the images of the eigenvalues of the reduced system.

We alternatively propose selecting a subset $S = \{s_1, \dots, s_m\}$ of the poles of G , e.g., the most dominant, interpolating at the images $\{-s_1, \dots, -s_m\}$ and using L to assign the eigenvalues of A_m^L at S . As mentioned earlier, since the pencil (A_m, C_{V_m}) is detectable (Proposition 2.2), eigenvalue placement of $A_m^L = A_m + LC_{V_m}$ can be obtained via a proper selection of L as described in [63]. Note that the computation of a subset of eigenvalues of A can be efficiently carried out using Krylov subspace techniques [80, 86]. The right choice of the interpolation points is an open research problem [33]. The suggestion to interpolate at the images of the most dominant eigenvalues of G , does not provide the best approximation amongst other models interpolating at alternative interpolation points. However, this should give an excellent approximation compared to any other approximation having the same poles, e.g. compared to the standard modal truncation.

5.3.3 Two-sided Arnoldi

One way of improving the accuracy of the approximation is to use W_m to interpolate $G(s)$ at an additional set of points, say by using the rational Arnoldi algorithm with A and C . Of course, if we interpolate at m points (counting multiplicities), then W_m is fixed and we have no degrees of freedom in the interpolating function. The problem is that the approximation may not preserve some system properties such as stability. Here, we suggest a procedure for interpolating $G(s)$ at p additional points. If $p < m$, we derive a parameterization of such interpolating functions.

Let $\tilde{S} = \{\sigma_1, \dots, \sigma_{\tilde{K}}\} \subset \mathcal{C}$ be a set of $\tilde{K}$ distinct interpolating points, with associated multiplicities $p_{\sigma_1}, \dots, p_{\sigma_{\tilde{K}}}$, respectively, and assume that $\tilde{S}$ is closed under conjugation and that none of the interpolating points are eigenvalues of A . Define $\mathcal{U}_p$ as the union of the Krylov subspaces

$$\mathcal{U}_p = \bigcup_{\substack{k=1 \\ \sigma_k \neq \infty}}^{\tilde{K}} \mathcal{K}_{p_{\sigma_k}}((\sigma_k I - A^T)^{-1}, (\sigma_k I - A^T)^{-1} C^T) \bigcup \mathcal{K}_{p_\infty}(A^T, C^T)$$

where $p = \sum_{k=1}^{\tilde{K}} \alpha_k p_{\sigma_k}$ and where $\alpha_k = 1$ if σ_k is real or infinite and $\alpha_k = 2$ if complex. Let $U_p \in \mathcal{R}^{n \times p}$ be an orthogonal basis of $\mathcal{U}_p$ (U_p may be constructed using the rational Arnoldi Algorithm 2.4) and assume that $p < m$.

Writing $G(s) = \tilde{F}(s)B$ where $\tilde{F}(s) = C(sI - A)^{-1}$, then $\tilde{F}_m^{W_m}(s) = CV_m(sI - W_m^T A V_m)^{-1} W_m^T$

matches the first p_{σ_k} moments of $\tilde{F}(s)$ at σ_k , for $k = 1, \dots, \tilde{K}$, that is,

$$\tilde{F}^{(i)}(\sigma_k) = (\tilde{F}_m^{W_m})^{(i)}(\sigma_k), \quad i = 0, \dots, p_{\sigma_k} - 1, \quad k = 1, \dots, \tilde{K} \quad (5.15)$$

provided that $W_m^T V_m = I$ and $U_p \subseteq \text{colsp}(W_m)$ (assuming that none of the interpolating points are eigenvalues of $W_m^T A V_m$). Now, $W_m^T V_m = I$ if and only if

$$W_m = V_m + v_{m+1} L^T + V_\perp L_\perp^T \quad (5.16)$$

for some L and $L_\perp$ and where $V_\perp$ is an orthogonal completion for V_{m+1} . Also, $U_p \subseteq \text{colsp}(W_m)$ if and only if

$$U_p = W_m B_{W_m}^T$$

for some full rank matrix $B_{W_m} \in \mathcal{R}^{p \times m}$. Thus we require

$$U_p = V_m B_{W_m}^T + v_{m+1} L^T B_{W_m}^T + V_\perp L_\perp^T B_{W_m}^T.$$

Since V_m and U_p are orthogonal, we require $B_{W_m} = U_p^T V_m$ and $B_{W_m} L = U_p^T v_{m+1} =: \hat{D}_0$. We assume that $U_p^T V_m$ is full rank, which is essentially equivalent to assuming no breakdown in a two-sided Arnoldi. Let $B_{W_m}^\dagger \in \mathcal{R}^{m \times p}$ be a right-inverse of B_{W_m} and let $B_{W_m}^\perp \in \mathcal{R}^{m \times (m-p)}$ span the null space of $B_{W_m}^T$ so that

$$B_{W_m} \begin{bmatrix} B_{W_m}^\dagger & B_{W_m}^\perp \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix}.$$

Thus the parameter L has the form

$$L = B_{W_m}^\dagger \hat{D}_0 + B_{W_m}^\perp \hat{D}$$

where $\hat{D} \in \mathcal{R}^{(m-p) \times 1}$ is a free parameter. It follows that all approximations of the form

$$G_m^{\hat{D}}(s) = C V_m (sI - A_m^{\hat{D}})^{-1} B_m^{\hat{D}},$$

where

$$A_m^{\hat{D}} = (A_m + B_{W_m}^\dagger \hat{D}_0 C V_m) + B_{W_m}^\perp \hat{D} C V_m, \quad B_m^{\hat{D}} = (B_m + B_{W_m}^\dagger \hat{D}_0 D V_m) + B_{W_m}^\perp \hat{D} D V_m$$

interpolate $G(s)$ at S (through V_m) and at $\tilde{S}$ (through U_p) [25].

It follows that we may use the degrees of freedom in $\hat{D}$ to ensure preservation of model properties such as stability, minimum-phase character, etc. Note, however, that the problem of selecting such a $\hat{D}$ becomes much more difficult. For example, if we want to preserve stability, we need to find $\hat{D}$ such that $A_m^{\hat{D}} = (A_m + B_{W_m}^\dagger \hat{D}_0 C_{V_m}) + B_{W_m}^\perp \hat{D} C_{V_m}$ is stable. This is equivalent to the problem of stabilization using static output feedback, which is known to be a hard problem. See [93] for a recent technique for solving this problem.

5.4 Numerical examples

In this section we present two numerical examples to illustrate our procedures.

5.4.1 Example 1

In this example we consider the RLC-system given in Figure 5.1, of order $n = 101$ (number of inductors is $n_L = 50$ and number of capacitors is $n_L + 1$) where the values of the resistors, capacitors and inductors may be selected randomly. We take the input signal to be the supplied voltage $u(t)$ while the output is the current $y(t)$. The states are the capacitor voltages $(x_1, x_3, \dots)$ and the inductor currents $(x_2, x_4, \dots)$. We took $R_1 = R_2 = 1$, $C_i = 0.1$ and $L_i = 1$ for all i . It is then easy to show that A is tridiagonal with the diagonal elements given as $A_{11} = -(R_1 C_1)^{-1} = -10$, $A_{nn} = -(R_2 C_{n_L+1})^{-1} = -10$, $A_{ii} = 0$, $i = 2, \dots, n-1$; the first super diagonal of A is given as $[C_1^{-1}, L_1^{-1}, C_2^{-1}, L_2^{-1}, \dots, C_{n_L}^{-1}, L_{n_L}^{-1}] = [10, 1, 10, 1, \dots, 10, 1]$ and the first sub-diagonal of A is given as $[-L_1^{-1}, -C_2^{-1}, -L_2^{-1}, -C_3^{-1}, \dots, -L_{n_L}^{-1}, -C_{n_L+1}^{-1}] = [-1, -10, -1, -10, \dots, -1, -10]$. Also, the last elements of b and C are given by $(R_2 C_{n_L+1})^{-1} = 10$ and $-R_2^{-1} = -1$, respectively, with all other elements of b and C being zeros. Finally, $d = R_2^{-1} = 1$.

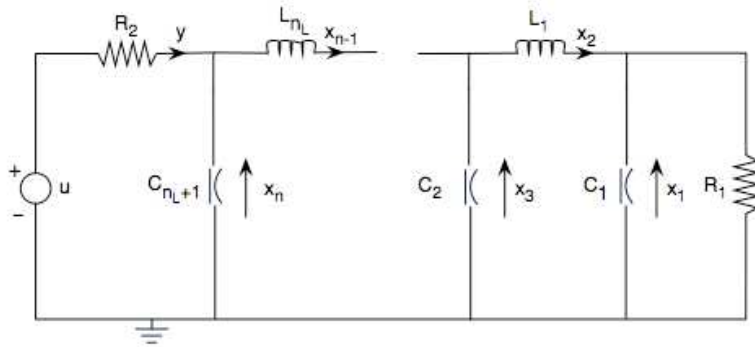


Figure 5.1: RLC circuit

It is not difficult to show that this system is stable, minimum-phase, passive and bounded real ($\|G\|_{\mathcal{H}_\infty} = 1$). Thus for each value of m we obtain an order m interpolating approximation $G_m^L(s)$ which is stable, minimum-phase, passive and positive real using the results of Section 5.2. We also use the interpolating minimum residual approach of Section 5.3.1 to enhance the performance. For our interpolation points, we use the one-sided adaptive rational Arnoldi algorithm (G_m) [31–33]. The adaptive rational Arnoldi procedure computes the interpolation points adaptively: at each iteration, the next interpolation points are computed to be the (complex) frequencies at which the $\mathcal{H}_\infty$ –norm of the residual error expression is achieved. The first two interpolating points are chosen to be at ∞ and 0.

