Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

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

Serkan Gugercin

***Abstract*— We propose a model reduction algorithm which combines the SVD and Krylov-based techniques. It is a two-sided projection method where one side carries the SVD (Gramian) information and the other side the Krylov information. While the SVD-side depends on the observability gramian, the Krylov-side is obtained via iterative rational Krylov steps. The reduced model is asymptotically stable and matches the moments of the original system at the mirror images of the reduced system poles; hence it is the best $\mathcal{H}_2$ approximation among all reduced models having the same reduced system poles. Numerical results proves the effectiveness of the proposed approach.**

I. INTRODUCTION

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

$$\begin{array}{rcl} \mathbf{G}(s) & : & \left\{ \begin{array}{rcl} \dot{\mathbf{x}}(t) & = & \mathbf{A}\mathbf{x}(t) + \mathbf{B}\,\mathbf{u}(t) \\ \mathbf{y}(t) & = & \mathbf{C}\,\mathbf{x}(t), \end{array} \right. \\ & \Updownarrow & \\ \mathbf{G}(s) & := & \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & 0 \end{array} \right], \end{array} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^n$, and $\mathbf{C}^T \in \mathbb{R}^n$. In (1), $\mathbf{x}(t) \in \mathbb{R}^n$ is the *state*, $\mathbf{u}(t) \in \mathbb{R}$ is the *input*, and $\mathbf{y}(t) \in \mathbb{R}$ is the *output* of $\mathbf{G}(s)$. The transfer function of $\mathbf{G}(s)$ is given by $\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$. Following traditional abuse of notation, we note that both the underlying dynamical system and its transfer function are denoted by $\mathbf{G}(s)$. In the sequel, we will assume that the full-order model $\mathbf{G}(s)$ is *asymptotically stable*, i.e. $\mathrm{Real}(\lambda_i(\mathbf{A})) < 0$ for $i = 1, \ldots, n$; and is *minimal*, i.e. both reachable and observable. On the other hand, we call a dynamical system *stable*, if it has no poles in the right-half plane but has poles on the imaginary axis.

In many applications; see [26] for a recent collection of such benchmark problems, the system dimension n is quite large, while the number of inputs m and outputs p usually satisfy $m, p \ll n$, making the computations infeasible due to memory, time limitations, and numerical ill-conditioning. The goal of model reduction is, then, to produce a much smaller order system $\mathbf{G}_r(s)$ with state-space form:

$$\begin{array}{rcl} \mathbf{G}_r(s) & : & \left\{ \begin{array}{l} \dot{\mathbf{x}}_r(t) = \mathbf{A}_r\mathbf{x}_r(t) + \mathbf{B}_r\mathbf{u}(t) \\ \mathbf{y}_r(t) = \mathbf{C}_r\mathbf{x}(t), \end{array} \right. \\ & \Updownarrow & \\ \mathbf{G}_r(s) & := & \left[\begin{array}{c|c} \mathbf{A}_r & \mathbf{B}_r \\ \hline \mathbf{C}_r & 0 \end{array} \right], \end{array} \tag{2}$$

S. Gugercin is with Department of Mathematics, Virginia Tech., Blacksburg, VA, USA, gugercin@math.vt.edu.

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

In the sequel, we will construct the reduced order models $\mathbf{G}_r(s)$ through projection. In other words, we construct matrices $\mathbf{V} \in \mathbb{R}^{n\times r}$ and $\mathbf{Z} \in \mathbb{R}^{n\times r}$ such that $\mathbf{Z}^T\mathbf{V} = \mathbf{I}_r$ and the reduced order model $\mathbf{G}_r(s)$ in (2) is then obtained as

$$\mathbf{A}_r = \mathbf{Z}^T\mathbf{A}\mathbf{V}, \quad \mathbf{B}_r = \mathbf{Z}^T\mathbf{B}, \quad \text{and} \quad \mathbf{C}_r = \mathbf{C}\mathbf{V}. \tag{3}$$

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

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

(a) SVD (Gramian) based methods,

(b) Krylov (moment matching) based methods, and

(c) SVD-Krylov based methods.

The Hankel singular values, which are the singular values of the Hankel operator associated with $\mathbf{G}(s)$, are the key elements of the SVD-based model reduction algorithms. They play the same role as that of the singular values in the optimal 2-norm approximation of constant matrices. *Balanced Truncation* [32], [31], is the most common approach in this category. When applied to asymptotically stable systems, it preserves asymptotic stability and provides an *a priori* bound on the approximation error. However, for large-scale settings, *exact balanced truncation* is expensive to implement because it requires dense matrix factorizations. The resulting computational complexity is $\mathcal{O}(n^3)$ and the storage requirements are $\mathcal{O}(n^2)$. Therefore, in large-scale settings, one uses approximate low-rank versions of balanced truncation; see, for example, [21], [33], [27], [7], which reduce the cost to $\mathcal{O}(n^2)$ and the storage to $\mathcal{O}(nr)$. For more detail on the efficient implementation of balancing related model reduction in the large-scale settings, see [21], [7], [8], [37], [1], [38]. *Hankel Norm Approximation* [15], and *Balanced Singular Perturbation Approximation* [28] are two other common model reduction techniques belonging to this category.

