

RATIONAL KRYLOV METHODS FOR OPTIMAL \mathcal{H}_2 MODEL REDUCTION

S. GUGERCIN*, A. C. ANTOULAS†, AND C. BEATTIE*

Abstract. We develop and describe an iteratively corrected rational Krylov algorithm for the solution of the optimal \mathcal{H}_2 model reduction problem. The formulation is based on finding a reduced order model that satisfies interpolation based first-order necessary conditions for \mathcal{H}_2 optimality and results in a method that is numerically effective and suited for large-scale problems. We provide a new elementary proof of the interpolation based condition that clarifies the importance of the mirror images of the reduced system poles. We also show that the interpolation based condition is equivalent to two types of first-order necessary conditions associated with Lyapunov-based approaches for \mathcal{H}_2 optimality. Under some technical hypotheses, local convergence of the algorithm is guaranteed for sufficiently large model order with a linear rate that decreases exponentially with model order. We illustrate the performance of the method with a variety of numerical experiments and comparisons with existing methods.

Key words. Model Reduction, Krylov Projection, \mathcal{H}_2 Approximation

AMS subject classifications. 34C20,41A05,49K15,49M05,93A15,93C05,93C15,

1. Introduction. Given a dynamical system described by a set of first order differential equations, the model reduction problem seeks to replace this original set of equations with a (much) smaller set of such equations so that the behavior of both systems is similar, in an appropriately defined sense. Such situations arise frequently when physical systems need to be simulated or controlled; the greater the level of detail that is required the greater is the number of resulting equations. In large-scale settings, computations become infeasible due to limitations on computational resources as well as growing inaccuracies due to numerical ill-conditioning. Examples of large-scale systems abound, ranging from the design of VLSI (Very Large Scale Integration) chips, to the simulation and control of MEMS (Micro Electro Mechanical System) devices. In all these cases the number of equations involved may range from a few hundred to a few million. For an overview of model reduction for large-scale dynamical systems we refer to the book [4]. See also [23] for a recent collection of large-scale benchmark problems.

In this paper, we consider single input/single output (SISO) linear systems represented in state-space form:

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

where $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$; n is the *dimension (order)* of the system while $\mathbf{x}(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$, are its *state*, *input*, and *output*, respectively. It will be assumed that the system is *stable*, that is, the eigenvalues of \mathbf{A} have negative real parts. The transfer function of the system is $G(s) = \mathbf{c}^T(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$. With a standard abuse of notation, we denote both the system and its transfer function as G . The model

*Dept. of Mathematics, Virginia Tech., Blacksburg, VA, USA ({gugercin,beattie}@vt.edu). The work of these authors was supported in part by the NSF through Grants DMS-050597 and DMS-0513542; and the AFOSR through Grant FA9550-05-1-0449.

†Dept. of Electrical and Computer Engineering, Rice University, Houston, TX, USA (aca@ece.rice.edu). The work of this author was supported in part by the NSF through Grants CCR-0306503 and ACI-0325081.

reduction process produces another system

$$\begin{aligned} \dot{\mathbf{x}}_r(t) &= \mathbf{A}_r \mathbf{x}_r(t) + \mathbf{b}_r u(t) \\ y_r(t) &= \mathbf{c}_r^T \mathbf{x}_r(t) \end{aligned} \Leftrightarrow G_r(s) = \left[\begin{array}{c|c} \mathbf{A}_r & \mathbf{b}_r \\ \hline \mathbf{c}_r^T & 0 \end{array} \right], \quad (1.2)$$

of smaller dimension $r < n$, with $\mathbf{A}_r \in \mathbb{R}^{r \times r}$ and $\mathbf{b}_r, \mathbf{c}_r \in \mathbb{R}^r$.

It is desirable that as many of the following properties are satisfied as possible.

1. Some appropriate measure of the approximation error $G - G_r$ is small.
2. Critical system properties, such as stability, are preserved.
3. The reduced system can be obtained by means of a computationally effective and numerically stable procedure.

In general, reduced order models are constructed by a Galerkin process. Let $\mathbf{V} \in \mathbb{R}^{n \times r}$ and $\mathbf{Z} \in \mathbb{R}^{n \times r}$ be given so that $\mathbf{Z}^T \mathbf{V} = \mathbf{I}_r$. Then with $\mathbf{x}_r(t) \in \mathbb{R}^r$, $\mathbf{V} \mathbf{x}_r(t) \in \mathbb{R}^n$ will approximate $\mathbf{x}(t)$ by forcing

$$\mathbf{Z}^T (\mathbf{V} \dot{\mathbf{x}}_r(t) - \mathbf{A} \mathbf{V} \mathbf{x}_r(t) - \mathbf{b} u(t)) = 0$$

The reduced order model in (1.2) is then obtained as follows:

$$\mathbf{A}_r = \mathbf{Z}^T \mathbf{A} \mathbf{V}, \quad \mathbf{b}_r = \mathbf{Z}^T \mathbf{b}, \quad \mathbf{c}_r^T = \mathbf{c}^T \mathbf{V}.$$

In this work, the columns of \mathbf{V} and \mathbf{Z} span Krylov subspaces that are chosen to minimize the deviation of $\mathbf{V} \mathbf{x}_r(t)$ from $\mathbf{x}(t)$ uniformly over a large class of inputs $u(t)$. The relevant concepts are detailed in the next section.

The problem of model reduction minimizing an \mathcal{H}_2 error criterion has been the object of many investigations; see for instance [6, 32, 30, 9, 20, 25, 22, 31, 24, 12] and references therein. Finding global minima is a hard task so the goal, as for many optimization problems, is to locate solutions that satisfy first-order necessary conditions for optimality. A reduced-order model that is a global minimizer for the \mathcal{H}_2 error criterion is guaranteed to exist in the single-input/single-output case. However, the existence of global minimizers in the multi-input/multi-output case is still an open question. Most methods that locate solutions that satisfy first-order optimality conditions require dense matrix operations, e.g., solving a series of Lyapunov equations, which rapidly becomes intractable as the dimension increases. Such methods are unsuitable even for medium scale problems. Here, we propose an iterative algorithm which is based on computationally effective use of rational Krylov subspaces. The proposed method is suitable for systems whose dimension n is of the order of many thousands of state variables.

The rest of the paper is organized as follows. In Section 2, we review the moment matching problem for model reduction and its solution by the rational Krylov method. Section 3 describes the main results of the paper and introduces the proposed approach. Sections 4 and 5 contain discussion and proofs of the principal results of the paper. Numerical examples are presented in Section 6. The paper concludes with final remarks and future directions in Section 7.

2. Background.

2.1. Moment matching and rational Krylov methods. Given the system (1.1), reduction by *moment matching* consists in finding a system (1.2) so that $G_r(s)$ interpolates the values of $G(s)$, and perhaps also derivative values as well, at selected points σ_k in the complex plane. For our purposes, simple Hermite interpolation suffices so our problem is to find \mathbf{A}_r , \mathbf{b}_r , and \mathbf{c}_r so that

$$G_r(\sigma_k) = G(\sigma_k) \quad \text{and} \quad G_r'(\sigma_k) = G'(\sigma_k)$$

for $k = 1, \dots, r$ or equivalently

$$\mathbf{c}^T(\sigma_k \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} = \mathbf{c}_r^T(\sigma_k \mathbf{I}_r - \mathbf{A}_r)^{-1} \mathbf{b}_r \quad \text{and} \quad \mathbf{c}^T(\sigma_k \mathbf{I} - \mathbf{A})^{-2} \mathbf{b} = \mathbf{c}_r^T(\sigma_k \mathbf{I}_r - \mathbf{A}_r)^{-2} \mathbf{b}_r$$

for $k = 1, \dots, r$. The quantity $\mathbf{c}^T(\sigma_k \mathbf{I} - \mathbf{A})^{-(j+1)} \mathbf{b}$ is called the j^{th} moment of $G(s)$ at σ_k .

If $\sigma_k = \infty$, the moments are called Markov parameters and are given by $\mathbf{c}^T \mathbf{A}^{j-1} \mathbf{b}$, $j > 0$. The corresponding moment matching problem is known as *partial realization*; see [1], [3] for details of its solution. Moment matching for finite $\sigma \in \mathbb{C}$, becomes *rational interpolation*, see for example [2]. Importantly, these problems can be solved in a recursive and numerically effective way, by means of the *Lanczos/Arnoldi* procedures for $\sigma = \infty$, and by means of *rational Lanczos/Arnoldi* procedures otherwise.

Rational interpolation by projection was first proposed by Skelton *et. al.* in [10, 33, 34]. Grimme [16] showed how one can obtain the required projection using the rational Krylov method of Ruhe [29].

We emphasize that Krylov-based methods, such as Lanczos, Arnoldi, are able to match moments without ever computing them explicitly. This is important since the computation of moments is in general ill-conditioned. This is a fundamental motivation behind the Krylov-based methods [11].

Grimme [16] showed the connection between multi-point rational interpolation and Krylov projections. We state a particular case of this here that is adequate for our purposes.

PROPOSITION 2.1. *Consider the system G defined by \mathbf{A} , \mathbf{b} , \mathbf{c} , distinct shifts given by σ_k , $k = 1, \dots, r$, and subspaces spanned by the columns of \mathbf{V} and \mathbf{Z} with $\mathbf{Z}^T \mathbf{V} = \mathbf{I}$, where*

$$\begin{aligned} \text{Ran}(\mathbf{V}) &= \text{span} \{ (\sigma_1 \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}, \dots, (\sigma_r \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \} \\ \text{and } \text{Ran}(\mathbf{Z}) &= \text{span} \{ (\sigma_1 \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}, \dots, (\sigma_r \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c} \} \end{aligned}$$

The reduced order system G_r defined by $\mathbf{A}_r = \mathbf{Z}^T \mathbf{A} \mathbf{V}$, $\mathbf{b}_r = \mathbf{Z}^T \mathbf{b}$, $\mathbf{c}_r^T = \mathbf{c}^T \mathbf{V}$, matches two moments of $G(s)$ at each of the interpolation points σ_k , $k = 1, \dots, r$.

Thus for Krylov-based model reduction, all one has to do is construct \mathbf{V} and \mathbf{Z} as above. Unlike gramian based model reduction methods such as balanced truncation (see [4]), Krylov-based model reduction requires only matrix-vector multiplications, some sparse linear solvers, and can be iteratively implemented; hence it is computationally effective; for details, see also [14, 15].

2.2. \mathcal{H}_2 -optimal model reduction. The \mathcal{H}_2 norm of the system defined by (1.1), assumed stable, is

$$\|G\|_{\mathcal{H}_2} := \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} |G(j\omega)|^2 d\omega \right)^{1/2} \quad (2.1)$$

Let \mathbf{P} be the reachability gramian of $G(s)$, i.e. \mathbf{P} is the unique positive definite solution to the Lyapunov equation $\mathbf{A} \mathbf{P} + \mathbf{P} \mathbf{A}^T + \mathbf{b} \mathbf{b}^T = 0$. Then, the \mathcal{H}_2 norm of $G(s)$ can be computed using the following formula:

$$\|G\|_{\mathcal{H}_2} = \sqrt{\mathbf{c}^T \mathbf{P} \mathbf{c}} \quad (2.2)$$

Given G as in (1.1) we seek to construct a *Krylov-based* reduced system G_r of

order r as in (1.2), which solves the optimal \mathcal{H}_2 model reduction problem, i.e.

$$\|G - G_r\|_{\mathcal{H}_2} = \min_{\substack{\dim(\hat{G}) = r \\ \hat{G}: \text{stable}}} \|G - \hat{G}\|_{\mathcal{H}_2}. \quad (2.3)$$

3. Main results.

3.1. Interpolation-based optimality conditions. In addition to widely known expressions for the \mathcal{H}_2 norm of a dynamical system, such as (2.1) and (2.2), recently Antoulas [4] obtained a new expression for $\|G\|_{\mathcal{H}_2}$ based on the poles and residues of $G(s)$. The result is stated below:

LEMMA 3.1. *Suppose $G(s)$ is a stable SISO system with simple poles $\lambda_1, \lambda_2, \dots, \lambda_n$. Let ϕ_i denote the associated residue of $G(s)$ at λ_i : $\phi_i = \lim_{s \rightarrow \lambda_i} G(s)(s - \lambda_i)$, $i = 1, \dots, n$. Then,*

$$\|G\|_{\mathcal{H}_2}^2 = \sum_{i=1}^n \phi_i G(-\lambda_i) \quad (3.1)$$

Lemma 3.1 immediately yields the following result regarding the \mathcal{H}_2 norm of the error system:

LEMMA 3.2. *Given the full-order model $G(s)$ and a reduced order model $G_r(s)$, let λ_i and $\hat{\lambda}_i$ be the poles of $G(s)$ and $G_r(s)$, respectively, and suppose that the poles of $G_r(s)$ are distinct. Let ϕ_i and $\hat{\phi}_i$ denote the residues of the transfer functions $G(s)$ and $G_r(s)$ at λ_i and $\hat{\lambda}_i$, respectively, i.e., ϕ_i as above and $\hat{\phi}_j = \lim_{s \rightarrow \hat{\lambda}_j} G_r(s)(s - \hat{\lambda}_j)$ for $j = 1, \dots, r$. Then the \mathcal{H}_2 norm of the error system, denoted by \mathcal{J} , is given by*

$$\begin{aligned} \mathcal{J} &:= \|G(s) - G_r(s)\|_{\mathcal{H}_2}^2 \\ &= \sum_{i=1}^n \phi_i \left(G(-\lambda_i) - G_r(-\lambda_i) \right) + \sum_{j=1}^r \hat{\phi}_j \left(G_r(-\hat{\lambda}_j) - G(-\hat{\lambda}_j) \right). \end{aligned} \quad (3.2)$$

