Inexact Solves in Krylov-based Model Reduction

Christopher A. Beattie and Serkan Gugercin

***Abstract*— We investigate the use of inexact solves in a Krylov-based model reduction setting and present the resulting perturbation effects on the underlying model reduction problem. We show that for a *good* selection of interpolation points, Krylov-based model reduction is robust with respect to the perturbations due to inexact solves. On the other hand, when the interpolation points are *poorly* selected, these perturbations are magnified through the model reduction process. Finally, we incorporate inexact solves for the Krylov-based optimal $\mathcal{H}_2$ approximation. The result is an effective optimal model reduction algorithm applicable in realistic large-scale settings.**

I. INTRODUCTION

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

$$\mathbf{H}(s) \quad : \quad \begin{cases} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{b}\,u(t) \\ y(t) &= \mathbf{c}^T\mathbf{x}(t), \end{cases} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$. $\mathbf{x}(t) \in \mathbb{R}^n$ is the *state*, $u(t) \in \mathbb{R}$ is the *input*, and $y(t) \in \mathbb{R}$ is the *output* of $\mathbf{H}(s)$. The transfer function of $\mathbf{H}(s)$ is given by $\mathbf{H}(s) = \mathbf{c}^T(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}$. Both the underlying dynamical system and its transfer function will be denoted by $\mathbf{H}(s)$.

In many applications, the system dimension n is too large for efficient simulation and control computation; see [15] for a recent collection of such benchmark problems. The goal of model reduction is, then, to produce a much smaller order system $\mathbf{H}_r(s)$ with state-space form:

$$\mathbf{H}_r(s) \quad : \quad \begin{cases} \dot{\mathbf{x}}_r(t) = \mathbf{A}_r\mathbf{x}_r(t) + \mathbf{b}_r u(t) \\ y_r(t) = \mathbf{c}_r^T\mathbf{x}(t), \end{cases} \tag{2}$$

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

We construct reduced order models through projection. That is, we construct matrices $\mathbf{V}_r \in \mathbb{R}^{n\times r}$ and $\mathbf{W}_r \in \mathbb{R}^{n\times r}$ such that $\mathbf{W}_r^T\mathbf{V}_r = \mathbf{I}_r$ and the reduced order model $\mathbf{H}_r(s)$ in (2) is then obtained as

$$\mathbf{A}_r = \mathbf{W}_r^T\mathbf{A}\mathbf{V}_r, \quad \mathbf{b}_r = \mathbf{W}_r^T\mathbf{b}, \quad \text{and} \quad \mathbf{c}_r^T = \mathbf{c}^T\mathbf{V}_r. \tag{3}$$

The corresponding oblique projector is given by $\mathbf{W}_r\mathbf{V}_r^T$.

II. KRYLOV-BASED MODEL REDUCTION

Given a set of interpolation points $\{\sigma_i\}_{i=1}^r$, Krylov-based model reduction requires solving the linear systems $(\sigma_j\mathbf{I} - \mathbf{A})\mathbf{x}_j = \mathbf{b}$, and $(\overline{\sigma_j}\mathbf{I} - \mathbf{A}^T)y_j = \mathbf{c}$ for $j = 1, \ldots, r$ and computing orthogonal bases for the subspaces span$\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ and span$\{\mathbf{y}_1, \cdots, \mathbf{y}_r\}$. Define the Krylov matrices:

$$\begin{aligned} \mathbf{K}_1 &= [\ \mathbf{x}_1, \ \cdots \ \mathbf{x}_r\ ] \\ &= \left[\ (\sigma_1\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}, \ \cdots, \ (\sigma_r\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}\ \right] \end{aligned} \tag{4}$$

and

$$\begin{aligned} \mathbf{K}_2 &= [\ y_1, \ \cdots, \ y_r\ ] \\ &= \left[(\overline{\sigma_1}\mathbf{I}-\mathbf{A}^T)^{-1}\mathbf{c}, \ \cdots, \ (\overline{\sigma_r}\mathbf{I}-\mathbf{A}^T)^{-1}\mathbf{c}\right] \end{aligned} \tag{5}$$

and let

$$\mathbf{K}_1 = \mathbf{V}_r\mathbf{R}_1 \quad \text{and} \quad \mathbf{K}_2 = \mathbf{W}_r\mathbf{R}_2 \tag{6}$$

be biorthogonal-decompositions of $\mathbf{K}_1$ and $\mathbf{K}_2$, respectively, such that $\mathbf{R}_1$ and $\mathbf{R}_2$ are upper triangular and $\mathbf{W}_r^T\mathbf{V}_r = \mathbf{I}$ (see [21]). Then, the reduced-model $\mathbf{H}_r(s) = \mathbf{c}_r^T(s\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{b}_r$ is obtained by projection as in (3) using $\mathbf{W}_r$ and $\mathbf{V}_r$ in (6). $\mathbf{H}_r(s)$ interpolates $\mathbf{H}(s)$ together with its first derivative (first moment in the time domain) at the selected interpolation points $\{\sigma_j\}_{j=1}^r$. This discussion can be generalized to allow interpolation of higher order moments/derivatives of $\mathbf{H}(s)$ as well, see [10], [12], [1], [8] and references therein.

Krylov-based model reduction as exemplified in [10], [7], has become the method of choice for large-scale problems since it does not require any dense matrix operations; the same advantages enjoyed by the Arnoldi[2], Lanczos [16], and rational Krylov methods [24]. This contrasts with Gramian-based model reduction approaches such as balanced truncation [20], [19], optimal Hankel norm approximation [9] and singular perturbation approximation [17]. Krylov-based model reduction has a computational complexity of order $\mathcal{O}(nr^2)$ compared to $\mathcal{O}(n^3)$ for gramian-based reduction. Moreover, Gugercin *et al.* [13] recently introduced a strategy for selecting optimal interpolation points σ_i for Krylov-based model reduction.

III. INEXACT KRYLOV-BASED MODEL REDUCTION

Krylov-based model reduction methods presume that the linear systems $(\sigma_j\mathbf{I} - \mathbf{A})\mathbf{x}_j = \mathbf{b}$ and $(\overline{\sigma_j}\mathbf{I} - \mathbf{A}^T)y_j = \mathbf{c}$ are solved exactly or nearly so, up to limits associated with machine accuracy. Often direct methods such as sparse-LU factorization are used. Since the need for more detail and accuracy in the modeling stage can drive the system dimension, n, to the order of millions, the use of direct

---

This work was supported in parts by the NSF Grants DMS-0505971 and DMS-0513542 and AFOSR Grant FA9550-05-1-0449.