For comparison purposes, we also give the results for the balanced truncation algorithm ($G_b(s)$) and the $\mathcal{H}_2$ model reduction method of [51] ($G_2(s)$). For the latter method, we chose a random set of initial interpolating points, then applied the Arnoldi algorithm to interpolate $G(s)$ and its derivative iteratively, where at each iteration we replace the interpolating points by the negatives of the poles of the resulting approximation. The algorithm converged quickly to a (local) minimum, although it is not possible to guarantee the resulting approximation is a global minimum.

Using the three model reduction approaches always resulted in reduced order models which satisfied all the requirements (stability, minimum-phase character and passivity). The following table gives the $\mathcal{H}_\infty$ and $\mathcal{H}_2$ norms for the approximation errors for a few values of m .

Table 5.1: Approximation accuracy for the RLC-system in Figure 5.1

m	$\ G - G_m^L\ _{\mathcal{H}_\infty}$	$\ G - G_2\ _{\mathcal{H}_\infty}$	$\ G - G_b\ _{\mathcal{H}_\infty}$	$\ G - G_m^L\ _{\mathcal{H}_2}$	$\ G - G_2\ _{\mathcal{H}_2}$	$\ G - G_b\ _{\mathcal{H}_2}$
4	0.53	0.47	0.52	0.43	0.28	0.41
8	0.53	0.49	0.52	0.41	0.28	0.41
12	0.51	0.49	0.52	0.36	0.28	0.41

Note that the given $G(s)$ is particularly difficult to approximate accurately as the step and frequency responses shown below demonstrate. In the figures, we also show the responses for our approximation ($G_m^L(s)$), the $\mathcal{H}_2$ approximation ($G_2(s)$) and the balanced truncation approximation ($G_b(s)$) for $m = 12$. For our approximation, the interpolating

points and the value of L were obtained as

$$S = \begin{bmatrix} \infty \\ 0 \\ \pm 0.4472i \\ \pm 0.4267i \\ \pm 0.5846i \\ \pm 0.6671i \\ \pm 0.2654i \end{bmatrix}, \quad L = \begin{bmatrix} -0.3882 \\ 0.0496 \\ -0.0650 \\ 0.1285 \\ -0.1059 \\ 0.2066 \\ 0.5648 \\ -0.0746 \\ -0.4274 \\ -0.0478 \\ -0.3589 \\ -0.0066 \end{bmatrix}$$

while for the $\mathcal{H}_2$ approximation, the interpolating points were

$$S = \begin{bmatrix} 8.0717 \pm 1.2469i \\ 3.6651 \pm 5.2193i \\ 1.3337 \pm 6.1622i \\ 0.3949 \pm 6.3056i \\ 0.0903 \pm 6.3218i \\ 0.0122 \pm 6.3239i \end{bmatrix}$$

Although Table 5.1 suggests that the $\mathcal{H}_2$ approximation $G_2(s)$ enjoys superior error bounds, the responses below show that it, together with the balanced truncation approximation, are unable to reproduce the step and frequency responses of $G(s)$, in particular, the low frequency and steady-state behavior of $G(s)$. This demonstrates that the $\mathcal{H}_2$ -norm is not an appropriate measure for approximating the model.

5.4.2 Example 2

As a second example, we use the stable model `heat-cont` [23]. This is a 200-state single-input single-output, stable, minimum-phase model with 67 zeros at ∞ and with symmetric A -matrix. It represents a semi-discretized model for a one-dimensional heat diffusion equation.

Figures 5.3 and 5.4 show the frequency responses for `heat-cont`, as well as the approxi-

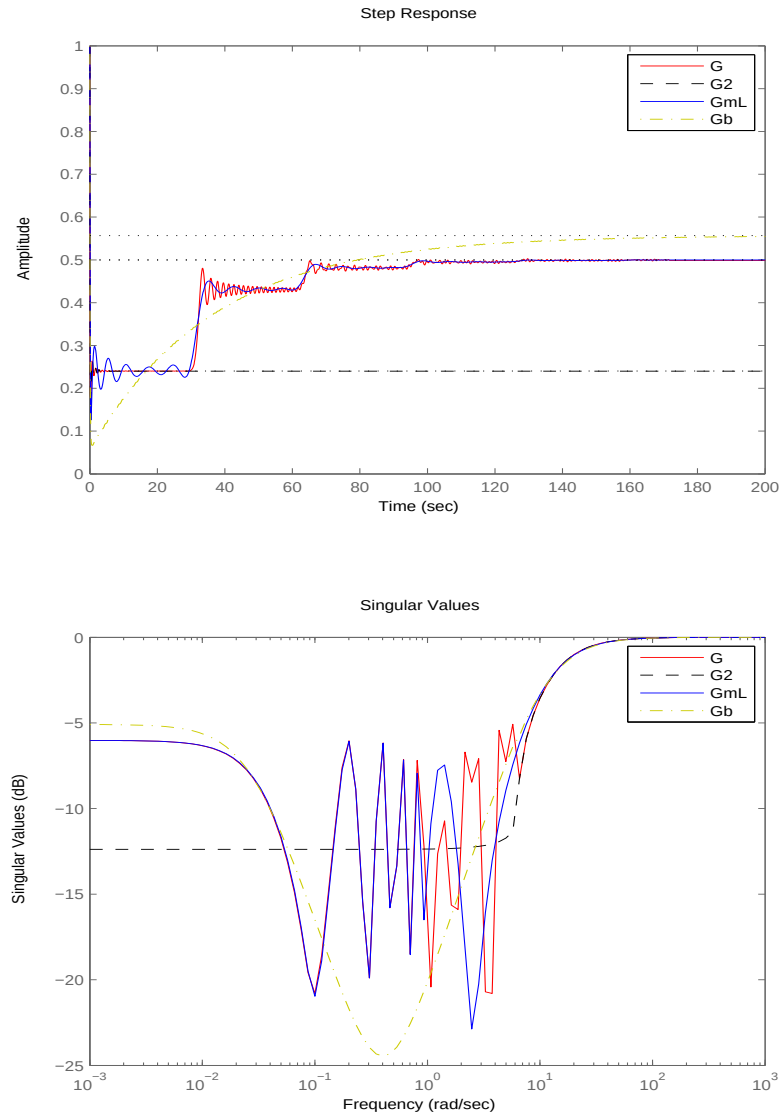


Figure 5.2: Step and frequency responses of the RLC-system in Figure 5.1 and its approximations for $m = 12$

mations $G_2(s)$, $G_m^L(s)$ and $G_m^l(s)$ for $m = 2$ and $m = 8$, respectively.

The approximation $G_2(s)$ uses the $\mathcal{H}_2$ model reduction method of [51]. The interpolating

points were obtained as

$$S = \begin{bmatrix} 0.0998 \\ 0.4112 \end{bmatrix} \quad (m = 2), \quad S = \begin{bmatrix} 8.6855 \pm 12.9920i \\ 6.3036 \pm 3.9163i \\ 2.8010 \\ 1.5421 \\ 0.3946 \\ 0.0987 \end{bmatrix} \quad (m = 8)$$

The resulting $G_2(s)$ was not minimum-phase and had no zeros at ∞ .

For $G_m^L(s)$, we used $-S$ defined above as our interpolating points and we chose L such that all the zeros of $G_m^L(s)$ are at ∞ using Proposition 5.2. Note that since the interpolating points are all finite, $D_{V_m} \neq 0$ from Lemma 2.4.2. We then used the remaining degree of freedom to ensure stability and minimize the dc error between $G_m^L(s)$ and $G(s)$. The values of L were obtained as

$$L = \begin{bmatrix} 0.0028 \\ -0.0035 \end{bmatrix} \quad (m = 2), \quad L = \begin{bmatrix} 0.3708 \\ -0.3820 \\ 0.2513 \\ 0.2435 \\ -0.1670 \\ -0.2185 \\ -0.0387 \\ -0.2572 \end{bmatrix} \quad (m = 8)$$

For $G_m^l(s)$, we also used $-S$ defined above as our interpolating points and we chose L such that A_m^L is symmetric using Proposition 5.5. We then used the remaining degree of freedom to ensure that $G_m^l(s)$ is stable, minimum-phase and minimizes the dc error between $G_m^l(s)$ and $G(s)$. The values of L were obtained as

$$L = \begin{bmatrix} -0.0004 \\ -0.0062 \end{bmatrix} \quad (m = 2), \quad L = \begin{bmatrix} 0.2122 \\ -0.0106 \\ 0.1345 \\ 0.1763 \\ -0.0478 \\ -0.0339 \\ -0.1156 \\ -0.2979 \end{bmatrix} \quad (m = 8)$$