PROOF: Let $\tilde{\phi}_i$ and $\tilde{\lambda}_i$ denote the residues and poles of the error system $G(s) - G_r(s)$, respectively, for $i = 1, \dots, n + r$. It readily follows that

$$\tilde{\phi}_i = \begin{cases} \phi_i, & i = 1, \dots, n \\ -\hat{\phi}_{i-n}, & i = n + 1, \dots, n + r, \end{cases} \quad \text{and} \quad \tilde{\lambda}_i = \begin{cases} \lambda_i, & i = 1, \dots, n \\ \hat{\lambda}_{i-n}, & i = n + 1, \dots, n + r. \end{cases} \quad (3.3)$$

Then, combining (3.3) with (3.1) yields the desired result. \square

REMARK 3.1. *The \mathcal{H}_2 error expression (3.2) generalizes the \mathcal{H}_2 result of [19, 18], proven only for model reduction by the Lanczos procedure, to the most general setting, valid for any reduced order model regardless of the underlying reduction technique.*

Lemma 3.2 has the system-theoretic interpretation that the \mathcal{H}_2 error is due to mismatch of the transfer functions $G(s)$ and $G_r(s)$ at the mirror images of the full-order poles λ_i and the reduced order poles $\hat{\lambda}_i$ ¹. Hence, this expression reveals that for a good \mathcal{H}_2 performance, $G_r(s)$ should approximate $G(s)$ well at $-\lambda_i$ and $-\hat{\lambda}_j$. Note that $\hat{\lambda}_i$ is not known *a priori*. Therefore, to minimize the \mathcal{H}_2 error, Gugercin

¹In the sequel, we will refer to $\hat{\lambda}_i$ as *Ritz values* as well as reduced order poles.

and Antoulas [19] proposed choosing $\sigma_i = -\lambda_i(\mathbf{A})$ where $\lambda_i(\mathbf{A})$ are the poles with big residuals ϕ_i . They have illustrated that this selection of interpolation points works quite well, see [18, 19]. However, as (3.2) illustrates, there is a second part of the \mathcal{H}_2 error due to the mismatch at $-\hat{\lambda}_j$. Indeed, as we will show below, interpolation at $-\hat{\lambda}_i$ is more important for the model reduction and is the necessary condition for the optimal \mathcal{H}_2 model reduction, i.e. $\sigma_i = -\hat{\lambda}_i$ is the optimal shift selection for the \mathcal{H}_2 model reduction.

THEOREM 3.3. *Given the full-order system $G(s) = \sum_{k=1}^n \frac{\phi_k}{s-\lambda_k}$, let $G(r) = \sum_{k=1}^r \frac{\hat{\phi}_k}{s-\hat{\lambda}_k}$ solve the optimal \mathcal{H}_2 problem (2.3). Then, $G_r(s)$ interpolates $G(s)$ and its first derivative at $-\hat{\lambda}_i$, $i = 1, \dots, r$, i.e.*

$$G_r(-\hat{\lambda}_k) = G(-\hat{\lambda}_k) \quad \text{and} \quad G'_r(-\hat{\lambda}_k) = G'(-\hat{\lambda}_k), \quad \text{for } i = 1, \dots, r. \quad (3.4)$$

The first order conditions (3.4), we refer to as Meier-Luenberger conditions, recognizing the work of [25]. We provide a new and simple proof of the necessity of (3.4) for \mathcal{H}_2 -optimality and show the equivalence to other commonly used first order necessary conditions in Section 4. We assume genericity in this paper, i.e., multiple poles are not considered. However in [4], one can find formula (3.1) for multiple poles which would be the starting point for including multiple poles in this framework resulting in higher derivatives at mirror images.

3.2. Iterated Interpolation. We propose an effective numerical algorithm which produces a reduced order model $G_r(s)$ that satisfies the interpolation-based first-order necessary conditions (3.4). Effectiveness of the proposed algorithm results from the fact that we use rational Krylov steps to construct $G_r(s)$ that meets the first-order conditions (3.4). No Lyapunov solvers or dense matrix decompositions are needed. Therefore, the method is suited for large-scale systems where $n \gg 1000$.

Several approaches have been proposed in the literature to compute a reduced order model which satisfies *some form* of a first-order necessary conditions; see [32, 30, 9, 20, 25, 22, 31, 24]. However, these approaches do not seem to be suitable for large-scale problems. The ones based on Lyapunov-based conditions, e.g. [31, 22, 30, 32], require solving a couple of Lyapunov equations at each step of the iteration. To our knowledge, the only methods which depend on interpolation-based necessary conditions have been proposed in [24] and [25]. The authors work with the transfer functions of $G(s)$ and $G_r(s)$; make an iteration on the denominator [24] or poles and residues [25] of $G_r(s)$; and explicitly compute $G(s)$, $G_r(s)$ and their derivatives at certain points in the complex plane. However, working with the transfer function, its values, and its derivative values explicitly is not desirable in large-scale settings. Indeed, one will most likely be given a state-space representation of $G(s)$ rather than the transfer function. And trying to compute the coefficients of the transfer function can be highly ill-conditioned. These approaches are similar to [27, 28] where interpolation is done by explicit usage of transfer functions. On the other hand, our approach, which is detailed below, is based on the connection between interpolation and effective rational Krylov iteration, and is therefore numerically effective and stable.

Let $\boldsymbol{\sigma}$ denote the set of interpolation points $\{\sigma_1, \dots, \sigma_r\}$; use these interpolation points to construct a reduced order model, $G_r(s)$, that interpolates both $G(s)$ and $G'(s)$ at $\{\sigma_1, \dots, \sigma_r\}$; let $\boldsymbol{\lambda}(\boldsymbol{\sigma})$ denote the resulting reduced order poles of $G_r(s)$. Define the function $\mathbf{g}(\boldsymbol{\sigma}) = \boldsymbol{\lambda}(\boldsymbol{\sigma}) + \boldsymbol{\sigma}$. Aside from issues related to the ordering of the reduced order poles, $\mathbf{g}(\boldsymbol{\sigma}) = 0$ is equivalent to (3.4) and, hence, is a necessary condition for \mathcal{H}_2 -optimality of the reduced order model, $G_r(s)$. Thus one can formulate a search

for optimal \mathcal{H}_2 reduced order systems by considering the root finding problem $\mathbf{g}(\boldsymbol{\sigma}) = \mathbf{0}$. Many plausible approaches to this problem originate with Newton's method, which appears as

$$\boldsymbol{\sigma}^{(k+1)} = \boldsymbol{\sigma}^{(k)} - (\mathbf{I}_r + \mathbf{J})^{-1} \left(\boldsymbol{\sigma}^{(k)} + \boldsymbol{\lambda} \left(\boldsymbol{\sigma}^{(k)} \right) \right) \quad (3.5)$$

In (3.5), \mathbf{J} is the Jacobian of $\boldsymbol{\lambda}(\boldsymbol{\sigma})$ with respect to $\boldsymbol{\sigma}$. In the discrete time case, the root-finding problem becomes $\mathbf{g}(\boldsymbol{\sigma}) = \boldsymbol{\Sigma}\boldsymbol{\lambda}(\boldsymbol{\sigma}) - \mathbf{e}$, where $\mathbf{e}^T = [1, 1, \dots, 1]$ and $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma})$; the associated Newton step is

$$\boldsymbol{\sigma}^{(k+1)} = \boldsymbol{\sigma}^{(k)} - (\mathbf{I}_r + \boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}\mathbf{J})^{-1} \left(\boldsymbol{\sigma}^{(k)} - \boldsymbol{\Lambda}^{-1}\mathbf{e} \right)$$

where $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$.

3.3. Proposed Algorithm. We seek a reduced-order transfer function $G_r(s)$ that interpolates $G(s)$ at the mirror images of the poles of $G_r(s)$ by solving the equivalent root finding problem, say by a variant of (3.5). It is often the case that in the neighborhood of an \mathcal{H}_2 -optimal shift set, the entries of the Jacobian matrix become small (also see Section 5) and simply setting $\mathbf{J} = 0$ might serve as a relaxed iteration strategy. This leads to a successive substitution framework: $\sigma_i \leftarrow -\lambda_i(\mathbf{A}_r)$; successive interpolation steps using a rational Krylov method are used so that at the $(i+1)^{\text{st}}$ step interpolation points are chosen as the mirror images of the Ritz values from the i^{th} step. Despite its simplicity, this appears to be a very effective strategy in many circumstances.

Here is a sketch of the proposed algorithm:

ALGORITHM 3.1. An Iterative Rational Krylov Algorithm (IRKA):

1. Make an initial selection of σ_i , for $i = 1, \dots, r$
2. Choose \mathbf{V} and \mathbf{Z} so that $\text{Ran}(\mathbf{V}) = \text{span} \{ (\sigma_1 \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}, \dots, (\sigma_r \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \}$
 $\text{Ran}(\mathbf{Z}) = \text{span} \{ (\sigma_1 \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}, \dots, (\sigma_r \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c} \}$ and $\mathbf{Z}^T \mathbf{V} = \mathbf{I}_r$.
3. while (not converged)
 - (a) $\mathbf{A}_r = \mathbf{Z}^T \mathbf{A} \mathbf{V}$,
 - (b) Assign $\sigma_i \leftarrow -\lambda_i(\mathbf{A}_r)$ for $i = 1, \dots, r$
 - (c) Update \mathbf{V} and \mathbf{Z} so $\text{Ran}(\mathbf{V}) = \text{span} \{ (\sigma_1 \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}, \dots, (\sigma_r \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \}$
 $\text{Ran}(\mathbf{Z}) = \text{span} \{ (\sigma_1 \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}, \dots, (\sigma_r \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c} \}$ and $\mathbf{Z}^T \mathbf{V} = \mathbf{I}_r$.
4. $\mathbf{A}_r = \mathbf{Z}^T \mathbf{A} \mathbf{V}$, $\mathbf{b}_r = \mathbf{Z}^T \mathbf{b}$, $\mathbf{c}_r^T = \mathbf{c}^T \mathbf{V}$

Upon convergence, the first-order necessary conditions (3.4) for \mathcal{H}_2 optimality will be satisfied. Notice that Step 3(b) could be replaced with some variant of a Newton step (3.5). For discrete time systems, Step 3b becomes $\sigma_i \leftarrow 1/\lambda_i(\mathbf{A}_r)$ for $i = 1, \dots, r$.

We have implemented the above algorithm for many different large-scale systems. In each of our numerical examples, the algorithm worked very effectively: It has always converged after a small number of steps, and resulted in stable reduced systems. For some standard test problems where the global optimum is known, Algorithm 3.1 has converged to this global optimum.

It should be noted that the solution is obtained via Krylov projection methods only and its computation is suitable for large-scale systems. To our knowledge, this is the first numerically effective approach for the optimal \mathcal{H}_2 reduction problem.

We know that the reduced model $G_r(s)$ resulting from the above algorithm will satisfy the first-order conditions. However, the question which arises is whether this reduced order model is globally optimal in some sense. The next result provides a

clue in this direction. It is an extension to continuous time of Theorem 3, p. 86 in Gaier's monograph [13]. It can be also found in [25].

THEOREM 3.4. *Given a stable SISO transfer function $G(s)$, and fixed stable reduced poles $\alpha_1, \dots, \alpha_r$, define*

$$G_r(s) := \frac{\beta_0 + \beta_1 s + \dots + \beta_r s^r}{(s - \alpha_1) \dots (s - \alpha_r)}.$$

Then $\|G - G_r\|_{\mathcal{H}_2}$ is minimized if and only if

$$G(s) = G_r(s) \quad \text{for} \quad s = -\bar{\alpha}_1, -\bar{\alpha}_2, \dots, -\bar{\alpha}_r. \quad (3.6)$$

Note that (3.6) can be rewritten as

$$G(s) = G_r(s) \quad \text{for} \quad s = -\alpha_1, -\alpha_2, \dots, -\alpha_r.$$

since the poles, $\{\alpha_i\}$ occur in complex conjugate pairs.

Theorem 3.4 states that if $G_r(s)$ interpolates $G(s)$ at the mirror images of the poles of $G_r(s)$, then $G_r(s)$ is guaranteed to be an *optimal* approximation of $G(s)$ with respect to the \mathcal{H}_2 norm among all reduced order systems having the same reduced system poles $\{\alpha_i\}$, $i = 1, \dots, r$. Hence, the following corollary holds:

COROLLARY 3.5. *Let $G_r(s)$ be the reduced model resulting from Algorithm 3.1. Then, $G_r(s)$ is the optimal approximation of $G(s)$ with respect to the \mathcal{H}_2 norm among all reduced order systems having the same reduced system poles as $G_r(s)$; therefore, Algorithm 3.1 generates a reduced model $G_r(s)$ which is the optimal solution for a restricted \mathcal{H}_2 problem.*

3.4. Initial Shift Selection. For the proposed algorithm, the final reduced model can depend on the initial shift selection. Nonetheless for most of the cases, a random initial shift selection resulted in satisfactory reduced model. For small order benchmark examples taken from [22, 24, 32, 30], the algorithm converged to the global minimizer. For larger problems, the results were as good as those obtained by balanced truncation. Therefore, while staying within a numerically effective Krylov projection framework, we have been able to produce results close to or better than those obtained by balanced truncation (which requires the solution of two large-scale Lyapunov equations).