Christopher A. Beattie and Serkan Gugercin are with Department of Mathematics, Virginia Tech,, 460 McBryde Hall, Blacksburg, VA, 24061-0123, USA `{beattie,gugercin}@math.vt.edu`

solvers for the linear system $(\sigma \mathbf{I}_n - \mathbf{A})\mathbf{x} = \mathbf{b}$ can become infeasible and iterative methods that terminate with approximate solutions must be employed. In this section, we examine the perturbative effects of employing approximate solutions in (4) and (5) on the final reduced model.

For simplicity, we first consider one-sided Krylov-based reduction of $\mathbf{H}(s) = \mathbf{c}^T(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$. Hence, only approximate solutions to (4) are used and in (3), $\mathbf{W}_r = \mathbf{V}_r$ with $\mathbf{V}_r^T\mathbf{V}_r = \mathbf{I}_r$. The reduced model is obtained via an orthogonal projection $\mathbf{\Pi}_r = \mathbf{V}_r\mathbf{V}_r^T$ as

$$\mathbf{A}_r = \mathbf{V}_r^T\mathbf{A}\mathbf{V}_r, \quad \mathbf{b}_r = \mathbf{V}_r^T\mathbf{b}, \quad \mathbf{c}_r^T = \mathbf{c}^T\mathbf{V}_r. \tag{7}$$

Let $\hat{\mathbf{x}}_j$ be an inexact solution to the j^{th} linear system $(\sigma_j\mathbf{I} - \mathbf{A})\mathbf{x}_j = \mathbf{b}$ giving rise to a residual $\delta\mathbf{b}_j$,

$$\delta\mathbf{b}_j = (\sigma_j\mathbf{I} - \mathbf{A})\hat{\mathbf{x}}_j - \mathbf{b} \tag{8}$$

The associated solution error is

$$\delta\mathbf{x}_j := \hat{\mathbf{x}}_j - \mathbf{x}_j = (\sigma_j\mathbf{I} - \mathbf{A})^{-1}\delta\mathbf{b}_j, \tag{9}$$

and the resulting inexact Krylov matrix is

$$\widehat{\mathbf{K}} := \mathbf{K}_1 + [\delta\mathbf{x}_1, \ \cdots, \ \delta\mathbf{x}_r]. \tag{10}$$

Then, the *inexact* Krylov-based reduced model will be obtained as in (7), but in this case, using an orthogonal basis, $\widehat{\mathbf{V}}_r$ generated from $\widehat{\mathbf{K}}$:

$$\widehat{\mathbf{K}} = \widehat{\mathbf{V}}_r\widehat{\mathbf{R}} \qquad \text{with} \qquad \widehat{\mathbf{V}}_r^T\widehat{\mathbf{V}}_r = \mathbf{I}_r. \tag{11}$$

The reduced model $\widehat{\mathbf{H}}_r(s) = \widehat{\mathbf{c}}_r^T(s\mathbf{I}_r - \widehat{\mathbf{A}}_r)^{-1}\widehat{\mathbf{b}}_r$ is then obtained from

$$\widehat{\mathbf{A}}_r = \widehat{\mathbf{V}}_r^T\mathbf{A}\widehat{\mathbf{V}}_r, \quad \widehat{\mathbf{b}}_r = \widehat{\mathbf{V}}_r^T\mathbf{b}, \quad \widehat{\mathbf{c}}_r^T = \mathbf{c}^T\widehat{\mathbf{V}}_r. \tag{12}$$

A. Interpolation Error

Inexactness in the solution of the linear systems produces a computed reduced order transfer function, $\widehat{\mathbf{H}}_r(s)$ that no longer interpolates $\mathbf{H}(s)$ at the points $\{\sigma_i\}_{i=1}^r$. That is, the reduced order system response will no longer exactly match the true system response at $\{\sigma_i\}_{i=1}^r$. We may associate the response error at the interpolation points with residuals produced by inexact solves.

Lemma 3.1: Let $\widehat{\mathbf{V}}_r$ be an orthonormal basis for an inexact Krylov basis as described in (10) and (11) which then yields a reduced order model as in (12). The response error at σ_j is $\widehat{\mathbf{H}}_r(\sigma_j) - \mathbf{H}(\sigma_j) = \mathbf{c}^T(\sigma_j\mathbf{I}_n - \mathbf{A})^{-1}\Delta\mathbf{b}_j$, where for $j = 1, \ldots, r$, $\Delta\mathbf{b}_j = (\mathbf{I}_n - \mathbf{P}_r)\,\delta\mathbf{b}_j$ and

$$\mathbf{P}_r = (\sigma_j\mathbf{I}_n - \mathbf{A})\widehat{\mathbf{V}}_r(\sigma_j\mathbf{I}_r - \widehat{\mathbf{A}}_r)^{-1}\widehat{\mathbf{V}}_r^T \tag{13}$$

is a skew projector onto $Ran(\sigma_j\mathbf{I}_n - \mathbf{A})\widehat{\mathbf{V}}_r$ along $Ran(\widehat{\mathbf{V}}_r)^\perp$.

Proof: From (8), $\hat{\mathbf{x}}_j = (\sigma_j\mathbf{I}_r - \mathbf{A})^{-1}(\mathbf{b} + \delta\mathbf{b}_j)$ and since $\widehat{\mathbf{V}}_r\widehat{\mathbf{V}}_r^T\hat{\mathbf{x}}_j = \hat{\mathbf{x}}_j$, also $\hat{\mathbf{x}}_j = \widehat{\mathbf{V}}_r(\sigma_j\mathbf{I}_r - \widehat{\mathbf{A}}_r)^{-1}\widehat{\mathbf{V}}_r^T(\mathbf{b} + \delta\mathbf{b}_j)$. Premultiplying by $\mathbf{c}^T$ and rearranging gives the first conclusion. That $\mathbf{P}_r^2 = \mathbf{P}_r$ can be verified directly. Thus, $\mathbf{P}_r$ is a projector with $Ran(\mathbf{P}_r) = Ran(\sigma_j\mathbf{I}_n - \mathbf{A})\widehat{\mathbf{V}}_r$ and $Ker(\mathbf{P}_r) = Ker(\widehat{\mathbf{V}}_r^T) = Ran(\widehat{\mathbf{V}}_r)^\perp$. ■

