

# An iterative SVD-Krylov based method for model reduction of large-scale dynamical systems\*

Serkan Gugercin

Department of Mathematics, Virginia Tech., Blacksburg, VA, USA, 24061-0123

gugercin@math.vt.edu

## Abstract

In this note, we propose a model reduction algorithm for approximation of large-scale linear time-invariant dynamical systems. The method is a two-sided projection combining features of the singular value decomposition (SVD) and Krylov based model reduction techniques. While the SVD-side of the projection depends on the observability gramian, the Krylov side is obtained via iterative rational Krylov steps. The reduced model is asymptotically stable, matches certain moments and solves a restricted  $\mathcal{H}_2$  minimization problem. We present modifications to the proposed approach for employing low-rank gramians in the reduction step and also for reducing discrete-time systems. Several numerical examples from various disciplines verify the effectiveness of the proposed approach. It consistently yields smaller  $\mathcal{H}_2$  error norm than balanced truncation and a satisfactory  $\mathcal{H}_\infty$  performance, better than or close to that of balanced truncation. Moreover, the method proves to be robust with respect to perturbations due to usage of approximate gramians.

## 1 Introduction

Dynamical systems are the basic framework for modeling and control of an enormous variety of complex systems. Examples include heat transfer, temperature control in various media, signal propagation and interference in electric circuits, wave propagation and vibration suppression in large structures; and behavior of micro-electro-mechanical systems. Direct numerical simulation of the associated models has been one of the few available means for studying complex underlying physical phenomena. However, the ever increasing need for improved accuracy requires the inclusion of ever more detail in the modeling stage, leading inevitably to ever larger-scale, ever more complex dynamical systems. Simulations in such large-scale settings often lead to unmanageably large demands on computational resources, which is the main motivation for model reduction. The goal is to produce a much lower dimensional system with input/output behavior close to the original one.

In this paper, we consider a single-input/single-output (SISO) linear time invariant (LTI) system  $\mathbf{G}(s)$  given in state space form as:

$$\mathbf{G}(s) : \begin{cases} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{c}\mathbf{x}(t), \end{cases} \quad \Leftrightarrow \quad \mathbf{G}(s) := \left[ \begin{array}{c|c} \mathbf{A} & \mathbf{b} \\ \hline \mathbf{c} & 0 \end{array} \right], \quad (1.1)$$

where  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{b} \in \mathbb{R}^n$ , and  $\mathbf{c}^T \in \mathbb{R}^n$ . In (1.1),  $\mathbf{x}(t) \in \mathbb{R}^n$  is the *state*,  $\mathbf{u}(t) \in \mathbb{R}$  is the *input*, and  $\mathbf{y}(t) \in \mathbb{R}$  is the *output* of  $\mathbf{G}(s)$ . The transfer function of  $\mathbf{G}(s)$  is given by  $\mathbf{G}(s) = \mathbf{c}(s\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{b}$ . Following traditional abuse of notation, we note that both the underlying dynamical system and its transfer function are denoted by  $\mathbf{G}(s)$ . In the sequel, we will assume that the full-order model  $\mathbf{G}(s)$  is *asymptotically stable*, i.e.  $\text{Real}(\lambda_i(\mathbf{A})) < 0$  for  $i = 1, \dots, n$ ; and is *minimal*, i.e. reachable and observable. We call a dynamical system *stable*, if  $\text{Real}(\lambda_i(\mathbf{A})) \leq 0$  for  $i = 1, \dots, n$ ; that is, unlike an *asymptotically stable* dynamical system, a *stable* one might have a pole on the imaginary axis.

---

\*This work is supported in part by the NSF through Grants DMS-050597 and DMS-0513542, and the AFOSR through Grant FA9550-05-1-0449..

The goal of model reduction, in this setting, is to produce a much smaller order system  $\mathbf{G}_r(s)$  with state-space form:

$$\mathbf{G}_r(s) : \begin{cases} \dot{\mathbf{x}}_r(t) = \mathbf{A}_r \mathbf{x}_r(t) + \mathbf{b}_r \mathbf{u}(t) \\ \mathbf{y}_r(t) = \mathbf{c}_r \mathbf{x}(t), \end{cases} \Leftrightarrow \mathbf{G}_r(s) := \left[ \begin{array}{c|c} \mathbf{A}_r & \mathbf{b}_r \\ \hline \mathbf{c}_r & 0 \end{array} \right], \quad (1.2)$$

where  $\mathbf{A}_r \in \mathbb{R}^{r \times r}$ ,  $\mathbf{b}_r \in \mathbb{R}^r$ , and  $\mathbf{c}_r^T \in \mathbb{R}^r$  (with  $r \ll n$ ), such that the reduced system  $\mathbf{G}_r(s)$  will have approximately the same response (output) as the original system to any given input  $\mathbf{u}(t)$ , i.e.  $\mathbf{y}_r(t)$  approximates  $\mathbf{y}(t)$  well.

In this note, we will construct the reduced order models  $\mathbf{G}_r(s)$  through projection:  $\mathbf{G}_r(s)$  in (1.2) will be obtained as

$$\mathbf{A}_r = \mathbf{Z}^T \mathbf{A} \mathbf{V}, \quad \mathbf{b}_r = \mathbf{Z}^T \mathbf{b}, \quad \text{and} \quad \mathbf{c}_r = \mathbf{c} \mathbf{V}. \quad (1.3)$$

where  $\mathbf{V} \in \mathbb{R}^{n \times r}$  and  $\mathbf{Z} \in \mathbb{R}^{n \times r}$  such that  $\mathbf{Z}^T \mathbf{V} = \mathbf{I}_r$ . The corresponding oblique projector is given by  $\mathbf{V} \mathbf{Z}^T$ .

The model reduction algorithms we will consider can be put under three categories, namely

- (a) **SVD (Gramian) based methods,**
- (b) **Krylov (moment matching) based methods,** and
- (c) **SVD-Krylov based methods.**

In SVD-based model reduction, the projection  $\mathbf{V} \mathbf{Z}^T$  depends on the reachability and/or observability gramians. The Hankel singular values, singular values of the Hankel operator associated with  $\mathbf{G}(s)$ , are the key ingredients in this category and play a similar role to that of the singular values in the optimal 2-norm approximation of constant matrices. *Balanced Truncation* [37, 36], is the most common SVD-based method. When applied to asymptotically stable systems, it preserves asymptotic stability and provides an *a priori* bound on the approximation error. However, since *exact balanced truncation* requires dense matrix factorizations, its computational complexity is  $\mathcal{O}(n^3)$  and is expensive to implement in large-scale settings. Hence, in such cases, one uses approximate low-rank versions of balanced truncation [22, 38, 32, 10]. For more detail on efficient implementations of balancing related model reduction in large-scale settings, see [22, 10, 44, 1, 38]. *Optimal Hankel Norm Approximation* [16], and *Balanced Singular Perturbation Approximation* [33] are two other common SVD-based techniques.

The main ingredients for the Krylov based methods are the *moments* of  $\mathbf{G}(s)$ . The  $k^{\text{th}}$  moment of  $\mathbf{G}(s)$  at a point  $s_0 \in \mathbb{C}$  is the  $k^{\text{th}}$  derivative of  $\mathbf{G}(s)$  at  $s_0$ . Krylov-based model reduction constructs a reduced model  $\mathbf{G}_r(s)$  that *interpolates* certain number of moments of  $\mathbf{G}(s)$  at selected interpolation points. Under this category, we list the Arnoldi [6] and Lanczos procedures [34], and rational Krylov method [17, 41, 14, 12]. Compared to the SVD-based methods, these methods are numerically more reliable and can be implemented iteratively; the number of computations is of  $\mathcal{O}(nr^2)$  and the storage requirement is of  $\mathcal{O}(nr)$ . Also, the asymptotic stability of the reduced model can be obtained through restarting [18]. But there exists no *a priori* error bounds. However, recently in [21, 23], a global *error expression* has been developed for the Krylov-based methods.

Recently much research has been done to obtain a model reduction algorithm which connects the SVD and Krylov based methods; see, for example, [1, 2, 39, 15, 22, 24]. The goal of these works is to *combine* the theoretical features of the SVD based methods such as stability, global error bounds, with efficient numerical implementation of the Krylov-based methods. In this paper, we propose a model reduction algorithm which achieves this goal. The method is a two-sided projection method where one side reflects the Krylov part of the algorithm, and the other side reflects the SVD (Gramian) part. The reduced model is asymptotically stable, solves a restricted  $\mathcal{H}_2$  minimization problem and matches certain moments.

The rest of the paper is organized as follows: In Section 2, we review the basic facts related to model reduction problem in our setting. Section 3 describes the proposed method and presents the main results of the paper followed by numerical examples in Sections 4. Conclusions are given in Section 5.

## 2 Some Preliminaries

As stated above, the proposed method carries both gramian (SVD) and moment matching (Krylov) information. Hence, in this section, we review some basic facts related to these concepts.

## 2.1 $\mathcal{H}_\infty$ and $\mathcal{H}_2$ Norm of a Dynamical System

The two main system norms to measure how well  $\mathbf{G}_r(s)$  approximates  $\mathbf{G}(s)$  are the  $\mathcal{H}_\infty$  and  $\mathcal{H}_2$  norms:

**Definition 2.1** Let  $\mathbf{G}(s)$  be an asymptotically stable SISO system as in (1.1). The  $\mathcal{H}_\infty$  norm of  $\mathbf{G}(s)$  is defined as  $\|\mathbf{G}\|_{\mathcal{H}_\infty} := \sup_{w \in \mathbb{R}} |\mathbf{G}(jw)|$ . On the other hand, the  $\mathcal{H}_2$  norm is given by  $\|\mathbf{G}\|_{\mathcal{H}_2} := \left( \int_{-\infty}^{+\infty} |\mathbf{G}(jw)|^2 dw \right)^{1/2}$ .

The  $\mathcal{H}_\infty$  norm of the error system  $\mathbf{G}_e(s) := \mathbf{G}(s) - \mathbf{G}_r(s)$  is a measure of the *worst case*  $\mathcal{L}_2$  output error  $\|\mathbf{y}(t) - \mathbf{y}_r(t)\|$  over all unit energy inputs; whereas  $\mathcal{H}_2$  norm of the error system measures the energy of the output error aggregated over an orthogonal family of unit energy inputs. In Section 4, we will compare the performance of the proposed approach with other model reduction techniques via both  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  error measures.

## 2.2 Lyapunov Equations, System Gramians and Balanced Truncation

Given  $\mathbf{G}(s)$  as in (1.1),  $\mathbf{P}$  and  $\mathbf{Q}$ , solutions to the following Lyapunov equations

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{b}\mathbf{b}^T = 0, \quad \mathbf{A}^T\mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{c}^T\mathbf{c} = 0, \quad (2.1)$$

are called the *reachability* and *observability gramians*, respectively. Under the assumption that  $\mathbf{G}(s)$  is asymptotically stable and minimal,  $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{n \times n}$  are unique symmetric positive definite matrices. Gramians play a crucial role in SVD-based model reduction such as in applying balanced truncation as explained below. The square roots of the eigenvalues of the product  $\mathbf{P}\mathbf{Q}$  are the singular values of the Hankel operator associated with  $\mathbf{G}(s)$  and are called the Hankel singular values  $\sigma_i(\mathbf{G}(s))$  of  $\mathbf{G}(s)$ :  $\sigma_i(\mathbf{G}(s)) = \sqrt{\lambda_i(\mathbf{P}\mathbf{Q})}$ . The Hankel singular values and the rate they decay are the critical components of SVD-based model reduction. In most cases,  $\sigma_i(\mathbf{G}(s))$  decay very rapidly. The faster they decay the easier to reduce  $\mathbf{G}(s)$ . For more discussion on this issue, see [5, 4].

Let  $\mathbf{P} = \mathbf{U}\mathbf{U}^T$  and  $\mathbf{Q} = \mathbf{L}\mathbf{L}^T$ . Also, let  $\mathbf{U}^T\mathbf{L} = \mathbf{W}\mathbf{S}\mathbf{Y}^T$  be the singular value decomposition with  $\mathbf{S} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ . Let  $\mathbf{S}_r = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ ,  $r < n$ . Construct

$$\mathbf{Z} = \mathbf{L}\mathbf{Y}_r\mathbf{S}_r^{-1/2} \quad \text{and} \quad \mathbf{V} = \mathbf{U}\mathbf{W}_r\mathbf{S}_r^{-1/2} \quad (2.2)$$

where  $\mathbf{W}_r$  and  $\mathbf{Y}_r$  denote the leading  $r$  columns of  $\mathbf{W}$  and  $\mathbf{Y}$  respectively; hence  $\mathbf{Z}, \mathbf{V} \in \mathbb{R}^{n \times r}$ . Then the  $r^{\text{th}}$  order reduced order model  $\mathbf{G}_r(s) = \mathbf{c}_r(\mathbf{s}\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{b}_r$  via *balanced truncation* is obtained by reducing  $\mathbf{G}(s)$  using  $\mathbf{Z}$  and  $\mathbf{V}$  in (2.2), i.e.  $\mathbf{A}_r = \mathbf{Z}^T\mathbf{A}\mathbf{V}$ ,  $\mathbf{b}_r = \mathbf{Z}^T\mathbf{b}$ ,  $\mathbf{c}_r = \mathbf{c}\mathbf{V}_r$ . The reduced system  $\mathbf{G}_r(s)$  due to balanced truncation is asymptotically stable and the  $\mathcal{H}_\infty$  norm of the error system satisfies

$$\|\mathbf{G}(s) - \mathbf{G}_r(s)\|_{\mathcal{H}_\infty} \leq 2(\sigma_{r+1} + \dots + \sigma_n). \quad (2.3)$$

For details, see [5].