Krylov based methods are aimed at matching the so-called *moments* of $\mathbf{G}(s)$ at certain points in the complex plane. The k^{th} moment of $\mathbf{G}(s)$ at a point $\sigma_0 \in \mathbb{C}$ is the k^{th} derivative of $\mathbf{G}(s)$ at σ_0. Hence the goal is to construct a reduced model $\mathbf{G}_r(s)$ that matches certain number of moments of $\mathbf{G}(s)$ at selected interpolation points. Under this category, we list the Arnoldi [5] and Lanczos procedures [29], and rational Krylov method [16], [34], [12],

0-7803-9568-9/05/$20.00 ©2005 IEEE

[13]. Compared to the SVD-based methods, these methods are numerically more reliable and can be implemented iteratively; the number of computations is of $\mathcal{O}(nr^2)$ and the storage requirement is of $\mathcal{O}(nr)$. Also, the asymptotic stability of the reduced model can be obtained through restarting [17], [25]. But there exists no *a priori* error bounds. However, recently in [20], [22], a global *error expression* has been developed for Krylov-based methods.

Recently much research has been done to obtain a model reduction algorithm which connects the SVD and Krylov based methods; see, for example, [1], [25], [2], [14], [21], [23]. The goal of these works is to *combine* the theoretical features of the SVD based methods such as stability, global error bounds, with the efficient numerical implementation of the Krylov-based methods. In this paper, we propose a model reduction algorithm which achieves this goal. The method is a two-sided projection method where one side reflects the Krylov part of the algorithm, and the other reflects the SVD (Gramian) part. The reduced model is asymptotically stable, solves a restricted $\mathcal{H}_2$ minimization problem and matches certain moments. Numerical results prove the effectiveness of the algorithm.

II. Some Preliminaries

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

A. System Gramians

Given a dynamical system $\mathbf{G}(s)$ as in (1), $\mathbf{P}$ and $\mathbf{Q}$, the solutions to the following continuous-time Lyapunov equations

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0, \tag{4}$$

$$\mathbf{A}^T\mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{C}^T\mathbf{C} = 0 \tag{5}$$

are called *reachability* and *observability gramians*, respectively. Under the assumption that $\mathbf{G}(s)$ is asymptotically stable and minimal, $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{n\times n}$ are unique symmetric positive definite matrices. Gramians play a crucial role in model reduction as balanced truncation amounts to finding a state space transformation so that in this new co-ordinate system, the two gramians are diagonal and equal. In this transformed basis, states that are difficult to reach are simultaneously difficult to observe. Then the reduced model is simply obtained by truncating the states which are both difficult to reach and observe.

B. Krylov-based model reduction

Given $\mathbf{G}(s)$ as in (1), the moment matching problem is to find a reduced model $\mathbf{G}_r(s)$ that interpolates $\mathbf{G}(s)$ as well as a certain number of its derivatives (called moments) at some selected points σ_k in the complex plane. In other words, the problem is to find $\mathbf{A}_r$, $\mathbf{B}_r$, and $\mathbf{C}_r$ so that

$$\frac{(-1)^j}{j!}\frac{d^j\mathbf{G}(s)}{ds^j}\bigg|_{s=\sigma_k} = \mathbf{C}(\sigma_k\mathbf{I}-\mathbf{A})^{-(j+1)}\mathbf{B} =$$

$$\mathbf{C}_r(\sigma_k\mathbf{I}_r-\mathbf{A}_r)^{-(j+1)}\mathbf{B}_r = \frac{(-1)^j}{j!}\frac{d^j\mathbf{G}_r(s)}{ds^j}\bigg|_{s=\sigma_k}$$

for $k = 1,\ldots,K$ and for $j = 1,\ldots,J$. Here K is the number of interpolation points σ_k and J is the number of moments to be matched at each σ_k. The quantity $\mathbf{C}(\sigma_k\mathbf{I}-\mathbf{A})^{-(j+1)}\mathbf{B}$ is called the j^{th} moment of $\mathbf{G}(s)$ at σ_k. If $\sigma_k = \infty$, the moments are called Markov parameters and are given by $\mathbf{C}\mathbf{A}^j\mathbf{B}$ for $j = 0,1,2,\ldots$. The problem described above is sometimes called the multi-point rational interpolation problem. In the projection framework, i.e. $\mathbf{G}_r(s)$ is obtained as in (3), the problem was first treated by Skelton *et. al.* in [10], [41], [40]. Grimme [16] showed how one can obtain the required projection in a numerically efficient way using the rational Krylov method of Ruhe [34], hence showed how to solve moment matching (multi-point rational interpolation) problem using *Krylov projection* methods in an effective way.