Remark 1: The interpolation/response error at σ_j may be expressed as

$$\|\delta\mathbf{b}_j\| \cdot \mathbf{c}^T(\sigma_j\mathbf{I}_n - \mathbf{A})^{-1}(\mathbf{I}_n - \mathbf{P}_r)\,\frac{\delta\mathbf{b}_j}{\|\delta\mathbf{b}_j\|},$$

allowing one to isolate the effect of the residual norm, $\|\delta\mathbf{b}_j\|$. The expression $(\sigma_j\mathbf{I}_n - \mathbf{A})^{-1}(\mathbf{I}_n - \mathbf{P}_r) = \left[(\sigma_j\mathbf{I}_n - \mathbf{A})^{-1} - \widehat{\mathbf{V}}_r(\sigma_j\mathbf{I}_r - \mathbf{A}_r)^{-1}\widehat{\mathbf{V}}_r^T\right]$ has size related to how well $\widehat{\mathbf{V}}_r(\sigma_j\mathbf{I}_r - \mathbf{A}_r)^{-1}\widehat{\mathbf{V}}_r^T$ approximates $(\sigma_j\mathbf{I}_n - \mathbf{A})^{-1}$, which in turn is associated to the quality of the Ritz approximation $\widehat{\mathbf{V}}_r\mathbf{A}_r\widehat{\mathbf{V}}_r^T$ to $\mathbf{A}$. If $\widehat{\mathbf{V}}_r$ generates a good Ritz approximation to $\mathbf{A}$ then the interpolation error may be small (independent of $\mathbf{c}$) even when the residual norms are not. This depends to a great extent on the selection of interpolation points σ_j and leads to the observation that for a good selection of interpolation points, Krylov-based model reduction can be expected to be robust with respect to the magnitude of residual norms. On the other hand, if interpolation points are poorly selected, even small magnitude residuals can produce a magnified effect through the model reduction process. Numerical examples in Section V illustrate these observations.

B. Backward error

The interpolation error resulting from inexact solves can be viewed as producing perturbations of the original system. At each interpolation point, σ_j, $\widehat{\mathbf{H}}_r(s)$ will interpolate a nearby system having the same state matrix $\mathbf{A}$, the same state-to-output matrix $\mathbf{c}^T$, but a perturbed input-to-state matrix $\mathbf{b} + \delta\mathbf{b}_j$.

Lemma 3.2: Let $\widehat{\mathbf{V}}_r$ be an orthonormal basis for an inexact Krylov basis as described in (10) and (11) which yields a reduced order model as in (12). Then, for each $j = 1, \ldots, r$

$$\widehat{\mathbf{H}}_r(\sigma_j) = \mathbf{H}^{[j]}(\sigma_j) \tag{14}$$

where $\mathbf{H}^{[j]}(s) = \mathbf{c}^T(s\mathbf{I}_n - \mathbf{A})^{-1}(\mathbf{b} + \Delta\mathbf{b}_j)$ is the transfer function of a perturbed system, $\Delta\mathbf{b}_j$ and $\mathbf{P}_r$ are defined as in Lemma 3.1.

Remark 2: Just as for the interpolation error, the backward error $\Delta\mathbf{b}_j$ will be small either if the residual norms are small or if the resulting Ritz approximation of $\mathbf{A}$ is good.

C. The two-sided case

For two-sided projections as in (3), similar but stronger estimates can be made. Suppose that the dynamical system $\mathbf{H}(s) = \mathbf{b}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$ is given and r interpolation points σ_i, $i = 1, \ldots, r$ for Krylov-based model reduction. Exact Krylov reduction is applied by constructing biorthogonal bases for the two matrix Krylov subspace $\mathbf{K}_1$ and $\mathbf{K}_2$ defined in (4) and (5), respectively.

For the inexact Krylov setting, let (8)-(10) hold as for the one sided case with $\sigma_1, \ldots, \sigma_r$ and define

$$\widehat{\mathbf{K}}_1 := \left[(\sigma_1\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + \delta\mathbf{x}_1, \ \cdots, \ (\sigma_r\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} + \delta\mathbf{x}_r\right]. \tag{15}$$

Similarly, let $\hat{\mathbf{y}}_j$ be an inexact solution for the j^{th} linear system $(\overline{\sigma_j}\mathbf{I} - \mathbf{A}^T)y_j = \mathbf{c}$ with an associated residual $\delta\mathbf{c}_j$:

$$(\overline{\sigma_j}\mathbf{I} - \mathbf{A}^T)\hat{y}_j - \mathbf{c} = \delta\mathbf{c}_j^T \tag{16}$$

Analogous to the one-sided case, define

$$\delta\mathbf{y}_j := \hat{\mathbf{y}}_j - \mathbf{y}_j = (\overline{\sigma_j}\mathbf{I} - \mathbf{A}^T)^{-1}\delta\mathbf{c}_j, \tag{17}$$

and a second inexact Krylov matrix $\widehat{\mathbf{K}}_2$ as

$$\begin{aligned}\widehat{\mathbf{K}}_2 \;\; := \;\; & \big[(\overline{\sigma_1}\mathbf{I} - \mathbf{A}^T)^{-1}\mathbf{c} + \delta\mathbf{y}_1, \\ & \qquad \cdots, \; (\overline{\sigma_r}\mathbf{I} - \mathbf{A}^T)^{-1}\mathbf{c} + \delta\mathbf{y}_r\big]\,. \end{aligned} \tag{18}$$

A two-sided version of Lemma 3.1 is straightforward to show:

Lemma 3.3: Let $\widehat{\mathbf{V}}_r$ and $\widehat{\mathbf{W}}_r$ be biorthogonal bases for $\widehat{\mathbf{K}}_1$ and $\widehat{\mathbf{K}}_2$, respectively, defined as in (15) and (18) with the property $\widehat{\mathbf{V}}_r^T\widehat{\mathbf{W}}_r = \mathbf{I}_r$. Define the computed reduced order system by $\widehat{\mathbf{H}}_r(s) = \mathbf{c}_r^T(s\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{b}$ with $\mathbf{A}_r = \mathbf{W}_r^T\mathbf{A}\mathbf{V}_r$, $\mathbf{b}_r = \mathbf{W}_r^T\mathbf{b}$, and $\mathbf{c}_r^T = \mathbf{c}^T\mathbf{V}_r$.