## 2.3 Krylov (Moment Matching) based model reduction

In model reduction by moment matching (also called multi-point rational interpolation), one seeks a reduced model  $\mathbf{G}_r(s)$  that interpolates  $\mathbf{G}(s)$  as well as a certain number of its derivatives (called moments) at selected points  $s_k$  in the complex plane. In other words, the goal is to find the reduced system matrices  $\mathbf{A}_r$ ,  $\mathbf{b}_r$ , and  $\mathbf{c}_r$  so that

$$\left. \frac{(-1)^j}{j!} \frac{d^j \mathbf{G}(s)}{ds^j} \right|_{s=s_k} = \mathbf{c}(s_k \mathbf{I}_n - \mathbf{A})^{-(j+1)} \mathbf{b} = \mathbf{c}_r(s_k \mathbf{I}_r - \mathbf{A}_r)^{-(j+1)} \mathbf{b}_r = \left. \frac{(-1)^j}{j!} \frac{d^j \mathbf{G}_r(s)}{ds^j} \right|_{s=s_k}$$

for  $k = 1, \dots, K$  and for  $j = 1, \dots, J$  where  $K$  and  $J$  denote, respectively, the number of interpolation points  $s_k$  and the number of moments to be matched at each  $s_k$ . The quantity  $\mathbf{c}(s_k \mathbf{I}_n - \mathbf{A})^{-(j+1)} \mathbf{b}$  is the  $j^{\text{th}}$  moment of  $\mathbf{G}(s)$  at  $s_k$ . If  $s_k = \infty$ , the moments are called Markov parameters and are given by  $\mathbf{c}\mathbf{A}^j\mathbf{b}$  for  $j = 0, 1, 2, \dots$ . In the projection framework, the problem was first treated by Skelton *et. al.* in [12, 47, 46]. Grimme [17] showed how one can obtain the required projection in a numerically efficient way using the rational Krylov method of Ruhe [41], hence showed how to solve moment matching (multi-point rational interpolation) problem using *Krylov projection* methods in an effective way. Before we state this result, we define the Krylov subspace of index  $j$  for a matrix  $\mathbf{F} \in \mathbb{C}^{n \times n}$ , a vector  $\mathbf{g} \in \mathbb{C}^n$ , and a point  $s \in \mathbb{C}$ :

$$\begin{aligned} \mathcal{K}_j(\mathbf{F}, \mathbf{g}; s) &:= \text{Im}([\mathbf{g} \ \mathbf{F}\mathbf{g} \ \mathbf{F}^2\mathbf{g} \ \dots \ \mathbf{F}^{j-1}\mathbf{g}]) && \text{if } s = \infty \\ \mathcal{K}_j(\mathbf{F}, \mathbf{g}; s) &:= \text{Im}([(s\mathbf{I}_n - \mathbf{F})^{-1}\mathbf{g} \ \dots \ (s\mathbf{I}_n - \mathbf{F})^{-j}\mathbf{g}]) && \text{if } s \neq \infty \end{aligned}$$

**Theorem 2.1** [17] *If*

$$\text{Ran}(\mathbf{V}) \supseteq \text{Im}[\mathcal{K}_{j_1}(\mathbf{A}, \mathbf{b}; s_1), \dots, \mathcal{K}_{j_K}(\mathbf{A}, \mathbf{b}; s_K)] \quad \text{and} \quad (2.4)$$

$$\text{Ran}(\mathbf{Z}) \supseteq \text{Im}[\mathcal{K}_{j_{K+1}}(\mathbf{A}^T, \mathbf{c}^T; s_{K+1}), \dots, \mathcal{K}_{j_{2K}}(\mathbf{A}^T, \mathbf{c}^T; s_{2K})], \quad (2.5)$$

with  $\mathbf{Z}^T \mathbf{V} = \mathbf{I}_r$ , then the reduced order model  $\mathbf{G}_r(s)$  in (1.2) matches  $j_k$  number of moments of  $\mathbf{G}(s)$  at the interpolation point  $s_k$  for  $k = 1, \dots, 2K$ , i.e.  $\mathbf{G}_r(s)$  interpolates  $\mathbf{G}(s)$  and its first  $j_k - 1$  derivatives at  $s_k$ .

Theorem [17] states that to solve the multi-point rational interpolation problem by Krylov projection, one needs to construct matrices  $\mathbf{V}$  and  $\mathbf{Z}$  spanning the required rational Krylov subspaces as shown above. Choosing *good/optimal* interpolation points is the most important question in Krylov-based model reduction and until very recently this choice has been usually done in an ad-hoc way. Gugercin and Antoulas [23, 21] introduced a systematic, but not optimal, way of choosing the shifts and showed that this selection strategy worked quite efficiently. In a very recent paper, Gugercin *et al.* [26] has proposed an optimal shift selection strategy for solving optimal  $\mathcal{H}_2$  model reduction problem in a solely Krylov-based setting and has largely resolved the shift selection issue. For more details on Krylov-based model reduction, see [14, 17, 21, 3, 5].

### 3 The Proposed Method

In this section, we propose a SVD-Krylov based model reduction algorithm that produces a reduced model  $\mathbf{G}_r(s)$  by projection as in (1.3) with the matrix  $\mathbf{Z}$  having the specific form:

$$\mathbf{Z} := \mathbf{QV}(\mathbf{V}^T \mathbf{QV})^{-1}, \quad (3.1)$$

where  $\mathbf{Q}$  is the observability gramian as defined in (2.1) and  $\mathbf{V}$  spans a *rational Krylov subspace* as in (2.4). The specific choice of  $\mathbf{V}$  will be explained below. Clearly,  $\mathbf{Z}^T \mathbf{V} = \mathbf{I}_r$  and  $\mathbf{G}_r(s) = \mathbf{c}_r(s\mathbf{I}_r - \mathbf{A}_r)^{-1} \mathbf{b}_r$  is given by

$$\mathbf{G}_r(s) = \left[ \begin{array}{c|c} \mathbf{A}_r & \mathbf{b}_r \\ \mathbf{c}_r & 0 \end{array} \right] = \left[ \begin{array}{c|c} (\mathbf{V}^T \mathbf{QV})^{-1} \mathbf{V}^T \mathbf{QAV} & (\mathbf{V}^T \mathbf{QV})^{-1} \mathbf{V}^T \mathbf{Qb} \\ \mathbf{cV} & 0 \end{array} \right] \quad (3.2)$$

In the reduction step (3.2),  $\mathbf{V}$  reflects the Krylov-side of the algorithm and  $\mathbf{Z}$  reflects the SVD-side. With the choice of  $\mathbf{Z}$  as in (3.1), the quality of the approximant  $\mathbf{G}_r(s)$  critically depends on modeling subspace  $\mathbf{V}$ ; consequently the interpolation points  $s_k$  used to form  $\mathbf{V}$ . In this note, in constructing the rational Krylov subspace  $\mathbf{V}$ , we will choose the interpolation points in an (sub)optimal way based on the following theorem, an extension of Gaier's result [13] to continuous time.

**Theorem 3.1** *Given a stable SISO transfer function  $\mathbf{G}(s)$  as in (1.1), and fixed stable reduced poles  $\alpha_1, \dots, \alpha_r$ , define  $\mathbf{G}_r(s) := \frac{\beta_0 + \beta_1 s + \dots + \beta_{r-1} s^{r-1}}{(s - \alpha_1) \dots (s - \alpha_r)}$ . Then  $\|\mathbf{G} - \mathbf{G}_r\|_{\mathcal{H}_2}$  is minimized if and only if*

$$\mathbf{G}(s) = \mathbf{G}_r(s) \quad \text{for } s = -\bar{\alpha}_1, -\bar{\alpha}_2, \dots, -\bar{\alpha}_r. \quad (3.3)$$

Since the poles,  $\{\alpha_i\}$  occur in complex conjugate pairs, (3.3) can be rewritten as  $\mathbf{G}(s) = \mathbf{G}_r(s)$  for  $s = -\alpha_1, \dots, -\alpha_r$ . Theorem 3.1 states that if  $\mathbf{G}_r(s)$  interpolates  $\mathbf{G}(s)$  at the mirror images of the poles of  $\mathbf{G}_r(s)$ , then  $\mathbf{G}_r(s)$  is guaranteed to be an *optimal* approximation of  $\mathbf{G}(s)$  with respect to the  $\mathcal{H}_2$  norm among all reduced order systems having the same reduced system poles  $\{\alpha_i\}$ ,  $i = 1, \dots, r$ . For the shift selection for *general*, unconstrained (without fixed reduced poles) Krylov-based optimal  $\mathcal{H}_2$  approximation problem, see the recent paper by Gugercin *et al.* [26].

Theorem 3.1 classifies an *optimal* shift selection strategy as the mirror images of the poles of  $\mathbf{G}_r(s)$ , i.e. as the mirror images of the eigenvalues of  $\mathbf{A}_r$ . However, since these reduced poles are not known *a priori*, one *cannot* simply set  $s_i = -\lambda_i(\mathbf{A}_r)$  and successive rational Krylov steps are needed. Hence, inspired by [26], we propose to run iterative rational Krylov steps where at the  $(k+1)^{\text{st}}$  step, interpolation points are chosen as the mirror images of the eigenvalues of  $\mathbf{A}_r$  from the  $k^{\text{th}}$  step. This forms the matrix  $\mathbf{V}$  at each step. Then, the corresponding  $\mathbf{Z}$  matrix is obtained from the formula (3.1). Here is a sketch of the proposed algorithm:

**Algorithm 3.1 An Iterative SVD-Rational Krylov Based Model Reduction Method (ISRK):**

1. Make an initial shift selection  $s_i$ , for  $i = 1, \dots, r$ .
2.  $\text{Ran}(\mathbf{V}) = \text{Span} \{ (s_1 \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{b}, \dots, (s_r \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{b} \}$  with  $\mathbf{V}^T \mathbf{V} = \mathbf{I}_r$ .
3.  $\mathbf{Z} = \mathbf{QV}(\mathbf{V}^T \mathbf{QV})^{-1}$
4. while (not converged)
  - (a)  $\mathbf{A}_r = \mathbf{Z}^T \mathbf{AV}$ ,
  - (b)  $s_i \leftarrow -\lambda_i(\mathbf{A}_r)$  for  $i = 1, \dots, r$
  - (c)  $\text{Ran}(\mathbf{V}) = \text{Span} \{ (s_1 \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{b}, \dots, (s_r \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{b} \}$  with  $\mathbf{V}^T \mathbf{V} = \mathbf{I}_r$ .
  - (d)  $\mathbf{Z} = \mathbf{QV}(\mathbf{V}^T \mathbf{QV})^{-1}$
5.  $\mathbf{A}_r = \mathbf{Z}^T \mathbf{AV}$ ,  $\mathbf{b}_r = \mathbf{Z}^T \mathbf{b}$ ,  $\mathbf{c}_r = \mathbf{cV}$
6.  $\mathbf{G}_r(s) = \mathbf{c}_r (s \mathbf{I}_r - \mathbf{A}_r)^{-1} \mathbf{b}_r$ .

It follows that upon convergence,  $s_i = -\lambda_i(\mathbf{A}_r)$ , for  $i = 1, \dots, r$ ; and hence  $\mathbf{G}_r(s)$  interpolates  $\mathbf{G}(s)$  at the mirror images of the reduced poles, as desired. We note that orthogonalization of  $\mathbf{V}$  in Steps 2 and 4(b) above are for numerical purposes only. Instead, one can simply set  $\mathbf{V} = [(s_1 \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{b}, \dots, (s_r \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{b}]$ .