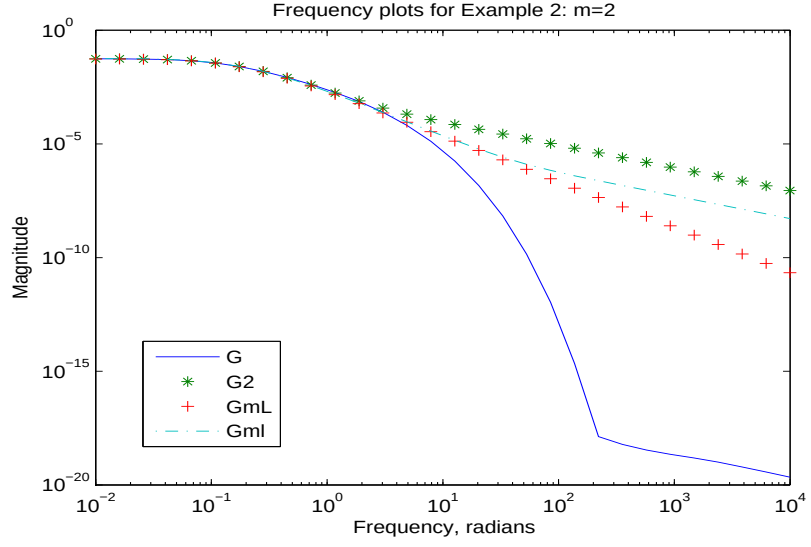


Figure 5.3: Frequency responses for heat-cont and its approximations for $m = 2$

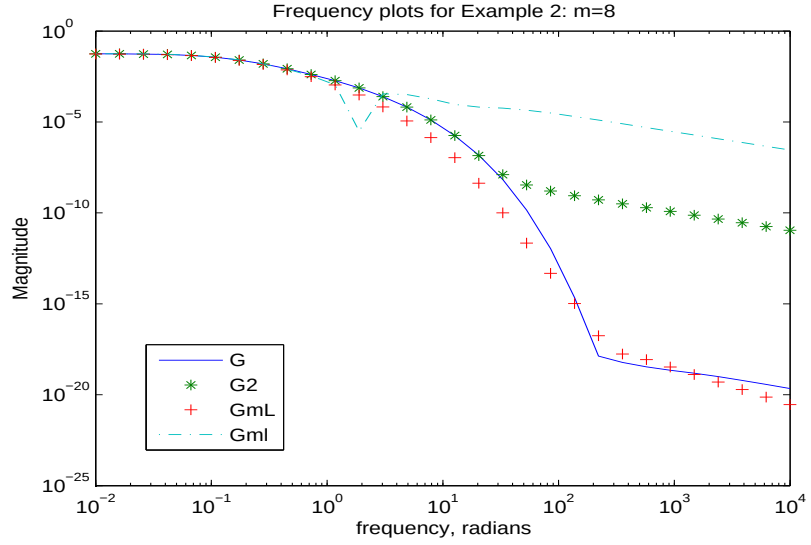


Figure 5.4: Frequency responses for heat-cont and its approximations for $m = 8$

Note from Figures 5.3 and 5.4 that $G(s)$ has a steep initial roll-off rate due to the zeros at ∞ . This is not well captured by $G_2(s)$ since it has no zeros at ∞ , while it is better captured by $G_m^L(s)$ since all its zeros are at ∞ . Note also that, while $G_m^l(s)$ is stable, minimum-phase and has a symmetric A -matrix, it does not perform well in capturing the frequency response of $G(s)$.

5.5 Conclusions

In this chapter the applicability of the parameterization derived in [30] to preserve stability is extended by considering other choices of L that preserve some other desirable properties such as minimum-phase character, passivity and bounded realness as well as symmetry. Using the Arnoldi equations, simple residual error expressions are derived and a minimum residual approximation method, optimal in the $\mathcal{H}_\infty$ -norm, is described. Also a procedure is given for finding the interpolating function with the least residual $\mathcal{H}_\infty$ -norm. Numerical results indicate that the approach is effective in preserving model properties. However, this is achieved at the expense of approximation accuracy since, by using the degrees of freedom in preserving model properties, we forego the opportunity of interpolating the model at additional points. Thus improving the approximation accuracy can most effectively be carried out by a proper choice of the interpolating points.

Chapter 6

Krylov Based Restart Scheme for the Solution of Linear Matrix Equations

This chapter utilizes the parameterization of rational interpolating approximation described in Chapter 5 to propose a restart scheme for the low rank Krylov based approximate solution of large-scale Sylvester equations. The proposed method is a novel restart scheme which improves further the computational efficiency and storage requirements of the standard Krylov subspace methods. The approach can also be used for solving large-scale Lyapunov equations, where it is shown that the approximation through the restart scheme approaches monotonically to the actual solution.

6.1 Restart Scheme

It is shown in Section 2.7 that a low rank approximate solution to the Sylvester/Lyapunov equation can be computed by using the rational Arnoldi method. This section propose a restart scheme that efficiently update the low rank approximate solution of the Sylvester/Lyapunov equation.

Suppose we perform m ($\hat{m}$) iterations of the rational Arnoldi algorithm (Algorithm 2.4) using A , B and S ($\hat{A}$, $\hat{B}$ and $\hat{S}$) to obtain A_m , B_m , C_{Vm} , D_{Vm} ($\hat{A}_{\hat{m}}$, $\hat{B}_{\hat{m}}$, $\hat{C}_{\hat{V}\hat{m}}$, $\hat{D}_{\hat{V}\hat{m}}$) such that the Arnoldi-like equations (4.2) are satisfied and suppose that an approximation

$\tilde{P}$ is computed by using (2.88). Let the interpolation points are selected such that S ($\hat{S}$) includes the first m eigenvalues of $-A$ ($-\hat{A}$) that are closest to the imaginary axis. Then the residual error associated with the approximate solution $\tilde{P}$ can be written as,

$$\begin{aligned}\tilde{R} &= A\tilde{P} + \tilde{P}\hat{A}^T + B\hat{B}^T, \\ &= AV_m X \hat{V}_{\hat{m}}^T + V_m X \hat{V}_{\hat{m}}^T \hat{A}^T + B\hat{B}^T.\end{aligned}$$

Using Arnoldi-like equations (5.5) in terms of free parameters L and $\hat{L}$ to replace AV_m , B , $\hat{A}\hat{V}_{\hat{m}}$ and $\hat{B}$ in the above equation, we get,

$$\tilde{R} = V_{m+1}^L \tilde{S} (\hat{V}_{\hat{m}+1}^{\hat{L}})^T, \quad (6.1)$$

where $V_{m+1}^L = [V_m \quad v_{m+1}^L]$, $\hat{V}_{\hat{m}+1}^{\hat{L}} = [\hat{V}_{\hat{m}} \quad \hat{v}_{\hat{m}+1}^{\hat{L}}]$ and

$$\tilde{S} = \begin{bmatrix} A_m^L X + X(\hat{A}_{\hat{m}}^{\hat{L}})^T + B_m^L (\hat{B}_{\hat{m}}^{\hat{L}})^T & X \hat{C}_{\hat{V}_{\hat{m}}}^T + B_m^L \hat{D}_{\hat{V}_{\hat{m}}}^T \\ C_{V_m} X + D_{V_m} (\hat{B}_{\hat{m}}^{\hat{L}})^T & D_{V_m} \hat{D}_{\hat{V}_{\hat{m}}}^T \end{bmatrix}, \quad (6.2)$$

where A_m^L , B_m^L are defined in Theorem 2.2 and similarly one can define $\hat{A}_{\hat{m}}^{\hat{L}}$, $\hat{B}_{\hat{m}}^{\hat{L}}$. Thus two free parameters L and $\hat{L}$ are introduced in the residual error expression. Now to update the approximation $\tilde{P}$, we perform a single restart that results in an updated approximation of the form,

$$\tilde{P}_1 = V_m X \hat{V}_{\hat{m}}^T + V_{m_1} X_1 \hat{V}_{\hat{m}_1}^T, \quad (6.3)$$

where $V_{m_1} \in \mathcal{R}^{n \times m_1}$ and $\hat{V}_{\hat{m}_1} \in \mathcal{R}^{\hat{n} \times \hat{m}_1}$ are the orthogonal matrices obtained by using m_1 and $\hat{m}_1$ iterations of the rational Arnoldi algorithm (Algorithm 2.4) with interpolation points as selected previously. Also $X_1 \in \mathcal{R}^{m_1 \times \hat{m}_1}$ while $\tilde{P}_1 \in \mathcal{R}^{n \times \hat{n}}$. In the above equation, the number of orthogonal columns in V_{m_1} and $\hat{V}_{\hat{m}_1}$ i.e., m_1 and $\hat{m}_1$ is small compared to m and $\hat{m}$. Note that once we obtain the product $V_m X \hat{V}_{\hat{m}}^T$, we are not storing the orthogonal matrices V_m and $\hat{V}_{\hat{m}}$ for further computations. That is V_{m_1} and $\hat{V}_{\hat{m}_1}$ are not orthogonalized with respect to V_m and $\hat{V}_{\hat{m}}$. This saves the computations of reorthogonalization and the storage of the orthogonal columns. The residual error

associated with the updated approximate solution can be written as,

$$\begin{aligned}
\tilde{R}_1 &= A\tilde{P}_1 + \tilde{P}_1\hat{A}^T + B\hat{B}^T \\
&= AV_mX\hat{V}_{\hat{m}}^T + V_mX\hat{V}_{\hat{m}}^T\hat{A}^T + B\hat{B}^T \\
&\quad + AV_{m_1}X_1\hat{V}_{\hat{m}_1}^T + V_{m_1}X_1\hat{V}_{\hat{m}_1}^T\hat{A}^T \\
&= AV_{m_1}X_1\hat{V}_{\hat{m}_1}^T + V_{m_1}X_1\hat{V}_{\hat{m}_1}^T\hat{A}^T + \tilde{R}
\end{aligned}$$

where $\tilde{R}$ is the residual error corresponding to the initial approximation, $\tilde{P}$. If we choose the free parameters L and $\hat{L}$ such that the residual error $\tilde{R}$ is expressed in the low rank form, preferably rank 1 then we can write,

$$\tilde{R} = B^1(\hat{B}^1)^T \quad (6.4)$$

where $B^1 \in \mathcal{R}^n$ and $\hat{B}^1 \in \mathcal{R}^{\hat{n}}$. The next section describes the choice of L and $\hat{L}$ for which the residual error is rank 1. Using (6.4), the residual error associated with the updated approximation becomes,

$$\tilde{R}_1 = AV_{m_1}X_1\hat{V}_{\hat{m}_1}^T + V_{m_1}X_1\hat{V}_{\hat{m}_1}^T\hat{A}^T + B^1(\hat{B}^1)^T. \quad (6.5)$$