The response error at σ_j is $\widehat{\mathbf{H}}_r(\sigma_j) - \mathbf{H}(\sigma_j) = \Delta\mathbf{c}_j^T(\sigma_j\mathbf{I}_n - \mathbf{A})^{-1}\Delta\mathbf{b}_j$, where for $j = 1, \ldots, r$, $\Delta\mathbf{b}_j = (\mathbf{I}_n - \mathbf{P}_r)\,\delta\mathbf{b}_j$ and $\Delta\mathbf{c}_j = (\mathbf{I}_n - \mathbf{Q}_r)\,\delta\mathbf{c}_j$.

$$\mathbf{P}_r = (\sigma_j\mathbf{I}_n - \mathbf{A})\widehat{\mathbf{V}}_r(\sigma_j\mathbf{I}_r - \widehat{\mathbf{A}}_r)^{-1}\widehat{\mathbf{W}}_r^T \tag{19}$$

is a skew projector onto $Ran(\sigma_j\mathbf{I}_n - \mathbf{A})\widehat{\mathbf{V}}_r$ along $Ran(\widehat{\mathbf{W}}_r)^{\perp}$.

$$\mathbf{Q}_r = \widehat{\mathbf{V}}_r(\sigma_j\mathbf{I}_r - \widehat{\mathbf{A}}_r)^{-1}\widehat{\mathbf{W}}_r^T(\sigma_j\mathbf{I}_n - \mathbf{A}) \tag{20}$$

is a skew projector onto $Ran(\widehat{\mathbf{V}}_r)$ along $Ran((\overline{\sigma_j}\mathbf{I}_n - \mathbf{A}^T)\widehat{\mathbf{W}}_r)^{\perp}$.

Remark 3: The interpolation/response error at σ_j may be rewritten as

$$\|\delta\mathbf{b}_j\| \cdot \|\delta\mathbf{c}_j\| \cdot \frac{\delta\mathbf{c}_j^T}{\|\delta\mathbf{c}_j\|}\mathbf{M}_j(\sigma_j\mathbf{I}_n - \mathbf{A})\mathbf{M}_j\frac{\delta\mathbf{b}_j}{\|\delta\mathbf{b}_j\|},$$

where $\mathbf{M}_j = \left[(\sigma_j\mathbf{I}_n - \mathbf{A})^{-1} - \widehat{\mathbf{V}}_r(\sigma_j\mathbf{I}_r - \mathbf{A}_r)^{-1}\widehat{\mathbf{W}}_r^T\right]$. As before, the size of the interpolation error is related not only to the magnitude of the residual norms, $\|\delta\mathbf{b}_j\|$ and $\|\delta\mathbf{c}_j\|$, but also to how well $\widehat{\mathbf{V}}_r(\sigma_j\mathbf{I}_r - \mathbf{A}_r)^{-1}\widehat{\mathbf{W}}_r^T$ approximates $(\sigma_j\mathbf{I}_n - \mathbf{A})^{-1}$, which in turn is associated to the quality of the Galerkin approximation $\widehat{\mathbf{V}}_r\mathbf{A}_r\widehat{\mathbf{W}}_r^T$ to $\mathbf{A}$. The quadratic character of the error is an added feature in this case.

Remark 4: We have assumed that the same interpolation points were used to generate both left and right modeling spaces. In general, one may chose left and right interpolation points independently, however the cost of generating the reduced order model is likely to double and one loses the quadratic character of the interpolation bound in Lemma 3.3.

Remark 5: One may derive backward error expressions similar to what was found for one-sided projections.

IV. OPTIMAL APPROXIMATION BY KRYLOV PROJECTION

The lack of criteria for selection of interpolation points and the lack of a guarantee on global $\mathcal{H}_2$ and $\mathcal{H}_\infty$ performance of the resulting reduced model have been the main prior disadvantages of Krylov-based model reduction. Recently, however, Gugercin *et al.* [13] have shown that an optimal shift selection strategy exists for the optimal $\mathcal{H}_2$ approximation problem, and proposed an Iterative Rational Krylov Algorithm (**IRKA**) for model reduction that exploits it. We briefly review this method and then discuss how inexact solves can be effectively employed in that setting.

A. Iterative Rational Krylov Algorithm for Optimal $\mathcal{H}_2$ Approximation

Given an n^{th} order dynamical system $\mathbf{H}(s) = \mathbf{c}^T(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$ as in (1), the goal of optimal $\mathcal{H}_2$ approximation is to find a stable r^{th} order reduced system $\mathbf{H}_r(s) = \mathbf{c}_r^T(s\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{b}_r$ with $r < n$, such that $\mathbf{H}_r(s)$ minimizes the $\mathcal{H}_2$ error, i.e.

$$\mathbf{H}_r(s) = \arg \min_{\deg(\hat{\mathbf{H}})=r} \left\| \; \mathbf{H}(s) - \hat{\mathbf{H}}(s) \; \right\|_{\mathcal{H}_2}. \tag{21}$$

where

$$\|\mathbf{H}\|_{\mathcal{H}_2} := \left(\int_{-\infty}^{+\infty} \mid \mathbf{H}(\jmath w) \mid^2 dw\right)^{1/2}.$$

Many researchers have worked on this problem; see [27], [25], [6], [18], [14], [26] and references therein. Since obtaining a global minimum is a hard task, the general approach is to find a reduced-order model that satisfies first-order conditions for (21). The main drawback of such methods is that they require solving large-scale Lyapunov equations (possibly many of them) and dense matrix operations such as inversion. These approaches rapidly become intractable as the dimension increases in large-scale settings. Based on a result of Meier and Luenberger [18], Gugercin *et al.* [13] proposed a Krylov-based $\mathcal{H}_2$ approximation method that does not require the solution of any Lyapunov equations:

Theorem 1: Let $\mathbf{H}_r(s)$ solve the optimal $\mathcal{H}_2$ problem and let $\hat{\lambda}_i$ denote the eigenvalues of $\mathbf{A}_r$, i.e. $\hat{\lambda}_i$ are the Ritz values. Assume (for simplicity) that $\mathbf{A}_r$ is nonderogatory. Then, the first-order necessary conditions for $\mathcal{H}_2$ optimality are

$$\left.\frac{d^k}{ds^k}\mathbf{H}(s)\right|_{s=-\hat{\lambda}_i} = \left.\frac{d^k}{ds^k}\mathbf{H}_r(s)\right|_{s=-\hat{\lambda}_i}, \quad k = 0, 1. \tag{22}$$

Theorem 1 states that $\mathbf{G}_r(s)$ has to *interpolate* $\mathbf{H}(s)$ and its first derivative at the mirror images of the Ritz values. Thus, first-order conditions are given in the framework of interpolation. The method of Gugercin *et al.* [13] produces a reduced order model $\mathbf{H}_r(s)$ satisfying the interpolation-based first-order necessary conditions of (22) and exploiting the connection between the Krylov-based reduction and interpolation. But since the optimal interpolation points of (22) depend on the final reduced model and are not known *a priori*, [13] uses rational Krylov steps to iteratively correct the reduced-order model $\mathbf{H}_r(s)$ so that the next (corrected) reduced-order model interpolates the full-order model at mirrored Ritz values $-\lambda_i(\mathbf{A}_r)$ from the previous reduced-order model. This continues until Ritz values from consecutive reduced-order models stagnate. Below, we give a sketch of this algorithm:

