IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 9, 2019, accepted May 14, 2019, date of publication May 23, 2019, date of current version June 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2918722

Inexact Linear Solves in Model Reduction of Bilinear Dynamical Systems

RAJENDRA CHOUDHARY AND KAPIL AHUJA
Computational Science and Engineering Laboratory, IIT Indore, Indore 453552, India
Corresponding author: Rajendra Choudhary (rajendracse46@gmail.com)

ABSTRACT The bilinear iterative rational Krylov algorithm (BIRKA) is a very popular, standard, and mathematically sound algorithm for reducing bilinear dynamical systems that arise commonly in science and engineering. This reduction process is termed as a model order reduction (MOR) and leads to a faster simulation of such systems. An efficient variant of the BIRKA, Truncated BIRKA (TBIRKA) has also been recently proposed. Like for any MOR algorithm, these two algorithms also require solving multiple linear systems as part of the model reduction process. For reducing the MOR time, these linear systems are often solved by an iterative solver, which introduces approximation errors (implying inexact solves). Hence, stability analysis of the MOR algorithms with respect to inexact linear solves is important. In our past work, we have shown that under mild conditions, the BIRKA is stable. Here, we look at the stability of the TBIRKA in the same context. Besides deriving the conditions for a stable TBIRKA, our other novel contribution is the more intuitive methodology for achieving this. The stability analysis techniques that we propose here can be extended to many other methods for doing the MOR of bilinear dynamical systems, e.g., using balanced truncation or the ADI methods.

INDEX TERMS Backward stability, bilinear dynamical systems, interpolatory projection, model order reduction, perturbation analysis, stability analysis, volterra series interpolation.

I. INTRODUCTION

A dynamical system, usually represented by a set of differential equations, can be linear or non-linear [1]–[3]. Linear dynamical systems have been studied more than the non-linear ones because of the obvious ease in working with them. Bilinear dynamical systems form a good bridge between the linear and the non-linear cases, and are usually approximated by a varying degree of bilinearity [4]–[6]. In this manuscript, we focus on bilinear dynamical systems.

Dynamical systems coming from different real world applications are very large in size. Thus, simulation and computation with such systems is prohibitively expensive in time. Model Order Reduction (MOR) techniques provide a smaller system that besides being cheaper to work with, also replicates the input-output behavior of the original system to a great extent [7]–[16]. Since, bilinear dynamical systems have been recently studied, the techniques for reducing them are also recent.

Out of the many methods available for performing bilinear MOR [14], [15], [17]–[22], we focus on a commonly used interpolatory projection method. BIRKA (Bilinear Iterative Rational Krylov Algorithm) [15] is a very popular algorithm based upon this technique for reducing *first-order* bilinear dynamical systems.[1] BIRKA's biggest drawback is that it does not scale well in time (with respect to increase in the size of the input dynamical system). A cheaper variant of BIRKA, called TBIRKA (Truncated Bilinear Iterative Rational Krylov Algorithm) [21], [22] has also been proposed.

Like in any other MOR algorithm, in BIRKA and TBIRKA also, people often use direct methods like LU-factorization, etc., to solve the arising linear systems, which have a high time complexity ($\mathcal{O}(n^3)$, where n is the original system size) [23], [24]. A common solution to this scaling problem is to use iterative methods like the Krylov subspace methods, etc.,[2] which have a reduced time complexity (i.e., $\mathcal{O}(n \times nnz)$, where nnz is the number of nonzeros in the system matrix) [23], [25]. Although iterative methods are cheap, they are inexact too. Hence, studying stability of the underlying

The associate editor coordinating the review of this manuscript and approving it for publication was Feiqi Deng.

[1] *First-order* implies that the highest derivative of the state variable in the dynamical system is one. *Second-order* and *higher-orders* are similarly defined.

[2] Here, a new iterative method for solving "Super Large Scale Systems", with $n = 1,000,000$, can also be used [35].

MOR algorithm (here BIRKA and TBIRKA) with respect to such approximate (inexact) linear solves becomes important [26], [27].

One of the first works that performed such a stability analysis focused on popular MOR algorithms for *first-order* linear dynamical systems [28]. Here, the authors briefly mention that their analysis would be easily carried from the *first-order* to the *second-order* case. A detailed stability analysis focusing on reducing *second-order* linear dynamical systems has been done in [29]. A different kind of stability analysis for MOR of *second-order* linear dynamical systems has been done in [30]. In this, the authors first show that the SOAR algorithm (second order Arnoldi) is unstable with respect to the machine precision errors (and not inexact linear solves). Then, they propose a Two-level orthogonal Arnoldi (TOAR) algorithm that cures this instability of SOAR. An extended stability analysis for BIRKA (as above, a popular MOR algorithm for *first-order* bilinear dynamical systems) has been recently done in [31]. For rest of this manuscript, whenever stability analysis is referred, we mean it with respect to inexact linear solves.