This residual error now has the form of the Sylvester equation and can be solved by using Krylov subspace techniques. Using the Arnoldi-like equations (5.5) to replace AV_{m_1} , B^1 , $\hat{A}\hat{V}_{\hat{m}_1}$ and $\hat{B}^1$, we can express the new residual error $\tilde{R}_1$ in the form of (6.1) introducing two new free parameters L_1 and $\hat{L}_1$. These new parameters can be selected such that the new residual error $\tilde{R}_1$ has rank 1,

$$\tilde{R}_1 = B^2(\hat{B}^2)^T. \quad (6.6)$$

This shows that the approximation $\tilde{P}$ can be updated such that the residual error associated with each updated approximation can be expressed in terms of a rank 1 residual. Let the algorithm perform $K > 1$ restarts,

$$\tilde{P}_K = V_mX\hat{V}_{\hat{m}}^T + \sum_{i=1}^K V_{m_i}X_i\hat{V}_{\hat{m}_i}^T. \quad (6.7)$$

Then the corresponding residual errors can be written as

$$\tilde{R}_K = B^K(\hat{B}^K)^T + AV_{m_K}X_K\hat{V}_{\hat{m}_K}^T + V_{m_K}X_K\hat{V}_{\hat{m}_K}^T\hat{A}^T. \quad (6.8)$$

The complete restart algorithm for the solution of the Sylvester equation is summarized in Algorithm 6.1. The algorithm performs restarts by expressing the residual error in the low rank form using parameters L and $\hat{L}$. This means that at the last update we can obtain the updated approximation without using the parameters L_K and $\hat{L}_K$ and can compute the associated residual error (not necessarily low rank). It is found that the approximation computed through the restart algorithm without using parameters L_K and $\hat{L}_K$ at the final restart approaches the approximation to the solution of the Sylvester equation without restart.

Algorithm 6.1 Restart scheme for the solution of Sylvester equation

- 1: **Inputs:** $A, \hat{A}, B, \hat{B}, m, \hat{m}, S = \{s_1, s_2, \dots, s_m\}$ and $\hat{S} = \{\hat{s}_1, \hat{s}_2, \dots, \hat{s}_{\hat{m}}\}$.
 - 2: Initialize: K, m_r and $\hat{m}_r$
 - 3: Use A, B, S and the rational Arnoldi algorithm 2.4 to obtain V_m and v_{m+1} . Compute $A_m, B_m, C_{V_m}, D_{V_m}$ such that the Arnoldi-like equations (4.2) are satisfied.
 - 4: Use $\hat{A}, \hat{B}, \hat{S}$ and the rational Arnoldi algorithm 2.4 to obtain $\hat{V}_{\hat{m}}$ and $\hat{v}_{\hat{m}+1}$. Compute $\hat{A}_{\hat{m}}, \hat{B}_{\hat{m}}, \hat{C}_{\hat{V}_{\hat{m}}}$ and $\hat{D}_{\hat{V}_{\hat{m}}}$ such that the Arnoldi-like equations similar to (4.2) are satisfied.
 - 5: Calculate X by solving the nonsymmetric algebraic Riccati equation (6.11) and find the parameters L and $\hat{L}$ from (6.9) and (6.10)
 - 6: Calculate the approximation $\tilde{P}$ and the corresponding residual error $\tilde{R}$ using (2.88) and (6.1) respectively
 - 7: Choose $B^1 = v_{m+1}^L D_{V_m}$ and $\hat{B}^1 = \hat{v}_{\hat{m}+1} \hat{D}_{\hat{V}_{\hat{m}}}$
 - 8: **for** $i = 1$ to K **do**
 - 9: Use $A, \hat{A}, B^i, \hat{B}^i, S_i = \{s_{m+1}, s_{m+2}, \dots, s_{m+m_r}\}, \hat{S}_i = \{\hat{s}_{\hat{m}+1}, \hat{s}_{\hat{m}+2}, \dots, \hat{s}_{\hat{m}+\hat{m}_r}\}$ and the rational Arnoldi algorithm to obtain $A_{m_i}, B_{m_i}, C_{V_{m_i}}, D_{V_{m_i}}, \hat{A}_{\hat{m}_i}, \hat{B}_{\hat{m}_i}, \hat{C}_{\hat{V}_{\hat{m}_i}}, \hat{D}_{\hat{V}_{\hat{m}_i}}, \hat{V}_{\hat{m}_i}, \hat{v}_{\hat{m}_i+1}, V_{m_i}, v_{m_i+1}$
 - 10: Calculate X_i by solving the updated equation (6.11) and find the corresponding parameters L_i and $\hat{L}_i$ from (6.9) and (6.10).
 - 11: Calculate the updated solution $\tilde{P}_i = \tilde{P}_{i-1} + V_{m_i} X_i \hat{V}_{\hat{m}_i}^T$, where $\tilde{P}_0 = \tilde{P}$
 - 12: Compute the corresponding low rank residual error $\tilde{R}_i = B^{i+1}(\hat{B}^{i+1})^T$.
 - 13: Set $m = m + m_r$ and $\hat{m} = \hat{m} + \hat{m}_r$.
 - 14: **end for**
 - 15: **Output:** $\tilde{P}_K$
-

6.2 Low Rank Residual Error Analysis

In this section, we describe the choice of the parameters L and $\hat{L}$ to obtain the low rank residual error expression. It follows from (6.2) that a suitable choice of L , $\hat{L}$ and X is one that makes $\tilde{S} = D_{Vm} \hat{D}_{\hat{Vm}}^T$. Expressions for parameters L and $\hat{L}$ in terms of unknown X can be obtained by equating the (1,2) and (2,1) blocks of $\tilde{S}$ to zero,

$$L = -(B_m \hat{D}_{\hat{Vm}}^T + X \hat{C}_{\hat{Vm}}^T)(D_{Vm} \hat{D}_{\hat{Vm}}^T)^{-1}, \quad (6.9)$$

$$\hat{L} = -(\hat{B}_{\hat{Vm}} D_{Vm}^T + X^T C_{Vm}^T)(\hat{D}_{\hat{Vm}} D_{Vm}^T)^{-1}, \quad (6.10)$$

provided that D_{Vm} and $\hat{D}_{\hat{Vm}}$ are nonzero, which we will assume. Note that $D_{Vm} \neq 0$ and $\hat{D}_{\hat{Vm}} \neq 0$ provided that ∞ is not chosen to be an interpolating point for (A, B) and $(\hat{A}, \hat{B})$. Using the expressions of L and $\hat{L}$ in the (1,1) block of $\tilde{S}$, we get a low order nonsymmetric algebraic Riccati equation given by,

$$A_i X + X \hat{A}_i^T - X \hat{C}_i^T C_i X = 0 \quad (6.11)$$

where $A_i = A_m - B_m D_{Vm}^{-1} C_{Vm}$, $\hat{A}_i = \hat{A}_{\hat{m}} - \hat{B}_{\hat{m}} \hat{D}_{\hat{Vm}}^{-1} \hat{C}_{\hat{Vm}}$, $C_i = D_{Vm}^{-1} C_{Vm}$ and $\hat{C}_i = \hat{D}_{\hat{Vm}}^{-1} \hat{C}_{\hat{Vm}}$. The low order Riccati equation can be solved for X by using one of the techniques described in Subsection 6.2.1 below. The unknown parameters L and $\hat{L}$ given by (6.9) and (6.10) can then be computed by using the value of X . For these values of L , $\hat{L}$ and X , the residual error can be written as,

$$\tilde{R} = v_{m+1}^L D_{Vm} \hat{D}_{\hat{Vm}}^T (\hat{v}_{m+1}^{\hat{L}})^T. \quad (6.12)$$