For a matrix $\mathbf{F} \in \mathbb{C}^{n\times n}$, a vector $\mathbf{g} \in \mathbb{C}^n$, and a point $\sigma \in \mathbb{C}$, we define the Krylov space of index j:

$$\mathcal{K}_j(\mathbf{F},\mathbf{g};\sigma) := \text{Im}([\ \mathbf{g}\ \ \mathbf{F}\mathbf{g}\ \ \mathbf{F}^2\mathbf{g}\ \ \cdots\ \ \mathbf{F}^{j-1}\mathbf{g}\ ]) \quad \text{if} \quad \sigma = \infty$$

$$\mathcal{K}_j(\mathbf{F},\mathbf{g};\sigma) := \text{Im}([\ (\sigma\mathbf{I}-\mathbf{F})^{-1}\mathbf{g}\ \cdots\ (\sigma\mathbf{I}-\mathbf{F})^{-j}\mathbf{g}\ ]) \quad \text{if} \quad \sigma \neq \infty$$

The following theorem [16] shows how to solve the multi-point rational interpolation problem by Krylov projection:

Theorem 1: [16] If

$$\text{Im}(\mathbf{V}) = \text{span}\,[\mathcal{K}_{j_1}(\mathbf{A},\mathbf{B};\sigma_1)\ \ \cdots\ \ \mathcal{K}_{j_K}(\mathbf{A},\mathbf{B};\sigma_K)]$$

$$\text{and}\quad \text{Im}(\mathbf{Z}) =$$

$$\text{Im}\left[\mathcal{K}_{j_{K+1}}(\mathbf{A}^T,\mathbf{C}^T;\sigma_{K+1})\ \ \cdots\ \ \mathcal{K}_{j_{2K}}(\mathbf{A}^T,\mathbf{C}^T;\sigma_{2K})\right]$$

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

Theorem 1 states that for Krylov-based model reduction, all one has to do is to construct matrices $\mathbf{V}$ and $\mathbf{Z}$ as above. Efficient implementation of the rational Krylov method can be found in [16]; for more details on Krylov-based model reduction, see [12], [16], [20], [3], [4].

III. An Iterative SVD-Krylov Algorithm

As mentioned above, we will construct the reduced model $\mathbf{G}_r(s)$ using an oblique projection as in (3). For the proposed algorithm, the matrix $\mathbf{Z}$ will have the following specific form:

$$\mathbf{Z} := \mathbf{Q}\mathbf{V}(\mathbf{V}^T\mathbf{Q}\mathbf{V})^{-1} \tag{6}$$

where $\mathbf{Q}$ is the observability gramian as defined in (5) and $\mathbf{V}$ spans a certain *rational Krylov subspace* as in Theorem 1.

Choice of $\mathbf{V}$ will be explained below. Clearly, $\mathbf{Z}^T\mathbf{V} = \mathbf{I}_r$, as desired. Note that since $\mathbf{Q}$ is symmetric, $(\mathbf{V}^T\mathbf{Q}\mathbf{V})^{-1} = (\mathbf{V}^T\mathbf{Q}\mathbf{V})^{-T}$. Hence $\mathbf{G}_r(s)$ is given by:

$$\mathbf{G}_r(s) = \left[\begin{array}{c|c} \mathbf{A}_r & \mathbf{B}_r \\ \hline \mathbf{C}_r & 0 \end{array}\right] = \left[\begin{array}{c|c} (\mathbf{V}^T\mathbf{Q}\mathbf{V})^{-1}\mathbf{V}^T\mathbf{Q}\mathbf{A}\mathbf{V} & (\mathbf{V}^T\mathbf{Q}\mathbf{V})^{-1}\mathbf{V}^T\mathbf{Q}\mathbf{B} \\ \hline \mathbf{C}\mathbf{V} & 0 \end{array}\right] \quad (7)$$

Obviously, the quality of the approximant $\mathbf{G}_r(s)$ critically depends on the choice of the rational Krylov subspace $\mathbf{V}$. What interpolation points (shifts) σ_k should one choose to form $\mathbf{V}$? Choice of interpolation points is the most crucial question in Krylov-based model reduction. Until very recently, selection of interpolation points has been usually an ad-hoc process. A systematic way of choosing the shifts has been proposed by Gugercin and Antoulas [22], [20]. Even though this selection strategy worked quite efficiently as shown in [22], [20], it is not an optimal choice.

In this note, we will choose the interpolation points and, hence, construct $\mathbf{V}$ based on the following important theorem, an extension of Theorem 3, p. 86 in Gaier's monograph [11] to continuous time.