We follow the stability analysis framework of BIRKA from [31] and propose equivalent theorems for TBIRKA. The approach here is slightly different, which forms our most *novel* contribution. Norm of the dynamical system plays an important role in stability analysis (the kind of norm is discussed later). In BIRKA stability analysis, a single expression for bilinear dynamical system norm is used (involving a Volterra series). In TBIRKA stability analysis, a similar single expression (involving a truncated Volterra series) leads to complications. Alternatively, in TBIRKA, because of truncation, the bilinear dynamical system can be represented by a finite set of functions. This was not possible in BIRKA where infinite such functions were needed. Thus, in TBIRKA stability analysis, we use norm of all such functions leading to easier derivations. Our stability analysis, as done for BIRKA earlier and for TBIRKA here, can be easily extended to other MOR algorithms for bilinear dynamical systems, e.g., projection based [14], implicit Volterra series [17], balanced truncation [18], gramian based [20], etc.

The rest of the paper is divided into three more parts. In Section II, we first give a brief overview of MOR for bilinear dynamical systems using a projection method. Next, we review the stability analysis of BIRKA from [31]. Stability analysis of TBIRKA is discussed in Section III. In Section IV, we give conclusions and discuss the future work. For the rest of this paper, we use the terms and notations as listed below.

a. The H_2-norm is a functional norm defined as [21], [22], [28]

$$\|H_k\|_{H_2}^2 = \left(\frac{1}{2\pi}\right)^k \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \times \|H_k\left(i\omega_1, \ldots, i\omega_k\right)\|_F^2 \, d\omega_1 \ldots d\omega_k, \quad (1)$$

where i denotes $\sqrt{-1}$. Here, we assume that all H_2-norms computed further exist. In other words, the improper integrals defined by the H_2-norm give finite value. This is a reasonable assumption because this happens often in practice (see [28] and [31], where stability analysis of IRKA and BIRKA is done, respectively).

b. The $H_\infty-$norm is also a functional norm, defined as [21], [22], [28]

$$\|H_k\|_{H_\infty} = \underset{\omega_1,\ldots,\,\omega_k \in \mathbb{R}}{max} \|H_k\left(i\omega_1, \ldots, i\omega_k\right)\|_2 .$$

c. The Kronecker product between two matrices P (of size $m \times n$) and Q (of size $s \times t$) is defined as

$$P \otimes Q = \begin{bmatrix} p_{11}Q & \cdots & p_{1n}Q \\ \vdots & \ddots & \vdots \\ p_{m1}Q & \cdots & p_{mn}Q \end{bmatrix},$$

where $p_{ij} \in P$ and order of $P \otimes Q$ is $ms \times nt$.

d. *vec* operator on a matrix P is defined as

$$vec(P) = \left(p_{11}, \ldots, p_{m1}, p_{12}, \ldots, p_{m2}, \ldots, p_{1n}, \ldots, p_{mn}\right)^T.$$

e. Also, I_n denotes an identity matrix of size $n \times n$ and $\mathbb{R}$ denotes the set of real numbers.

II. BACKGROUND

A *first-order* bilinear dynamical system is usually represented as [14], [15]

$$\zeta : \begin{cases} \dot{x}(t) = Ax(t) + Nx(t)u(t) + bu(t), \\ y(t) = cx(t), \end{cases} \quad (2)$$

where $A, N \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^{n\times 1}$, and $c \in \mathbb{R}^{1\times n}$. Also, $u(t) : \mathbb{R} \to \mathbb{R}$, $y(t) : \mathbb{R} \to \mathbb{R}$, and $x(t) : \mathbb{R} \to \mathbb{R}^n$, represent input, output, and state of the bilinear dynamical system, respectively. This is a Single Input Single Output (SISO) system, which we have chosen for ease of our analysis. We plan to look at Multiple Input Multiple Output (MIMO) systems as part of our future work.