This residual error has rank 1 and can be expressed in the form of (6.4) by using $B^1 = v_{m+1}^L D_{Vm}$ and $\hat{B}^1 = \hat{v}_{m+1}^{\hat{L}} \hat{D}_{\hat{Vm}}$ where v_{m+1}^L . We can also derive the above low rank residual error expression by equating the Schur complement of S given in (2.90) to zero and introducing the free parameters L and $\hat{L}$ as expressed in (6.9) and (6.10). Let (2.90) is represented as $\tilde{S} = \begin{bmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ \tilde{S}_{21} & \tilde{S}_{22} \end{bmatrix}$, then the parameters L and $\hat{L}$ can be written as $L = -\tilde{S}_{12} \tilde{S}_{22}^{-1}$ and $\hat{L} = -\tilde{S}_{21}^T \tilde{S}_{22}^{-T}$ and thus,

$$\begin{bmatrix} I & L \\ 0 & 1 \end{bmatrix} S \begin{bmatrix} I & 0 \\ \hat{L}^T & 1 \end{bmatrix} = \begin{bmatrix} \tilde{S}_{11} - \tilde{S}_{12} \tilde{S}_{22}^{-1} \tilde{S}_{21} & 0 \\ 0 & \tilde{S}_{22} \end{bmatrix},$$

where $\tilde{S}_{11} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21}$ is the Schur complement of S . Using the above equation and the inverse of the unit triangular matrices, the residual error expression can be written as,

$$\begin{aligned}\tilde{R} &= V_{m+1} \begin{bmatrix} I & -L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{S}_{11} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} & 0 \\ 0 & \tilde{S}_{22} \end{bmatrix} \\ &\quad \begin{bmatrix} I & 0 \\ -\hat{L}^T & 1 \end{bmatrix} \hat{V}_{\hat{m}+1}^T \\ &= V_{m+1}^L \begin{bmatrix} \tilde{S}_{11} - \tilde{S}_{12}\tilde{S}_{22}^{-1}\tilde{S}_{21} & 0 \\ 0 & \tilde{S}_{22} \end{bmatrix} (\hat{V}_{\hat{m}+1}^L)^T\end{aligned}\tag{6.13}$$

where $V_{m+1}^L = [V_m \quad v_{m+1}^L]$ and $\hat{V}_{\hat{m}+1}^L = [\hat{V}_{\hat{m}} \quad \hat{v}_{\hat{m}+1}^L]$. Now if the Schur complement of $\tilde{S}$ is zero then the residual error is low rank and is given by equation (6.12).

6.2.1 Solution of Nonsymmetric Algebraic Riccati Equation

It is clear that the low rank residual error expression requires the solution of the reduced order Riccati equation (6.11). In this subsection, we consider two possible methods for the solution of the Riccati equation.

The first method is a standard approach for solving algebraic Riccati equation (see Theorem 7.1.2 in [64]). The technique involves the use of the following $(m + \hat{m}) \times (m + \hat{m})$ Hamiltonian matrix associated with the algebraic Riccati equation (6.11),

$$H = \begin{bmatrix} \hat{A}_i^T & -C_i\hat{C}_i^T \\ 0 & -A_i \end{bmatrix}.\tag{6.14}$$

Suppose a matrix T transforms H into a Jordan form $HT = TJ$, where J is the Jordan matrix with the eigenvalues of H as the diagonal entries and T contains the $m + \hat{m}$ eigenvectors of H . If T is partitioned as $T = [T_1 \quad T_2]$ where $T_1 \in \mathcal{R}^{(m+\hat{m}) \times m}$ contains the first m eigenvectors of H and $T_2 \in \mathcal{R}^{(m+\hat{m}) \times \hat{m}}$ contains the remaining $\hat{m}$ eigenvectors, then we may write

$$HT_1 = T_1\Lambda\tag{6.15}$$

where Λ contain the eigenvalues of H corresponding to the eigenvectors in T_1 . It is easy to show that each solution to the Quadratic Sylvester equation can be expressed as, $X = T_{21}T_{11}^{-1}$, where $T_{11} \in \mathcal{R}^{m \times m}$ is assumed to be nonsingular and $[T_{11}^T \quad T_{21}^T]^T = T_1$. This method works for any choice of Λ and an interesting issue is to investigate an optimal choice of Λ , which will be considered in future work.

The second method transforms the nonsymmetric algebraic Riccati equation (6.11) in a low order Sylvester equation if the matrix X is square and invertible. It is easy to see that if $Y = X^{-1}$ exists, then pre-multiplication and post-multiplication of (6.11) by Y implies (6.11) results in,

$$\hat{A}_i^T Y + Y A_i - \hat{C}_i^T C_i = 0. \quad (6.16)$$

The above equation can be solved for Y by using direct methods. The required solution is then, $X = Y^{-1}$.

6.3 Solution of Lyapunov Equation

The next result shows that the proposed restart scheme converges monotonically to the actual solution of the Lyapunov equation.

Theorem 6.1. *Let $Q \in \mathcal{R}^{n \times n}$ be the positive definite solution of the Lyapunov equation,*

$$AQ + QA^T + BB^T = 0, \quad (6.17)$$

where $A \in \mathcal{R}^{n \times n}$ is stable and $B \in \mathcal{R}^n$. Suppose $Q_m = V_m X V_m^T$, is the initial rational Arnoldi approximation to Q and $Q_{m_i}^i = V_{m_i}^i X_i (V_{m_i}^i)^T$, for $i = 1, \dots, K$ are the first K updates of Q_m obtained by using the restart algorithm (Algorithm 6.1 with $\hat{A} = A$ and $\hat{B} = B$). Then the restart scheme converges monotonically to the actual solution Q , that is,

$$Q \geq Q_{m_K}^K \geq Q_{m_{K-1}}^{K-1} \geq \dots \geq Q_{m_1}^1 \geq Q_m. \quad (6.18)$$

Proof. Since Q_m is the initial approximate solution of the Lyapunov equation (6.17) using the proposed technique,

$$AQ_m + Q_m A^T + BB^T - B^1 (B^1)^T = 0, \quad (6.19)$$

where $B^1 (B^1)^T$ is the residual error associated with the approximate solution Q_m . Subtracting the above expression from (6.17), we get,

$$A(Q - Q_m) + (Q - Q_m)A^T + B^1 (B^1)^T = 0. \quad (6.20)$$

As $Q - Q_m$ is the solution of a Lyapunov equation with a stable A , we have $Q \geq Q_m$. Now after the first restart of the algorithm, the updated approximate solution is of the form,

$$Q_{m_1}^1 = Q_m + V_{m_1}^1 X^1 V_{m_1}^{1T}. \quad (6.21)$$

As $X^1 \geq 0$, the updated approximation satisfies $Q_{m_1}^1 \geq Q_m$. Similarly it can be shown that for the K -th restart $Q_{m_K}^K \geq Q_{m_{K-1}}^{K-1} \geq \dots \geq Q_m$. Now to show that $Q \geq Q_{m_K}^K$, observe that each updated approximate solution has a low rank residual error expression and can be written in terms of the Lyapunov equation of the form (6.20), which proves the theorem. $\square$

6.4 Numerical Results

In this section, we compare the proposed restart scheme and the standard Krylov method for the approximation of the solution to the Sylvester equation. The approximate solution obtained by the proposed scheme converges to the one obtained by the standard Krylov technique with two important advantages; low computational cost and less storage requirements. These advantages are due to the fact that the proposed scheme avoids the storage of large orthogonal matrices and it also avoids orthogonalization of the bases vectors with previously computed bases. A complete comparison of the method in terms of computational cost, storage requirement and convergence rates with other existing methods is part of the future work.

No. of restarts	Size of X	$\ R_m\ _2$	$\ R_m^L\ _2$	$\ P - P_m\ _2$	$\ P - P_m^L\ _2$
0	10	71.8908	82.0524	9.3003	11.2505
1	14	49.5623	57.5872	3.4635	4.1599
2	18	13.7751	21.6624	0.7587	1.1038
3	22	4.8326	7.8826	0.1717	0.3621

Table 6.1: The 2-norm of the forward error and the residual error for the restart scheme and the standard Krylov technique to solve the Sylvester equation

6.4.1 Example 1

This example is taken from [88] where the matrices $A \in \mathcal{R}^{n \times n}$ and $\hat{A} \in \mathcal{R}^{\hat{n} \times \hat{n}}$ are in the form

$$A = \text{tridiag}\left(-1 + \frac{10}{n+1}, -2, -1 + \frac{10}{n+1}\right),$$
$$\hat{A} = \text{tridiag}\left(-1 + \frac{10}{\hat{n}+1}, -2, -1 + \frac{10}{\hat{n}+1}\right),$$

where $n = 300$ and $\hat{n} = 250$. The vectors B and $\hat{B}$ are randomly selected. Table 6.1 shows the norm of the forward error and the residual error for the proposed scheme and the standard Krylov technique for the solution of Sylvester equations. The two strategies are compared at three different projection sizes where the size of the projected Sylvester equation is taken as the size of matrix X . For all cases, the approximation computed through the restart scheme is less accurate as compared to the standard Krylov technique however the restart scheme has the advantage that it does not involve the orthogonalization cost and the storage of orthogonal columns.

6.5 Conclusions

In this chapter a computationally efficient Krylov based restart scheme has been developed for the approximate solution of the Sylvester and the Lyapunov equations. In case of Lyapunov equations, the updated approximations monotonically converge to the actual solution. The approach utilizes parameterized interpolating approximations and it is shown that low rank residual error expressions can be computed by a proper choice of the free parameters.

Chapter 7

Conclusions and Future Research Directions

This chapter summarizes the main contribution of all the work presented in the previous chapters. Some possible future directions and extensions of the research are also described.