Algorithm 1: [13] **An Iterative Rational Krylov Algorithm (IRKA):**

```
1) Make an initial shift selection $\sigma_i$ for $i = 1, \ldots, r$
2) $\mathbf{W} = \left[\, (\overline{\sigma_1}\mathbf{I} - \mathbf{A}^T)^{-1}\mathbf{c}, \cdots, (\overline{\sigma_r}\mathbf{I} - \mathbf{A}^T)^{-1}\mathbf{c} \,\right]$.
3) $\mathbf{V} = \left[\, (\sigma_1\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}, \cdots, (\sigma_r\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} \,\right]$
4) $\mathbf{W} = \mathbf{W}(\mathbf{W}^T\mathbf{V})^{-T}$    (to make $\mathbf{W}^T\mathbf{V} = \mathbf{I}_r$)
5) while (not converged)
    a) $\mathbf{A}_r = \mathbf{W}^T\mathbf{A}\mathbf{V}$,
    b) $\sigma_i \longleftarrow -\lambda_i(\mathbf{A}_r)$ for $i = 1, \ldots, r$
    c) $\mathbf{W} = \left[(\overline{\sigma_1}\mathbf{I} - \mathbf{A})^{-T}\mathbf{c}, \cdots, (\overline{\sigma_r}\mathbf{I} - \mathbf{A})^{-T}\mathbf{c}^T\right]$
    d) $\mathbf{V} = \left[\, (\sigma_1\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}, \cdots, (\sigma_r\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} \,\right]$
    e) $\mathbf{W} = \mathbf{W}(\mathbf{W}^T\mathbf{V})^{-T}$    (to make $\mathbf{W}^T\mathbf{V} = \mathbf{I}_r$)
6) $\mathbf{A}_r = \mathbf{W}^T\mathbf{A}\mathbf{V}$, $\mathbf{b}_r = \mathbf{W}^T\mathbf{b}$, $\mathbf{c}_r^T = \mathbf{c}^T\mathbf{V}$
```

Upon convergence, **IRKA** produces a reduced-order model $\mathbf{H}_r(s)$ that satisfies the desired interpolation conditions (22).

*B. Inexact-Iterative Rational Krylov Algorithm (***I-IRKA***)*

The main cost in a step of **IRKA** is dominated by solving $2r$ large linear systems. If **IRKA** converges after k steps then one will need to solve a total of $2rk$ linear systems. In cases where the system dimension n is on the order of millions, iterative linear system solvers become necessary and inexact solves must then be incorporated into **IRKA**. We will refer to the resulting model reduction method as the Inexact-Iterative Rational Krylov Algorithm (**I-IRKA**) for optimal $\mathcal{H}_2$ approximation.

Algorithm 2: **An Inexact Iterative Rational Krylov Algorithm (I-IRKA):**

```
1) Make an initial shift selection $\sigma_i$ for $i = 1, \ldots, r$
2) for $i = 1, \ldots, r$
    a) $\mathbf{x}_0 = \mathbf{0}$
    b) $\mathbf{x}_i = \mathbf{f}\left(\mathbf{A}, \mathbf{b}, \sigma_i, \mathbf{x}_0, \epsilon\right)$
    c) $y_i = \mathbf{f}\left(\mathbf{A}^T, \mathbf{c}^T, \sigma_i, \mathbf{x}_0, \epsilon\right)$
3) $\mathbf{W} = [\ y_1,\ y_2,\ \ldots,\ y_r\ ]$
4) $\mathbf{V} = [\ \mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_r\ ]$.
5) $\mathbf{W} = \mathbf{W}(\mathbf{W}^T\mathbf{V})^{-T}$    (to make $\mathbf{W}^T\mathbf{V} = \mathbf{I}_r$)
6) while (not converged)
    a) $\mathbf{A}_r = \mathbf{W}^T\mathbf{A}\mathbf{V}$,
    b) $\sigma_i \longleftarrow -\lambda_i(\mathbf{A}_r)$ for $i = 1, \ldots, r$
    c) for $i = 1, \ldots, r$
        i) $\mathbf{x}_i = \mathbf{f}\left(\mathbf{A}, \mathbf{b}, \sigma_i, \mathbf{x}_i, \epsilon\right)$
        ii) $y_i = \mathbf{f}\left(\mathbf{A}^T, \mathbf{c}^T, \sigma_i, y_i, \epsilon\right)$
    d) $\mathbf{W} = [\ y_1,\ y_2,\ \ldots,\ y_r\ ]$
    e) $\mathbf{V} = [\ \mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_r\ ]$.
    f) $\mathbf{W} = \mathbf{W}(\mathbf{W}^T\mathbf{V})^{-T}$    (to make $\mathbf{W}^T\mathbf{V} = \mathbf{I}_r$)
7) $\mathbf{A}_r = \mathbf{W}^T\mathbf{A}\mathbf{V}$, $\mathbf{b}_r = \mathbf{W}^T\mathbf{b}$, $\mathbf{c}_r^T = \mathbf{c}^T\mathbf{V}$
```

As discussed in [13], in most cases **IRKA** shows rapid convergence behavior; that is, the interpolation points at the k^{th} step $\{\sigma_j^{(k)}\}$ stagnate rapidly with respect to k. This observation leads to the following important observation: As k increases, the solution $\mathbf{x}_j^{(k)}$ of the linear system $(\sigma_j^{(\mathrm{k})}\mathbf{I} - \mathbf{A})\mathbf{x}^{(k)} = \mathbf{b}$ from the k^{th} step become very close to the solution $\mathbf{x}_j^{(k+1)}$ of the linear system $(\sigma_j^{(\mathrm{k+1})}\mathbf{I} - \mathbf{A})\mathbf{x}_j^{(k+1)} = \mathbf{b}$ at the $(k+1)^{\text{st}}$ step. This suggests that in **I-IRKA**, one use the solution $\mathbf{x}_j^{(k)}$ as an initial guess in solving $(\sigma_j^{(\mathrm{k+1})}\mathbf{I} - \mathbf{A})\mathbf{x}_j^{(k+1)} = \mathbf{b}$ in the next step. Due to observed rapid convergence [13] of the exact-Krylov based algorithm **IRKA**, one may anticipate a speed-up of the iterative solves in passing from step to step. In the outline of the resulting algorithm, the functon $\mathbf{f}(\mathbf{A}, \mathbf{b}, \sigma, \mathbf{x}_0, \epsilon)$ denotes an iterative solve for the linear system $(\sigma\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}$ with an initial guess $\mathbf{x}_0$ and a relative residual termination tolerance ϵ: $\frac{\|\delta\mathbf{b}_j\|}{\|\mathbf{b}\|} \leq \epsilon$ and $\frac{\|\delta\mathbf{c}_j\|}{\|\mathbf{c}\|} \leq \epsilon$.