The following theorem lists the properties of the proposed algorithm:

**Theorem 3.2** *Given an asymptotically stable and minimal dynamical system  $\mathbf{G}(s) = \mathbf{c}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$ , let the reduced model  $\mathbf{G}_r(s)$  be obtained by Algorithm 3.1. Then,  $\mathbf{G}_r(s)$  is asymptotically stable. Also, let  $\alpha_1, \dots, \alpha_r$  denote the poles of  $\mathbf{G}_r(s)$ .  $\mathbf{G}_r(s)$  interpolates  $\mathbf{G}(s)$  at  $-\alpha_i$ , for  $i = 1, \dots, r$ , and therefore minimizes the  $\mathcal{H}_2$  error  $\left\| \mathbf{G} - \tilde{\mathbf{G}} \right\|_{\mathcal{H}_2}$  among all  $r^{\text{th}}$  order reduced models  $\tilde{\mathbf{G}}(s)$  having the same poles  $\alpha_1, \dots, \alpha_r$ .*

PROOF: Without loss of generality, we can assume that  $\mathbf{Q} = \mathbf{I}_n$ . Then  $\mathbf{Z} = \mathbf{V}$  with  $\mathbf{V}^T \mathbf{V} = \mathbf{I}_r$ . Hence, the observability Lyapunov equation for  $\mathbf{G}(s)$  becomes

$$\mathbf{A}^* + \mathbf{A} + \mathbf{c}^* \mathbf{c} = 0 \quad (3.4)$$

where  $*$  denotes complex-conjugate transpose<sup>1</sup>. Multiplying (3.4) by  $\mathbf{V}^*$  from left and  $\mathbf{V}$  from right, we obtain

$$\mathbf{A}_r^* + \mathbf{A}_r + \mathbf{c}_r^* \mathbf{c}_r = 0. \quad (3.5)$$

(3.5) proves that  $\mathbf{I}_r$  is the observability gramian for  $\mathbf{G}_r(s)$ ; and consequently that  $\mathbf{A}_r$  is stable. Due to interpolation upon convergence of Algorithm 3.1, there exists a nonsingular matrix  $\mathbf{K}$  such that

$$\mathbf{AVK} + \mathbf{VKA}_r^* + \mathbf{bb}_r^* = 0. \quad (3.6)$$

Multiplying (3.6) by  $\mathbf{V}^*$  from left leads to

$$\mathbf{A}_r \mathbf{K} + \mathbf{KA}_r^* + \mathbf{b}_r \mathbf{b}_r^* = 0. \quad (3.7)$$

Recall that due to (3.4),  $\mathbf{A}_r$  is stable. To prove that  $\mathbf{A}_r$  is *asymptotically* stable, i.e., it has no poles on the imaginary axis, we use contradiction. Assume that  $\mathbf{A}_r$  has an eigenvalue on the imaginary axis, i.e.

$$\mathbf{z}^* \mathbf{A}_r = \mathbf{z}^* \lambda, \quad \text{where } \lambda = j\omega$$

Let  $\|\mathbf{z}\|_2 = 1$ . Then,

$$\mathbf{z}^* \mathbf{A}_r \mathbf{z} = \lambda \quad \text{and} \quad \mathbf{z}^* \mathbf{A}_r^* \mathbf{z} = \lambda^* = -\lambda.$$

---

<sup>1</sup>To unify the notation, through out the proof, we will use complex-conjugate transpose ( $*$ ) instead of transpose ( $T$ ) even for real parameters.

Multiplying (3.7) by  $\mathbf{z}^*$  and  $\mathbf{z}$  from left and right, respectively, one gets

$$\mathbf{b}_r^* \mathbf{z} = 0. \quad (3.8)$$

Same manipulations for (3.5) lead to

$$\mathbf{c}_r \mathbf{z} = 0. \quad (3.9)$$

Also, multiplying (3.5) by  $\mathbf{z}^*$  from left and using (3.9), we obtain

$$(\mathbf{A}_r + \lambda^* \mathbf{I}_r) \mathbf{z} = 0. \quad (3.10)$$

Similarly, multiplying (3.7) by  $\mathbf{z}$  from right and using (3.8) yields

$$(\mathbf{A}_r + \lambda^* \mathbf{I}_r) \mathbf{K} \mathbf{z} = 0. \quad (3.11)$$

(3.10) and (3.11) mean that

$$\mathbf{K} \mathbf{z} = \alpha \mathbf{z}, \quad \text{where } \alpha \in \mathbb{C}. \quad (3.12)$$

Finally, multiplying (3.6) by  $\mathbf{z}$  from right, we obtain

$$\mathbf{A} \mathbf{V} \mathbf{K} \mathbf{z} = \mathbf{V} \mathbf{K} \mathbf{z} \lambda.$$

This final expression reveals that  $\mathbf{V} \mathbf{K} \mathbf{z}$  is an eigenvector of  $\mathbf{A}$  with the corresponding eigenvalues  $\lambda$ . But

$$\mathbf{c} \mathbf{V} \mathbf{K} \mathbf{z} = \alpha \mathbf{c} \mathbf{V} \mathbf{z} = \alpha \mathbf{c}_r \mathbf{z} = 0$$

due to (3.9). Hence,  $\mathbf{A}$  has an eigenvector in the kernel of  $\mathbf{c}$ , which contradicts the fact that the pair  $(\mathbf{c}, \mathbf{A})$  is observable. Therefore,  $\mathbf{A}_r$  does not have an eigenvalue on the imaginary axis; and consequently  $\mathbf{G}_r(s)$  is asymptotically stable. The second part of the theorem follows from Theorem 3.1.  $\blacksquare$

Before presenting some remarks, we note that in the sequel, both Algorithm 3.1 and **ISRK** will be used to refer to the proposed method.

**Remark 3.1** *The reduced order models of the form similar to (3.2) have appeared in the work of Skelton et. al. in [12, 47, 46]. In [47] and [46], the dual projection is used where  $\mathbf{Q}$  is replaced by  $\mathbf{P}$  and  $\mathbf{V}$  is chosen as the observability matrix of order  $r$  leading to the so-called **q-cover realizations**. On the other hand, in [12], these results were generalized to the case where  $\mathbf{V}$  were replaced by a rational Krylov subspace. However, the proposed algorithm, **ISRK**, is different from these approaches in the specific way we construct  $\mathbf{V}$ , through iterative rational Krylov steps. In these works [12, 47, 46], the reduced model was guaranteed to be only stable, not asymptotically stable; i.e.  $\mathbf{G}_r(s)$  might have a pole on the imaginary axis even though the original model  $\mathbf{G}(s)$  does not. However, **ISRK** guarantees asymptotic stability of  $\mathbf{G}_r(s)$ . Moreover, optimality in the  $\mathcal{H}_2$  sense does not hold in [12, 47, 46], since this optimality requires the interpolation condition (3.3). The numerical examples in Section 4 illustrate that even though the proposed method has a similar structure to that of **q-cover** realization, it performs drastically better due to optimality of the interpolation points.*

**Remark 3.2** *In the discrete-time case, one can apply the projection (3.2) with replacing  $\mathbf{Q}$  by the observability gramian of the corresponding discrete-time systems. This leads to the **least-squares model reduction** approach of Gugercin and Antoulas [24]. Unlike the continuous-time case, regardless of the choice of  $\mathbf{V}$ ,  $\mathbf{G}_r(s)$  is guaranteed to be asymptotically stable. These are the precise reasons that [24] proposed, first, transforming a continuous-time system into discrete-time, applying the least-squares reduction in discrete-time, and then transforming back to continuous time. However, in our proposed approach, we will achieve asymptotic stability while staying in continuous time. In addition, we obtain the optimality in the  $\mathcal{H}_2$  sense due to (3.3), which does not hold for the least-model reduction method [24].*

**Remark 3.3** *Recently, Gugercin et. al [26] proposed a solely Krylov-based iterative algorithm for optimal  $\mathcal{H}_2$  reduction, that has a structure similar to that of Algorithm 3.1. However, the method in [26] does not use the gramian  $\mathbf{Q}$  unlike the proposed method; both  $\mathbf{Z}$  and  $\mathbf{V}$  are rational Krylov subspaces. The main difference is that while the starting point for **ISRK** is the projection structure in (3.1) and (3.2), [26] uses the interpolation based first-order optimality conditions of the optimal  $\mathcal{H}_2$  model reduction [35] as a starting point and generates a reduced model satisfying these conditions. Also, even though an unstable reduced model has been observed extremely rarely, [26] does not guarantee stability since the gramian  $\mathbf{Q}$  is not used in the reduction process.*

We have implemented Algorithm 3.1 for many different large-scale systems. In each of our numerical examples, the algorithm has *always* converged after a small number of steps. Even though we tried to force a convergence failure for the proposed method by making unrealistically poor initial shift selections, it has still succeeded to converge in small number of steps; see Section 4.4. However, despite this overwhelming numerical evidence, a convergence proof of **ISRK** has not been obtained yet and this issue is currently under investigation.

Even though in all of our simulations, *a random initial shift selection for ISRK resulted in a satisfactory reduced model*, in most cases better than those obtained by balanced truncation as shown in Section 4, here we briefly discuss the initialization issue. It is clear that one should make the initial shift selection in the region where the mirror images of the spectrum of  $\mathbf{A}$  lies. This comes from the fact that upon convergence the proposed algorithm yields interpolation points as the mirror images of the reduced system poles, and, as in the eigenvalue computations, these reduced poles will reflect the original pole distribution. One can easily find the eigenvalues of  $\mathbf{A}$  with the smallest and largest real and imaginary parts. Then we suggest choosing shifts in this region. We would like to note that the task of computing the eigenvalues of  $\mathbf{A}$  with the smallest/largest real and imaginary part can be achieved effectively using an implicitly restarted Arnoldi (IRA) algorithm [43]; see numerical examples in Section 4 for more discussion on this issue.

### 3.1 Newton Formulation

Although we have *never* observed a convergence failure for Algorithm 3.1 which uses successive substitution framework  $s_i \leftarrow -\lambda_i(\mathbf{A}_r)$ , in this section we will develop a Newton iteration framework which guarantees local convergence. We will heavily borrow from [26] in deriving the Newton framework.

Let  $\mathbf{s}$  denote the set of interpolation points  $\{s_1, \dots, s_r\}$  and  $\boldsymbol{\lambda}(\mathbf{s})$  denote the resulting reduced order poles  $\widehat{\lambda}_1, \dots, \widehat{\lambda}_r$  obtained by using the projection  $\boldsymbol{\Pi} = \mathbf{V}\mathbf{Z}^T$  where  $\mathbf{V}$  and  $\mathbf{Z}$  are as given in Algorithm 3.1. Since we require  $\mathbf{s} = -\boldsymbol{\lambda}(\mathbf{s})$  upon convergence of Algorithm 3.1, as observed in [26], one can convert the problem into a root finding problem for the function  $\mathbf{f}(\mathbf{s}) = \boldsymbol{\lambda}(\mathbf{s}) + \mathbf{s}$ . Hence, convergence of Algorithm 3.1 amounts to  $\mathbf{f}(\mathbf{s}) = \mathbf{0}$ , which, in return, implies  $\mathbf{s} = -\boldsymbol{\lambda}(\mathbf{s})$  as desired. By this observation, the Newton framework for **ISRK** can be directly obtained by replacing the successive substitution step, i.e. Step 4-(b), of Algorithm 3.1 with the following Newton step:

$$\boxed{\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} - (\mathbf{I}_r + \mathbf{J})^{-1} \left( \mathbf{s}^{(k)} + \boldsymbol{\lambda} \left( \mathbf{s}^{(k)} \right) \right)} \quad (3.13)$$

where  $\mathbf{s}^{(k)}$  and  $\mathbf{s}^{(k+1)}$  are the set of shifts at the  $k^{\text{th}}$  and  $(k+1)^{\text{th}}$  steps, respectively;  $\boldsymbol{\lambda}(\mathbf{s}^{(k)})$  are the reduced poles at  $k^{\text{th}}$  step due to  $\mathbf{s}^{(k)}$ ; and  $\mathbf{J}$  is the Jacobian and represents the sensitivity of  $\boldsymbol{\lambda}(\mathbf{s})$  with respect to  $\mathbf{s}$  at the  $k^{\text{th}}$  step. Note that Step 4-(b) of Algorithm 3.1 becomes equivalent to (3.13) by choosing  $\mathbf{J} = \mathbf{0}$ . Hence, the only missing component in the Newton formulation of **ISRK** is the computation of the Jacobian  $\mathbf{J}$ , which is discussed in the next section.