7.1 Summary

Numerical simulations are often based on large scale complex models that are used to measure and control the behavior of some output parameters with respect to a given set of inputs. The concept of model reduction is to approximate this input-output behavior by a much simple model that can predict the actual behavior. Different applications in various areas require model reduction for example, in electronics, model reduction is used to predict the behaviour of complicated interconnect systems. Various techniques have been proposed in the literature to solve the problem of model reduction and some of these techniques are reviewed in the first chapter of the thesis. The Balanced truncation and optimal Hankel-norm approximation methods guarantee stability of the actual system in the approximate model and have a priori error bound. However these techniques involve the computation of solution of Lyapunov equations and decomposition of matrices which make them computationally expensive for large scale systems and inefficient to utilize sparsity or any structure of large scale systems. This leads to Krylov subspace techniques which involve only matrix-vector multiplications in the large dimension and can achieve

moment matching without explicit computation of moments. These properties make them computationally efficient and a better choice for model reduction of large scale systems.

The standard Arnoldi and Lanczos algorithm are two well-known Krylov subspace techniques that compute orthogonal and bi-orthogonal bases respectively to a low dimensional Krylov subspace. In the second chapter of this thesis some projection processes are described that utilize these bases to compute reduced order approximate models for large linear systems. It is shown that the resulting reduced order approximation achieves moment matching at infinity. An important extension of the Arnoldi/Lanczos algorithm is the rational Arnoldi/Lanczos algorithm which compute orthogonal/bi-orthogonal basis to the union of low dimensional Krylov subspaces. The reduced order approximation obtained via projection onto the union of Krylov subspaces has the advantage that it achieves moment matching at a predefined set of interpolation points in contrast to the Arnoldi/Lanczos approximation which achieve moment matching only at infinity. Thus the rational Arnoldi/Lanczos approximations are highly dependant on the choice of interpolation points. Various techniques have been proposed in the literature for the selection of interpolation points in the rational Krylov method. Recently an iterative rational Krylov algorithm (IRKA) has been proposed in the literature which updates the interpolation points such that on convergence these points are the negative eigenvalues of the reduced order approximation, which is a necessary condition for $\mathcal{H}_2$ optimal approximation.

The IRKA method in each iteration utilizes rational Krylov steps to solve a sequence of shifted linear systems. It therefore requires matrix-vector multiplications for each shifted linear system. In Chapter 3 an alternative implementation is suggested to the IRKA method by proposing an iterative algorithm that utilizes a variant of the bi-conjugate gradient (Bi-CG) method such that for each shifted linear system the algorithm only requires matrix-vector multiplications associated with a single linear system. This means that if we have m shifted linear systems, the proposed algorithm saves $m - 1$ matrix-vector multiplications. These savings are especially useful for dense matrices since in that case the matrix-vector multiplications are more expensive to compute.

The IRKA method and the iterative bi-conjugate gradient method both converge to the same local minimizer of the $\mathcal{H}_2$ optimal model reduction problem, starting from an initial set of interpolation points. However the approximations obtained through these methods are not necessarily the global minimizers. In Chapter 4 some conditions are derived that allow us to compute all first and second order local minimizers and hence identify the global minimizer to the problem. In case of first order approximations, the main steps are to compute (i) all positive zeros of a linear system and (ii) all first order approximations,

each interpolating at a positive zero. The conditions are slightly complicated in case of second order approximation as it involves all simultaneous solutions of two bivariate polynomial equations. For a general m -th order approximation, the condition becomes the simultaneous solution of m multivariate polynomial equations. The idea of implicitly computing all the local minimizers to identify the global minimizer may not be useful and an important future extension would be to explicitly compute the local minimizers through the derived conditions. Some new error expressions in terms of the $\mathcal{H}_2$ norm are also derived for the general case in Chapter 4. These error expressions show a link of the $\mathcal{H}_2$ optimal model reduction problem to the residual error associated with approximate solution of the Sylvester/Lyapunov equations obtained via the standard Krylov methods.

An important issue with the rational Krylov method is that the reduced order interpolating approximation does not guarantee the useful properties such as stability, passivity, minimum-phase and bounded real character as well as structure of the actual system. Recently a parameterization of rational interpolating approximation in terms of a free low dimensional parameter was introduced that selects the free parameter such that for a given stable system, the reduced order approximation is guaranteed to be stable. However the retention of other properties was not considered. In Chapter 5, the idea of parameterization is extended to derive conditions on the free parameter such that the reduced order interpolating approximation guarantees any of the useful properties mentioned above as well as combinations of them. In addition to the use of parameterization for the retention of important properties, the free parameter can also be used to improve the accuracy of the approximation. It is shown that the free parameter can be selected to have a minimal residual approximation, $\mathcal{H}_2$ optimal approximation and a two-sided rational Krylov approximation.

The rational Krylov method has also been used for the low rank approximate solution of the Sylvester/Lyapunov equations. The approach involves the computation of two set of basis vectors in which each vector is orthogonalized against all previous vectors. This orthogonalization becomes computationally expensive and requires high storage capacity as the number of basis vectors increases. In Chapter 6, a restart scheme is proposed which restarts without requiring that the new vectors are orthogonal to the previous vectors. Instead, two new orthogonal basis vectors are computed. This saves the computational burden of orthogonalization and the requirement of storage capacity. The restart scheme is due to the parameterization of rational interpolating approximation that allows us to derive low rank residual error expressions associated with the approximate solution of Sylvester and Lyapunov equations.

7.2 Thesis Contribution

The key contribution of this thesis are given in the following.

1. **Iterative Bi-CG method for $\mathcal{H}_2$ optimal model reduction:** A computationally efficient iterative algorithm is proposed in Chapter 3 that converges to a local minimizer of the $\mathcal{H}_2$ optimal model reduction problem.
2. **Error expressions in terms of $\mathcal{H}_2$ norm:** Some new forward error expressions are derived in terms of $\mathcal{H}_2$ norm for the approximation that satisfy the necessary conditions for $\mathcal{H}_2$ optimality. These expressions are given in Chapter 4 and some of them have been published in [3, 4].
3. **First and second order global minimizers for the $\mathcal{H}_2$ optimal model reduction problem:** Important conditions are derived in Chapter 4 for the first order and second order interpolating approximations that allow us to identify global minimizer to the $\mathcal{H}_2$ optimal model reduction problem. This work has been published in [3, 4].
4. **Retention of useful properties of the actual system in the parameterized rational interpolating approximation:** The concept of parameterization is extended in Chapter 5 to preserve useful properties such as passivity, minimum phase and bounded-real character as well as structure of the actual system in the approximate system. The work has been submitted in SIAM journal of matrix analysis and is currently under review.
5. **Minimal residual, $\mathcal{H}_2$ optimal and Two-sided rational Arnoldi approximations using parameterization:** The concept of parameterization is introduced in Chapter 5 to have minimal residual approximation, $\mathcal{H}_2$ optimal approximation and two-sided rational Arnoldi approximation.
6. **Restart scheme for low rank approximate solution of the Sylvester and Lyapunov equations:** A Krylov based restart algorithm is proposed in Chapter 6 for the approximate solution of the Sylvester as well as Lyapunov equation that saves the computational expenses associated with the reorthogonalization process of the standard Krylov methods. The result of this work have appeared in [5].

7.3 Future Work

Several research issues arise from the work presented in this thesis. Some of these issues are described in the following.

- **Extensions for the optimality conditions:** An important and common issue with the rational Krylov method is the selection of interpolation points. In Chapter 4 it is shown that these interpolation points are the zeroes of a linear system for a first order $\mathcal{H}_2$ optimal approximation. In case of a second order $\mathcal{H}_2$ optimal approximation, the problem becomes the simultaneous solution of two bivariate polynomial equations. Extension to higher order $\mathcal{H}_2$ optimal approximation will involve exhaustive computations. However what might be interesting is to analyse any structure or identify any link of the first and second order $\mathcal{H}_2$ optimal approximations to the existing model reduction techniques.
- **Application of the iterative algorithms:** The iterative rational Krylov algorithm (IRKA) and the proposed iterative bi-conjugate gradient (IBi-CG) method compute a local minimizer to the $\mathcal{H}_2$ optimal model reduction problem. The conditions derived in Chapter 4 compute all local minimizers including the one obtained from the IRKA/IBi-CG method. It will be nice to see if the information obtained from the IRKA method about one of the local minimizer can be useful to compute all other local minimizers from the derived conditions in an efficient way.
- **Preconditioning of the iterative bi-conjugate gradient algorithm:** The iterative bi-conjugate gradient algorithm for $\mathcal{H}_2$ optimal model reduction proposed in Algorithm 3.5 requires the storage of a complete Krylov subspace. Therefore, we need to limit the dimensions of the Krylov subspace. This can be done by utilizing preconditioners [27] for the Bi-CG method. However the preconditioned system should correspond to the shift invariant property of the Krylov subspace. This means that the problem requires the design of special preconditioner which does not effect the collinear relationship between the residual errors, given by (3.7).
- **Relation between $\mathcal{H}_2$ optimal approximation and low rank solution of Lyapunov equation:** The error expression (4.10) i.e., $\|G - G_m\|_{\mathcal{H}_2}^2 = \|C(sI - A)^{-1}v_{m+1}D_{V_m}\|_{\mathcal{H}_2}^2$ derived in Chapter 4 shows that the $\mathcal{H}_2$ optimal approximation is linked to the residual error $\tilde{R} = v_{m+1}D_{V_m}D_{V_m}^T v_{m+1}^T$ associated with the approximate solution of the Lyapunov equation, $AP + PA^T + BB^T = 0$. The only unknown term in the error expression is the vector $v_{m+1}D_{V_m}$ and its product with its transpose vector gives the residual error $\tilde{R}$. This suggest us to investigate the

possibility of deriving conditions on the residual error $\tilde{R}$ that guarantee the $\mathcal{H}_2$ optimal reduced order approximation.