A bilinear dynamical system can also be represented by a infinite set of transfer functions [21]. That is,

$$\zeta = \underset{k\to\infty}{Lim}\, \zeta^k, \quad (3)$$

where $\zeta^k = \{H_1(s_1), H_2(s_1, s_2), \ldots, H_k(s_1, s_2, \ldots, s_k)\}$. Here, $H_k(s_1, s_2, \ldots, s_k)$ is called the k^{th} order transfer function of the bilinear dynamical system and is defined as

$$H_k\left(s_1, \ldots, s_k\right) = c\left(s_k I - A\right)^{-1} N \left(s_{k-1}I - A\right)^{-1} \ldots N\left(s_1 I - A\right)^{-1} b. \quad (4)$$

After reduction, the bilinear dynamical system (2) can be represented as

$$\zeta_r : \begin{cases} \dot{x}_r(t) = A_r x_r(t) + N_r x_r(t)u(t) + b_r u(t), \\ y_r(t) = c_r x_r(t), \end{cases} \quad (5)$$

where $A_r,\ N_r \in \mathbb{R}^{r\times r}$, $b_r \in \mathbb{R}^{r\times 1}$, and $c_r \in \mathbb{R}^{1\times r}$ with $r \ll n$. The main goal of model reduction is to approximate ζ by ζ_r in an appropriate norm, such that for all admissible inputs, $y_r(t)$ is nearly same to $y(t)$.

As mentioned earlier, we use interpolatory projection for performing model reduction. The two common and standard algorithms here, BIRKA and TBIRKA, use a Petrov-Galerkin projection. Let $\mathcal{V}_r$ and $\mathcal{W}_r$ be the two r-dimensional subspaces, such that $\mathcal{V}_r = Range(V_r)$ and $\mathcal{W}_r = Range(W_r)$, where $V_r \in \mathbb{R}^{n\times r}$ and $W_r \in \mathbb{R}^{n\times r}$ are matrices. Also, let $\left(W_r^T V_r\right)$ be invertible. [3] Applying the projection $x(t) = V_r x_r(t)$, and enforcing the Petrov-Galerkin conditions [15], [22] on the original bilinear dynamical system (2), we get the reduced system as

$$\begin{aligned} &W_r^T\left(V_r\dot{x}_r(t) - AV_r x_r(t) - NV_r x_r(t)u(t) - bu(t)\right) = 0,\\ &y(t) = cV_r x_r(t). \end{aligned}$$

Comparing the above two equations with their respective equations in (5), we get a relation between the original system matrices and the reduced system matrices, i.e.,

$$\begin{aligned} &A_r = \left(W_r^T V_r\right)^{-1} W_r^T A V_r, \quad N_r = \left(W_r^T V_r\right)^{-1} W_r^T N V_r,\\ &b_r = \left(W_r^T V_r\right)^{-1} W_r^T b, \quad \text{and } c_r = cV_r. \end{aligned}$$

One way of obtaining subspaces $\mathcal{V}_r$ and $\mathcal{W}_r$ is to use Volterra series interpolation. Further, to decide where to interpolate so as to obtain an optimal reduced model, an H_2-optimization problem is commonly solved (Theorem 4.7 from [21]).

Theorem 1 [21]: Let ζ be a bilinear system of order n. Let ζ_r be an H_2-optimal approximation of order r. Then, ζ_r satisfies the following multi-point Volterra series interpolation conditions:

$$\begin{aligned} &\sum_{k=1}^{\infty}\sum_{l_1=1}^{r}\cdots\sum_{l_k=1}^{r}\phi_{l_1,\, l_2,\, \ldots,\, l_k}\\ &\quad\times H_k\left(-\lambda_{l_1},\ -\lambda_{l_2},\ \ldots,\ -\lambda_{l_k}\right)\\ &= \sum_{k=1}^{\infty}\sum_{l_1=1}^{r}\cdots\sum_{l_k=1}^{r}\phi_{l_1,\, l_2,\, \ldots,\, l_k}\\ &\quad\times H_{r_k}\left(-\lambda_{l_1},\ -\lambda_{l_2},\ \ldots,\ -\lambda_{l_k}\right), \end{aligned}$$

and

$$\begin{aligned} &\sum_{k=1}^{\infty}\sum_{l_1=1}^{r}\cdots\sum_{l_k=1}^{r}\phi_{l_1,\, l_2,\, \ldots,\, l_k}\\ &\quad\times\left(\sum_{j=1}^{k}\frac{\partial}{\partial s_j}H_k\left(-\lambda_{l_1},\ -\lambda_{l_2},\ \ldots,\ -\lambda_{l_k}\right)\right)\\ &= \sum_{k=1}^{\infty}\sum_{l_1=1}^{r}\cdots\sum_{l_k=1}^{r}\phi_{l_1,\, l_2,\, \ldots,\, l_k}\\ &\quad\times\left(\sum_{j=1}^{k}\frac{\partial}{\partial s_j}H_{r_k}\left(-\lambda_{l_1},\ -\lambda_{l_2},\ \ldots,\ -\lambda_{l_k}\right)\right), \end{aligned}$$

where $\phi_{l_1,\, l_2,\, \ldots,\, l_k}$ and $\lambda_{l_1},\ \lambda_{l_2},\ \ldots,\ \lambda_{l_k}$ are residues and poles of the transfer function H_{r_k} associated with ζ_r, respectively.