#### 3.1.1 Jacobian Computation

Since the Jacobian,  $\mathbf{J}$ , measures the sensitivity of  $\boldsymbol{\lambda}(\mathbf{s})$  with respect to  $\mathbf{s}$ , the  $(i, j)^{\text{th}}$  component of  $\mathbf{J}$  is given by  $\frac{\partial \widehat{\lambda}_i}{\partial s_j}$  where  $\widehat{\lambda}_i$  denotes the  $i^{\text{th}}$  reduced order pole, i.e. the  $i^{\text{th}}$  eigenvalue of  $\mathbf{A}_r$ . We assume that  $\widehat{\lambda}_i$  has multiplicity one. The general case follows similarly. Let  $\mathbf{W}$  and  $\mathbf{V}$  be the left and right reducing subspaces for model reduction, not necessarily satisfying  $\mathbf{W}^T \mathbf{V} = \mathbf{I}_r$ . Hence the reduced matrix  $\mathbf{A}_r$  is given by  $\mathbf{A}_r = (\mathbf{W}^T \mathbf{V})^{-1} \mathbf{W}^T \mathbf{A} \mathbf{V}$ . Let  $\widehat{\mathbf{x}}_i$  be an eigenvector of  $\mathbf{A}_r$  with unit length, corresponding to  $\widehat{\lambda}_i$ , i.e.

$$\mathbf{W}^T \mathbf{A} \mathbf{V} \widehat{\mathbf{x}}_i = \widehat{\lambda}_i \mathbf{W}^T \mathbf{V} \widehat{\mathbf{x}}_i \quad (3.14)$$

Then, based on the above set-up, it was shown in [26] that

$$\frac{\partial \widehat{\lambda}_i}{\partial s_j} = \frac{\widehat{\mathbf{x}}_i^T (\partial_j \mathbf{W}^T) \left( \mathbf{A} \mathbf{V} \widehat{\mathbf{x}}_i - \widehat{\lambda}_i \mathbf{V} \widehat{\mathbf{x}}_i \right) + \left( \widehat{\mathbf{x}}_i^T \mathbf{W}^T \mathbf{A} - \widehat{\lambda}_i \widehat{\mathbf{x}}_i^T \mathbf{W}^T \right) (\partial_j \mathbf{V}) \widehat{\mathbf{x}}_i}{\widehat{\mathbf{x}}_i^T \mathbf{W}^T \mathbf{V} \widehat{\mathbf{x}}_i} \quad (3.15)$$

where  $\partial_j \mathbf{W}^T = \frac{\partial}{\partial s_j} \mathbf{W}^T$  and  $\partial_j \mathbf{V} = \frac{\partial}{\partial s_j} \mathbf{V}$ . (3.15) can be easily used for the proposed method by noting that in the **ISRK** framework,  $\mathbf{V} = [(s_1 \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{b}, \dots, (s_r \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{b}]$  and  $\mathbf{W}^T = \mathbf{V}^T \mathbf{Q}$ . Hence, in our case, the partials

$\partial_j \mathbf{W}^T$  and  $\partial_j \mathbf{V}$  are given by

$$\partial_j \mathbf{W}^T = \mathbf{e}_j \mathbf{b}^T (s_j \mathbf{I}_n - \mathbf{A}^T)^{-2} \mathbf{Q} \quad \text{and} \quad \partial_j \mathbf{V} = (s_j \mathbf{I}_n - \mathbf{A})^{-2} \mathbf{b} \mathbf{e}_j^T, \quad (3.16)$$

where  $\mathbf{e}_j$  denotes the  $i^{\text{th}}$  unit vector. Then the Jacobian for the Newton formulation is obtained by combining (3.15) with (3.16). As this analysis illustrates, the Jacobian computation requires solving a small  $r \times r$  generalized eigenvalue problem to compute  $\hat{\lambda}_i$  and  $\hat{\mathbf{x}}_i$ , and  $r$  additional linear solves to compute  $(s_i \mathbf{I}_n - \mathbf{A})^{-2} \mathbf{b}$ . However, since constructing  $\mathbf{V}$  already requires computing  $(s_i \mathbf{I}_n - \mathbf{A})^{-1} \mathbf{b}$ , Jacobian computation does not require additional factorizations, only some additional triangular solves are needed.

## 3.2 Implementation issues in large-scale settings: Use of low-rank gramians

It follows from (3.2) and Algorithm 3.1 that the proposed model reduction method requires computing matrices  $\mathbf{Z}$  and  $\mathbf{V}$  such that

$$\text{Ran}(\mathbf{V}) = \text{Span} \{ (s_1 \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}, \dots, (s_r \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \} \quad \text{and} \quad \text{Ran}(\mathbf{Z}) = \text{Ran}(\mathbf{Q}\mathbf{V}) \quad \text{with} \quad \mathbf{Z}^T \mathbf{V} = \mathbf{I}_r. \quad (3.17)$$

The rational Krylov subspace  $\mathbf{V}$  can be effectively constructed using numerically efficient rational Krylov method [17]. However, in large-scale settings, computing an exact gramian is a numerically ill conditioned problem. Hence, in this section we discuss the effect of replacing the exact gramian  $\mathbf{Q}$  by a low-rank approximation in construction of  $\mathbf{Z}$ .

### 3.2.1 Low-rank approximation to $\mathbf{Q}$

In large-scale settings, computing an exact, full-rank observability gramian  $\mathbf{Q}$  using the standard approach by Bartels-Stewart [7] method as modified by Hammarling [28] is numerically infeasible since this method requires the computation of a Schur decomposition. Therefore, iterative low-rank schemes have been developed in the literature including [29, 30, 31, 38, 22]. The goal is to find a low-rank approximation  $\hat{\mathbf{Q}} = \mathbf{L}\mathbf{L}^T$  to  $\mathbf{Q}$  where  $\mathbf{L} \in \mathbb{R}^{n \times k}$  and  $\hat{\mathbf{Q}} = \mathbf{L}\mathbf{L}^T \approx \mathbf{Q}$ . Effectiveness of the low-rank schemes stems from the fact that solutions to the Lyapunov equations associated with dynamical systems have often low numerical rank [4, 40].

In large-scale settings, we propose to replace  $\mathbf{Q}$  by an effective low-rank approximation  $\mathbf{L}\mathbf{L}^T$  in computing  $\mathbf{G}_r(s)$ . In addition to reducing the computational cost, using a low-rank approximation will reduce the memory requirements drastically as well by avoiding the storage of the dense  $n \times n$  matrix  $\mathbf{Q}$ . Instead, we will store only the low-rank factor  $\mathbf{L} \in \mathbb{R}^{n \times k}$ . With the low-rank approximation  $\hat{\mathbf{Q}} = \mathbf{L}\mathbf{L}^T$ , the approximate  $\mathbf{Z}$ , denoted by  $\hat{\mathbf{Z}}$ , can be computed as

$$\begin{aligned} \hat{\mathbf{Z}} &= \hat{\mathbf{Q}}\mathbf{V}(\mathbf{V}^T \hat{\mathbf{Q}}\mathbf{V})^{-1} \\ &= \underbrace{\mathbf{V}^T \mathbf{L}\mathbf{L}^T}_{:=\mathbf{T}} (\mathbf{V}^T \mathbf{L}\mathbf{L}^T \mathbf{V})^{-1} = \mathbf{T}\mathbf{L}^T (\mathbf{T}\mathbf{T}^T)^{-1} \end{aligned} \quad (3.18)$$

where  $\mathbf{T} \in \mathbb{R}^{r \times k}$ . The low-rank formulation (3.18) illustrates how to compute the reduced-order model  $\mathbf{G}_r(s)$  without ever computing and storing a dense  $n \times n$  matrix.

**Remark 3.4 Effect of using a low-rank gramian on the stability of the reduced model:** *As shown in Theorem 3.2, upon convergence, the proposed algorithm generates an asymptotically stable reduced order model. However, stability of the reduced system is not always guaranteed when the exact gramian  $\mathbf{Q}$  is replaced by the low-rank approximation  $\hat{\mathbf{Q}} = \mathbf{L}\mathbf{L}^T$ . This is similar to the case of approximate balanced truncation [22, 38, 25] where approximate low-rank gramians are used to balance the system and the stability is no longer guaranteed. However, as in the case of approximate balanced truncation [22, 38, 25], in practice, the stability does not seem to be an issue when a low-rank gramian  $\hat{\mathbf{Q}}$  is used. For every numerical example where  $\mathbf{Q}$  is replaced by  $\hat{\mathbf{Q}}$ , we have always obtained an asymptotically stable reduced model. These considerations are illustrated by numerical examples in Section 4.*

Next, we state the properties of the reduced order model resulting from Algorithm 3.1 when  $\mathbf{Q}$  is replaced by a low-rank approximation:

**Theorem 3.3** *Given an asymptotically stable and minimal dynamical system  $\mathbf{G}(s) = \mathbf{c}(s\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{b}$ , let the reduced model  $\widehat{\mathbf{G}}_r(s)$  be obtained by Algorithm 3.1 by replacing the exact gramian  $\mathbf{Q}$  by a low-rank approximation  $\widehat{\mathbf{Q}}$ . Also, let  $\widehat{\alpha}_1, \dots, \widehat{\alpha}_r$  denote the poles of  $\widehat{\mathbf{G}}_r(s)$ . Then  $\widehat{\mathbf{G}}_r(s)$  interpolates  $\mathbf{G}(s)$  at  $-\widehat{\alpha}_i$ , for  $i = 1, \dots, r$ . Moreover, if  $\widehat{\mathbf{G}}_r(s)$  is asymptotically stable, it minimizes the  $\mathcal{H}_2$  error  $\left\| \mathbf{G} - \widetilde{\mathbf{G}} \right\|_{\mathcal{H}_2}$  among all  $r^{\text{th}}$  order reduced models  $\widetilde{\mathbf{G}}(s)$  having the same poles  $\widehat{\alpha}_1, \dots, \widehat{\alpha}_r$ .*

PROOF: Note that the rational Krylov subspace  $\mathbf{V}$  used in computing  $\widehat{\mathbf{G}}_r(s)$  does not depend on the gramian  $\mathbf{Q}$ . Therefore, upon convergence, interpolation at  $-\widehat{\alpha}_i$ , for  $i = 1, \dots, r$  still holds due to Theorem 2.1. The second part of the theorem, i.e.  $\mathcal{H}_2$  optimality, is a direct consequence of Theorem 3.1. ■

Theorem 3.3 states that interpolation property of the proposed method is not affected from using an approximate gramian. Moreover, restricted  $\mathcal{H}_2$  optimality still holds provided that the reduced model is asymptotically stable.

### 3.2.2 Perturbation effects of using low-rank gramians

In case of approximate gramians, an important error measure to quantify is the deviation from the exact reduced model. Below, we will give an exact expression for this error and then discuss the robustness of the proposed method with respect to using low-rank gramians.

**Theorem 3.4** *Given  $\mathbf{G}(s) = \mathbf{c}(s\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{b}$ , let  $\mathbf{Q}$  and  $\widehat{\mathbf{Q}}$  denote, respectively, the exact and approximate observability gramian. For a selection of  $r$  interpolation points  $\{s_i\}_{i=1}^r$ , let the reduced-model  $\mathbf{G}_r(s) = \mathbf{c}_r(s\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{b}_r$  be obtained as in (3.2), i.e.  $\mathbf{A}_r = \mathbf{Z}^T\mathbf{A}\mathbf{V}$ ,  $\mathbf{b}_r = \mathbf{Z}^T\mathbf{b}$ , and  $\mathbf{c}_r = \mathbf{c}\mathbf{V}$  with  $\mathbf{V} = [(s_1\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{b}, \dots, (s_r\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{b}]$ , and  $\mathbf{Z} = \mathbf{Q}\mathbf{V}(\mathbf{V}^T\mathbf{Q}\mathbf{V})^{-1}$ . On the other hand, let  $\widehat{\mathbf{G}}_r(s) = \widehat{\mathbf{c}}_r(s\mathbf{I}_r - \widehat{\mathbf{A}}_r)^{-1}\widehat{\mathbf{b}}_r$  be obtained similarly as  $\widehat{\mathbf{A}}_r = \widehat{\mathbf{Z}}^T\mathbf{A}\mathbf{V}$ ,  $\widehat{\mathbf{b}}_r = \widehat{\mathbf{Z}}^T\mathbf{b}$ , and  $\widehat{\mathbf{c}}_r = \mathbf{c}\mathbf{V}$  where  $\mathbf{V}$  is as before and  $\widehat{\mathbf{Z}} = \widehat{\mathbf{Q}}\mathbf{V}(\mathbf{V}^T\widehat{\mathbf{Q}}\mathbf{V})^{-1}$ , i.e.  $\widehat{\mathbf{Z}}$ , and consequently  $\widehat{\mathbf{G}}_r(s)$ , is constructed via approximate gramian. Then, the error between the two reduced systems is given by*

$$\mathbf{G}_r(s) - \widehat{\mathbf{G}}_r(s) = \mathbf{c}_r(s\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{Z}^T [\mathbf{I}_n - (s\mathbf{I}_n - \mathbf{A})\mathbf{V}(s\mathbf{I}_r - \widehat{\mathbf{A}}_r)^{-1}\widehat{\mathbf{Z}}^T] \mathbf{b}. \quad (3.19)$$

**Proof:** The result follows from observing that  $\mathbf{c}_r = \widehat{\mathbf{c}}_r$ ,  $\mathbf{Z}^T\mathbf{V} = \widehat{\mathbf{Z}}^T\mathbf{V} = \mathbf{I}_r$  and factoring out  $(s\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{Z}^T$  from left and  $\mathbf{b}$  from right in  $\mathbf{G}_r(s) - \widehat{\mathbf{G}}_r(s)$ . ■

The error  $\mathbf{G}_r(s) - \widehat{\mathbf{G}}_r(s)$  in (3.19) has size related to how well  $\mathbf{V}(s\mathbf{I}_r - \widehat{\mathbf{A}}_r)^{-1}\widehat{\mathbf{Z}}^T$  approximates  $(s\mathbf{I}_n - \mathbf{A})^{-1}$ . This is related to the quality of the Ritz approximation  $\mathbf{V}\widehat{\mathbf{A}}_r\widehat{\mathbf{Z}}^T$  to  $\mathbf{A}$ , which greatly depends on the selection of interpolation points  $s_j$ . This means that for a good/optimal selection of interpolation points, we expect  $\mathbf{V}(s\mathbf{I}_r - \widehat{\mathbf{A}}_r)^{-1}\widehat{\mathbf{Z}}^T$  to approximate  $(s\mathbf{I}_n - \mathbf{A})^{-1}$  well, hence consequently,  $\widehat{\mathbf{G}}_r(s)$  to be close to  $\mathbf{G}_r(s)$ . Since the proposed method **ISRK** leads to an (sub)optimal shift selection upon convergence, we believe that it will be robust with respect to perturbations due to using approximate gramian  $\widehat{\mathbf{Q}}$ . These considerations are strongly supported by the two numerical examples in Section 4.3 illustrating that in terms of both  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  error measures, **ISRK** is the most robust one among various methods with respect to low-rank gramian effects. We also note that both  $\mathbf{G}_r(s)$  and  $\widehat{\mathbf{G}}_r(s)$  interpolate  $\mathbf{G}(s)$  at the interpolation points  $\{s_i\}_{i=1}^r$  irrespective of usage of low-rank gramians as shown in Theorem 3.3; hence the error  $\mathbf{G}(s) - \widehat{\mathbf{G}}_r(s)$  in (3.19) is zero for  $s = s_1, \dots, s_r$ . However, additionally, we expect optimal interpolation points to represent a satisfactory global behavior besides exact local matching, consequently leading to small  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  error norms for (3.19) as discussed above and as illustrated in Section 4.3. A similar phenomenon occurs in a different setting in employing inexact solves for solving the the linear systems  $(s_i\mathbf{I}_n - \mathbf{A})\mathbf{x} = \mathbf{b}$  in Krylov-based model reduction as recently examined by Beattie and Gugercin in [9, 27]. [9] illustrates that for *good* (optimal) interpolation points, the reduced model obtained via inexact solves is very close to the reduced model obtained via exact solves, in terms of both  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  behavior. On the other hand, for poorly selected interpolation points, inexact solves lead to considerably high  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  errors in the resulting reduced order model. Hence, good/optimal interpolation points provide not only effective reduced order models, but also robustness to perturbations due to either low-rank gramians as in **ISRK** or inexact solves as in [9].