Theorem 2: Given a stable SISO transfer function $\mathbf{G}(s)$ as in (1), and fixed stable reduced poles $\alpha_1, \ldots, \alpha_r$, define

$$\mathbf{G}_r(s) := \frac{\beta_0 + \beta_1 s + \cdots + \beta_{r-1}s^{r-1}}{(s-\alpha_1)\ldots(s-\alpha_r)}.$$

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

$$\mathbf{G}(s) = \mathbf{G}_r(s) \quad \text{for} \quad s = -\bar{\alpha}_1, -\bar{\alpha}_2, \ldots, -\bar{\alpha}_r. \quad (8)$$

Note that (8) can be rewritten as

$$\mathbf{G}(s) = \mathbf{G}_r(s) \quad \text{for} \quad s = -\alpha_1, -\alpha_2, \ldots, -\alpha_r.$$

since the poles, $\{\alpha_i\}$ occur in complex conjugate pairs. Theorem 2 states that if $\mathbf{G}_r(s)$ interpolates $\mathbf{G}(s)$ at the mirror images of the poles of $\mathbf{G}_r(s)$, then $\mathbf{G}_r(s)$ is guaranteed to be an *optimal* approximation of $\mathbf{G}(s)$ with respect to the $\mathcal{H}_2$ norm among all reduced order systems having the same reduced system poles $\{\alpha_i\}$, $i = 1, \ldots, r$. See [39], [24], [35], [30] and references there in, for *general* optimal $\mathcal{H}_2$ approximatiom problem.

Theorem 2 classifies an *optimal* shift selection strategy: Interpolation points as the mirror images of the poles of $\mathbf{G}_r(s)$, i.e. as the mirror images of the eigenvalues of $\mathbf{A}_r$. However, one *cannot* simply set $\sigma_i = -\lambda_i(\mathbf{A}_r)$ since these reduced poles are not known *a priori*. To achieve this goal, we propose to run the rational Krylov method successively; that is we run iterative rational Krylov steps where at the $(k+1)^{\text{st}}$ step the interpolation points are chosen as the mirror images of the eigenvalues of $\mathbf{A}_r$ from the k^{th} step. This forms the matrix $\mathbf{V}$ at each step. Then, the corresponding $\mathbf{Z}$ matrix is obtained from the formula (6). Here is a sketch of the proposed algorithm:

Algorithm 1: **An Iterative SVD-Krylov Based Model Reduction Method:**

1) Make an initial shift selection σ_i, for $i = 1, \ldots, r$.
2) $\mathbf{V} = \text{Im}\left[(\sigma_1\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}, \cdots, (\sigma_r\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\right]$
3) $\mathbf{Z} = \mathbf{Q}\mathbf{V}(\mathbf{V}^T\mathbf{Q}\mathbf{V})^{-1}$
4) while (not converged)
   a) $\mathbf{A}_r = \mathbf{Z}^T\mathbf{A}\mathbf{V}$,
   b) $\sigma_i \longleftarrow -\lambda_i(\mathbf{A}_r)$ for $i = 1, \ldots, r$
   c) $\mathbf{V} = \text{Im}\left[(\sigma_1\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}, \cdots, (\sigma_r\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\right]$
   d) $\mathbf{Z} = \mathbf{Q}\mathbf{V}(\mathbf{V}^T\mathbf{Q}\mathbf{V})^{-1}$
5) $\mathbf{A}_r = \mathbf{Z}^T\mathbf{A}\mathbf{V}$, $\mathbf{B}_r = \mathbf{Z}^T\mathbf{B}$, $\mathbf{C}_r = \mathbf{C}\mathbf{V}$
6) $\mathbf{G}_r(s) = \left[\begin{array}{c|c} \mathbf{A}_r & \mathbf{B}_r \\ \hline \mathbf{C}_r & 0 \end{array}\right]$.

It is clear that upon convergence, there holds $\sigma_i = -\lambda_i(\mathbf{A}_r)$, for $i = 1, \ldots, r$; and hence $\mathbf{G}_r(s)$ interpolates $\mathbf{G}(s)$ at the mirror images of the reduced poles, as desired.

The following theorem lists the important properties of the proposed algorithm:

Theorem 3: Given an asymptotically stable, reachable and observable dynamical system $\mathbf{G}(s) = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & 0 \end{array}\right]$, let the reduced model $\mathbf{G}_r(s)$ be obtained by Algorithm 1, Iterative SVD-Krylov Based Model Reduction Method. Then, $\mathbf{G}_r(s)$ is asymptotically stable. Also, let $\alpha_1, \ldots, \alpha_r$ denote the poles of $\mathbf{G}_r(s)$. $\mathbf{G}_r(s)$ interpolates $\mathbf{G}(s)$ at $-\alpha_i$, for $i = 1, \ldots, r$, and therefore minimizes the $\mathcal{H}_2$ error $\left\|\mathbf{G} - \hat{\mathbf{G}}\right\|_{\mathcal{H}_2}$ among all r^{th} order reduced models $\hat{\mathbf{G}}(s)$ having the same poles $\alpha_1, \ldots, \alpha_r$.