We outline some initialization strategies which can be expected to improve the results. Recall that at convergence, interpolation points are mirror images of the eigenvalues of \mathbf{A}_r . The eigenvalues of \mathbf{A}_r might be expected to approximate the eigenvalues of \mathbf{A} . Hence, at convergence, interpolation points will lie in the mirror spectrum of \mathbf{A} . Therefore, one could choose initial shifts randomly distributed within a region containing the mirror image of the numerical range of \mathbf{A} . The boundary of the numerical range can be estimated by computing the eigenvalues of \mathbf{A} with the smallest and largest real and imaginary parts using numerically effective tools such as implicitly restarted Arnoldi (IRA) algorithm.

The starting point for another initialization strategy is the \mathcal{H}_2 expression presented in Lemma 3.2. Based on this expression, it is appropriate to initiate the proposed algorithm with $\sigma_i = -\lambda_i(\mathbf{A})$ where $\lambda_i(\mathbf{A})$ are the poles with big residuals ϕ_i for $i = 1, \dots, r$. The main disadvantage of this approach is that it requires a modal state-space decomposition for $G(s)$, which will be numerically expensive for large-scale problems. However, there might be some applications where the original

state-space representation is in the modal-form and ϕ_i might be directly read from the entries of the matrices \mathbf{b} and \mathbf{c}^T .

REMARK 3.2. *Unstable reduced order models are not acceptable candidates for optimal \mathcal{H}_2 reduction. Nonetheless stability of a reduced model is not guaranteed a priori and might depend on the initial shift selection. We have observed that if one avoids making extremely unrealistic initial shift selections, stability will be preserved. In our simulations we have never generated an unstable system when the initial shift selection was not drastically different from the mirror spectrum of \mathbf{A} , but otherwise random. We were able to produce an unstable reduced order system, however this occurred for a case where the real parts of the eigenvalues of \mathbf{A} were between -1.5668×10^{-1} and -2.0621×10^{-3} yet we chose initial shifts bigger than 50. We believe that with a good starting point, stability will not be an issue. These considerations are illustrated for many numerical examples in Section 6.*

4. First-order \mathcal{H}_2 optimality conditions. In this section, we give a new proof of the Meier-Luenberger (interpolation based) first-order conditions (3.4) using the new \mathcal{H}_2 error expression (3.2). Besides (3.4), Lyapunov-based first-order conditions for the optimal \mathcal{H}_2 problem exist as well. The equivalence of (3.4) and Lyapunov-based conditions is not *a priori* evident; we briefly review the Lyapunov-based conditions of Wilson [31] and Hyland-Bernstein [22] for the \mathcal{H}_2 problem; and prove that both Lyapunov-based frameworks are equivalent to (3.4).

4.1. Meier - Luenberger conditions: A new proof. We first list some expressions that are used in the proof. Let partial fraction expansions of $G(s)$ and $G_r(s)$ be as follows:

$$G(s) = \sum_{k=1}^n \frac{\phi_k}{s - \lambda_k} \quad \text{and} \quad G_r(s) = \sum_{k=1}^r \frac{\hat{\phi}_k}{s - \hat{\lambda}_k} \quad (4.1)$$

Then, following equalities directly follow from (4.1):

$$\frac{\partial G_r(-\lambda_i)}{\partial \hat{\phi}_m} = \frac{-1}{\lambda_i + \hat{\lambda}_m}, \quad \frac{\partial G_r(-\hat{\lambda}_j)}{\partial \hat{\phi}_m} = \frac{-1}{\hat{\lambda}_j + \hat{\lambda}_m}, \quad \frac{\partial G_r(-\lambda_i)}{\partial \hat{\lambda}_m} = \frac{\hat{\phi}_m}{(\lambda_i + \hat{\lambda}_m)^2}, \quad (4.2)$$

$$\frac{\partial G(-\hat{\lambda}_j)}{\partial \hat{\lambda}_m} = \begin{cases} 0 & j \neq m \\ \sum_{k=1}^n \frac{\phi_k}{(\hat{\lambda}_m + \lambda_k)^2} & j = m \end{cases}, \quad \text{and} \quad (4.3)$$

$$\frac{\partial G_r(-\hat{\lambda}_j)}{\partial \hat{\lambda}_m} = \begin{cases} \frac{\hat{\phi}_m}{(\hat{\lambda}_j + \hat{\lambda}_m)^2} & j \neq m \\ \sum_{k=1, k \neq m}^r \frac{\hat{\phi}_k}{(\hat{\lambda}_m + \hat{\lambda}_k)^2} + \frac{\hat{\phi}_m}{2\hat{\lambda}_m^2} & j = m \end{cases}, \quad (4.4)$$

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

PROOF OF THEOREM 3.3: Note that, $G_r(s)$ depends on $2r$ free parameters $\hat{\lambda}_i$ and $\hat{\phi}_i$ for $i = 1, \dots, r$. Hence, the first-order necessary conditions (3.4) can be obtained by taking the derivative of the error function \mathcal{J} in (3.2) with respect to these variables.

Using (4.2), the derivative of \mathcal{J} with respect to $\hat{\phi}_m$ yields

$$\frac{\partial \mathcal{J}}{\partial \hat{\phi}_m} = \underbrace{\sum_{k=1}^n \phi_i \left(\frac{-1}{-\lambda_i - \hat{\lambda}_m} \right)}_{=-G(-\hat{\lambda}_m)} + G_r(-\hat{\lambda}_m) - G(-\hat{\lambda}_m) + \underbrace{\sum_{j=1}^r \hat{\phi}_j \frac{1}{-\hat{\lambda}_j - \hat{\lambda}_m}}_{=G_r(-\hat{\lambda}_m)} \quad (4.5)$$

$$= -2G(-\hat{\lambda}_m) + 2G_r(-\hat{\lambda}_m), \quad m = 1, \dots, r. \quad (4.6)$$

Setting $\frac{\partial \mathcal{J}}{\partial \hat{\phi}_m} = 0$ leads to $G(-\hat{\lambda}_m) = G_r(-\hat{\lambda}_m)$ for $m = 1, \dots, r$.

Using (4.2)-(4.4), the derivative of \mathcal{J} with respect $\hat{\lambda}_m$ can be computed as

$$\begin{aligned} \frac{\partial \mathcal{J}}{\partial \hat{\lambda}_m} &= -\hat{\phi}_m \left[\underbrace{\sum_{i=1}^n \frac{\phi_i}{(\lambda_i + \hat{\lambda}_m)^2}}_{=-G'(-\hat{\lambda}_m)} + \underbrace{\sum_{k=1}^n \frac{\phi_k}{(\hat{\lambda}_m + \lambda_k)^2}}_{=-G'(-\hat{\lambda}_m)} + \underbrace{\sum_{k=1}^r \frac{\hat{\phi}_k}{(\hat{\lambda}_m + \hat{\lambda}_k)^2}}_{=-G'_r(-\hat{\lambda}_m)} + \underbrace{\sum_{j=1}^r \frac{\hat{\phi}_j}{(\hat{\lambda}_m + \hat{\lambda}_j)^2}}_{=-G'_r(-\hat{\lambda}_m)} \right] \\ &= 2\hat{\phi}_m \left(G'(-\hat{\lambda}_m) - G'_r(-\hat{\lambda}_m) \right) \end{aligned}$$

Setting $\frac{\partial \mathcal{J}}{\partial \hat{\phi}_m} = 0$ leads to $G'(-\hat{\lambda}_m) = G'_r(-\hat{\lambda}_m)$ for $m = 1, \dots, r$ as desired. Note

that without loss of generality, we assume that $\hat{\phi}_m \neq 0$ since $\hat{\phi}_m = 0$ implies to saying that the reduced order is less than r . \square

4.2. Wilson conditions. Let $G_r(s)$ defined by \mathbf{A}_r , \mathbf{b}_r , and \mathbf{c}_r^T solve the optimal \mathcal{H}_2 problem (2.3). Define the error system

$$G_e(s) := G(s) = G_r(s) := \left[\begin{array}{c|c} \mathbf{A}_e & \mathbf{b}_e \\ \mathbf{c}_e^T & 0 \end{array} \right] := \left[\begin{array}{c|c} \left[\begin{array}{cc} \mathbf{A} & 0 \\ 0 & \mathbf{A}_r \end{array} \right] & \left[\begin{array}{c} \mathbf{b} \\ \mathbf{b}_r \end{array} \right] \\ \hline \left[\begin{array}{cc} \mathbf{c}^T & -\mathbf{c}_r^T \end{array} \right] & 0 \end{array} \right] \quad (4.7)$$

Let \mathbf{P}_e and \mathbf{Q}_e be the gramians for the error system $G_e(s)$, i.e. \mathbf{P}_e and \mathbf{Q}_e solve

$$\mathbf{A}_e \mathbf{P}_e + \mathbf{P}_e \mathbf{A}_e^T + \mathbf{b}_e \mathbf{b}_e^T = 0 \quad (4.8)$$

$$\mathbf{Q}_e \mathbf{A}_e + \mathbf{A}_e^T \mathbf{Q}_e + \mathbf{c}_e \mathbf{c}_e^T = 0 \quad (4.9)$$

Partition \mathbf{P}_e and \mathbf{Q}_e :

$$\mathbf{P}_e = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{12}^T & \mathbf{P}_{22} \end{bmatrix} \quad \mathbf{Q}_e = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{12}^T & \mathbf{Q}_{22} \end{bmatrix} \quad (4.10)$$

Then, the first-order necessary conditions imply that

$$\mathbf{P}_{12}^T \mathbf{Q}_{12} + \mathbf{P}_{22} \mathbf{Q}_{22} = 0, \quad (4.11)$$

$$\mathbf{Q}_{12}^T \mathbf{b} + \mathbf{Q}_{22} \mathbf{b}_r = 0 \quad (4.12)$$

$$\mathbf{c}_r^T \mathbf{P}_{22} - \mathbf{c}^T \mathbf{P}_{12} = 0. \quad (4.13)$$

It follows that, the reduced order model

$$G_r(s) = \left[\begin{array}{c|c} \mathbf{A}_r & \mathbf{b}_r \\ \mathbf{c}_r^T & 0 \end{array} \right] = \left[\begin{array}{c|c} \mathbf{Z}^T \mathbf{A} \mathbf{V} & \mathbf{Z}^T \mathbf{b} \\ \hline \mathbf{c}^T \mathbf{V} & 0 \end{array} \right], \quad (4.14)$$

$$\text{where } \mathbf{V} := \mathbf{P}_{12} \mathbf{P}_{22}^{-1} \text{ and } \mathbf{Z} = -\mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} \quad (4.15)$$

obtained by a projection satisfies the first-order conditions of the optimal \mathcal{H}_2 problem. It was also shown in [31] that $\mathbf{Z}^T \mathbf{V} = \mathbf{I}_r$.

4.3. Hyland – Bernstein conditions. Suppose $G_r(s)$ defined by \mathbf{A}_r , \mathbf{b}_r and \mathbf{c}_r^T solves the optimal \mathcal{H}_2 problem. Then there exist positive nonnegative matrices $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{n \times n}$ such that

$$\mathbf{P}\mathbf{Q} = \mathbf{V}\mathbf{M}\mathbf{Z}^T, \mathbf{Z}^T \mathbf{V} = \mathbf{I}_r, \quad (4.16)$$

where \mathbf{M} is similar to a positive definite matrix. Then $G_r(s)$ is given by

$$G_r(s) = \left[\begin{array}{c|c} \mathbf{A}_r & \mathbf{b}_r \\ \hline \mathbf{c}_r^T & 0 \end{array} \right] = \left[\begin{array}{c|c} \mathbf{Z}^T \mathbf{A} \mathbf{V} & \mathbf{Z}^T \mathbf{b} \\ \hline \mathbf{c}^T \mathbf{V} & 0 \end{array} \right],$$

such that, with $\mathbf{\Pi} = \mathbf{V}\mathbf{Z}^T$, the following conditions are satisfied:

$$\text{rank}(\mathbf{P}) = \text{rank}(\mathbf{Q}) = \text{rank}(\mathbf{P}\mathbf{Q}) \quad (4.17)$$

$$\mathbf{\Pi} [\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{b}\mathbf{b}^T] = 0 \quad (4.18)$$

$$[\mathbf{A}^T \mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{c}\mathbf{c}^T] \mathbf{\Pi} = 0 \quad (4.19)$$

REMARK 4.1. *Note that in both [31] and [22], the first-order necessary conditions are given in terms of (coupled) Lyapunov equations. Both [31] and [22] proposed iterative algorithms to obtain a reduced order model satisfying these Lyapunov-based first-order conditions. However, the main drawback in each case is that both approaches require solving two large-scale Lyapunov equations at each step of the algorithm. [35] discusses computational issues related to solving associated linearized problems within each step.*

REMARK 4.2. *We note that even though in both cases the optimal reduced model is obtained via projection, the projection framework was not enforced on the reduced system and followed from the structure of the problem.*

4.4. Equivalence of the first order conditions. Above, we have briefly reviewed three different forms of first-order conditions for the optimal \mathcal{H}_2 model reduction problem. While the Meier–Luenberger conditions are framed in terms of interpolation, the Frameworks of [31] and [22] are in terms of Lyapunov equations. Equivalence between the two Lyapunov frameworks [31] and [22] has been proved in [22]. However the equivalence between [31]-[22] and the Meier–Luenberger conditions has not been reported in the literature yet. In this section, we state this equivalence; In other words, *we connect the Lyapunov-based projection framework of [31] and [22] to the Meier–Luenberger conditions.* As expected, the key point is model reduction via rational Krylov projection. This connection allows us to tackle the optimal \mathcal{H}_2 problem in a numerically effective Krylov projection framework rather than a computationally expensive Lyapunov framework.