## 3.3 Discrete-time case

In this section, we will examine the implementation of the proposed method in the discrete-time case. Given an asymptotically stable, SISO discrete-time dynamical system  $\mathbf{G}(z) = \mathbf{c}(z\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{b}$  where  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{b} \in \mathbb{R}^n$ ,

and  $\mathbf{c}^T \in \mathbb{R}^n$ , the goal is to construct a reduced order discrete dynamical system  $\mathbf{G}_r(z) = \mathbf{c}_r(z\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{b}_r$  where  $\mathbf{A}_r \in \mathbb{R}^{r \times r}$ ,  $\mathbf{b}_r \in \mathbb{R}^r$ , and  $\mathbf{c}_r^T \in \mathbb{R}^r$  via a discrete-time implementation of Algorithm 3.1. In discrete-time, asymptotical stability of  $\mathbf{G}(z)$  means that  $|\lambda_i(\mathbf{A})| < 1$ . On the other hand, the  $h_2$  norm of  $\mathbf{G}(z)$  is defined  $\|\mathbf{G}(z)\|_{h_2}^2 = \frac{1}{2\pi} \int_0^{2\pi} |\mathbf{G}(e^{j\theta})|^2 d\theta$ . The necessary modification to Algorithm 3.1 for the discrete-time case implementation is clear from the discrete-time version of Theorem 3.1 due to Gaier[13]: For a discrete-time system  $\mathbf{G}(z)$ , one should apply Algorithm 3.1 by replacing the updating step, i.e. Step 4-(b) with

$$\boxed{s_i \leftarrow \frac{1}{\lambda_i(\mathbf{A}_r)}, \text{ for } i = 1, \dots, r.} \quad (3.20)$$

The resulting reduced discrete-time system has the similar properties as those in the continuous-time case as listed in the next result:

**Corollary 3.1** *Given an asymptotically stable and minimal discrete-time dynamical system  $\mathbf{G}(z) = \mathbf{c}(z\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{b}$ , let the reduced model  $\mathbf{G}_r(z)$  be obtained by Algorithm 3.1 by updating Step 4-(b) as in (3.20). Then,  $\mathbf{G}_r(z)$  is asymptotically stable. Also, let  $\alpha_1, \dots, \alpha_r$  denote the poles of  $\mathbf{G}_r(z)$ .  $\mathbf{G}_r(z)$  interpolates  $\mathbf{G}(z)$  at  $\frac{1}{\alpha_i}$ , for  $i = 1, \dots, r$ , and therefore minimizes the  $h_2$  error  $\|\mathbf{G} - \tilde{\mathbf{G}}\|_{h_2}$  among all  $r^{\text{th}}$  order reduced models  $\tilde{\mathbf{G}}(z)$  having the same poles  $\alpha_1, \dots, \alpha_r$ .*

## 4 Numerical Examples

In this section, we apply the proposed algorithm to several dynamical systems and compare its performance with Balanced Truncation [37, 36], Rational Krylov Method [17], Least-squares Model Reduction [24], and (rational) **q-cover** Projection Method [12]. We note that we have not used the Newton formulation in the numerical examples, since Algorithm 3.1 has always converged after a very small number of iterations.

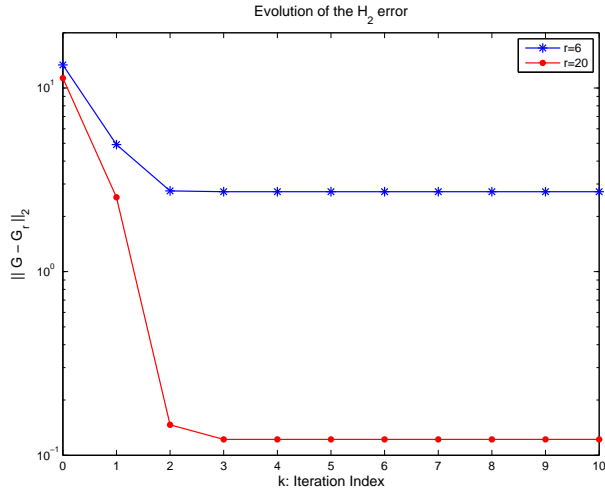
### 4.1 CD Player Model

The original model, obtained by finite elements, describes the dynamics between the lens actuator and the radial arm position of a portable CD player. The model has 120 states, i.e.,  $n=120$ , with a single input and a single output. For more details on this system, see [21, 3, 19, 11].

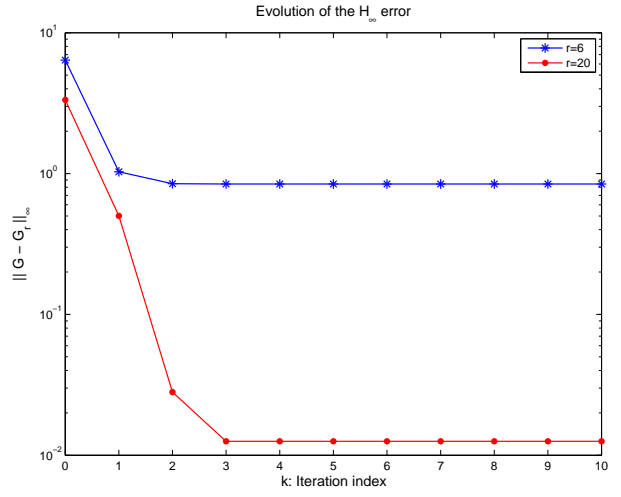
First, we examine convergence behavior of the proposed method. Towards this goal, we reduce the order to  $r = 6$  and  $r = 20$  using Algorithm 3.1. Initial shifts are complex and *randomly* selected in the rectangular region over the complex plane with the real parts of the shifts bounded by  $[-\max_i(\text{Real}(\lambda_i(\mathbf{A}))), -\min_i(\text{Real}(\lambda_i(\mathbf{A})))]$  and the imaginary parts by  $[\min_i(\text{Imag}(\lambda_i(\mathbf{A}))), \max_i(\text{Imag}(\lambda_i(\mathbf{A})))]$ , for  $i = 1, \dots, n$ , reflecting the mirror spectrum of  $\mathbf{A}$ . At each step of the iteration, we compute the  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  errors due to the current estimate and plot this error vs iteration index  $k$ . The results are shown in Figure 1. The figure illustrates that for both cases  $r = 6$  and  $r = 20$  at each step of the iteration, both  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  error norms are reduced. Moreover, in each case, the algorithm converges after 3 steps even for randomly selected initial interpolation points.

#### 4.1.1 Comparison with Balanced Truncation

Since balanced truncation is well known to yield small  $\mathcal{H}_\infty$  and  $\mathcal{H}_2$  error norms; see [3, 19], in this section, we present a detailed comparison between Algorithm 3.1 and balanced truncation in terms of both  $\mathcal{H}_\infty$  and  $\mathcal{H}_2$  error measures. Using both balanced truncation and the proposed approach, we reduce the order to  $r$  as  $r$  varies from 2 to 20 with increments of 2; and for each  $r$  value, we compare the  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  error norms due to balanced truncation and due to **ISRK** for which the initial interpolation points are chosen randomly as explained above. The results are depicted in Figures 2 and 3 below. In both figures,  $\mathbf{G}(s)$  denotes the full-order model. Moreover,  $\mathbf{G}_{\text{ISRK}}(s)$  and  $\mathbf{G}_{\text{bal}}(s)$  denote the reduced models due to **ISRK** and balanced truncation, respectively. While Figure 2-(a) shows the  $\mathcal{H}_2$  error norm vs the reduced order  $r$ , Figure 2-(b) depicts the difference between the  $\mathcal{H}_2$  error norms due to two algorithms, i.e. depicts  $\|\mathbf{G}(s) - \mathbf{G}_{\text{bal}}(s)\|_{\mathcal{H}_2} - \|\mathbf{G}(s) - \mathbf{G}_{\text{ISRK}}(s)\|_{\mathcal{H}_2}$  vs  $r$ . As Figures 2 (a) and (b) illustrate, the proposed algorithm consistently leads to a smaller  $\mathcal{H}_2$  error and outperforms balanced truncation for all  $r$  values. We would like note that this is achieved by a random initial shift selection and solving only one Lyapunov equation. Since the iteration converges in a small number of steps, the cost due to the Krylov side is small; hence overall cost of the proposed method is about half the cost of balanced truncation.

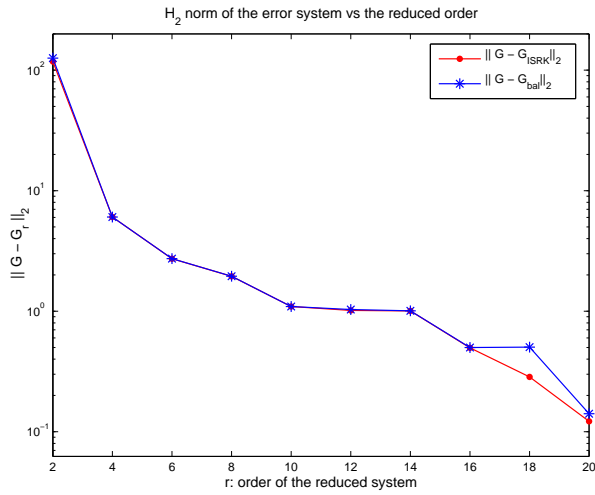


(a)

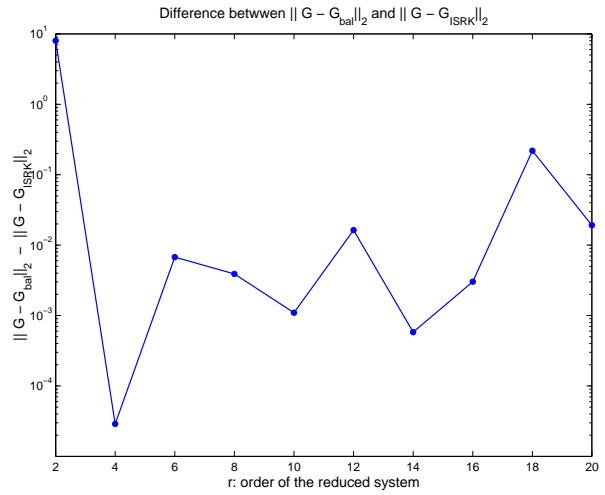


(b)

Figure 1: (a)  $\mathcal{H}_2$  error norm vs the number of iterations (b)  $\mathcal{H}_\infty$  error norm vs the number of iterations



(a)



(b)

Figure 2: (a)  $\mathcal{H}_2$  norm of the error system vs  $r$  (b)  $\|\mathbf{G}(s) - \mathbf{G}_{\text{bal}}(s)\|_{\mathcal{H}_2} - \|\mathbf{G}(s) - \mathbf{G}_{\text{ISRK}}(s)\|_{\mathcal{H}_2}$  vs  $r$

Next, we present the same analysis for the  $\mathcal{H}_\infty$  error.  $\mathcal{H}_\infty$  error norm vs  $r$  for both methods are plotted in Figure 3-(a), and the difference between the  $\mathcal{H}_\infty$  errors, i.e.  $\|\mathbf{G}(s) - \mathbf{G}_{\text{bal}}(s)\|_{\mathcal{H}_\infty} - \|\mathbf{G}(s) - \mathbf{G}_{\text{ISRK}}(s)\|_{\mathcal{H}_\infty}$  is plotted in Figure 3-(b). These figures reveal that the proposed approach yields satisfactory  $\mathcal{H}_\infty$  performance as well. However, unlike the  $\mathcal{H}_2$  case, for some  $r$  values, balanced truncation is slightly better. Although the proposed approach has optimality in the  $\mathcal{H}_2$  sense as shown in Theorem 3.2, good  $\mathcal{H}_\infty$  performance is not surprising as Beattie [8] has recently shown that moment matching at the mirror images of the reduced system poles is *asymptotically* the optimal choice for the  $\mathcal{H}_\infty$  performance as well.