C. Effect of Inexact Solves in the **I-IRKA** *Setting*

A crucial question in the **I-IRKA** setting is how much the resulting optimal model obtained via **I-IRKA** (denoted by $\mathbf{H}_{\mathbf{I-IRKA}}$) will deviate from one obtained via **IRKA** (denoted by $\mathbf{H}_{\mathbf{IRKA}}$). As we discussed in Section III, for a good selection of interpolation points, Krylov-based reduction is robust with respect to perturbations due to inexact solves. Hence, if one feeds the resulting optimal interpolation points from **IRKA** into **I-IRKA**, we expect that $\mathbf{H}_{\mathbf{I-IRKA}}$ will be close to $\mathbf{H}_{\mathbf{IRKA}}$. However, the final interpolation points are not known initially and **I-IRKA** must inevitably be started with a nonoptimal initial shift selection. At the early stages of the iteration, if the initial shift selection is poor, perturbations due to inexact solves might be magnified by this poor shift selection as discussed in Section III. Therefore, one can prevent this from happening by using a small termination threshold ϵ in the early steps of **I-IRKA**, and then gradually increase ϵ as the iteration starts to converge, i.e. as the $\{\sigma_i\}$ starts to converge to an optimal shift selection. Notably, in all of our numerical experiments using random initial interpolation points, **I-IRKA** performed very efficiently and yielded a reduced model $\mathbf{H}_{\mathbf{I-IRKA}}$ very close to $\mathbf{H}_{\mathbf{IRKA}}$ in both the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ sense. Several effective initialization strategies were proposed for **IRKA** and studied in [13].

Remark 6: We have also studied in some detail consequences related to the iterative solution of linear systems in the context of **I-IRKA**, such as preconditioning, effective restart strategies, and system response-based termination criteria. We omit them in this document because of space constraints. They will be included in the full paper. For example,

1) Since each linear system is in the form of $(\sigma\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}$, the shift invariance of Krylov subspaces provides advantages to Krylov solvers such as GMRES, MINRES, and SYMMLQ since one need only generate a single Krylov subspace and use this single subspace, for all linear systems. Restarts and preconditioning complicate the picture, however, since shift invariance is lost.
2) In the model reduction setting, we are not interested in the particular solution vectors, $\hat{\mathbf{x}}_j$, so much as the subspace that $\{\hat{\mathbf{x}}_j\}_{i=1}^{r}$ span. This suggests that one could

afford a less accurate solution for $\hat{\mathbf{x}}_j$ if this direction is already included in the subspace $\mathrm{span}\{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{j-1}\}$.

3) One can exploit the fact $\{\sigma_j\}$ is rapidly convergent in order to construct effective preconditioners for $\sigma_j\mathbf{I}-\mathbf{A}$. Since $\sigma_j^{(k+1)}$ converges to $\sigma_j^{(k+1)}$, one may choose to re-use the preconditioner from a previous step.

These considerations will be discussed and several numerical experiments will be presented in detail in the full paper.

V. EXAMPLES: A SEMI-DISCRETIZED HEAT TRANSFER PROBLEM FOR OPTIMAL COOLING OF STEEL PROFILES

This problem arises during a cooling process in a rolling mill and is modeled as boundary control of a two dimensional heat equation. A finite element discretization results in a descriptor system of the form

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

where $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times 7}$, $\mathbf{C} \in \mathbb{R}^{6\times n}$ with n depending on the maximum mesh width $w_{\max}$. We consider the full-order SISO system relating the sixth input of this system to the second output. Note that in this case $\mathbf{E} \neq \mathbf{I}_n$ is a positive definite matrix, and the previous discussion is still applicable replacing $\sigma\mathbf{I}$ by $\sigma\mathbf{E}$ and making obvious modifications. For details regarding the modelling, discretization, optimal control design, and model reduction, see [23], [4], [5]. We consider two different cases. In one case, $w_{\max} = 1.3820 \times 10^{-2}$ resulting in $n = 20,209$ and in the other case $w_{\max} = 6.9100 \times 10^{-3}$ resulting in $n = 79,841$.

A. *Case 1:* $w_{\max} = 1.3820 \times 10^{-2}$ *and* $n = 20,209$

We compare exact Krylov reduction with inexact Krylov reduction by reducing the order to $r = 6$ for both poor and good shift selections. As can be seen from the Bode plot of original model $\mathbf{H}(s)$ in Figure 1, $\sigma_i = \mathrm{logspace}(-8, -4, 6)$ is a bad selection of interpolation points since this selection omits a dominant part of the frequency range. The resulting sixth order models obtained by exact Krylov and inexact Krylov reduction are denoted by $\mathbf{H}_1(s)$ and $\mathbf{H}_2(s)$ respectively. The inexact case uses GMRES with relative residual termination tolerance of 1×10^{-5}. Amplitude Bode plots of $\mathbf{H}(s)$, $\mathbf{H}_1(s)$ and $\mathbf{H}_2(s)$ are shown in Figure 1 below. Figure 1 reveals two facts: First of all, both $\mathbf{H}_1(s)$ and $\mathbf{H}_2(s)$ are poor approximations due to poor selection of σ_1. But more importantly within the context of this paper, $\mathbf{H}_2(s)$ deviates from significantly from $\mathbf{H}_1(s)$ as well. This is exactly what we expected: Poor selection of σ_i magnifies the error coming from inexact solves and inexact Krylov reduction deviates from exact Krylov reduction. The exact $\mathcal{H}_\infty$ error norms are given below:

$$\|\mathbf{H}-\mathbf{H}_1\|_{\mathcal{H}_\infty} = 2.18\times 10^{-3}, \|\mathbf{H}-\mathbf{H}_2\|_{\mathcal{H}_\infty} = 2.42\times 10^{-3},$$
$$\|\mathbf{H}_1-\mathbf{H}_2\|_{\mathcal{H}_\infty} = 9.18\times 10^{-4},$$

Relative and absolute errors in the computed first and second moments due to inexact solves are depicted in 2. Even with a relative error tolerance of 1×10^{-5}, the relative error in some of the computed first moments are on the order of 10^{-2}. This is again due to the fact that the error $\left\|(\sigma_j\mathbf{I}_n - \mathbf{A})^{-1} - \mathbf{V}(\sigma_j\mathbf{I}_r - \mathbf{A}_r)^{-1}\mathbf{V}^{\mathrm{T}}\right\|$ is large due to poor shift selection and this magnifies the residual error $\|\delta\mathbf{b}_j\|$.