- **Complete computational analysis of the restart algorithm:** The restart algorithm proposed in Chapter 6 constructs two new orthogonal matrices in each restart without requiring them to be orthogonal to the previous orthogonal matrices. This saves the computational burden of orthogonalizing vectors with all the previous orthogonal vectors. However the restart scheme involves some extra computations such as the solution of a reduced order algebraic Riccati equation and the computation of the free parameters. It will be interesting to compare the overall computations of the restart scheme with the one involving no restarts.

7.4 Publications Arising from this Work

The following publications have been written during the course of this thesis:

Journal papers

M. Frangos, Mian I. Ahmad and Imad M. Jaimoukha, “A Parameterization of Interpolating Approximations of Large Scale Systems via the Rational Arnoldi Algorithm”. *Under Review in SIAM Journal of Matrix Analysis*, 2011.

Conference papers

Mian I. Ahmad, Imad M. Jaimoukha and M. Frangos, “Krylov based techniques for large scale linear matrix equations”. In *Proc. of American Control Conference, ACC 2010, Baltimore, MD, USA*.

Mian I. Ahmad, M. Frangos and Imad M. Jaimoukha. “ $\mathcal{H}_2$ optimal model reduction of linear dynamical systems”. In *Proc. of 49-th IEEE conference on decision and control, CDC 2010, Atlanta, Georgia, USA*.

Mian I. Ahmad, M. Frangos and Imad M. Jaimoukha. “Second order $\mathcal{H}_2$ optimal model reduction of linear dynamical systems”. In *Proc. of 18-th World Congress of the International Federation of Automatic Control, IFAC 2011, Milan, Italy*.

Bibliography

- [1] V. Adamjan, D. Arov, and M. Krein. Analytic properties of schmidt pairs for a hankel operator and the generalized schur-takagi problem. *Math. USSR Sbornik*, 15:31–73, 1971.
- [2] V. Adamjan, D. Arov, and M. Krein. Infinite block hankel matrices and related extension problems. *Amer. Math. Soc. Trans.*, 111:133–156, 1978.
- [3] M. Ahmad, I. Jaimoukha, and M. Frangos. H_2 optimal model reduction of linear dynamical systems. In *49th IEEE Conference on Decision and Control*, 2010.
- [4] M. I. Ahmad, M. Frangos, and I. Jaimoukha. Second order $\mathcal{H}_2$ optimal approximation of linear dynamical systems. In *IFAC world congress*, 2011.
- [5] M. I. Ahmad, I. Jaimoukha, and M. Frangos. Krylov subspace restart scheme for solving large scale sylvester equations. In *American Control Conference, ACC*, 2010.
- [6] A. Antoulas. On the scalar rational interpolation problem. *IMA J. Mathematical Control and Information*, 3:61–88, 1986.
- [7] A. Antoulas, D. Sorensen, and S. Gugercin. A survey of model reduction methods for large-scale systems. *AMS*, 280:193–219, 2001.
- [8] A. C. Antoulas. *Approximation of large-scale dynamical systems*. SIAM Publications, 2005.
- [9] W. Arnoldi. The principle of minimized iterations in the solution of matrix eigenvalue problem. *Quart. Appl. Math*, 9:17–29, 1951.
- [10] A. Astolfi. A new look at model reduction by moment matching for linear systems. In *In Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA, USA*, Dec. 2007.

- [11] Z. Bai. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics*, 43:944, 2002.
- [12] Z. Bai, P. M. Dewilde, and R. W. Freund. Reduced-order modeling. In *Handbook of Numerical Analysis. Numerical Methods in Electromagnetics, Schilder WHA, ter Maten JW (eds) Elsevier: Amsterdam*, XIII:825–895, 2005.
- [13] G. Baker. *Essentials of Padé approximation*. New York NY: Academic Press, 1975.
- [14] L. Baratchart, M. Cardelli, and M. Olivi. Identification and rational l^2 approximation: a gradient algorithm,. *Automatica*, 27:413–418., 1991.
- [15] R. Bartels and G. Stewart. Solution of the matrix equation $AX + XA = C$: Algorithm 432. *Communications of the ACM*, 15:820–826, 1972.
- [16] C. Beattie and S. Gugercin. Krylov based minimization for optimal $\mathcal{H}_2$ model reduction. In *46th IEEE Conference on Decision and Control*, 2007.
- [17] P. Benner. Factorized solution of Sylvester equations with applications in control. In *Proceedings of International Symposium on Mathematical Theory of Networks and Systems (MTNS2004)*, 2004.
- [18] S. Boyd, E. Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in Systems and Control Theory*. SIAM Publications, 1991.
- [19] J. Brands. Projection methods for oversized linear algebra problems. *Nieuw Arch. Wiskd.*, 5:264–277, 2000.
- [20] A. Bultheel and B. D. Moor. Rational approximation in linear systems and control. *SIAM J. Matrix Anal. Appl.*, 121:355–378, 2000.
- [21] A. Bunse-Gerstner, D. Kubalinska, G. Vossen, and D. Wilczek. h_2 -norm optimal model reduction for large scale discrete dynamical mimo systems. *Journal of Computational and Applied Mathematics*, 233:1202–1216, 2010.
- [22] D. Calvetti and L. Reichel. Application of ADI iterative methods to the restoration of noisy images. *SIAM Journal on Matrix Analysis and Applications*, 17:165–186, 1996.
- [23] Y. Chahlaoui and P. V. Dooren. Youns chahlaoui and paul van dooren: A collection of benchmark examples for model reduction of linear time invariant dynamical systems;slicot working note 2002-2: February 2002. Technical report, SLICOT Working Note, 2002.

- [24] J. Cullum and T. Zhang. Two-sided arnoldi and nonsymmetric lanczos algorithms. *SIAM Journal on Matrix Analysis & Applications*, 24:303–319, 2002.
- [25] S. de Hillerin. *Model reduction of large-scale systems based on Krylov subspace methods*. PhD thesis, MSc thesis, Imperial College, London, 2008.
- [26] U. B. Desai and D. Pal. A transformation approach to stochastic model reduction. *IEEE Transactions on Automatic Control*, AC-29:1097–1100, 1984.
- [27] S. C. Eisenstat. Efficient implementation of a class of preconditioned conjugate gradient methods. *SIAM J. Sci. Stat. Comput.*, 2:1–4, 1981.
- [28] D. Enns. Model reduction with balanced realizations: An error bound and a frequency weighted generalization. In *In Proceedings of the 23rd IEEE Conference on Decision and Control, Las Vegas, Nevada, USA.*, 1984.
- [29] R. Fletcher. Conjugate gradient methods for indefinite systems. *Numerical Analysis. Lecture Notes in Mathematics (Springer Berlin / Heidelberg)*, 506:7389, 1976.
- [30] M. Frangos. *Adaptive Krylov Subspace Methods for Model Reduction of Large Scale Systems*. PhD thesis, Imperial College London, London, UK, May 2003.
- [31] M. Frangos and I. Jaimoukha. Adaptive rational Krylov algorithms for model reduction. In *European control conference*, pages 4179–4186, July 2007.
- [32] M. Frangos and I. Jaimoukha. Rational interpolation: Modified rational Arnoldi algorithm and Arnoldi-like equations. In *46th IEEE Conference on Decision and Control*, December 2007.
- [33] M. Frangos and I. Jaimoukha. Adaptive rational interpolation: Arnoldi and Lanczos-like equations. *European Journal of Control*, 14(4):342 – 354, 2008.
- [34] R. Freund and N. Nachtigal. Qmr: A quasi-minimal residual method for non-hermitian linear systems. *Numer. Math.*, 60:315–339, 1991.
- [35] R. W. Freund. Model reduction methods based on krylov subspaces. *Acta Numerica*, 12:267–319, 2003.
- [36] A. Frommer. Bicgstab(l) for families of shifted linear systems. *Computing*, 70:87109, 2003.
- [37] P. Fulcheri and M. Olivi. Matrix rational h_2 approximation: A gradient algorithm based on schur analysis. *SIAM J. Control Optim.*, 36:2103–2127, 1998.
- [38] K. Gallivan, E. Grimme, and P. V. Dooren. A rational Lanczos algorithm for model reduction. *Numerical Algorithms*, 12:33–64, 1996.

- [39] K. Gallivan, A. Vandendorpe, and P. V. Dooren. Sylvester equations and projection-based model reduction. *Journal of Computational and Applied Mathematics*, 162:213229, 2004.
- [40] S. Gilbert and F. George. *An Analysis of The Finite Element Method*. Prentice Hall, 1973.
- [41] K. Glover. All optimal hankel-norm approximations of linear multivariable systems and their $\mathcal{L}_\infty$ -error bounds. *International Journal of Control*, 39 (6):1115–1193, 1984.
- [42] G. H. Golub and C. F. V. Loan. *Matrix computations*. Johns Hopkins University Press, 3rd edition, 1996.
- [43] W. Gragg and A. Lindquist. On the partial realization problem. *Linear Algebra Appl*, 50:277–319, 1983.
- [44] A. Greenbaum. *Iterative methods for solving linear systems*. SIAM, Philadelphia, 1997.
- [45] E. Grimme. *Krylov projection methods for model reduction*. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, Illinois, 1997.
- [46] A. E. Guennouni, K. Jbilou, and A. J. Riquet. Block Krylov subspace methods for solving large Sylvester equations. *Numerical Algorithms*, 29:75–96, 2002.
- [47] S. Gugercin. *Projection Methods for Model Reduction of Large-Scale Dynamical Systems*. PhD thesis, Rice University, Houston, TX, 2002.
- [48] S. Gugercin and A. Antoulas. An h_2 error expression for the lanczos procedure. In *42nd IEEE Conference on Decision and Control*, 2003.
- [49] S. Gugercin and A. Antoulas. Model reduction of large-scale systems by least squares. *Linear Algebra and its Applications*, 415:290–321, 2006.
- [50] S. Gugercin and A. C. Antoulas. A survey of model reduction by balanced truncation and some new results. *INT. J. Control*, 8:748766, 2004.
- [51] S. Gugercin, A. C. Antoulas, and C. Beattie. $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM J. Matrix Anal. Appl.*, 30:609–638, 2008.
- [52] W. W. Hager. Updating the inverse of a matrix. *SIAM Review*, 31:221–239, 1989.
- [53] Y. Halevi. Frequency weighted model reduction via optimal projection. *IEEE Transactions on Automatic Control*, 37:1537–1542, 1992.