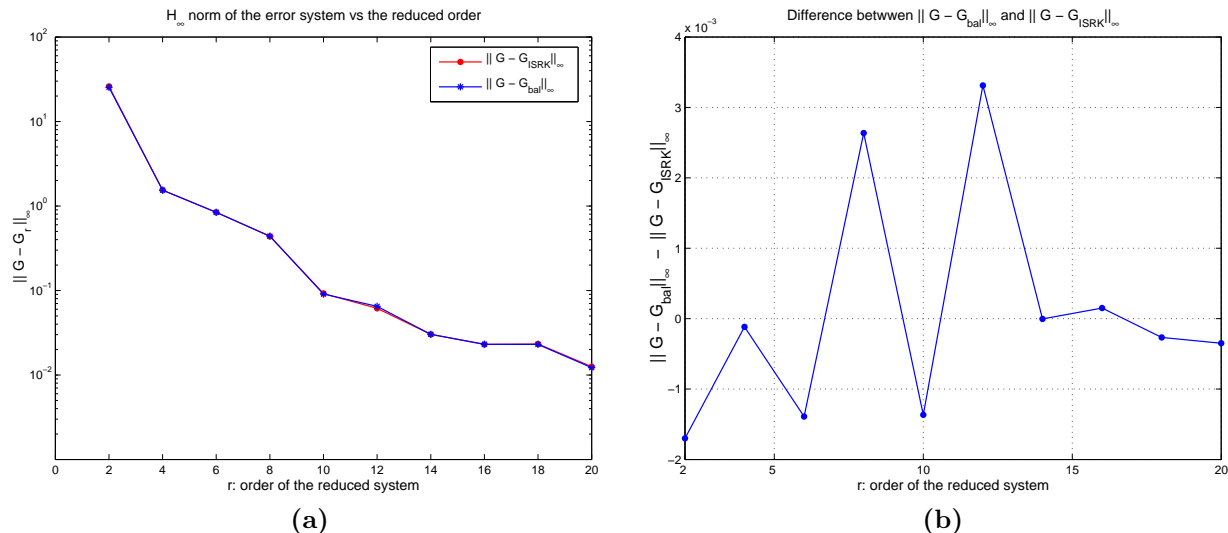


Figure 3: (a)  $\mathcal{H}_\infty$  norm of the error system vs  $r$  (b)  $\|\mathbf{G}(s) - \mathbf{G}_{\text{bal}}(s)\|_{\mathcal{H}_\infty} - \|\mathbf{G}(s) - \mathbf{G}_{\text{ISRK}}(s)\|_{\mathcal{H}_\infty}$  vs  $r$

#### 4.1.2 Comparison with Other Methods

In this section, using the CD player model, we compare the proposed approach (**ISRK**) with the standard Rational Krylov Method (**RK**) [17], Least Squares Model Reduction Algorithm (**LS**) [24], and (Rational) **q-cover** Projection Method<sup>2</sup> [12]. As above, we reduce the order to  $r = 2 : 2 : 20$  and compute both the  $\mathcal{H}_\infty$  and  $\mathcal{H}_2$  error norms for all methods. Interpolation points for **RK**, **LS** and **q-cover** methods are selected randomly as above and these same interpolation points are used as initial guesses for **ISRK**. Figures 4-(a) and 4-(b) below depict the resulting  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  error norms, respectively. The missing data points for **RK** in Figure 4-(a) mean that the reduced models due to **RK** are unstable.

Both figures show that even though all of these four methods have interpolation property, **RK** is very sensitive to the shift selection; the random selection results in a very poor performance and in an unstable reduced model for all cases. On the other hand, with the inclusion of the gramian information **Q** in the reduction step, **LS** and **q-cover** methods solve the stability issue and produce significantly better results compared to **RK**. However, their performance is still far away from that of the proposed method: **ISRK** clearly outperforms **LS** and **q-cover** methods, hence consequently **RK**. This illustrates that even though the method is interpolation based, **ISRK** is much less sensitive to the initial shift selection and successfully converge to the suboptimal interpolation points through the iterative steps.

## 4.2 International Space Station (ISS) 1R Module

The full-order system is a model of stage 1R of the International Space Station. It has 270 states, 3 inputs and 3 outputs. We consider a single-input/single-output (SISO) subsystem of this model corresponding to the first

<sup>2</sup>As mentioned in Remark 3.1, in the original formulation of the **q-cover** realization [46, 47],  $\mathbf{G}_r(s)$  is obtained as in (3.2) where **Q** is replaced by **P** and **V** is taken as the leading  $r$  columns of the observability matrix. However, in our examples, we used a rational Krylov subspace for **V** instead as suggested in [12] which yielded significantly better results than the original **q-cover** realization formulation.

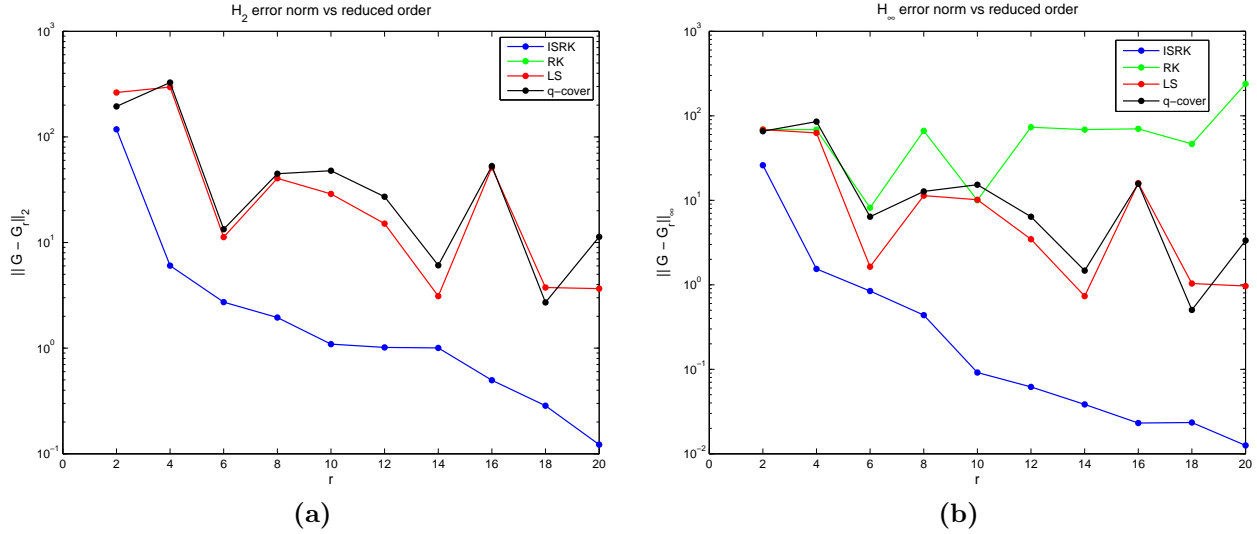


Figure 4: (a)  $\mathcal{H}_2$  norm of the error system vs  $r$  (b)  $\mathcal{H}_\infty$  norm of the error system vs  $r$

input and the first output. For more details on this system, see [20]. We follow the same analysis presented as in Example 4.1.

For convergence behavior, we reduce the order to  $r = 6$  and  $r = 26$  using Algorithm 3.1 with *randomly* selected shifts in the rectangular region bounded by  $[-\max_i(\text{Real}(\lambda_i(\mathbf{A}))), -\min_i(\text{Real}(\lambda_i(\mathbf{A})))]$  and  $[\min_i(\text{Imag}(\lambda_i(\mathbf{A}))), \max_i(\text{Imag}(\lambda_i(\mathbf{A})))]$ , for  $i = 1, \dots, n$ . The  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  errors at each iteration step of the algorithm are shown in Figure 5-(a) and 5-(b), respectively. As in the previous example, the algorithm converges after a very small number of steps and both  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  error norms are reduced throughout the iteration.

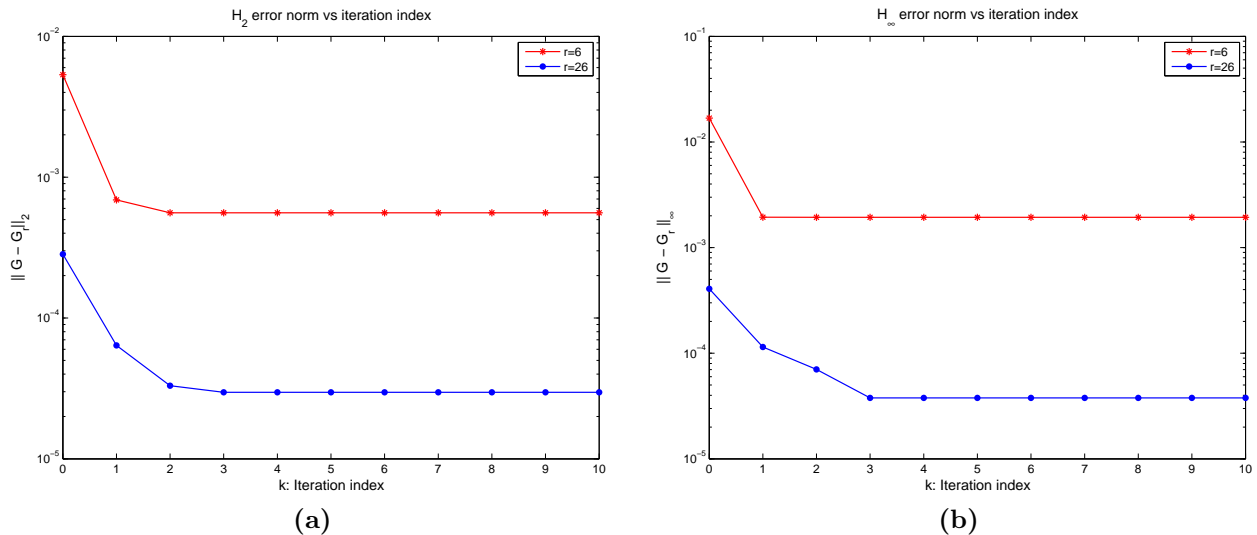


Figure 5: (a)  $\mathcal{H}_2$  error norm vs the number of iterations (b)  $\mathcal{H}_\infty$  error norm vs the number of iterations

#### 4.2.1 Comparison with Balanced Truncation

Using both balanced truncation and the proposed approach, we reduce the order to  $r = 2 : 2 : 40$  and compute the corresponding  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  error norms. The initial interpolation points for **ISRK** are chosen as before. Figure 6-(a) depicts the  $\mathcal{H}_2$  error norm vs the reduced order  $r$  (upper plot) and also the difference  $\|\mathbf{G}(s) - \mathbf{G}_{\text{bal}}(s)\|_{\mathcal{H}_2} -$

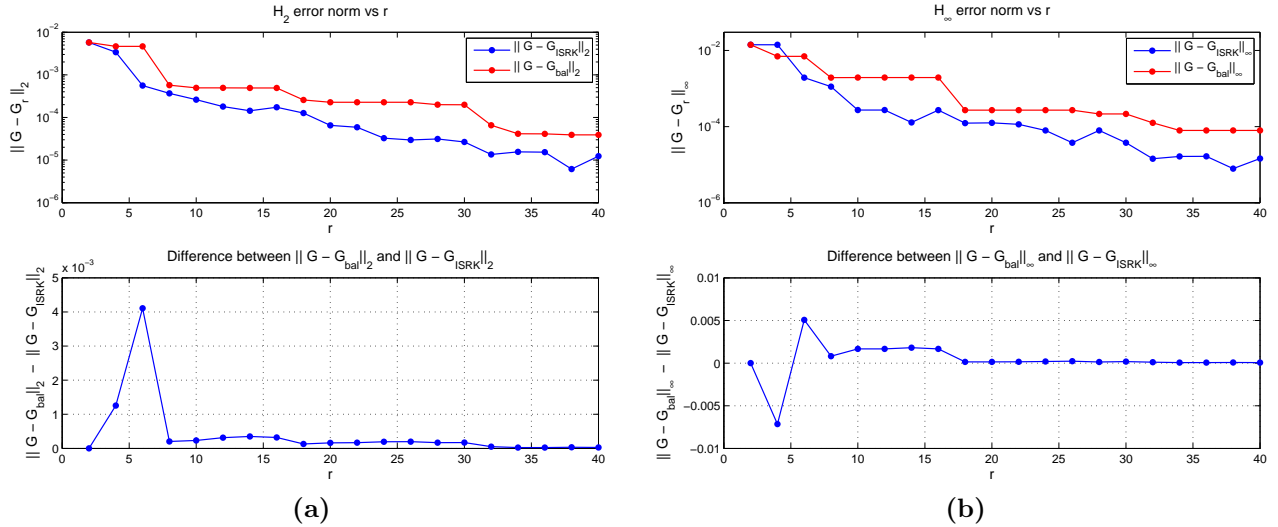


Figure 6: (a)  $\mathcal{H}_2$  norm of the error system vs  $r$  (b)  $\|\mathbf{G}(s) - \mathbf{G}_{\text{bal}}(s)\|_{\mathcal{H}_\infty} - \|\mathbf{G}(s) - \mathbf{G}_{\text{ISRK}}(s)\|_{\mathcal{H}_\infty}$  vs  $r$