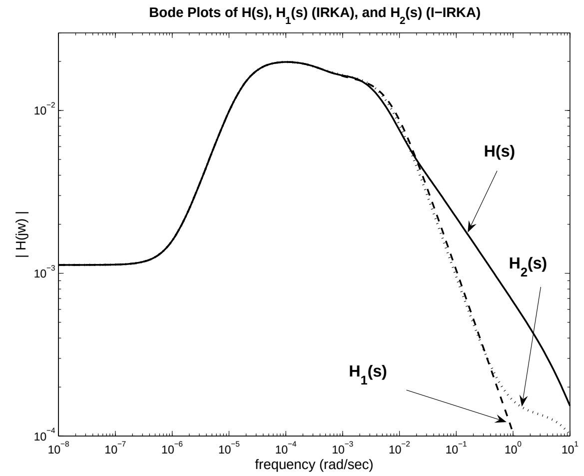

Fig. 1. Bode Plots for $\mathbf{H}(s)$, $\mathbf{H}_1(s)$, and $\mathbf{H}_2(s)$

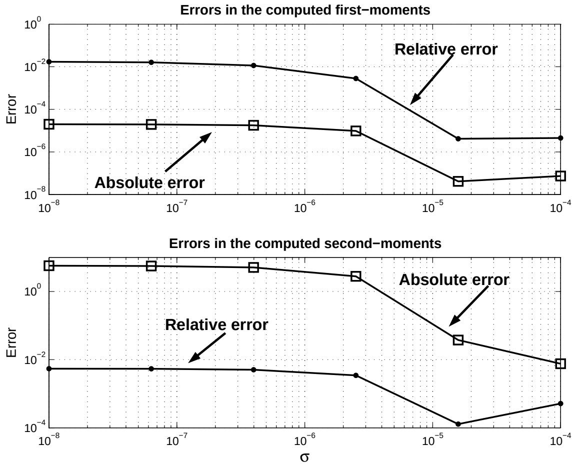

Fig. 2. Computed first and second moments

Next, we use six interpolation points obtained from running **IRKA** on this model [13]; we use an optimal shift selection. We perform the same operations as above and reduce the order to $r = 6$. In this case, we use a higher error tolerance value of 1×10^{-4} for GMRES. The resulting Bode plots, and relative and absolute errors in the computed moments are shown in Figures 3 and 4, respectively. Figure 3 reveals that not only that $\mathbf{H}_1(s)$ and $\mathbf{H}_2(s)$ are good approximation to $\mathbf{H}(s)$, but also they do not differ from each other. The reason is that since the shift selection is good (optimal), the model reduction error does **not** magnify the perturbation due to inexact solves. This results in $\mathbf{H}_2(s)$ being very close $\mathbf{H}_1(s)$. The exact $\mathcal{H}_\infty$ error norms in this case are

$$\begin{aligned}\|\mathbf{H}-\mathbf{H}_1\|_{\mathcal{H}_\infty} = \|\mathbf{H}-\mathbf{H}_2\|_{\mathcal{H}_\infty} &= 1.56\times 10^{-3},\\ \|\mathbf{H}_1-\mathbf{H}_2\|_{\mathcal{H}_\infty} &= 1.82\times 10^{-5}.\end{aligned}$$

We note that these smaller error numbers are achieved even with a relaxed error tolerance in GMRES. Figure 4 illustrate the same result for the moments as well. Both relative and absolute error are much smaller in this case. These examples support our expectations and discussion that Krylov-based reduction is robust with respect to inexact solution residuals when interpolation points are good.

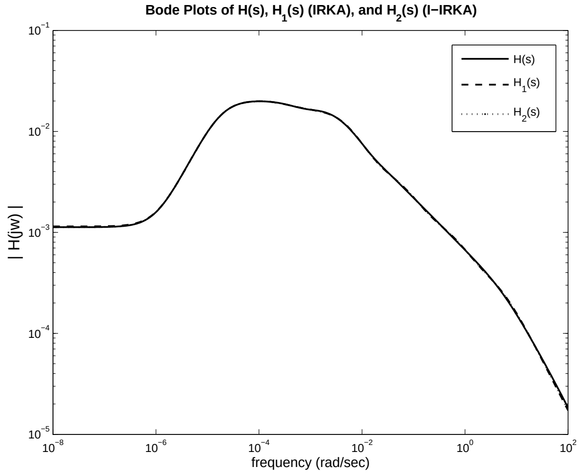

Fig. 3. Bode Plots for $\mathbf{H}(s)$, $\mathbf{H}_1(s)$, and $\mathbf{H}_2(s)$

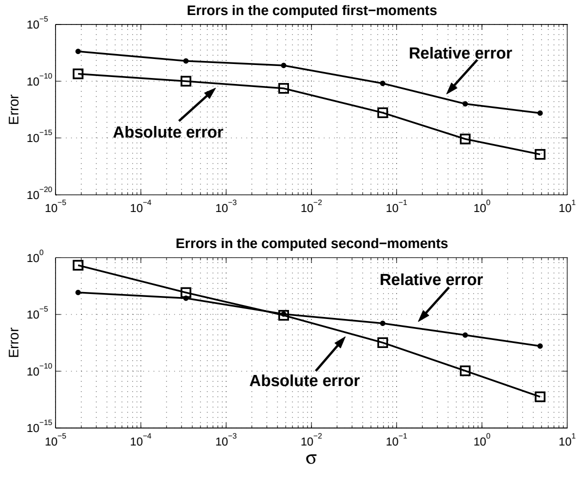

Fig. 4. Computed first and second moments

B. Case 2: $w_{\max} = 6.9100\times 10^{-3}$ and $n = 79,841$

We now illustrate the discussion of Section 2, and compare **IRKA** with **I-IRKA** using the same model with $n = 79,841$.

We reduce the order to $r = 6$ using **IRKA** and (**I-IRKA**) where each linear system is solved using GMRES with a relative residual termination threshold of 6×10^{-5}. Amplitude Bode plots of $\mathbf{H}(s)$, $\mathbf{H}_1(s)$ (due to **IRKA**), and $\mathbf{H}_2(s)$ (due to **I-IRKA**) are shown in Figure 5. The first observation is that both $\mathbf{H}_1(s)$ and $\mathbf{H}_2(s)$ match $\mathbf{H}(s)$ very well, indicating once more the effectiveness of model reduction with interpolation at the mirror images of the Ritz values. *The second observation is that $\mathbf{H}_1(s)$ almost perfectly replicates $\mathbf{H}_2(s)$; inexact solves do not degrade the final optimal reduced model even for a random initial shift selection.* The $\mathcal{H}_\infty$ error between the two reduced models is $\|\mathbf{H}_1(s)-\mathbf{H}_2(s)\|_{\mathcal{H}_\infty} = 3.01\times 10^{-5}$.