PROOF: For brevity, we give only a sketch of the proof. Obviously, without loss of generality, we can assume that $\mathbf{Q} = \mathbf{I}$. Then, $\mathbf{Z} = \mathbf{V}$ with $\mathbf{V}^T\mathbf{V} = \mathbf{I}_r$. Multiplying (5) by $\mathbf{V}^T$ from left and $\mathbf{V}$ from right proves that $\mathbf{A}_r$ is stable. To prove that $\mathbf{A}_r$ is *asymptotically* stable, i.e., it has no poles on the imaginary axis, we use contradiction. After some manipulations, it follows that the assumption that $\mathbf{A}_r$ has a pole on the imaginary axis leads to $\mathbf{G}(s)$ being unreachable, hence a contradiction. The key observation in proving this result is the fact $\mathbf{G}_r(s)$ interpolates $\mathbf{G}(s)$ at $-\lambda_i(\mathbf{A}_r)$, i.e. the result depends on the specific way of obtaining $\mathbf{V}$ through iterative rational Krylov steps in the proposed algorithm. The second part of the theorem follows from Theorem 2.

Some remarks are in order.

Remark 1: The reduced order models of the form (7) have appeared in the work of Skelton *et. al.* in [10], [41], [40]. In [41] and [40], the dual projection is used where $\mathbf{Q}$ is replaced by $\mathbf{P}$ and $\mathbf{V}$ is chosen as the observability matrix of order r leading to the so-called **q-cover realizations**. On the other hand, in [10], these results were generalized to the case where $\mathbf{V}$ were replaced by a rational Krylov subspace. However, the proposed algorithm is different from these approaches in the specific way we construct $\mathbf{V}$, through iterative rational Krylov steps. Indeed, in these works [10], [41], [40], the reduced model was *only* guaranteed to be stable, *not* asymptotically stable; i.e. $\mathbf{G}_r(s)$ might have a

pole on the imaginary axis even though the original model $\mathbf{G}(s)$ does not. However, the proposed approach *guarantees* asymptotic stability of $\mathbf{G}_r(s)$. Moreover, optimality in the $\mathcal{H}_2$ sense does not hold in [10], [41], [40], since this optimality requires the interpolation condition (8).

Remark 2: In the discrete-time case, one can apply the same projection in (7) with replacing $\mathbf{Q}$ by the observability gramian of the corresponding discrete-time systems, i.e. in the discrete-time case $\mathbf{Q}$ will be the solution to the Stein equation $\mathbf{A}^T\mathbf{Q}\mathbf{A} + \mathbf{C}^T\mathbf{C} = \mathbf{Q}$. This leads to the *least-squares model reduction* approach of Gugercin and Antoulas [23]. Unlike the continuous-time case, regardless of the choice of $\mathbf{V}$, $\mathbf{G}_r(s)$ is guaranteed to be asymptotically stable, i.e., in this case, $\mid \lambda_i(\mathbf{A}_r) \mid < 1$. These are the precise reasons that [23] proposed, first, transforming a continuous-time system into discrete-time, applying the least-squares reduction in discrete-time, and then transforming back to continuous time. However, in our proposed approach, we will achieve asymptotic stability while staying in continuous time. In addition, we will have the optimality in the $\mathcal{H}_2$ sense due to (8).

We have implemented Algorithm 1 for many different large-scale systems. In each of our numerical examples, the algorithm worked very efficiently and has always converged after a small number of steps. Even though these results suggested that Algorithm 1 is both efficient and reliable, conditions that guarantee convergence are not yet fully understood. This issue is currently under investigation.

A. Initial Shift Selection

For the proposed algorithm, the rate of convergence and the final reduced model depend on the initial shift selection. In this section, we discuss this issue. But, first, we would like to state that for almost all of our simulations, *a random initial shift selection resulted in a satisfactory reduced model.*

It is clear that one should make the initial shift selection in the region where the mirror images of the spectrum of $\mathbf{A}$ lies. This comes from the fact that at the end, we will have interpolation points as the mirror images of the reduced system poles, and, as in the eigenvalue computations, these reduces poles will somehow reflect the original pole distribution. One can easily find the eigenvalues of $\mathbf{A}$ with the smallest and largest real and imaginary parts. Then we suggest choosing shifts in this region. For example, let the eigenvalues of $\mathbf{A}$ be all real and lie in the interval $[-10, -0.1]$. Then one should make an initial shift selection inside the region $[0.1, 10]$. We would like to note that the task of computing the eigenvalues of $\mathbf{A}$ with the smallest/largest real and imaginary part can be achieved effectively using an implicitly restarted Arnoldi (IRA) algorithm [36].

Another initialization strategy is the shift selection strategy proposed by Gugercin and Antoulas in [22], [20]. Based on a $\mathcal{H}_2$ error expression, it was proposed choosing $\sigma_i = -\lambda_i(\mathbf{A})$ where $\lambda_i(\mathbf{A})$ are the original poles with big residuals. It is shown in [22], [20] that this selection of interpolation points works quite well, hence one can initiate Algorithm 1 with this selection. However, once more, we state that, with an initial random choice, the proposed method leads to reduced order models better than those obtained by balanced truncation as shown in Section IV.