LEMMA 4.1. Equivalence of Lyapunov and Interpolation Frameworks: *The first-order necessary conditions of both [22] as given in (4.17)-(4.19) and [31] as given in (4.14) and (4.15) are equivalent to those of [25] as given in (3.4); in other words the Lyapunov-based first-order conditions [31, 22] for optimal \mathcal{H}_2 problem are equivalent to the interpolation-based Meier–Luenberger conditions.*

PROOF OF EQUIVALENCE OF MEIER–LUENBERGER AND WILSON CONDITIONS: First, we show that (4.14) and (4.15) imply (3.4): (1,2) block of the Lyapunov equation (4.8) yields

$$\mathbf{A}\mathbf{P}_{12} + \mathbf{P}_{12}\mathbf{A}_r^T + \mathbf{b}\mathbf{b}_r^T = 0. \quad (4.20)$$

Since \mathbf{V} in (4.15) is given by $\mathbf{V} = \mathbf{P}_{12}\mathbf{P}_{22}^{-1}$, it follows that $\text{Ran}(\mathbf{V}) = \text{Ran}(\mathbf{P}_{12})$. Also, since \mathbf{P}_{12} solves the Sylvester equation (4.20), we obtain

$$\text{Ran}(\mathbf{V}) = \text{span} \left\{ (-\hat{\lambda}_1 \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}, \dots, (-\hat{\lambda}_r \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \right\} \quad (4.21)$$

where $\hat{\lambda}_i$ are the eigenvalues of \mathbf{A}_r . Similarly, (1,2) block of (4.8) yields

$$\mathbf{A}^T \mathbf{Q}_{12} + \mathbf{Q}_{12} \mathbf{A}_r - \mathbf{c} \mathbf{c}_r^T = 0. \quad (4.22)$$

Since \mathbf{Z} in (4.15) is $\mathbf{Z} = -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}$, it follows that $\text{Im}(\mathbf{Z}) = -\text{Im}(\mathbf{Q}_{12})$. Due to the same reasons as above,

$$\text{Ran}(\mathbf{Z}) = \text{span} \left\{ (-\hat{\lambda}_1 \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}, \dots, (-\hat{\lambda}_r \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c} \right\}. \quad (4.23)$$

Then it directly follows from (4.21) and (4.23) that the reduced model $G_r(s)$ in (4.14) interpolates the first two moments of the $G(s)$ at $-\hat{\lambda}_i$; hence (3.4) results.

Now, we prove that (3.4) implies (4.14) and (4.15): Let $G_r = \mathbf{c}_r^T (s\mathbf{I}_r - \mathbf{A}_r)^{-1} \mathbf{b}_r$ be a reduced order model satisfying the interpolation conditions in (4.14). Let $\hat{\lambda}_i$ denote the eigenvalues of \mathbf{A}_r as above. It follows from the theory of Krylov-based projection that there exist matrices \mathbf{V} and \mathbf{Z} such that

$$\text{Ran}(\mathbf{V}) = \text{span} \left\{ (-\hat{\lambda}_1 \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}, \dots, (-\hat{\lambda}_r \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \right\}, \quad (4.24)$$

$$\text{Ran}(\mathbf{Z}) = \text{span} \left\{ (-\hat{\lambda}_1 \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}, \dots, (-\hat{\lambda}_r \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c} \right\}. \quad (4.25)$$

and the reduced model $G_r(s)$ is obtained by an oblique projection, i.e.

$$G_r = \left[\begin{array}{c|c} \mathbf{A}_r & \mathbf{b}_r \\ \mathbf{c}_r^T & 0 \end{array} \right] = \left[\begin{array}{c|c} \mathbf{Z}^T \mathbf{A} \mathbf{V} & \mathbf{Z}^T \mathbf{b} \\ \mathbf{c}^T \mathbf{V} & 0 \end{array} \right] \quad \text{where } \mathbf{Z}^T \mathbf{V} = \mathbf{I}_r. \quad (4.26)$$

Define the error-system gramians \mathbf{P}_e and \mathbf{Q}_e as in (4.8) and (4.9), respectively, and partition as in (4.10). Then \mathbf{P}_{12} and \mathbf{P}_{22} satisfy

$$\mathbf{A} \mathbf{P}_{12} + \mathbf{P}_{12} \mathbf{A}_r^T + \mathbf{b} \mathbf{b}_r^T = 0. \quad (4.27)$$

Combining (4.27) with (4.24) leads to

$$\mathbf{V} = \mathbf{P}_{12} \mathbf{K}$$

where $\mathbf{K} \in \mathbb{R}^{r \times r}$ is a nonsingular matrix. A similar discussion reveals that

$$\mathbf{Z} = -\mathbf{Q}_{12} \mathbf{L}$$

where $\mathbf{L} \in \mathbb{R}^{r \times r}$ is a nonsingular matrix. Using these last two equalities in (4.26), we obtain

$$G_r = \left[\begin{array}{c|c} \mathbf{A}_r & \mathbf{b}_r \\ \mathbf{c}_r^T & 0 \end{array} \right] = \left[\begin{array}{c|c} -\mathbf{L}^T \mathbf{Q}_{12}^T \mathbf{A} \mathbf{P}_{12} \mathbf{K} & -\mathbf{L}^T \mathbf{Q}_{12}^T \mathbf{b} \\ \mathbf{c}^T \mathbf{P}_{12} \mathbf{K} & 0 \end{array} \right] \quad (4.28)$$

where $-\mathbf{L}^T \mathbf{Q}_{12}^T \mathbf{P}_{12} \mathbf{K} = \mathbf{I}_r$. Multiplying (4.27) by \mathbf{Q}_{12}^T from left and using the equalities $\mathbf{Q}_{12}^T \mathbf{A} \mathbf{P}_{12} = -\mathbf{L}^{-T} \mathbf{A}_r \mathbf{K}^{-1}$, $\mathbf{Q}_{12}^T \mathbf{b} = -\mathbf{L}^{-T} \mathbf{b}_r$ yield

$$-\mathbf{L}^{-T} \mathbf{A}_r \mathbf{K}^{-1} + \mathbf{Q}_{12}^T \mathbf{P}_{12} \mathbf{A}_r^T - \mathbf{L}^{-T} \mathbf{b}_r \mathbf{b}_r^T = 0.$$

Multiplying the last expression by $-\mathbf{L}^T$ from left and noticing that due to (4.26), $-\mathbf{L}^T \mathbf{Q}_{12}^T \mathbf{P}_{12} = \mathbf{K}^{-1}$, one obtains

$$\mathbf{A}_r \mathbf{K}^{-1} + \mathbf{K}^{-1} \mathbf{A}_r^T + \mathbf{b}_r \mathbf{b}_r^T = 0.$$

Since the optimal reduced model is stable, we obtain

$$\mathbf{P}_{22} = \mathbf{K}^{-1}.$$

A similar argument yields

$$\mathbf{Q}_{22} = \mathbf{L}^{-1}.$$

Finally, using these last two equalities in (4.26) yields precisely the first-order conditions of [31], namely (4.11)-(4.15). This completes the proof. \square

PROOF OF EQUIVALENCE OF MEIER-LUENBERGER AND HYLAND-BERNSTEIN CONDITIONS: First, we prove that (4.17)-(4.18) imply (3.4): It was shown in [22] that $\mathbf{\Pi P} = \mathbf{P}$ for $\mathbf{\Pi}$ and \mathbf{P} as in (4.17)-(4.18). Then, (4.17) can be re-written as

$$\mathbf{\Pi}[\mathbf{A}\mathbf{\Pi P} + \mathbf{\Pi P A}^T + \mathbf{b b}^T] = 0 \quad (4.29)$$

$$\implies \mathbf{V}[\mathbf{Z}^T \mathbf{A V Z}^T \mathbf{P} + \mathbf{Z}^T \mathbf{P A}^T + \mathbf{Z}^T \mathbf{b b}^T] = 0 \quad (4.30)$$

Since \mathbf{V} is full-rank, the last equality leads to

$$\mathbf{A P Z} + \mathbf{P Z A}_r^T + \mathbf{b b}_r^T = 0 \quad (4.31)$$

Because \mathbf{A}_r is stable,

$$\mathbf{P Z} = \mathbf{P}_{12} \quad (4.32)$$

where \mathbf{P}_{12} as defined in (4.10), i.e. \mathbf{P}_{12} is the (1,2) block of \mathbf{P}_e , reachability gramian for the error system. Using the fact that $\mathbf{\Pi P} = \mathbf{P}$ in (4.31) and multiplying this expression by \mathbf{Z}^T from left gives

$$\mathbf{Z}^T \mathbf{P Z} = \mathbf{P}_{22} \quad (4.33)$$

Combining (4.32) and (4.33) together with $\mathbf{Z}^T \mathbf{V} = \mathbf{I}_r$ yields that

$$\begin{aligned} \text{Ran}(\mathbf{V}) &= \text{Ran}(\mathbf{P}_{12}) \\ &= \text{span} \left\{ (-\hat{\lambda}_1 \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}, \dots, (-\hat{\lambda}_r \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \right\}. \end{aligned} \quad (4.34)$$

A similar argument leads to

$$\begin{aligned} \text{Ran}(\mathbf{Z}) &= \text{Ran}(-\mathbf{Q}_{12}) \\ &= \text{span} \left\{ (-\hat{\lambda}_1 \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}, \dots, (-\hat{\lambda}_r \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c} \right\}. \end{aligned} \quad (4.35)$$

Therefore, the interpolation conditions in (3.4) hold.

To prove that (3.4) imply (4.17)-(4.19), we reverse the above steps. Let $G_r = \mathbf{c}_r^T (s \mathbf{I}_r - \mathbf{A}_r)^{-1} \mathbf{b}_r$ be a reduced order model satisfying the interpolation conditions in (4.14). The interpolation conditions in (3.4) reveals that \mathbf{V} and \mathbf{Z} satisfy (4.34) and

(4.35), respectively. More precisely, there exist non-singular matrices \mathbf{K} and \mathbf{L} such that

$$\mathbf{P}_{12} = \mathbf{V}\mathbf{E} \quad \text{and} \quad \mathbf{Q}_{12} = -\mathbf{Z}\mathbf{F} \quad (4.36)$$

Then, $\mathbf{A}\mathbf{P}_{12} + \mathbf{P}_{12}\mathbf{A}_r^T + \mathbf{b}\mathbf{b}_r^T = 0$ becomes

$$\mathbf{A}\mathbf{V}\mathbf{E} + \mathbf{V}\mathbf{E}\mathbf{V}^T\mathbf{A}^T\mathbf{Z} + \mathbf{b}\mathbf{b}^T\mathbf{Z}^T = 0.$$

Transposing this expression followed by a multiplication by \mathbf{V} from left leads to

$$\mathbf{V}\mathbf{Z}^T[\mathbf{A}\mathbf{V}\mathbf{E}^T\mathbf{V} + \mathbf{V}\mathbf{E}^T\mathbf{V}\mathbf{A}^T + \mathbf{b}\mathbf{b}^T] = 0,$$

which is equivalent to (4.18) by defining $\mathbf{P} := \mathbf{V}\mathbf{E}^T\mathbf{V}^T$ and the projection $\mathbf{\Pi} := \mathbf{V}\mathbf{Z}^T$. A similar argument on $\mathbf{A}_r^T\mathbf{Q}_{12} + \mathbf{Q}_{12}\mathbf{A} + \mathbf{c}_r\mathbf{c}^T = 0$ yields

$$[\mathbf{A}^T\mathbf{Z}\mathbf{F}\mathbf{Z}^T + \mathbf{Z}\mathbf{F}\mathbf{Z}^T\mathbf{A} + \mathbf{c}^T\mathbf{c}]\mathbf{V}\mathbf{Z}^T = 0,$$

which is equivalent to (4.19) with $\mathbf{Q} = \mathbf{Z}\mathbf{F}\mathbf{Z}^T$. Moreover, (4.17) holds by construction. We also need to check that the decomposition in (4.16) is also valid. Using the same discussion as in the proof of Lemma 4.1, one can show that

$$\mathbf{E} = \mathbf{P}_{22} \quad \text{and} \quad \mathbf{F} = \mathbf{Q}_{22}.$$

Therefore,

$$\mathbf{P}\mathbf{Q} = \mathbf{V}\mathbf{E}^T\mathbf{F}\mathbf{Z}^T = \mathbf{V}\mathbf{P}_{22}\mathbf{Q}_{22}\mathbf{Z}^T,$$

and (4.16) holds with $\mathbf{M} = \mathbf{P}_{22}\mathbf{Q}_{22}$. Note that \mathbf{M} is similar to a positive definite matrix since \mathbf{P}_{22} and \mathbf{Q}_{22} are positive definite. Finally, $\mathbf{\Pi}\mathbf{P} = \mathbf{P}$ and $\mathbf{Q}\mathbf{\Pi} = \mathbf{Q}$ hold as in the framework of ([22]). This completes the proof. \square

5. Analysis of Rational Krylov Methods. We are better able to understand the behaviour of the algorithms proposed here in light of a few basic features related to rational Krylov methods that are not widely known. We collect some of these facts here.