$\|\mathbf{G}(s) - \mathbf{G}_{\text{ISRK}}(s)\|_{\mathcal{H}_2}$  vs  $r$  (lower plot). Once again, the proposed method outperforms balanced truncation in terms of  $\mathcal{H}_2$  performance. On the other hand, Figure 6 (a) illustrates the  $\mathcal{H}_\infty$  error norms vs the reduced order  $r$  (upper plot) and the difference  $\|\mathbf{G}(s) - \mathbf{G}_{\text{bal}}(s)\|_{\mathcal{H}_\infty} - \|\mathbf{G}(s) - \mathbf{G}_{\text{ISRK}}(s)\|_{\mathcal{H}_\infty}$  vs  $r$  (lower plot). The figure reveals that except for the  $r = 4$  case, **ISRK** is better than balanced truncation in terms of  $\mathcal{H}_\infty$  performance as well. Balanced truncation is well known to yield small  $\mathcal{H}_\infty$  errors since it uses the both gramians  $\mathbf{P}$  and  $\mathbf{Q}$ , that are closely related to the  $\mathcal{H}_\infty$  performance. However, the fact that **ISRK** yields a better  $\mathcal{H}_\infty$  performance than balancing is not surprising since, as mentioned earlier, interpolation at the mirror images of the reduced system poles is shown to be the asymptotically optimal choice for  $\mathcal{H}_\infty$  model reduction [8] as well.

#### 4.2.2 Comparison with Other Methods

As in Section 4.1.2, we compare **ISRK** with **RK**, **LS**, and **q-cover** methods. Using all four methods, we reduce the order to  $r = 2 : 2 : 40$  and compute both the  $\mathcal{H}_\infty$  and  $\mathcal{H}_2$  error norms. Interpolation points for **RK**, **LS** and **q-cover** methods are selected randomly and the same points are used as an initial guess for **ISRK**. The resulting  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  error norms are illustrated in Figures 7-(a) and 7-(b), respectively. As before, for most  $r$  values, **RK** generated an unstable reduced model leading to missing data points in Figure 7-(a).

Once more, **RK** is the worst among these four methods. By reaching the optimal interpolation points upon convergence, **ISRK** again performs considerably better than **LS** and **q-cover** even though all three methods use a similar SVD-Krylov based projection.

### 4.3 Perturbation Effects of Low-rank Gramians

In this section, we examine the sensitivity of the proposed method to using a low-rank approximation for the exact gramian in large-scale settings. The same sensitivity analysis will be performed for balanced truncation **BT**, least-squares method **LS** and the **q-cover** method as well, and the four methods will be compared. For each method, the exact gramians ( $\mathbf{P}$  and/or  $\mathbf{Q}$ ) will be replaced by a low-rank approximation  $\mathbf{L}\mathbf{L}^T$  where  $\mathbf{L} \in \mathbb{R}^{n \times k}$ . Then, the difference between the two reduced models due to exact and low-rank gramians for the corresponding method will be measured in terms of *relative*  $\mathcal{H}_\infty$  and  $\mathcal{H}_2$  error norms.

#### 4.3.1 Penzl's Model

This example is from [39]. The FOM is a dynamical system of order 1006 with single-input and single-output. State-space realization can be found in [39, 24, 22]. Using the modified Smith method of [22], both gramians are approximated by approximate gramians of rank 14, i.e.  $\mathbf{P} \approx \hat{\mathbf{P}} = \mathbf{U}\mathbf{U}^T$  and  $\mathbf{Q} \approx \hat{\mathbf{Q}} = \mathbf{L}\mathbf{L}^T$  where  $\mathbf{U}, \mathbf{L} \in \mathbb{R}^{1006 \times 14}$ .

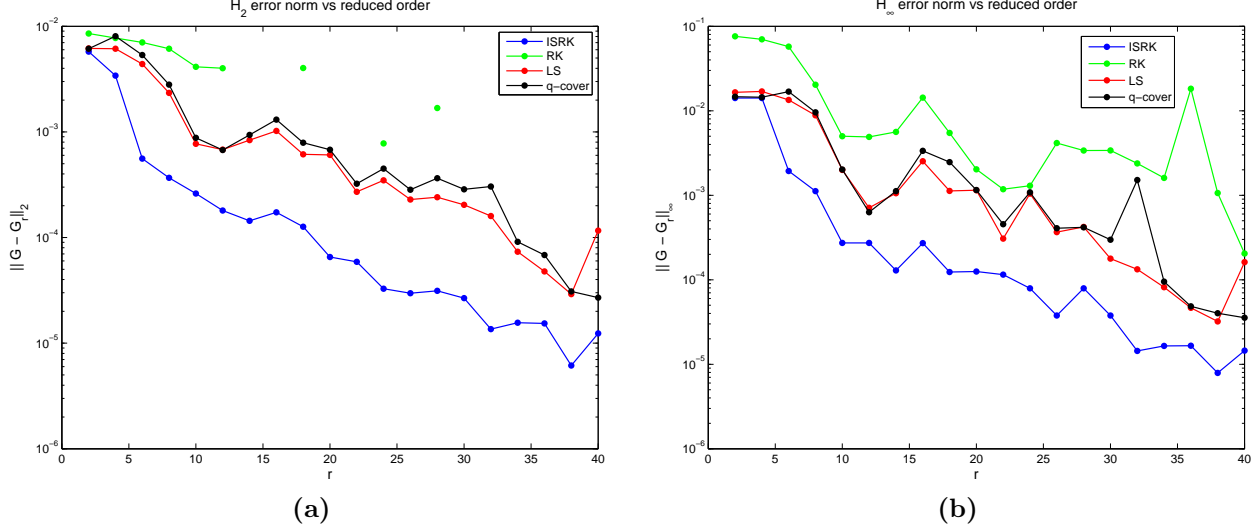


Figure 7: (a)  $\mathcal{H}_\infty$  norm of the error system vs  $r$  (b)  $\|\mathbf{G}(s) - \mathbf{G}_{\text{bal}}(s)\|_{\mathcal{H}_\infty} - \|\mathbf{G}(s) - \mathbf{G}_{\text{ISRK}}(s)\|_{\mathcal{H}_\infty}$  vs  $r$

Recall that  $\hat{\mathbf{P}}$  and  $\hat{\mathbf{Q}}$  are not explicitly computed and stored; only  $\mathbf{U}$  and  $\mathbf{L}$  are stored as discussed in Section 3.2.1. The relative error in the approximate gramians are as follows:

$$\frac{\|\mathbf{P} - \hat{\mathbf{P}}\|_2}{\|\mathbf{P}\|_2} = 5.1522 \times 10^{-6} \quad \frac{\|\mathbf{Q} - \hat{\mathbf{Q}}\|_2}{\|\mathbf{Q}\|_2} = 5.1521 \times 10^{-6}$$

Then, using both exact and low-rank gramians, we reduce the order of the model to  $r = 12$  via **ISRK**, **BT**, **LS** and **q-cover** methods to obtain  $\mathbf{G}_r(s)$  and  $\hat{\mathbf{G}}_r(s)$ , exact and approximate reduced models, for each method. Table 1 below tabulates the relative  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  errors between  $\mathbf{G}_r(s)$  and  $\hat{\mathbf{G}}_r(s)$  for each method:

	<b>ISRK</b>	<b>BT</b>	<b>LS</b>	<b>q-cover</b>
$\mathcal{H}_2$	$3.3364 \times 10^{-9}$	$7.2144 \times 10^{-9}$	$1.5863 \times 10^{-7}$	$3.9515 \times 10^{-7}$
$\mathcal{H}_\infty$	$1.4842 \times 10^{-9}$	$2.2167 \times 10^{-9}$	$4.7013 \times 10^{-8}$	$1.8389 \times 10^{-7}$

Table 1: Relative  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  errors due to low-rank gramians

Table 1 illustrates that among these four methods, **ISRK** is the least sensitive one to using low-rank gramians. On the other hand, the **q-cover** method is the most sensitive one. The relative errors in **BT** are approximately twice as those in **ISRK**. This is expected since in **BT**, the low-rank gramians affect both side of the projection due to  $\mathbf{P}$  and  $\mathbf{Q}$  unlike in **ISRK** where the low-rank gramian affects only one side of the projection. The more important observation is that even though both **LS** and **q-cover** use the same projection structure as **ISRK**, their relative errors are two order of magnitude higher than those of **ISRK**. These results are in strong agreement with the theoretical discussion of Section 3.2.2: Since, upon convergence, **ISRK** leads to (sub)optimal interpolation points, in the error expression (3.19),  $\mathbf{V}(s\mathbf{I}_r - \hat{\mathbf{A}}_r)^{-1}\hat{\mathbf{Z}}^T$  approximates  $(s\mathbf{I}_n - \mathbf{A})^{-1}$  well and  $\hat{\mathbf{G}}_r(s)$  is close to  $\mathbf{G}_r(s)$ ; hence **ISRK** proves to be robust to using low-rank gramians. However, since **LS** and **q-cover** do not have this property regarding the interpolation points, they are more sensitive to these perturbation effects. We note that for all methods, the approximate reduced models  $\hat{\mathbf{G}}_r(s)$  have been asymptotically stable; low-rank gramians have *not* caused instability as discussed in Remark 3.4.

Finally, for **ISRK**, we examine the maximum deviation from the exact optimal interpolation points due to usage of low-rank gramians. Towards this goal, we compute the ratio  $\max_{i=1, \dots, r} \frac{|s_i - \hat{s}_i|}{|s_i|}$ , where  $s_i$  and  $\hat{s}_i$  denote the  $i^{\text{th}}$  interpolation point due to **ISRK** upon convergence, implemented using  $\mathbf{Q}$  and  $\hat{\mathbf{Q}}$ , respectively:

$$\max_{i=1, \dots, r} \frac{|s_i - \hat{s}_i|}{|s_i|} = 2.1224 \times 10^{-7}$$

As this number illustrates, all the interpolation points are correct to, at least, 7 significant digits. This, once more, reveals that employing low-rank gramians in the implementation of **ISRK** has not perturbed the resulting optimal interpolation points; consequently neither the resulting reduced order model.

### 4.3.2 International Space Station (ISS) 12A Module

The full-order system is a model of stage 12A of the International Space Station with 1412 states, 3 inputs and 3 outputs. We consider a single-input/single-output (SISO) subsystem of this model corresponding to the first input and the first output. For more details, see [20].

Both gramians are approximated by low-rank gramians of rank 300, i.e.  $\mathbf{P} \approx \widehat{\mathbf{P}} = \mathbf{U}\mathbf{U}^T$  and  $\mathbf{Q} \approx \widehat{\mathbf{Q}} = \mathbf{L}\mathbf{L}^T$  where  $\mathbf{U}, \mathbf{L} \in \mathbb{R}^{1412 \times 300}$ . The relative error in the approximate gramians are

$$\frac{\|\mathbf{P} - \widehat{\mathbf{P}}\|_2}{\|\mathbf{P}\|_2} = 2.0881 \times 10^{-6} \quad \frac{\|\mathbf{Q} - \widehat{\mathbf{Q}}\|_2}{\|\mathbf{Q}\|_2} = 3.3337 \times 10^{-6}$$

As in the previous case, using both exact and low-rank gramians, we reduce the order of the model to  $r = 14$  via **ISRK**, **BT**, **LS** and **q-cover** methods to obtain  $\mathbf{G}_r(s)$  and  $\widehat{\mathbf{G}}_r(s)$ , exact and approximate reduced models, for each method. The relative  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  errors between  $\mathbf{G}_r(s)$  and  $\widehat{\mathbf{G}}_r(s)$  for each method is listed in Table 2:

	<b>ISRK</b>	<b>BT</b>	<b>LS</b>	<b>q-cover</b>
$\mathcal{H}_2$	$3.4300 \times 10^{-6}$	$4.6224 \times 10^{-5}$	$4.1120 \times 10^{-3}$	$1.6714 \times 10^{-1}$
$\mathcal{H}_\infty$	$4.6249 \times 10^{-6}$	$4.8187 \times 10^{-5}$	$5.4288 \times 10^{-3}$	$2.8166 \times 10^{-1}$

Table 2: Relative  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  errors due to low-rank gramians

Table 2 shows that **ISRK** is the most robust one with respect to low-rank gramian perturbations, for this example as well. **ISRK** is one order of magnitude better than balanced truncation, three orders of magnitude better than **LS** and five orders of magnitude better than **q-cover**, in both  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  error measures. As before, all approximate reduced models are asymptotically stable. For **ISRK**, the maximum deviation from the optimal interpolation point is

$$\max_{i=1, \dots, r} \frac{|s_i - \widehat{s}_i|}{|s_i|} = 4.6065 \times 10^{-7}.$$

Hence, for this example as well, all the interpolation points are correct to, at least, 7 significant digits.