IV. Numerical Examples

In this section, we apply the proposed algorithm to a dynamical system and compare its performance with that of balanced truncation. In this example, $\mathbf{G}(s)$ denotes the FOM, $\mathbf{G}_{\mathrm{bal}}(s)$ the reduced model due to balanced truncation and $\mathbf{G}_{\mathrm{IQRK}}(s)$ the reduced model due to the proposed algorithm, Algorithm 1.

A. CD Player Model

The original model, obtained by finite elements, describes the dynamics between the lens actuator and the radial arm position of a portable CD player. The model has 120 states, i.e., n=120, with a single input and a single output. For more details on this system, see [20], [3], [18], [9].

First, we examine convergence behavior of Algorithm 1. Towards this goal, we reduce the order to $r = 6$ and $r = 20$ using Algorithm 1. Initial shifts are complex and selected *randomly* in the rectangular region over the complex plane with real part of the shifts bounded by $[-\max_i(\mathrm{Real}(\lambda_i(\mathbf{A}))), -\min_i(\mathrm{Real}(\lambda_i(\mathbf{A})))]$ and the imaginary parts of the shitfs bounded by $[\min_i(\mathrm{Imag}(\lambda_i(\mathbf{A}))), \max_i(\mathrm{Imag}(\lambda_i(\mathbf{A})))]$, for $i = 1, \ldots, n$. At each step of the iteration, we compute the $\mathcal{H}_2$ error due to the current estimate and plot this error vs iteration index. The results are shown in Figure 1. The figure illustrates that for both cases $r = 6$ and $r = 20$ **(1)** at each step of the iteration, the $\mathcal{H}_2$ norm of the error is reduced and **(2)** The algorithm converges after a small number of step, 3 steps for these cases. We note that these properties seem to be valid in general, and as stated before are under investigation.

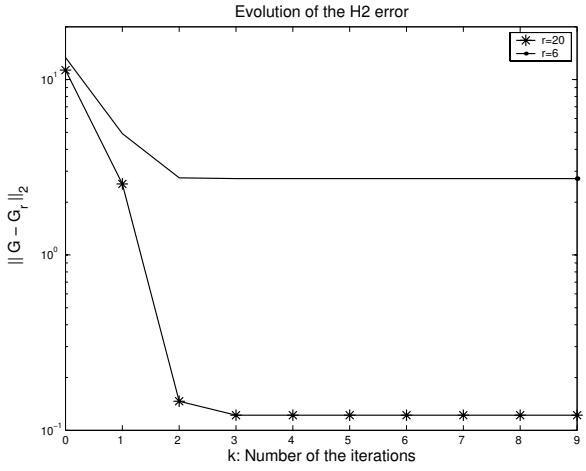

Fig. 1. $\mathcal{H}_2$ norm of the error system vs the number of iterations for CD Player Model

Next, we compare the performance of Algorithm 1 with that of balanced truncation. Recall that balanced truncation is well known as leading to small $\mathcal{H}_\infty$ and $\mathcal{H}_2$ error norms, see [3], [18]. Using both balanced truncation and the proposed approach, we reduce the order to r as r varies from 2 to 30; and for each r value, we compare the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ error norms due to balanced truncation and due to Algorithm 1. For Algorithm 1, the initial interpolation points are chosen randomly as explained above. Figure 2 shows the $\mathcal{H}_2$ errors vs r. It is clear that after $r = 18$, the proposed algorithm leads to a smaller $\mathcal{H}_2$ error and hence outperforms balanced truncation. Indeed, this is true for all the r values as shown in Figure 3. This figure depicts the difference between the $\mathcal{H}_2$ errors due to two algorithms, i.e. depicts $\|\mathbf{G}(s) - \mathbf{G}_{\rm bal}(s)\|_{\mathcal{H}_2} - \|\mathbf{G}(s) - \mathbf{G}_{\rm IQRK}(s)\|_{\mathcal{H}_2}$ vs r. As one can see, the $\mathcal{H}_2$ error for balanced truncation is always bigger. Hence, the proposed algorithm consistently leads to smaller $\mathcal{H}_2$ error. We would like note that this is achieved by a random initial shift selection. We would like to also note that we achieve this better performance by solving only one Lyapunov equation. Since the iteration converges in a small number of steps, the cost due to the Krylov side is small; overall cost of the proposed method is about the half of the cost of the balanced truncation.