5.1. A canonical rational Krylov decomposition. Suppose $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$. Fix an index $1 \leq r \leq n$. Suppose that σ_i are distinct points in \mathbb{C} none of which are eigenvalues of \mathbf{A} , and define the complex r -tuple $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \dots, \sigma_r]^T \in \mathbb{C}^r$ together with related matrices:

$$\mathbf{V}(\boldsymbol{\sigma}) = [(\sigma_1\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} \quad (\sigma_2\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} \quad \dots \quad (\sigma_r\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}] \in \mathbb{C}^{n \times r}$$

and

$$\mathbf{W}^T(\boldsymbol{\sigma}) = \begin{bmatrix} \mathbf{c}^T(\sigma_1\mathbf{I} - \mathbf{A})^{-1} \\ \mathbf{c}^T(\sigma_2\mathbf{I} - \mathbf{A})^{-1} \\ \vdots \\ \mathbf{c}^T(\sigma_r\mathbf{I} - \mathbf{A})^{-1} \end{bmatrix} \in \mathbb{C}^{r \times n}.$$

We normally suppress the dependence on $\boldsymbol{\sigma}$ and write $\mathbf{V}(\boldsymbol{\sigma}) = \mathbf{V}$ and $\mathbf{W}(\boldsymbol{\sigma}) = \mathbf{W}$.

LEMMA 5.1. Let $\omega_r(z) = (z - \sigma_1)(z - \sigma_2) \dots (z - \sigma_r)$ be the nodal polynomial associated with $\sigma_1, \sigma_2, \dots, \sigma_r$. Then for any monic polynomial $p_r \in \mathcal{P}_r$,

$$\mathbf{A}\mathbf{V} - \mathbf{V}\mathbf{A}_r = -p_r(\mathbf{A})[\omega_r(\mathbf{A})]^{-1}\mathbf{b}\mathbf{e}^T \quad (5.1)$$

and

$$\mathbf{W}^T \mathbf{A} - \mathbf{A}_r^T \mathbf{W}^T = -\mathbf{e}\mathbf{c}^T p_r(\mathbf{A})[\omega_r(\mathbf{A})]^{-1}, \quad (5.2)$$

where $\mathbf{A}_r = \Sigma_r - \mathbf{q}\mathbf{e}^T$ with $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$ and $q_i = \frac{p_r(\sigma_i)}{\omega_r'(\sigma_i)}$ for $i = 1, \dots, r$. Moreover, $p_r(z)$ is the characteristic polynomial for \mathbf{A}_r : $p_r(z) = \det(z\mathbf{I} - \mathbf{A}_r)$.

PROOF: Pick any index $1 \leq k \leq r$ and consider $f_k(z) = p_r(z) - z \cdot \prod_{i \neq k} (z - \sigma_i)$. Evidently, $f_k \in \mathcal{P}_{r-1}$ and so the Lagrange interpolant on $\sigma_1, \sigma_2, \dots, \sigma_r$ is exact:

$$f_k(z) = \sum_{i=1}^r f_k(\sigma_i) \frac{\omega_r(z)}{(z - \sigma_i)\omega_r'(\sigma_i)}$$

Divide by $\omega_r(z)$ and rearrange to get

$$\frac{z}{\sigma_k - z} - \sum_{i=1}^r \left(-\frac{f_k(\sigma_i)}{\omega_r'(\sigma_i)} \right) \frac{1}{\sigma_i - z} = -\frac{p_r(z)}{\omega_r(z)} \quad (5.3)$$

Let Γ be a Jordan curve that separates \mathbb{C} into two open, simply-connected sets, $\mathcal{C}_1, \mathcal{C}_2$ with \mathcal{C}_1 containing all the eigenvalues of \mathbf{A} and \mathcal{C}_2 containing both the point at ∞ and the shifts $\{\sigma_1, \dots, \sigma_r\}$. For any function $f(z)$ that is analytic in a compact set containing \mathcal{C}_1 , $f(\mathbf{A})$ can be defined as

$$f(\mathbf{A}) = \frac{1}{2\pi i} \int_{\Gamma} f(z) (z\mathbf{I} - \mathbf{A})^{-1} dz$$

Applying this to (5.3) gives

$$\mathbf{A}(\sigma_k \mathbf{I} - \mathbf{A})^{-1} - \sum_{i=1}^r \left(-\frac{f_k(\sigma_i)}{\omega_r'(\sigma_i)} \right) (\sigma_i \mathbf{I} - \mathbf{A})^{-1} = -p_r(\mathbf{A})[\omega_r(\mathbf{A})]^{-1}$$

Postmultiplication by \mathbf{b} provides the k^{th} column of (5.1), while premultiplication by \mathbf{c}^T (and since $(\sigma_k \mathbf{I} - \mathbf{A})^{-1}$ commutes with \mathbf{A}) provides the k^{th} row of (5.2).

The last statement follows by observing the alternative factorizations,

$$\begin{aligned} \begin{bmatrix} z\mathbf{I} - \Sigma_r & \mathbf{q} \\ \mathbf{e}^T & -1 \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & -\mathbf{q} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} z\mathbf{I} - \mathbf{A}_r & \mathbf{0} \\ \mathbf{0}^T & -1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{e}^T & 1 \end{bmatrix} \\ \begin{bmatrix} z\mathbf{I} - \Sigma_r & \mathbf{q} \\ \mathbf{e}^T & -1 \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{e}^T (z\mathbf{I} - \Sigma_r)^{-1} & 1 \end{bmatrix} \begin{bmatrix} z\mathbf{I} - \Sigma_r & \mathbf{0} \\ \mathbf{0}^T & -a(z) \end{bmatrix} \begin{bmatrix} \mathbf{I} & (z\mathbf{I} - \Sigma_r)^{-1} \mathbf{q} \\ \mathbf{0} & 1 \end{bmatrix} \end{aligned}$$

where $a(z) = 1 + \mathbf{e}^T (z\mathbf{I} - \Sigma_r)^{-1} \mathbf{q}$. Then we have that

$$\begin{aligned} \det(z\mathbf{I} - \mathbf{A}_r) &= \det(z\mathbf{I} - \Sigma_r) \cdot a(z) \\ &= \omega_r(z) \cdot (1 + \mathbf{e}^T (z\mathbf{I} - \Sigma_r)^{-1} \mathbf{q}) \\ &= \omega_r(z) + \sum_{i=1}^r p_r(\sigma_i) \left(\frac{\omega_r(z)}{\omega_r'(\sigma_i) (z - \sigma_i)} \right) = p_r(z) \end{aligned}$$

where the last equality follows by observing that the penultimate expression describes a monic polynomial of degree r that interpolates p_r at $\sigma_1, \sigma_2, \dots, \sigma_r$. \square

COROLLARY 5.2. *The matrices $\mathbf{W}^T \mathbf{A} \mathbf{V}$ and $\mathbf{W}^T \mathbf{V}$ are symmetric, though not necessarily Hermitian. If $\mathbf{W}^T(\boldsymbol{\sigma}) \mathbf{V}(\boldsymbol{\sigma})$ is nonsingular and*

$$p_r(z) = \det(z \mathbf{W}^T \mathbf{V} - \mathbf{W}^T \mathbf{A} \mathbf{V}) / \det(\mathbf{W}^T \mathbf{V})$$

then (5.1) and (5.2) hold with $\mathbf{A}_r = (\mathbf{W}^T \mathbf{V})^{-1} \mathbf{W}^T \mathbf{A} \mathbf{V}$. In particular, the Ritz matrix, $(\mathbf{W}^T \mathbf{V})^{-1} \mathbf{W}^T \mathbf{A} \mathbf{V}$ is a rank-one perturbation of the diagonal matrix of shifts, $\boldsymbol{\sigma}$.

PROOF: The symmetry of $\mathbf{W}^T \mathbf{A} \mathbf{V}$ and $\mathbf{W}^T \mathbf{V}$ can be checked directly. Now, choose a monic polynomial $\hat{p}_r \in \mathcal{P}_r$ so that $\mathbf{W}^T \hat{p}_r(\mathbf{A}) [\omega_r(\mathbf{A})]^{-1} \mathbf{b} = 0$. Then (5.1) and (5.2) hold with an associated $\mathbf{A}_r = \Sigma_r - \mathbf{q} \mathbf{e}^T$ as given in (5.1). But then applying \mathbf{W}^T to (5.1) leads to $\mathbf{A}_r = (\mathbf{W}^T \mathbf{V})^{-1} \mathbf{W}^T \mathbf{A} \mathbf{V}$. This in turn implies $\hat{p}_r(z) = p_r(z)$. \square

5.2. Shift Sensitivity. The matrices $\mathbf{W}^T \mathbf{A} \mathbf{V}$ and $\mathbf{W}^T \mathbf{V}$ are each continuously differentiable functions of $\boldsymbol{\sigma}$ for $\boldsymbol{\sigma}$ throughout the right halfplane. If $\mathbf{W}^T \mathbf{V}$ is invertible for shifts $\boldsymbol{\sigma}_c$ then $\mathbf{W}^T \mathbf{V}$ will be invertible and $(\mathbf{W}^T \mathbf{V})^{-1} \mathbf{W}^T \mathbf{A} \mathbf{V}$ will be continuously differentiable for $\boldsymbol{\sigma}$ in sufficiently small neighborhoods of $\boldsymbol{\sigma}_c$. Likewise if $\mathbf{A}_r = (\mathbf{W}^T \mathbf{V})^{-1} \mathbf{W}^T \mathbf{A} \mathbf{V}$ has simple eigenvalues at $\boldsymbol{\sigma}_c$ (the generic case) then the eigenvalues of \mathbf{A}_r , $\boldsymbol{\lambda} = \{\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_r\}$, viewed as functions of $\boldsymbol{\sigma} = \{\sigma_1, \sigma_2, \dots, \sigma_r\}$, $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\sigma})$, are continuously differentiable in a sufficiently small neighborhood of $\boldsymbol{\sigma}_c$.

The entries of the Jacobian matrix $\left(\frac{\partial \hat{\lambda}_i}{\partial \sigma_j}\right)$ provide a measure of the sensitivity of the reduced order poles, $\boldsymbol{\lambda}(\boldsymbol{\sigma})$, to perturbations of $\boldsymbol{\sigma}$. Write $\hat{\lambda}$ for $\hat{\lambda}_i$ and let $\hat{\mathbf{x}}$ be a unit eigenvector of $\mathbf{A}_r = (\mathbf{W}^T \mathbf{V})^{-1} \mathbf{W}^T \mathbf{A} \mathbf{V}$ associated with $\hat{\lambda}$, so

$$(a) \mathbf{W}^T \mathbf{A} \mathbf{V} \hat{\mathbf{x}} = \hat{\lambda} \mathbf{W}^T \mathbf{V} \hat{\mathbf{x}} \quad \text{and} \quad (b) \hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{A} \mathbf{V} = \hat{\lambda} \hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V}. \quad (5.4)$$

(5.4b) is obtained by transposition of (5.4a). $\hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V}$ is a left eigenvector for \mathbf{A}_r associated with $\hat{\lambda}_i$. Differentiate (5.4a) with respect to σ_j , premultiply with $\hat{\mathbf{x}}^T$, and simplify using (5.4b):

$$\hat{\mathbf{x}}^T \partial_j \mathbf{W}^T (\mathbf{A} \mathbf{V} \hat{\mathbf{x}} - \hat{\lambda} \mathbf{V} \hat{\mathbf{x}}) + \left(\hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{A} - \hat{\lambda} \hat{\mathbf{x}}^T \mathbf{W}^T\right) \partial_j \mathbf{V} \hat{\mathbf{x}} = \left(\frac{\partial \hat{\lambda}}{\partial \sigma_j}\right) \hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V} \hat{\mathbf{x}}$$

where $\partial_j \mathbf{W}^T = \frac{\partial}{\partial \sigma_j} \mathbf{W}^T = \mathbf{e}_j \mathbf{c}^T (\sigma_j \mathbf{I} - \mathbf{A})^{-2}$ and $\partial_j \mathbf{V} = \frac{\partial}{\partial \sigma_j} \mathbf{V} = (\sigma_j \mathbf{I} - \mathbf{A})^{-2} \mathbf{b} \mathbf{e}_j^T$. Notice that $\mathbf{V} \hat{\mathbf{x}}$ and $\hat{\mathbf{x}}^T \mathbf{W}^T$ are Galerkin approximations to right and left eigenvectors, respectively, of \mathbf{A} . $(\mathbf{A} \mathbf{V} \hat{\mathbf{x}} - \hat{\lambda} \mathbf{V} \hat{\mathbf{x}})$ and $(\hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{A} - \hat{\lambda} \hat{\mathbf{x}}^T \mathbf{W}^T)$ associated residuals so we might expect that reduced order poles become less sensitive to perturbations of shifts as the corresponding Ritz vectors become more accurate. However, other mechanisms can predict diminished sensitivity.

Suppose Ω is a set containing the eigenvalues of \mathbf{A} and define $\kappa(\Omega)$ as the smallest positive number so that

$$\|f(\mathbf{A})\| \leq \kappa(\Omega) \max_{z \in \Omega} |f(z)|$$

holds uniformly for all functions f analytic on Ω . Evidently, the value of $\kappa(\Omega)$ depends on the particular choice of Ω . $\kappa(\Omega)$ is monotone decreasing with respect to set inclusion on Ω . Indeed, if $\Omega_1 \subseteq \Omega_2$, then for each function f analytic on Ω_2 ,

$$\frac{\|f(\mathbf{A})\|}{\max\{|f(z)| : z \in \Omega_1\}} \geq \frac{\|f(\mathbf{A})\|}{\max\{|f(z)| : z \in \Omega_2\}}.$$

Thus, $\Omega_1 \subseteq \Omega_2$ implies $\kappa(\Omega_1) \geq \kappa(\Omega_2)$.