## 4.4 Poor initial shift selection

In this example, we illustrate that even though we try to force Algorithm 3.1 to diverge by making unrealistically poor initial shift selections, the algorithm succeeds to converge. We use the CD player example of Section 4.1 and reduce the order to  $r = 2$ . For the reduced-order  $r = 2$ , **ISRK** yields optimal shifts as  $\sigma_{1,2} = 1.0979 \times 10^1 \pm 3.0285 \times j10^2$ . For three different initial shift selections  $\sigma_0^{(i)}$  for  $i = 1, 2, 3$ , we initiate the algorithm far away from the optimal solution. Since shifts occur in complex-conjugate pairs, we will illustrate only one of the shifts. Three initialization points are chosen as

$$\sigma_0^{(1)} = 10^8 + j10^8, \quad \sigma_0^{(2)} = 10^6 + j10^6, \quad \text{and} \quad \sigma_0^{(3)} = 10^4 + j,$$

Clearly, these selections are away from the optimal solution. Also, since the mirror spectrum of  $\mathbf{A}$  is bounded by  $2.43 \times 10^{-2}$  and  $8 \times 10^2$  in the real axis and by 2.43 and  $4.33 \times 10^4$  in the imaginary axis, all of these three selections are away from the mirror spectrum of  $\mathbf{A}$  as well, once more showing that they are bad candidates as initial shifts. Figure 8 below depicts the convergence behavior of **ISRK** for these three choices by plotting the evolution of both real and imaginary part of the interpolation point: The figure reveals that, in all cases, the proposed method converges to the optimal solution in a small number of steps even for poor initialization points.

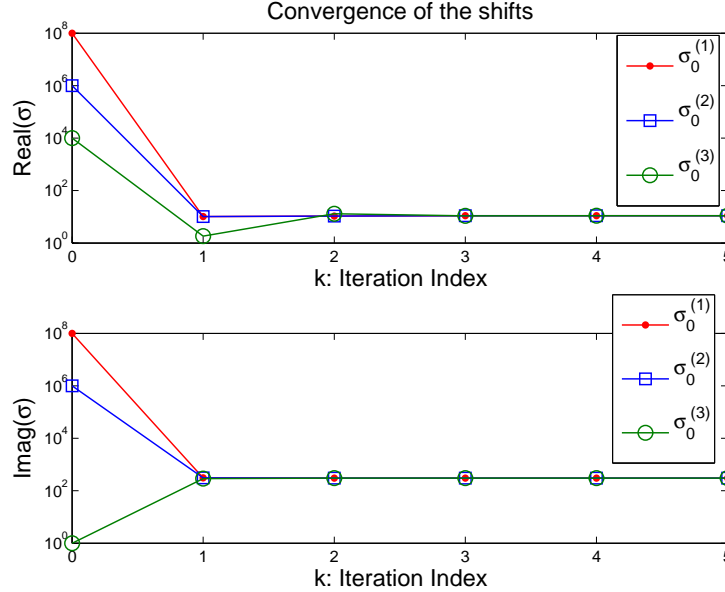


Figure 8: Convergence behavior of **ISRK** for poor initial shift selection

## 5 Conclusions

We have proposed a model reduction algorithm which combines the SVD and Krylov-based methods. It is a two-sided projection method where the SVD-side of the projection depends on the observability gramian and Krylov side is obtained via iterative rational Krylov steps. The reduced model is asymptotically stable and matches the moments of the original system at the mirror images of the reduced system poles; hence it is the best  $\mathcal{H}_2$  approximation among all reduced models having the same reduced system poles. Several numerical example verify the effectiveness of the proposed approach.

## References

- [1] A.C. Antoulas and D.C. Sorensen, *Projection methods for balanced model reduction*, Technical Report ECE-CAAM Depts, Rice University, September 1999.
- [2] A.C. Antoulas and D.C. Sorensen, *The Sylvester equation and approximate balanced reduction*, Fourth Special Issue on Linear Systems and Control, Edited by V. Blondel, D. Hinrichsen, J. Rosenthal, and P.M. van Dooren, *Linear Algebra and Its Applications*, **351-352**: 671-700, 2002.
- [3] A. C. Antoulas, D. C. Sorensen, and S. Gugercin, *A survey of model reduction methods for large scale systems*, *Contemporary Mathematics*, AMS Publications, **280**: 193-219, 2001.
- [4] A.C. Antoulas, D.C. Sorensen, and Y.K. Zhou, *On the decay rate of Hankel singular values and related issues*, to appear in *Systems and Control Letters*, 2002.
- [5] A.C. Antoulas, *Lectures on the approximation of linear dynamical systems*, *Advances in Design and Control DC-06*, SIAM, Philadelphia, 2005.
- [6] W.E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, *Quarterly of Applied Mathematics* **9**, pp: 17-29, 1951.
- [7] R. H. Bartels and G. W. Stewart, *Solution of the matrix equation  $AX + XA = C$ : Algorithm 432*, *Comm. ACM*, **15**: 820-826, 1972.
- [8] C.A. Beattie, *Projection methods for model reduction of dynamical systems*, *SIAM Conference on Computational Science and Eng*, Orlando, USA, February 2005.
- [9] C.A. Beattie and S. Gugercin, *Inexact solves in Krylov-based model reduction*, to appear in the *Proceedings of the 45th IEEE CDC*, San Diego, CA, December 2006.

- [10] P. Benner and E. S. Quintana-Orti. Solving stable generalized Lyapunov equation with the matrix sign function. *Numerical Algorithms*, Vol. 20, pp. 75-100, 1999.
- [11] Y. Chahlaoui and P. Van Dooren, *A collection of benchmark examples for model reduction of linear time invariant dynamical systems*, SLICOT Working Note 2002-2: February 2002. Available from <http://www.win.tue.nl/niconet/NIC2/benchmarks.html>
- [12] C. De Villemagne and R. Skelton, *Model reduction using a projection formulation*, International Jour. of Control, Vol. 40, 2141-2169, 1987.
- [13] D. Gaier, *Lectures on Complex Approximation*, Birkhauser, 1987.
- [14] K. Gallivan, E. Grimme, and P. Van Dooren, *A rational Lanczos algorithm for model reduction*, Numerical Algorithms, 2(1-2):33-63, April 1996.
- [15] K. Gallivan, A. Vandendorpe and P. Van Dooren, *Sylvester equations and projection-based model reduction*, J. Comp. Appl. Math., 162: 213-229, 2004.
- [16] K. Glover. All Optimal Hankel-norm Approximations of Linear Multivariable Systems and their  $L^\infty$ -error Bounds. *Int. J. Control*, 39: 1115-1193, 1984.
- [17] E.J. Grimme, *Krylov Projection Methods for Model Reduction*, Ph.D. Thesis, ECE Dept., U. of Illinois, Urbana-Champaign, 1997.
- [18] E.J. Grimme, D.C. Sorensen, and P. Van Dooren, *Model reduction of state space systems via an implicitly restarted Lanczos method*, Numerical Algorithms, 12: 1-31 (1995).
- [19] S. Gugercin and A. C. Antoulas, *A comparative study of 7 model reduction algorithms*, Proceedings of the 39th IEEE Conference on Decision and Control, Sydney, Australia, December 2000.
- [20] S. Gugercin, A. C. Antoulas and N. Bedrossian, *Approximation of the International Space Station 1R and 12A models*, in the Proceedings of the 40<sup>th</sup> CDC, December 2001.
- [21] S. Gugercin, *Projection methods for model reduction of large-scale dynamical systems*, Ph.D. Dissertation, ECE Dept., Rice University, December 2002.
- [22] S. Gugercin, D.C. Sorensen and A.C. Antoulas, *A modified low-rank Smith method for large-scale Lyapunov equations*, Numerical Algorithms, Vol. 32, Issue 1, pp. 27-55, January 2003.
- [23] S. Gugercin and A.C. Antoulas, *An  $\mathcal{H}_2$  error expression for the Lanczos procedure*, Proceedings of the 42nd IEEE Conference on Decision and Control, December 2003.
- [24] S. Gugercin and A.C. Antoulas, *Model reduction of large scale systems by least squares*, Vol: 415/2-3. pp. 290-321, 2006.
- [25] S. Gugercin and J.R. Li, *Smith-Type Methods for Balanced Truncation of Large Sparse Systems*, in P. Benner, G. Golub, V. Mehrmann, and D.C. Sorensen, editors, Dimension Reduction of Large-Scale Systems, Lecture Notes in Computational Science and Engineering, Springer-Verlag, Berlin/Heidelberg, Vol. 45, (ISBN 3-540-24545-6), 2005.
- [26] S. Gugercin, A.C. Antoulas and C.A. Beattie, *An iterative rational Krylov approach to optimal  $\mathcal{H}_2$  approximation*, Submitted to SIMAX, 2006.
- [27] S. Gugercin and C.A. Beattie, *Inexact solves in Krylov-based model reduction of large-scale dynamical systems*, Ninth Copper Mountain Conference on Iterative Methods,, Copper Mountain, CO, April 2-7, 2006.
- [28] S. Hammarling, *Numerical solution of the stable, non-negative definite Lyapunov equation*, IMA J. Numer. Anal., 2, pp: 303-323, 1982.
- [29] A. S. Hodel, K.P. Poola, and B. Tenison. *Numerical solution of the Lyapunov equation by approximate power iteration*, Linear Algebra Appl., 236: 205-230, 1996.
- [30] D. Y. Hu and L. Reichel, *Krylov subspace methods for the Sylvester equation*, Linear Algebra Appl., 172: 283-313, 1992.
- [31] I. M. Jaimoukha and E. M. Kasenally, *Krylov subspace methods for solving large Lyapunov equations*, SIAM J. Numerical Anal., 31: 227:251, 1994.
- [32] J.-R. Li and J. White. Low rank solution of Lyapunov equations. *SIAM J. Matrix Anal. Appl.*, Vol. 24, No. 1, 2002.
- [33] Y. Liu and B. D. O. Anderson, *Singular Perturbation Approximation of Balanced Systems*, *Int. J. Control*, 50: 1379-1405, 1989.
- [34] C. Lanczos, *An iteration method for the solution of the eigenvalue problem for linear differential and integral operators*, J. Res. Nat. Bur. Standards, 45, pp. 255-282, 1950.
- [35] L. Meier and D.G. Luenberger, *Approximation of Linear Constant Systems*, IEE. Trans. Automat. Contr., Vol. 12, pp. 585-588, 1967.

- [36] B. C. Moore, *Principal Component Analysis in Linear System: Controllability, Observability and Model Reduction*, IEEE Transactions on Automatic Control, AC-26:17-32, 1981.
- [37] C. T. Mullis and R. A. Roberts, *Synthesis of minimum roundoff noise fixed point digital filters*, IEEE Trans. on Circuits and Systems, **CAS-23**, pp: 551-562, 1976.
- [38] T. Penzl. A cyclic low-rank Smith method for large sparse Lyapunov equations. *SIAM J. Sci. Comput.*, Vol. 21, No. 4, pp: 1401-1418, 2000.
- [39] T. Penzl, *Algorithms for model reduction of large dynamical systems*, Technical Report SFB393/99-40, Sonderforschungsbereich 393 Numerische Simulation auf massiv parallelen Rechnern, TU Chemnitz, 09107, FRG, 1999. Available from <http://www.tu-chemnitz.de/sfb393/sfb99pr.html>.
- [40] T. Penzl, *Eigenvalue Decay Bounds for Solutions of Lyapunov Equations: The Symmetric Case*, Systems and Control Letters, **40**: 139-144 (2000).
- [41] A. Ruhe, *Rational Krylov algorithms for nonsymmetric eigenvalue problems II: matrix pairs*, Linear Alg. Appl., **197**:283-295, (1994).
- [42] J.T. Spanos, M.H. Millman, and D.L. Mingori, *A new algorithm for  $L_2$  optimal model reduction*, Automatics, pp. 897-909, 1992.
- [43] D.C. Sorensen, *Implicit application of polynomial filters in a  $k$ -step Arnoldi method*, SIAM J. Matrix Anal. Applic., **13**: 357-385 (1992).
- [44] A. Varga, *Model reduction software in the SLICOT library*, Applied and Computational Control, Signals, and Circuits, Ed. B. Datta", Kluwer Academic Publishers, Boston, 2001.
- [45] D.A. Wilson, *Optimum solution of model reduction problem*, in Proc. Inst. Elec. Eng., pp. 1161-1165, 1970
- [46] A. Yousouff, D. A. Wagie, and R. E. Skelton, *Linear system approximation via covariance equivalent realizations*, Journal of Math. Anal. and App., Vol. **196**, 91-115, 1985.
- [47] A. Yousouff and R.E. Skelton, *Covariance equivalent realizations with applications to model reduction of large-scale systems*, in Control and Dynamic Systems, C.T. Leondes ed., Academic Press, vol. 22, pp. 273-348, 1985.