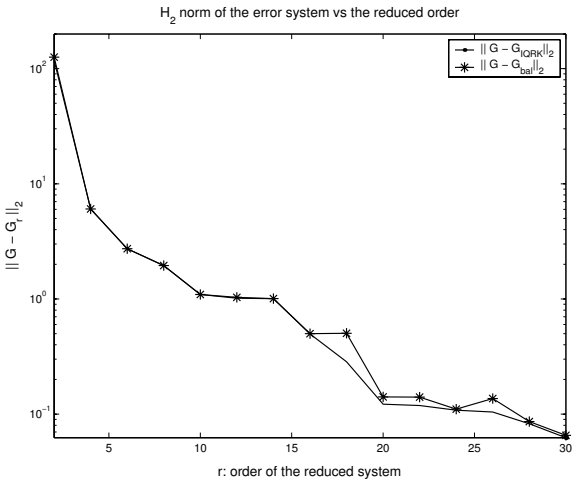

Fig. 2. $\mathcal{H}_2$ norm of the error system vs r for CD Player Model

We make the same analysis for the $\mathcal{H}_\infty$ error as well. $\mathcal{H}_\infty$ error vs the reduced order r for both methods are plotted in Figure 4, and the difference between the $\mathcal{H}_\infty$ errors, i.e. $\|\mathbf{G}(s) - \mathbf{G}_{\rm bal}(s)\|_{\mathcal{H}_\infty} - \|\mathbf{G}(s) - \mathbf{G}_{\rm IQRK}(s)\|_{\mathcal{H}_\infty}$ is plotted in Figure 5. These figures show that Algorithm 1 yields satisfactory $\mathcal{H}_\infty$ performance as well, . However, unlike the $\mathcal{H}_2$ case, there are some r values for which balanced truncation is slightly better. Even though the algorithm has optimality in the $\mathcal{H}_2$ sense, good $\mathcal{H}_\infty$ performance is expected since as recently shown in [6], moment matching at the mirror images of the reduced system poles is the right choice for the $\mathcal{H}_\infty$ performance as well.

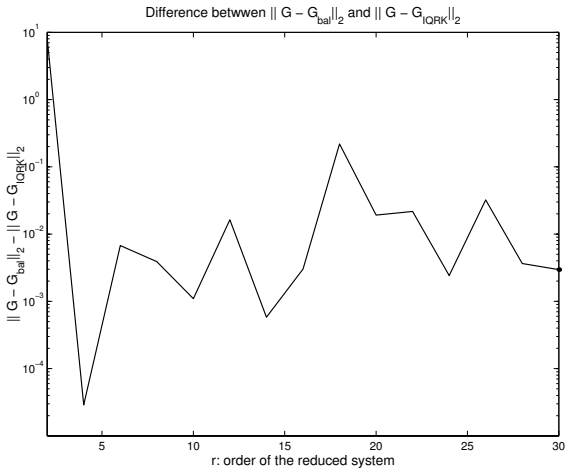

Fig. 3. $\|\mathbf{G}(s) - \mathbf{G}_{\rm bal}(s)\|_{\mathcal{H}_2} - \|\mathbf{G}(s) - \mathbf{G}_{\rm IQRK}(s)\|_{\mathcal{H}_2}$ vs r for CD Player Model

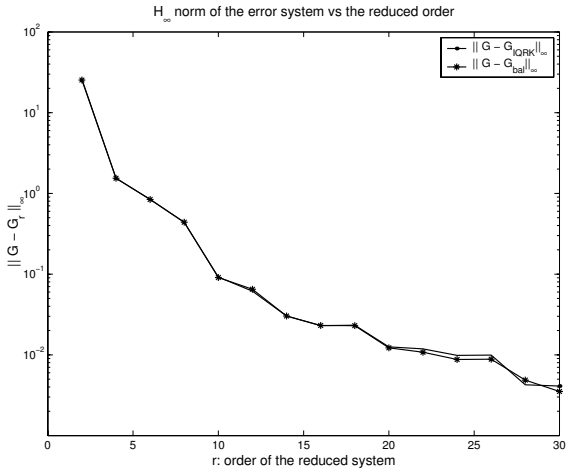

Fig. 4. $\mathcal{H}_\infty$ norm of the error system vs r for CD Player Model

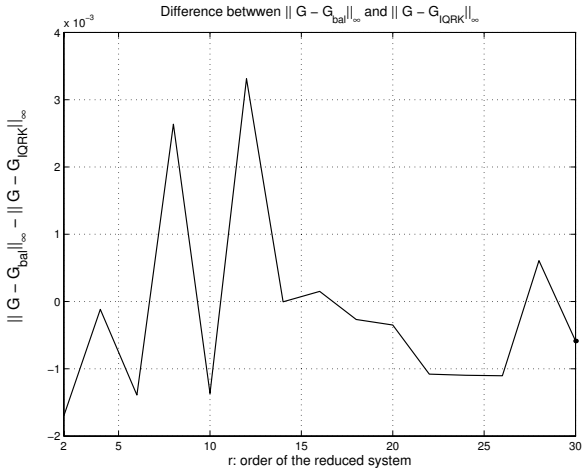

Fig. 5. $\|\mathbf{G}(s) - \mathbf{G}_{\rm bal}(s)\|_{\mathcal{H}_\infty} - \|\mathbf{G}(s) - \mathbf{G}_{\rm IQRK}(s)\|_{\mathcal{H}_\infty}$ vs r for CD Player Model

V. CONCLUSIONS

We have proposed a model reduction algorithm which combines the SVD and Krylov-based methods. It is a two-sided projection method where one side carries the SVD (Gramian) information and the other side the Krylov information. Krylov part of the projection is obtained via iterative rational Krylov steps. The reduced model is asymptotically stable and matches the moments of the original system at the mirror images of the reduced system poles; hence it is the best $\mathcal{H}_2$ approximation among all reduced models having the same poles. Numerical results proves that the method is very effective and yields results better than balanced truncation. Even though the proposed approach has always been observed to converge, this issue is still under investigation. Also, a modified version of the proposed approach is currently under investigation to solve the optimal $\mathcal{H}_2$ problem, *not only* the restricted $\mathcal{H}_2$ problem.