To speed-up **I-IRKA**, we used solution vectors from one step as an initial guess for the linear system in the next step as proposed in Section IV-B. The lower-plot in Figure 5 shows the total number of GMRES steps required to compute $\mathbf{V}_r$ and $\mathbf{W}_r$ at each step of **I-IRKA**, i.e., the total number of GMRES steps needed to solve the $2r$ linear systems. This figure clearly illustrates that the approach works very effectively at reducing the computational cost considerably at each iteration.

The important point of this example is the following: Until now, no reliable shift selection strategy was known for the rational Krylov algorithm; shift selection was generally *ad hoc*. Here, for a system of order $n = 79,841$, our algorithm *effectively and efficiently searches for an optimal shift selection and, consequently yields, at least, a locally optimal reduced order model with no tuning or user intervention.*

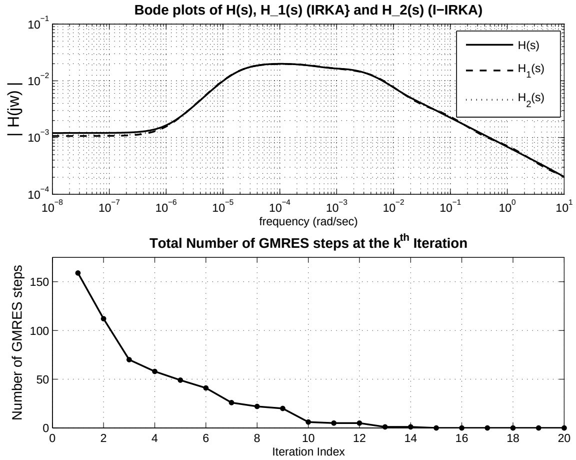

Fig. 5. **IRKA** vs **I-IRKA**

REFERENCES

[1] A.C. Antoulas, *Lectures on the approximation of linear dynamical systems*, Advances in Design and Control, SIAM, Philadelphia, 2004.

[2] 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.

[3] C.A. Beattie and S. Gugercin, *Projection methods for model reduction of dynamical systems*, SIAM Conference on Computational Science and Eng, Orlando, USA, February 2005.

[4] P. Benner, *Solving Large-Scale Control Problems*, IEEE Control Systems Magazine, Vol. 24, No. 1, pp. 44-59, 2004.

[5] P. Benner and J. Saak, *Efficient numerical solution of the LQR-problem for the heat equation*, submitted to Proc. Appl. Math. Mech., 2004.

[6] A.E. Bryson and A. Carrier, *Second-order algorithm for optimal model order reduction*, J. Guidance Contr. Dynam., pp-887-892, 1990

[7] K. Gallivan, E. Grimme, and P. Van Dooren, *A rational Lanczos algorithm for model reduction*, Numerical Algorithms, 2(1-2):33-63, April 1996.

[8] K. Gallivan, A. Vandendorpe, and P. Van Dooren, *Model reduction via truncation: An interpolation point of view*, Linear Algebra and Its Applications, Vol. 375, pp. 115-134, 2003.

[9] K. Glover. All Optimal Hankel-norm Approximations of Linear Mutilvariable Systems and their L^∞-error Bounds. *Int. J. Control*, **39**: 1115-1193, 1984.

[10] E.J. Grimme, *Krylov Projection Methods for Model Reduction*, Ph.D. Thesis, ECE Dept., U. of Illinois, Urbana-Champaign, 1997.

[11] 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).

[12] S. Gugercin, *Projection methods for model reduction of large-scale dynamical systems*, Ph.D. Dissertation, ECE Dept., Rice University, December 2002.

[13] S. Gugercin, A.C. Antoulas and C.A. Beattie, *First-order conditions and an Iterative Rational Krylov Algorithm for optimal $\mathcal{H}_2$ model reduction*, Technical Report, Department of Mathematics, Virginia Tech., September 2004.

[14] 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*, IEE. Trans. Automat. Contr., Vol. 30, No. 12, pp. 1201-1211, 1985.

[15] J.G. Korvink and E.B. Rudyni, *Oberwolfach Benchmark Collection*, In P. Benner, G. Golub, V. Mehrmann, and D. Sorensen, editors, Dimension Reduction of Large-Scale Systems, Springer-Verlag, Lecture Notes in Computational Science and Engineering, Vol. 45 (ISBN 3-540-24545-6) Berlin/Heidelberg, Germany, 2005.

[16] 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.

[17] Y. Liu and B. D. O. Anderson, Singular Perturbation Approximation of Balanced Systems, *Int. J. Control*, **50**: 1379-1405, 1989.

[18] L. Meier and D.G. Luenberger, *Approximation of Linear Constant Systems*, IEE. Trans. Automat. Contr., Vol. 12, pp. 585-588, 1967.

[19] B. C. Moore, *Principal Component Analysis in Linear System: Controllability, Observability and Model Reduction*, IEEE Transactions on Automatic Control, AC-26:17-32, 1981.

[20] 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.

[21] B. N. Parlett, *Reduction to tridiagonal form and minimal realizations*, SIAM J. Matrix Anal. Appl., **13**, pp: 567–593, 1992.

[22] 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.

[23] Thilo Penzl, *Algorithms for model reduction of large dynamical systems*, Technical Report SFB393/99-40, Sonderforschungsbereich 393 Numerische Simulation auf massiv parallelen Rechern, TU Chemnitz, 09107, FRG, 1999. Available from `http://www.tu-chemnitz.de/sfb393/sfb99pr.html`.

[24] A. Ruhe, *Rational Krylov algorithms for nonsymmetric eigenvalue problems II: matrix pairs*, Linear Alg. Appl., **197**:283-295, (1994).

[25] J.T. Spanos, M.H. Millman, and D.L. Mingori, *A new algorithm for L_2 optimal model reduction*, Automatics, pp. 897-909, 1992.

[26] D.A. Wilson, *Optimum solution of model reduction problem*, in Proc. Inst. Elec. Eng., pp. 1161-1165, 1970

[27] W-Y. Yan and J. Lam, *An approximate approach to $\mathcal{H}_2$ optimal model reduction*, IEEE Transactions on Automatic Control, **AC-44**, pp. 1341-1358, 1999.