Since constant functions are always among the available analytic functions on Ω , $\kappa(\Omega) \geq 1$. If \mathbf{A} is normal, $\kappa(\Omega) = 1$. If \mathbf{A} is defective, then some choices of Ω will not yield a finite value for $\kappa(\Omega)$.

LEMMA 5.3. *Let $H_r^{(opt)}(s)$ be an \mathcal{H}_2 -optimal stable r^{th} order approximant to $H(s)$ with r (reduced order) poles at $\{\widehat{\lambda}_1, \widehat{\lambda}_2, \dots, \widehat{\lambda}_r\}$ contained in the right halfplane. Let Ω be a compact subset in the left halfplane containing the system poles of both $H(s)$ and $H_r^{(opt)}(s)$. Define*

$$\mathbf{D} = \text{diag}_j(\|(\sigma_j \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}\| \|(\sigma_j \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}\|), \quad C_r = \max_{i,k} \left(\frac{|\sigma_k + \overline{\sigma}_i|}{\text{Re}(\sigma_k)} \frac{\widehat{\mathbf{x}}_i^* \mathbf{D} \widehat{\mathbf{x}}_i}{|\widehat{\mathbf{x}}_i^T \mathbf{W}^T \mathbf{V} \widehat{\mathbf{x}}_i|} \right)$$

and $\delta = \max_{\substack{z \in \Omega \\ i=1, \dots, r}} \frac{|z - \widehat{\lambda}_i|}{|z + \lambda_i|}$. Then the Jacobian matrix, $\mathbf{J}(\boldsymbol{\sigma}_*)$ of $\boldsymbol{\lambda}(\boldsymbol{\sigma}_*)$ is bounded by

$$\|\mathbf{J}(\boldsymbol{\sigma}_*)\|_\infty \leq \kappa(\Omega) C_r \delta^{2r-1}$$

In particular, if the numerical range of \mathbf{A} is contained within Ω and if, as the model order r increases, C_r remains modestly bounded, then for sufficiently large r , the fixed point mapping $\sigma_i \leftarrow -\lambda_i(\mathbf{A}_r)$ can be expected to exhibit local linear convergence with a rate that accelerates exponentially with respect to model order.

PROOF: Optimality of $H_r^{(opt)}$ and the necessity for the reduced order system to be real implies that there is an ordering for $\{\widehat{\lambda}_1, \widehat{\lambda}_2, \dots, \widehat{\lambda}_r\}$ such that $\widehat{\lambda}_i = -\overline{\sigma}_i$. Thus (5.1) and (5.2) hold with $p_r(z) = \prod_{i=1}^r (z + \overline{\sigma}_i)$ and $\mathbf{A}_r = (\mathbf{W}^T \mathbf{V})^{-1} \mathbf{W}^T \mathbf{A} \mathbf{V}$.

Define

$$\rho(z) = \frac{p_r(z)}{\omega_r(z)} = \prod_{i=1}^r \frac{z + \overline{\sigma}_i}{z - \sigma_i}.$$

Then $\max_{z \in \Omega} |\rho(z)| \leq \delta^r$ and $\|\rho(\mathbf{A})\| \leq \kappa(\Omega) \delta^r$. Note that from Corollary (5.2),

$$\mathbf{A} \mathbf{V} \widehat{\mathbf{x}} - \widehat{\lambda} \mathbf{V} \widehat{\mathbf{x}} = -\rho(\mathbf{A}) \mathbf{b} \mathbf{e}^T \widehat{\mathbf{x}} \quad \text{and} \quad \widehat{\mathbf{x}}^T \mathbf{W}^T \mathbf{A} - \widehat{\lambda} \widehat{\mathbf{x}}^T \mathbf{W}^T = -\widehat{\mathbf{x}}^T \mathbf{e} \mathbf{c}^T \rho(\mathbf{A}).$$

Thus, with the j^{th} component of $\widehat{\mathbf{x}}$ denoted as \widehat{x}_j , we have

$$\begin{aligned} \left(\frac{\partial \widehat{\lambda}}{\partial \sigma_j} \right) \widehat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V} \widehat{\mathbf{x}} &= -\widehat{x}_j \mathbf{c}^T (\sigma_j \mathbf{I} - \mathbf{A})^{-2} \rho(\mathbf{A}) \mathbf{b} \mathbf{e}^T \widehat{\mathbf{x}} - \widehat{\mathbf{x}}^T \mathbf{e} \mathbf{c}^T \rho(\mathbf{A}) (\sigma_j \mathbf{I} - \mathbf{A})^{-2} \mathbf{b} \widehat{x}_j \\ &= -\widehat{x}_j \mathbf{c}^T (\sigma_j \mathbf{I} - \mathbf{A})^{-1} \rho(\mathbf{A}) (\sigma_j \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \mathbf{e}^T \widehat{\mathbf{x}} \\ &\quad - \widehat{\mathbf{x}}^T \mathbf{e} \mathbf{c}^T (\sigma_j \mathbf{I} - \mathbf{A})^{-1} \rho(\mathbf{A}) (\sigma_j \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \widehat{x}_j \\ &= -2 \mathbf{c}^T (\sigma_j \mathbf{I} - \mathbf{A})^{-1} \rho(\mathbf{A}) (\sigma_j \mathbf{I} - \mathbf{A})^{-1} \mathbf{b} (\widehat{x}_j \mathbf{e}^T \widehat{\mathbf{x}}) \end{aligned}$$

Comparing the j^{th} components of $(\Sigma_r - \mathbf{q} \mathbf{e}^T) \widehat{\mathbf{x}} = \widehat{\lambda} \widehat{\mathbf{x}}$, and rearranging gives

$$(\sigma_j + \overline{\sigma}_i) \widehat{x}_j = (\sigma_j - \widehat{\lambda}_i) \widehat{x}_j = q_j \mathbf{e}^T \widehat{\mathbf{x}} = \frac{p_r(\sigma_j)}{\omega'_r(\sigma_j)} \mathbf{e}^T \widehat{\mathbf{x}}.$$

Thus,

$$|\mathbf{e}^T \widehat{\mathbf{x}}| = \frac{|\sigma_j + \overline{\sigma}_i|}{2 \text{Re}(\sigma_j)} \left(\prod_{k \neq j} \frac{|\sigma_j - \sigma_k|}{|\sigma_j + \overline{\sigma}_k|} \right) |\widehat{x}_j| \leq \frac{|\sigma_j + \overline{\sigma}_i|}{2 \text{Re}(\sigma_j)} |\widehat{x}_j| \delta^{r-1}$$

Now,

$$\begin{aligned} \left| \frac{\partial \widehat{\lambda}}{\partial \sigma_j} \right| &\leq \frac{|\mathbf{c}^T (\sigma_j \mathbf{I} - \mathbf{A})^{-1} \rho(\mathbf{A}) (\sigma_j \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}| |\hat{x}_j|^2}{|\hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V} \hat{\mathbf{x}}|} \frac{|\sigma_j + \bar{\sigma}_i|}{\text{Re}(\sigma_j)} \delta^{r-1} \\ &\leq \|\rho(\mathbf{A})\| \frac{\|(\sigma_j \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}\| \|(\sigma_j \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}\| |\hat{x}_j|^2}{|\hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V} \hat{\mathbf{x}}|} \frac{|\sigma_j + \bar{\sigma}_i|}{\text{Re}(\sigma_j)} \delta^{r-1} \\ &\leq \kappa(\Omega) \frac{\|(\sigma_j \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}\| \|(\sigma_j \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}\| |\hat{x}_j|^2}{|\hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V} \hat{\mathbf{x}}|} \frac{|\sigma_j + \bar{\sigma}_i|}{\text{Re}(\sigma_j)} \delta^{2r-1} \end{aligned}$$

So, in particular,

$$\begin{aligned} \sum_{j=1}^r \left| \frac{\partial \widehat{\lambda}}{\partial \sigma_j} \right| &\leq \kappa(\Omega) \sum_{j=1}^r \left(\frac{\|(\sigma_j \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}\| \|(\sigma_j \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}\| |\hat{x}_j|^2}{|\hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V} \hat{\mathbf{x}}|} \frac{|\sigma_j + \bar{\sigma}_i|}{\text{Re}(\sigma_j)} \right) \delta^{2r-1} \\ &\leq \kappa(\Omega) \sum_{j=1}^r \left(\frac{\|(\sigma_j \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}\| \|(\sigma_j \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}\| |\hat{x}_j|^2}{|\hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V} \hat{\mathbf{x}}|} \frac{|\sigma_j + \bar{\sigma}_i|}{\text{Re}(\sigma_j)} \right) \delta^{2r-1} \\ &\leq \kappa(\Omega) \max_k \frac{|\sigma_k + \bar{\sigma}_i|}{\text{Re}(\sigma_k)} \sum_{j=1}^r \left(\frac{\|(\sigma_j \mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{c}\| \|(\sigma_j \mathbf{I} - \mathbf{A})^{-1} \mathbf{b}\| |\hat{x}_j|^2}{|\hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V} \hat{\mathbf{x}}|} \right) \delta^{2r-1} \\ &\leq \kappa(\Omega) \max_k \frac{|\sigma_k + \bar{\sigma}_i|}{\text{Re}(\sigma_k)} \frac{\hat{\mathbf{x}}^* \mathbf{D} \hat{\mathbf{x}}}{|\hat{\mathbf{x}}^T \mathbf{W}^T \mathbf{V} \hat{\mathbf{x}}|} \delta^{2r-1} \end{aligned}$$

Directly then we obtain

$$\|\mathbf{J}(\boldsymbol{\sigma}_*)\|_\infty = \max_i \sum_{j=1}^r \left| \frac{\partial \widehat{\lambda}_i}{\partial \sigma_j} \right| \leq \kappa(\Omega) \max_{i,k} \left(\frac{|\sigma_k + \bar{\sigma}_i|}{\text{Re}(\sigma_k)} \frac{\hat{\mathbf{x}}_i^* \mathbf{D} \hat{\mathbf{x}}_i}{|\hat{\mathbf{x}}_i^T \mathbf{W}^T \mathbf{V} \hat{\mathbf{x}}_i|} \right) \delta^{2r-1}$$

6. Numerical Examples. In this section, we first compare our approach with the existing ones [22, 24, 32] for a number of *low order* benchmark examples presented in these papers, and show that in each case, we attain the minimum; the main difference, however, is that we achieve this minimum in a numerically efficient way. Then we test our method in a large-scale setting.

6.1. Low-order Models and Comparisons. We consider the following 4 models:

- FOM-1: Example 6.1 in [22]. State-space representation of FOM-1 is given by

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & -150 \\ 1 & 0 & 0 & -245 \\ 0 & 1 & 0 & -113 \\ 0 & 0 & 1 & -19 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

We reduce the order to $r = 3, 2, 1$ using the proposed successive rational Krylov algorithm, denoted by **IRKA** and compare our results with the gradient flow method of [32], denoted by **GFM**; the orthogonal projection method of [22], denoted by **OPM**; and the balanced truncation method, denoted by **BTM**.

- FOM-2: Example in [24]. Transfer function of FOM-2 is given by

$$G(s) = \frac{2s^6 + 11.5s^5 + 57.75s^4 + 178.625s^3 + 345.5s^2 + 323.625s + 94.5}{s^7 + 10s^6 + 46s^5 + 130s^4 + 239s^3 + 280s^2 + 194s + 60}$$

We reduce the order to $r = 6, 5, 4, 3$ using **IRKA**, and compare our results with **GFM**; **OPM**; **BTM**; and the method proposed in [24], denoted by **LMPV**.

- FOM-3: Example 1 in [30]. Transfer function of FOM-3 is given by

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

We reduce the order to $r = 3, 2, 1$ using **IRKA**, and compare our results with **GFM**; **OPM**; **BTM**; and the method proposed in [30], denoted by **SMM**.

- FOM-4: Example 2 in [30]. Transfer function of FOM-4 is given by

$$G(s) = \frac{10000s + 5000}{s^2 + 5000s + 25}$$

We reduce the order to $r = 1$ **IRKA** and compare our results with **GFM**; **OPM**; **BTM**; and **SMM**.

For all these cases, the resulting relative \mathcal{H}_2 errors $\frac{\|G(s) - G_r(s)\|_{\mathcal{H}_2}}{\|G(s)\|_{\mathcal{H}_2}}$ are tabulated in Table 6.1 below.

Table 6.1 clearly illustrates that the proposed method is the only one which attains the minimum in each case. More importantly, the proposed method achieves this value in a numerically efficient way staying in the Krylov projection framework. No Lyapunov solvers or dense matrix decompositions are needed. The only arithmetic operations involved are LU decompositions and some linear solvers. Moreover, our method does not require starting from an initial balanced realization as suggested in [32] and [22]. In all these simulations, we have chosen a random initial shift selection and the algorithm converged in a small number of steps.