REFERENCES

[1] A.C. Antoulas and D.C. Sorensen, *Projection methods for balanced model reduction*, Technical Report ECE-CAAM Depts, Rice University, September 1999.

[2] A.C. Antoulas and D.C. Sorensen, *The Sylvester equation and approximate balanced reduction*, Fourth Special Issue on Linear Systems and Control, Edited by V. Blondel, D. Hinrichsen, J. Rosenthal, and P.M. van Dooren, Linear Algebra and Its Applications, **351-352**: 671-700, 2002.

[3] A. C. Antoulas, D. C. Sorensen, and S. Gugercin, *A survey of model reduction methods for large scale systems*, Contemporary Mathematics, AMS Publications, **280:** 193-219, 2001.

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

[5] 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.

[6] 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.

[7] P. Benner and E. S. Quintana-Orti. Solving stable generalized Lyapunov equation with the matrix sign function. *Numerical Algorithms*, Vol. 20, pp. 75-100, 1999.

[8] P. Benner, E. S. Quintana-Orti, and G. Quintana-Orti. Efficient Numerical Algorithms for Balanced Stochastic Truncation. *International Journal of Applied Mathematics and Computer Science*, Special Issue: Numerical Analysis and Systems Theory, Editor Stephen L. CAMPBELL, Vol. 11, No. 5, 2001.

[9] Y. Chahlaoui and P. Van Dooren, *A collection of benchmark examples for model reduction of linear time invariant dynamical systems*, SLICOT Working Note 2002-2: February 2002.

[10] C. De Villemagne and R. Skelton, *Model reduction using a projection formulation*, International Jour. of Control, Vol. **40**, 2141-2169, 1987.

[11] D. Gaier, *Lectures on Complex Approximation*, Birkhauser, 1987.

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

[13] K. Gallivan, E. Grimme and P. Van Dooren, *Model reduction of large-scale systems, rational Krylov versus balancing techniques*, in Error control and adaptivity in scientific computing, pp: 177-190, Kluwer Acad. Publ., Dordrecht, 1999.

[14] K. Gallivan, A. Vandendorpe and P. Van Dooren, *Sylvester equations and projection-based model reduction*, J. Comp. Appl. Math., 162: 213-229, 2004.

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

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

[17] 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).

[18] S. Gugercin and A. C. Antoulas, *A comparative study of 7 model reduction algorithms*, Proceedings of the 39th IEEE Conference on Decision and Control, Sydney, Australia, December 2000.

[19] S. Gugercin, A. C. Antoulas and N. Bedrossian, *Approximation of the International Space Station 1R and 12A models*, in the Proceedings of the 40^{th} CDC, December 2001.

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

[21] S. Gugercin, D.C. Sorensen and A.C. Antoulas, *A modified low-rank Smith method for large-scale Lyapunov equations*, Numerical Algorithms, Vol. 32, Issue 1, pp. 27-55, January 2003.

[22] 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.

[23] S. Gugercin and A.C. Antoulas, *Model reduction of large scale systems by least squares*, accepted to appear in Linear Algebra and its Applications, Special Issue on Order Reduction of Large-Scale Systems, 2005.

[24] 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.

[25] I.M. Jaimoukha, E.M. Kasenally, *Implicitly restarted Krylov subspace methods for stable partial realizations*, SIAM J. Matrix Anal. Appl., **18**: 633-652 (1997).

[26] 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.

[27] J.-R. Li and J. White. Low rank solution of Lyapunov equations. *SIAM J. Matrix Anal. Appl.*, Vol. 24, No. 1, 2002.

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

[29] 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.

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

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

[32] 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.

[33] 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.

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

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

[36] D.C. Sorensen, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Applic., **13**: 357-385 (1992).

[37] A. Varga, *Model reduction software in the SLICOT library*, Applied and Computational Control, Signals, and Circuits, Ed. B. Datta", Kluwer Academic Publishers, Boston, 2001.

[38] A. Varga, *Balancing-free square-root algorithm for computing singular perturbation approximations*, Proc. of 30th IEEE CDC, Brighton, UK, pp. 1062-1065, 1991.

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

[40] A. Yousouff, D. A. Wagie, and R. E. Skelton, *Linear system approximation via covariance equivalent realizations*, Journal of Math. Anal. and App., Vol. **196**, 91-115, 1985.

[41] A. Yousouff and R.E. Skelton, *Covariance equivalent realizations with applications to model reduction of large-scale systems*, in Control and Dynamic Systems, C.T. Leondes ed., Academic Press, vol. 22, pp. 273-348, 1985.