- [54] S. Hammarling. Numerical solution of the stable, non-negative definite Lyapunov equation. *IMA Journal of Numerical Analysis*, 2:303–323., 1982.
- [55] M. R. Hestenes and E. Stiefel. Methods of conjugate gradient for solving linear systems. *Journal of research of the national bureau of standards*, 49:409–436, 1952.
- [56] D. Hyland and D. Bernstein. The optimal projection equations for model reduction and the relationships among the methods of wilson, skelton, and moore. *IEEE Transactions on Automatic Control*, AC-30(12):1201–1211, 1985.
- [57] I. Jaimoukha. A general minimal residual Krylov subspace method for large scale model reduction. *IEEE Transactions on Automatic Control*, 42:1422 –1427, 1997.
- [58] I. Jaimoukha and E. Kasenally. Krylov subspace methods for solving large Lyapunov equations. *SIAM Journal on Numerical Analysis*, 31:227–251, 1994.
- [59] I. Jaimoukha and E. Kasenally. Oblique projection methods for large scale model reduction. *SIAM Journal of Matrix Analysis and Applications*, 16:602627, 1995.
- [60] I. M. Jaimoukha and E. M. Kasenally. Implicitly restarted krylov subspace methods for stable partial realizations. *SIAM J. Matrix Anal. Appl.*, 18:633–652, 1997.
- [61] K. Jbilou and A. Riquet. Projection methods for large Lyapunov matrix equations. *Linear Algebra and its Applications*, 415:344–358,, 2006.
- [62] B. Jegerlehner. Krylov space solvers for shifted linear systems. IUHET-353, 1996.
- [63] J. Kautsky and N. Nichols. Robust pole assignment in linear state feedback. *Int. J. Control*, 41:1129–1155, 1985.
- [64] P. Lancaster and L. Rodman. *Algebraic Riccati Equations*. Clarendon press, Oxford, 1995.
- [65] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential equation and integral operator. *J. Res. Nat. Bur. Stand*, 45:255–282, 1950.
- [66] C. Lanczos. Solution of systems of linear equations by minimized iterations. *J. Res. Nat. Bur. Stand*, 49:33–53, 1952.
- [67] R. B. Lehoucq, K. Maschhoff, D. Sorensen, and C. Yang. Arpack software package. <http://www.caam.rice.edu/software/ARPACK/>, 1996.
- [68] D. Limebeer and M. Green. *Linear Robust Control*. Prentice Hall, 1995.

- [69] C. Lin and T. Chiu. Model reduction via frequency weighted balanced realization. *Control Theory and Advanced Technology*, 8:341351, 1992.
- [70] L. Meier and D. Luenberger. Approxiamtion of linear constant systems. *IEEE Transactions on Automatic Control*, 12:585–588, 1967.
- [71] B. Moore. Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Automat. Control*, 26(1):17–32, 1981.
- [72] K. Morton and D. Mayers. *Numerical Solution of Partial Differential Equations, An Introduction*. Cambridge University Press, 2005.
- [73] P. C. Opdenacker and E. A. Jonckheere. A contraction mapping preserving balanced reduction scheme and its in?nity norm error bounds. *IEEE Transactions on Circuits and Systems*, 35:184189, 1988.
- [74] C. C. Paige and M. A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM Journal on Numerical Analysis*, 12:617–629, 1975.
- [75] V. Papakos and I. M. Jaimoukha. Implicitly restarted lanczos algorithm for model reduction. In *Proc. 40th IEEE Conf. Decision Control, Orlando, FL*, page 36713672, 2001.
- [76] B. N. Parlett, D. R. Taylor, and Z. A. Liu. A look-ahead lanczos algorithm for unsymmetric matrices. *Mathematics of Computation*, 44:105–124, 1985.
- [77] J. Partington. *An introduction to Hankel operators*. London Math. Soc., Stud. Texts 13, 1988.
- [78] T. Penzl. Algorithms for model reduction of large dynamical systems. *Linear Algebra and its Applications*, 415:322343, 2006.
- [79] L. Pillage and R. A. Rohre. Asymptotic waveform evaluation for timing analysis. *IEEE Trans. on Computer-Aided Design*, 9(4):352–366, 1990.
- [80] J. Rommes and N. Martins. Efficient computation of transfer function dominant poles using subspace acceleration. *IEEE Transactions on Power Systems*, 21:1218–1226, 2006.
- [81] M. Rosenblum. On the operator equation $bx?xa = q$. *Duke Math. J*, 23:263270, 1956.
- [82] A. Ruhe. Rational krylov algorithms for nonsymmetric eigenvalue problems. ii. matrix pairs. *Linear Algebra and its Applications*, 197-198:283–295, 1994.

- [83] A. Ruhe. The rational Krylov algorithms for nonsymmetric eigenvalue problems iii. complex shifts for real matrices. *BIT Numerical Mathematics*, 34:165–176, 1994.
- [84] Y. Saad. Krylov subspace methods for solving large unsymmetric linear systems. *Mathematics of computation*, 37:105–126, 1981.
- [85] Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Manchester University Press, Manchester, UK, 1992.
- [86] Y. Saad. *Iterative methods for sparse linear systems (2nd edition)*. PWS, 1996.
- [87] Y. Saad and M. Schultz. Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7:856–869, 1986.
- [88] D. Salkuyeh and F. Toutounian. New approaches for solving large Sylvester equations. *Applied Mathematics and Computations*, 173:9–18, 2006.
- [89] Y. Shamash. Stable reduced-order models using padé-type approximations. *IEEE Trans. Automat. Control*, 19(5):615–616, 1974.
- [90] V. Simoncini. A new iterative method for solving large-scale Lyapunov matrix equations. *SIAM Journal on Scientific Computing*, 29(3):1268–1288, 2008.
- [91] V. Simoncini and D. B. Szyld. Recent computational developments in krylov subspace methods for linear systems. *Numerical Linear Algebra with Applications*, 14:159, 2007.
- [92] J. T. Spanos, M. H. Milman, and D. L. Mingori. A new algorithm for l2 optimal model reduction. *Automatica*, Vol. 28:897–909, 1992.
- [93] J. Stefanovski and G. Dimirovski. A new approach to static output feedback stabilization of linear dynamic systems. *International Journal of Systems Science*, 37:643–662, 2006.
- [94] H. Tuan, P. Apkarian, and Y. Nakashima. A new lagrangian dual global optimization algorithm for solving bilinear matrix inequalities. *International Journal of Robust and Nonlinear Control*, 10:561–578, 2000.
- [95] H. van der Vorst. Krylov subspace iteration, in top 10 algorithms of the 20th century. *Comput. Sci. Engrg.*, 2:32–37, 2000.
- [96] A. Varga. Enhanced modal approach for model reduction. *Mathematical Modelling of Systems*, 1:91–105, 1995.

- [97] C. D. Villemagne and R. Skelton. Model reduction using a projection formulation. *Int. J. Control*, 40:2141–2169, 1987.
- [98] G. Wang, V. Sreeram, and W. Q. Liu. A new frequency weighted balanced truncation method and an error bound. *IEEE Transactions on Automatic Control*, 44:1734–1737, 1999.
- [99] D. Wilczek, A. Bunse-Gerstner, D. Kubalinska, and G. Vossen. Equivalence of necessary optimality conditions for $\mathcal{H}_2$ -optimal model reduction. Technical report, ZeTeM Technical report, University of Bremen, 2007.
- [100] J. Willems. Least squares stationary optimal control and the algebraic Riccati equation. *Automatic Control on IEEE Transactions*, 16:621–634, 1971.
- [101] D. A. Wilson. Optimum solution of model-reduction problem. *Proc. IEEE*, 117:1161–1165, 1970.
- [102] W. Yan and J. Lam. An approximate approach to H^2 optimal model reduction. *IEEE Transactions on Automatic Control*, 44:1341–1358, July 1999.
- [103] K. Zhou. Frequency-weighted $\mathcal{L}_\infty$ norm and optimal Hankel norm model reduction. *IEEE Transactions on Automatic Control*, 40:1687–1699, 1995.
- [104] K. Zhou, J. Doyle, and K. Glover. *Robust and optimal control*. Prentice-Hall, Upper Saddle River, NJ, 1996.
- [105] D. Zigic, L. Watson, and C. Beattie. Contragradient transformations applied to the optimal projection equations. *Linear Algebra Applications*, 188:665–676, 1993.