To illustrate the evolution of the \mathcal{H}_2 error throughout the iteration, consider the model FOM-2 with $r = 3$. The proposed method yields the following third order optimal reduced model:

$$G_3(s) = \frac{2.155s^2 + 3.343s + 33.8}{s^3 + 7.457s^2 + 10.51s + 17.57}$$

Poles of $G_3(s)$ are $\hat{\lambda}_1 = -6.2217$ and $\hat{\lambda}_{2,3} = -6.1774 \times 10^{-1} \pm j1.5628$; and it can be shown that $G_3(s)$ interpolates the first two moments of $G(s)$ at $-\hat{\lambda}_i$, for $i = 1, 2, 3$. Hence, the first-order interpolation conditions are satisfied. This also means that if we start Algorithm 3.1 with the mirror images of these Ritz values, the algorithm converges at the first step. However, we will try four random, but *bad*, initial selections. In other words, we start away from the optimal solution. We test the following four selections:

$$\mathcal{S}_1 = \{-1.01, -2.01, -30000\} \quad (6.1)$$

$$\mathcal{S}_2 = \{0, 10, 3\} \quad (6.2)$$

$$\mathcal{S}_3 = \{1, 10, 3\} \quad (6.3)$$

$$\mathcal{S}_4 = \{0.01, 20, 10000\} \quad (6.4)$$

Model	r	IRKA	GFM	OPM
FOM-1	1	4.2683×10^{-1}	4.2709×10^{-1}	4.2683×10^{-1}
FOM-1	2	3.9290×10^{-2}	3.9299×10^{-2}	3.9290×10^{-2}
FOM-1	3	1.3047×10^{-3}	1.3107×10^{-3}	1.3047×10^{-3}
FOM-2	3	1.171×10^{-1}	1.171×10^{-1}	Divergent
FOM-2	4	8.199×10^{-3}	8.199×10^{-3}	8.199×10^{-3}
FOM-2	5	2.132×10^{-3}	2.132×10^{-3}	Divergent
FOM-2	6	5.817×10^{-5}	5.817×10^{-5}	5.817×10^{-5}
FOM-3	1	4.818×10^{-1}	4.818×10^{-1}	4.818×10^{-1}
FOM-3	2	2.443×10^{-1}	2.443×10^{-1}	Divergent
FOM-3	3	5.74×10^{-2}	5.98×10^{-2}	5.74×10^{-2}
FOM-4	1	9.85×10^{-2}	9.85×10^{-2}	9.85×10^{-2}

Model	r	BTM	LMPV	SMM
FOM-1	1	4.3212×10^{-1}		
FOM-1	2	3.9378×10^{-2}		
FOM-1	3	1.3107×10^{-3}		
FOM-2	3	2.384×10^{-1}	1.171×10^{-1}	
FOM-2	4	8.226×10^{-3}	8.199×10^{-3}	
FOM-2	5	2.452×10^{-3}	2.132×10^{-3}	
FOM-2	6	5.822×10^{-5}	2.864×10^{-4}	
FOM-3	1	4.848×10^{-1}		4.818×10^{-1}
FOM-3	2	3.332×10^{-1}		2.443×10^{-1}
FOM-3	3	5.99×10^{-2}		5.74×10^{-2}
FOM-4	1	9.949×10^{-1}	9.985×10^{-2}	

TABLE 6.1
Comparison

With selection \mathcal{S}_1 , we have initiated the algorithm with some negative shifts close to system poles, and consequently with a relative \mathcal{H}_2 error bigger than 1. However, in all four cases including \mathcal{S}_1 , the algorithm converged in 5 steps to the same reduced model. The results are depicted in Figure 6.1.

Before testing the proposed method in large-scale settings, we investigate FOM-4 further. As pointed out in [30], since $r = 1$, the optimal \mathcal{H}_2 problem can be formulated as only a function of the reduced system pole. It was shown in [30] that there are two local minima: one corresponding to a reduced pole at -0.0052 and consequently a reduced order model $G_1^l(s) = \frac{1.0313}{s+0.0052}$ and a relative error of 0.9949; and one to a reduced pole at -4998 and consequently a reduced model $G_1^g = \frac{9999}{s+4998}$ with a relative error of 0.0985. It follows that the latter, i.e. $G_1^g(s)$ is the global minimum. The first order balanced truncation for FOM-4 can be easily computed as $G_1^b(s) = \frac{1.0308}{s+0.0052}$. Therefore, it is highly likely that if one starts from a balanced realization, the algorithm would converge to the local minimum $G_1^l(s)$. This was indeed the case as reported in [30]. **SMM** converged to the local minimum for all starting poles bigger than -0.47 . On the other hand, **SMM** converged to the global minimum when it was started with an initial pole smaller than -0.47 . We have observed exactly the same situation in our simulations. When we start from an initial shift selection smaller than 0.48, **IRKA** converged to the local minimum. However, when we start with any initial shift bigger than 0.48, the algorithm converged to the

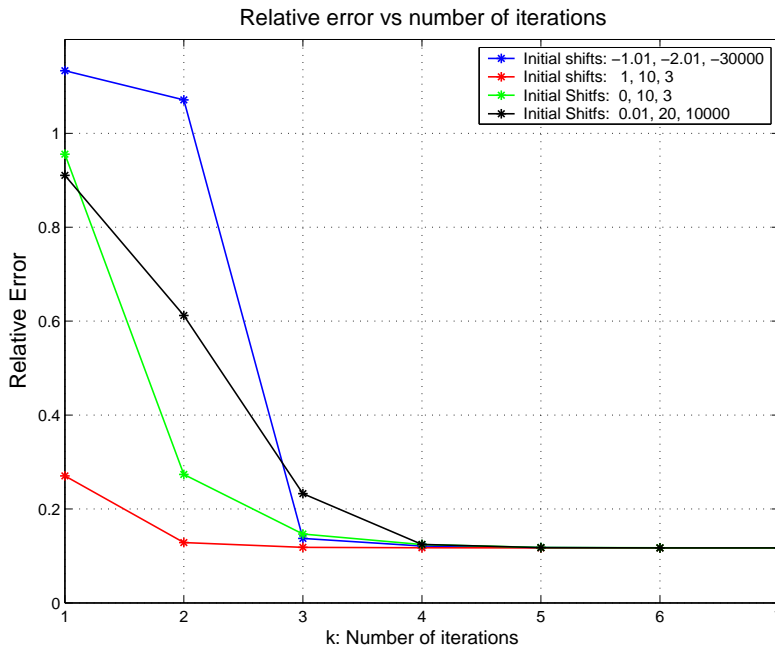


FIG. 6.1. \mathcal{H}_2 norm of the error system vs the number of iterations

global minimum in at most 3 steps. Therefore, for this example we were not able to avoid the local minimum if we start from a *bad* shift. These observations perfectly agree with the discussion of Section 3.4. Note that, transfer function of FOM-4 can be written as

$$G(s) = \frac{10000s + 5000}{s^2 + 5000s + 25} = \frac{0.99}{s + 0.0050} + \frac{9999}{s + 5000}.$$

The pole at -5000 is the one corresponding to the large residue of 9999. Therefore, a good initial shift is 5000. And if we start the proposed algorithm with an initial shift at 5000, or close, the algorithm converges to the global minimum.

6.2. CD Player Example. The original model describes the dynamics between a lens actuator and the radial arm position in a portable CD player. The model has 120 states, i.e., $n=120$, with a single input and a single output. As illustrated in [5], the Hankel singular values of this model do not decay rapidly and hence the model is hard to reduce. Moreover, even though the Krylov-based methods resulted in good local behavior, they are observed to yield large \mathcal{H}_∞ and \mathcal{H}_2 error compared to balanced truncation.

We compare the performance of the proposed method, Algorithm 3.1 with that of balanced truncation. Balanced truncation is well known as leading to small \mathcal{H}_∞ and \mathcal{H}_2 error norms, see [5, 17], mainly due to global information available through the usage of the two system gramians, the reachability and observability gramians, which are each solutions of a different Lyapunov equation. We reduce the order to r with r varying from 2 to 40; and for each r value, we compare the \mathcal{H}_2 error norms due to balanced truncation and due to Algorithm 3.1. For the proposed algorithm, two different selections have been tried for the initial shifts. **1.** Mirror images of the

eigenvalues corresponding to large residuals, and **2**. A random selection with real parts in the interval $[10^{-1}, 10^3]$ and the imaginary parts in the interval $[1, 10^5]$. To make this selection, we looked at the poles of $G(s)$ having the maximum/minimum real and imaginary parts. The results showing the relative \mathcal{H}_2 error for each r are depicted in Figure 6.2. The figure reveals that both selection strategies work quite well. Indeed, the random initial selection behaves better than the residual-based selection and outperforms balanced truncation for almost all the r values except $r = 2, 24, 36$. However, even for these r values, the resulting \mathcal{H}_2 error is not far away from the one due to balanced truncation. For the range $r = [12, 22]$, the random selection clearly outperforms the balanced truncation. We would like to emphasize that these results were obtained by a *random* shift selection and staying in the numerically effective Krylov projection framework *without* requiring any solutions to large-scale Lyapunov equations. This is the main difference of our algorithm with the existing methods and this makes the proposed algorithm numerically effective in large-scale settings.

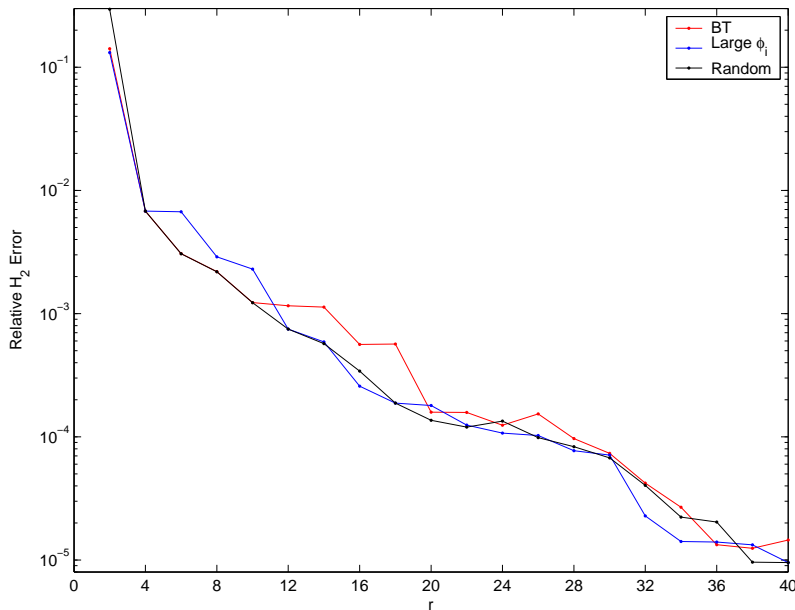


FIG. 6.2. Relative \mathcal{H}_2 norm of the error system vs r

To examine convergence behavior, we reduce the order to $r = 8$ and $r = 10$ using Algorithm 3.1. At each step of the iteration, we compute the \mathcal{H}_2 error due to the current estimate and plot this error vs the iteration index. The results are shown in Figure 6.3. The figure illustrates two important properties for both cases $r = 8$ and $r = 10$: **(1)** At each step of the iteration, the \mathcal{H}_2 norm of the error is reduced. **(2)** The algorithm converges after 3 steps. The resulting reduced models are stable for both cases.

6.3. A Semi-discretized Heat Transfer Problem for Optimal Cooling of Steel Profiles. This problem arises during a cooling process in a rolling mill when different steps in the production process require different temperatures of the raw material. To achieve high throughput, one seeks to reduce the temperature as

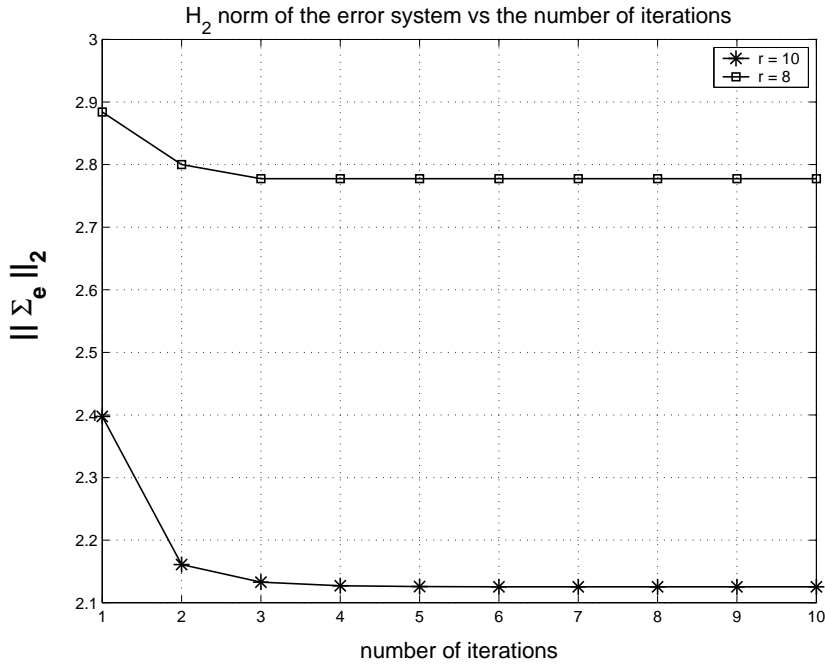


FIG. 6.3. \mathcal{H}_2 norm of the error system vs the number of iterations

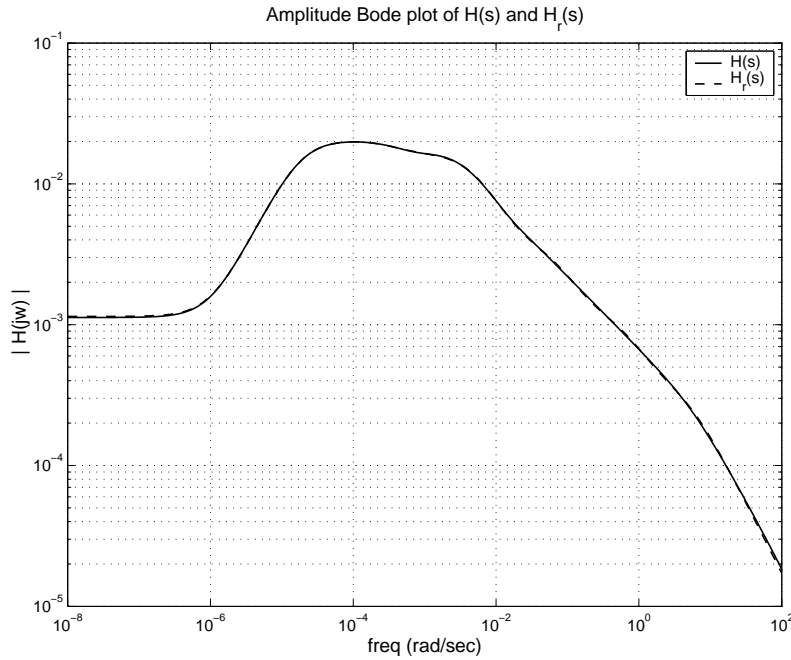
fast as possible to the required level before entering the next production phase. But the cooling process is realized by spraying cooling fluids on the surface and must be controlled so that material properties, such as durability or porosity, stay within given quality standards. The problem is modeled as boundary control of a two dimensional heat equation. A finite element discretization using two steps of mesh refinement with maximum mesh width of 1.382×10^{-2} results in a system of the form

$$\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t), \quad y(t) = \mathbf{c}^T \mathbf{x}(t).$$

with state-dimension $n = 20209$, i.e., $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{20209 \times 20209}$, $\mathbf{b} \in \mathbb{R}^{20209 \times 7}$, $\mathbf{c}^T \in \mathbb{R}^{6 \times 20209}$. Note that in this case $\mathbf{E} \neq \mathbf{I}_n$, but the algorithm works with the obvious modifications. For details regarding the modeling, discretization, optimal control design, and model reduction for this example, see [26, 7, 8]. We consider the full-order SISO system that associates the sixth input of this system with the second output. We apply our algorithm and reduce the order to $r = 6$. Amplitude Bode plots of $G(s)$ and $G_r(s)$ are shown in Figure 6.4. The output response of $G_r(s)$ is virtually indistinguishable from $G(s)$ in the frequency range considered. **IRKA** converged in 7 iteration steps in this case, although some interpolation points converged in the first 2-3 steps. The relative \mathcal{H}_∞ error obtained with our sixth order system was 7.85×10^{-3} . Note that in order to apply balanced truncation in this example, one would need to solve *two generalized* Lyapunov equations (since $\mathbf{E} \neq \mathbf{I}_n$) of order 20209.

6.4. Successive Substitution vs Newton Framework. In this section, we present two examples to show the effect of the Newton formulation for **IRKA** on two low-order examples.

The first example is **FOM-1** from Section 6.1. For this example, for reduction to $r = 1$, the optimal shift is $\sigma = 0.4952$. We initiate the both iterations, successive

FIG. 6.4. Amplitude Bode plots of $H(s)$ and $H_r(s)$

substitution and Newton frameworks, away from this optimal value with an initial selection $\sigma_0 = 10^4$. Figure 6.5 illustrates how each process converges. As the figure shows, even though it takes almost 15 iterations with oscillations for the successive substitution framework to converge, the Newton formulation reaches the optimal shift in 4 steps.

The second example in this section is a third order model with a transfer function

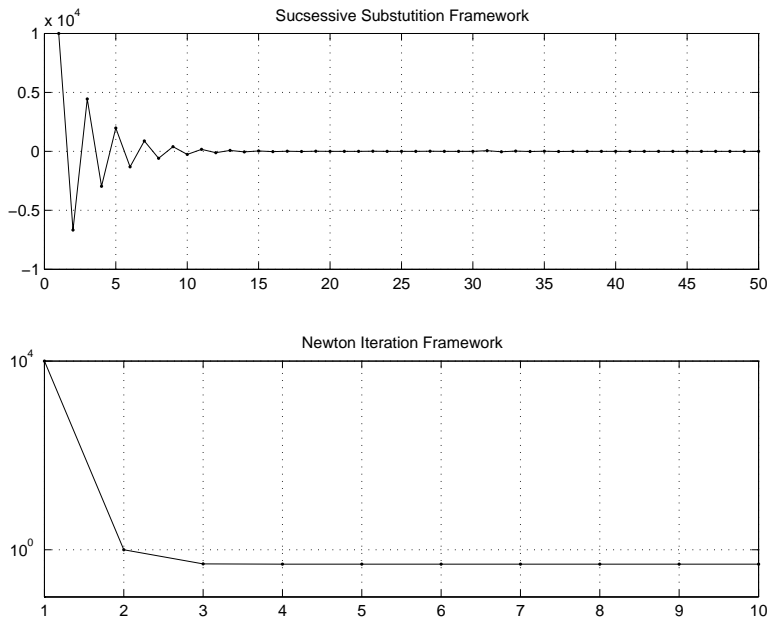
$$G = \frac{-s^2 + (7/4)s + 5/4}{s^3 + 2s^2 + (17/16)s + 15/32}.$$

One can exactly compute the optimal \mathcal{H}_2 reduced model for $r = 1$ as

$$G_r(s) = \frac{0.97197}{s + 0.2727272}$$

One can easily show that this reduced model interpolates $G(s)$ and its derivative at $\sigma = 0.2727272$. We initiate Algorithm 3.1 with $\sigma_0 = 0.27$, very close to the optimal shift. We initiate the Newton framework at $\sigma_0 = 2000$, far away from the optimal solution. Convergence behavior of both models is depicted in Figure 6.6 below. The figure shows that for this example, the successive substitution framework is divergent and indeed $\frac{\partial \lambda}{\partial \sigma} \approx 1.3728$. On the other hand, the Newton framework is able to converge to the optimal solution in a small number of steps.

7. Conclusions. We have developed an interpolation based rational Krylov algorithm that iteratively corrects interpolation locations until first-order \mathcal{H}_2 -optimality conditions are satisfied. The resulting method proves numerically effective and well suited for large-scale problems. A new derivation of the interpolation-based necessary

FIG. 6.5. Comparison for **FOM-1**

conditions are presented and shown to be equivalent to two other common frameworks for \mathcal{H}_2 -optimality. We offer a larger framework within which to understand rational Krylov methods and use this setting to provide asymptotic bounds to shift sensitivity.

REFERENCES

- [1] A.C. Antoulas, *On recursiveness and related topics in linear systems*, IEEE Trans. Automatic Control, **AC-31**: 1121-1135 (1986).
- [2] A.C. Antoulas and J.C. Willems, *A behavioral approach to linear exact modeling*, IEEE Trans. Automatic Control, **AC-38**: 1776-1802 (1993).
- [3] A.C. Antoulas, *Recursive modeling of discrete-time time series*, in IMA volume on Linear Algebra for Control, P. van Dooren and B.W. Wyman Editors, Springer Verlag, vol. **62**: 1-20 (1993).
- [4] A.C. Antoulas, *Approximation of large-scale dynamical systems*, Advances in Design and Control DC-06, SIAM, Philadelphia, 2005.
- [5] 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).
- [6] L. Baratchart, M. Cardelli and M. Olivi, *Identification and rational ℓ_2 approximation: a gradient algorithm*, Automatica, **27**: 413-418 (1991).
- [7] P. Benner, *Solving Large-Scale Control Problems*, IEEE Control Systems Magazine, Vol. 24, No. 1, pp. 44-59, 2004.
- [8] P. Benner and J. Saak, *Efficient numerical solution of the LQR-problem for the heat equation*, submitted to Proc. Appl. Math. Mech., 2004.
- [9] A.E. Bryson and A. Carrier, *Second-order algorithm for optimal model order reduction*, J. Guidance Contr. Dynam., pp-887-892, 1990
- [10] C. De Villemagne and R. Skelton, *Model reduction using a projection formulation*, International Journal of Control, **40**: 2141-2169 (1987).
- [11] P. Feldman and R.W. Freund, *Efficient linear circuit analysis by Padé approximation via a Lanczos method*, IEEE Trans. Computer-Aided Design, **14**: 639-649 (1995).
- [12] P. Fulcheri and M. Olivi, *Matrix rational \mathcal{H}_2 approximation: a gradient algorithm based on Schur analysis*, SIAM Journal on Control and Optimization, **36**: 2103-2127 (1998).
- [13] D. Gaier, *Lectures on Complex Approximation*, Birkhäuser, 1987.

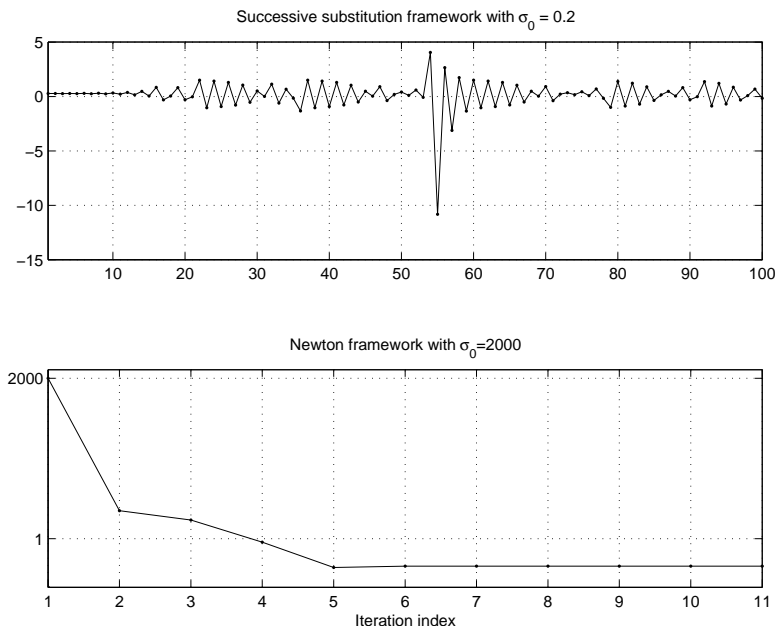


FIG. 6.6. Comparison for the random third order model

- [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, P. Van Dooren, and E. Grimme, *On some recent developments in projection-based model reduction*, in ENUMATH 97 (Heidelberg), World Sci. Publishing, River Edge, NJ, 1998, pp. 98–113, 1998.
- [16] E.J. Grimme, *Krylov Projection Methods for Model Reduction*, Ph.D. Thesis, ECE Dept., University of Illinois, Urbana-Champaign, 1997.
- [17] 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, December 2000.
- [18] S. Gugercin, *Projection methods for model reduction of large-scale dynamical systems*, Ph.D. Thesis, ECE Dept., Rice University, December 2002.
- [19] 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.
- [20] Y. Halevi, *Frequency weighted model reduction via optimal projection*, in Proc. IEEE Conf. Decision and Control, pp. 2906-2911, 1990.
- [21] D. Higham and N. Higham, *Structured backward error and condition of generalized eigenvalue problems*, SIAM Journal on Matrix Analysis and Applications, **20**: 493-512 (1998).
- [22] D.C. Hyland and D.S. Bernstein, *The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore*, IEEE Trans. Automatic Control, **30**: 1201-1211 (1985).
- [23] J.G. Korvink and E.B. Rudnyi, *Oberwolfach Benchmark Collection*, In P. Benner, G. Golub, V. Mehrmann, and D. Sorensen, editors, Dimension Reduction of Large-Scale Systems, Lecture Notes in Computational Science and Engineering, Vol. 45, Springer-Verlag, 2005.
- [24] A. Lepschy, G.A. Mian, G. Pinato and U. Viaro, *Rational L_2 approximation: A non-gradient algorithm*, Proceedings 30th IEEE CDC, 1991.
- [25] L. Meier and D.G. Luenberger, *Approximation of Linear Constant Systems*, IEEE Trans. Automatic Control, **12**: 585-588 (1967).
- [26] T. Penzl, *Algorithms for model reduction of large dynamical systems*, Technical Report SFB393/99-40, TU Chemnitz, 1999. Available from <http://www.tu-chemnitz.de/sfb393/sfb99pr.html>.
- [27] L.T. Pillage and R.A. Rohrer, *Asymptotic waveform evaluation for timing analysis*, IEEE Trans. Computer-Aided Design, **9**: 352-366 (1990).
- [28] V. Raghavan, R.A. Rohrer, L.T. Pillage, J.Y. Lee, J.E. Bracken, and M.M. Alaybeyi, *AWE*

- Inspired*, Proc. IEEE Custom Integrated Circuits Conference, May 1993.
- [29] A. Ruhe, *Rational Krylov algorithms for nonsymmetric eigenvalue problems II: matrix pairs*, Linear Alg. Applications, **197**: 283-295 (1994).
 - [30] J.T. Spanos, M.H. Milman, and D.L. Mingori, *A new algorithm for \mathcal{L}_2 optimal model reduction*, Automatica, **28**: 897-909 (1992).
 - [31] D.A. Wilson, *Optimum solution of model reduction problem*, in Proc. Inst. Elec. Eng., pp. 1161-1165, 1970.
 - [32] W-Y. Yan and J. Lam, *An approximate approach to \mathcal{H}_2 optimal model reduction*, IEEE Trans. Automatic Control, **AC-44**: 1341-1358 (1999).
 - [33] A. Yousoff 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.
 - [34] A. Yousoff, 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.
 - [35] D. Zigic, L. Watson, and C. Beattie, *Contragredient transformations applied to the optimal projection equations*. Linear Algebra and its Applications, **188/189**: 665-676 (1993)