BIRKA is designed in such a way that at convergence, the conditions of Theorem 1 are satisfied leading to a locally H_2-optimal reduced model. Algorithm 1 lists BIRKA.

Algorithm 1 BIRKA [15], [21], [22]

```
1: Given an input bilinear dynamical system $A,\ N,\ b,\ c$.
2: Select an initial guess for the reduced system as
   $\check{A},\ \check{N},\ \check{b},\ \check{c}$. Also select stopping tolerance $btol$.
3: While (relative change in eigenvalues of $\check{A} \geq btol$)
     a. $R\Lambda R^{-1} = \check{A},\ \check{\check{b}} = \check{b}^T R^{-T},\ \check{\check{c}} = \check{c}R,$
        $\check{\check{N}} = R^T\check{N}R^{-T}.$
     b. $vec\,(\mathbf{V}) =$
        $\left(-\Lambda \otimes I_n - I_r \otimes A - \check{\check{N}}^T \otimes N\right)^{-1}\left(\check{\check{b}}^T \otimes b\right).$
     c. $vec\,(\mathbf{W}) =$
        $\left(-\Lambda \otimes I_n - I_r \otimes A^T - \check{\check{N}} \otimes N^T\right)^{-1}\left(\check{\check{c}}^T \otimes c^T\right).$
     d. $\mathbf{V}_r = orth\,(\mathbf{V})\,,\ \mathbf{W}_r = orth\,(\mathbf{W}).$
     e. $\check{A} = (\mathbf{W}_r^T\mathbf{V}_r)^{-1}\mathbf{W}_r^T A\mathbf{V}_r,$
        $\check{N} = \left(\mathbf{W}_r^T\mathbf{V}_r\right)^{-1}\mathbf{W}_r^T N\mathbf{V}_r,$
        $\check{b} = \left(\mathbf{W}_r^T\mathbf{V}_r\right)^{-1}\mathbf{W}_r^T b, \quad \check{c} = c\mathbf{V}_r.$
4: $A_r = \check{A}, \quad N_r = \check{N}, \quad b_r = \check{b}, \quad c_r = \check{c}.$
```

TBIRKA is similar to BIRKA in most aspects except that it performs a truncated Volterra series interpolation. Here, instead of ζ in (2)-(3), they work with ζ^M, which is defined as

$$\begin{aligned} \zeta^M = \{&H_1\left(s_1\right),\ H_2\left(s_1,\ s_2\right),\ H_3\left(s_1,\ s_2,\ s_3\right),\\ &\ldots H_M\left(s_1,\ \ldots,\ s_M\right)\}, \end{aligned} \tag{6}$$

with $H_k\left(s_1,\ \ldots,\ s_k\right)$ for $k \in \{1,\ \ldots,\ M\}$ is given by (4). Equivalent of Theorem 1 above is as follows (Theorem 4.8 from [21]):

Theorem 2 [21]: Let $\zeta = (A, N, b, c)$ be an order n bilinear system and ζ^M be the polynomial system determined by ζ. Let $\zeta_r = (A_r, N_r, b_r, c_r)$ be a bilinear system of order r, and define ζ_r^M as the polynomial system determined by ζ_r. Suppose that ζ_r^M is an H_2-optimal approximation to ζ^M. Then ζ_r^M satisfies

$$\begin{aligned} &\sum_{k=1}^{M}\sum_{l_1=1}^{r}\cdots\sum_{l_k=1}^{r}\phi_{l_1,\, l_2,\, \ldots,\, l_k}\\ &\quad\times H_k\left(-\lambda_{l_1},\ -\lambda_{l_2},\ \ldots,\ -\lambda_{l_k}\right)\\ &= \sum_{k=1}^{M}\sum_{l_1=1}^{r}\cdots\sum_{l_k=1}^{r}\phi_{l_1,\, l_2,\, \ldots,\, l_k} \end{aligned}$$

[3] Obtaining such an invertible matrix is not difficult [15], [22].

$$\times H_{r_k}\left(-\lambda_{l_1},\, -\lambda_{l_2},\, \ldots,\, -\lambda_{l_k}\right),$$

and

$$\begin{aligned}
&\sum_{k=1}^{M}\sum_{l_1=1}^{r}\cdots\sum_{l_k=1}^{r}\phi_{l_1,\, l_2,\, \ldots,\, l_k} \\
&\qquad\times\left(\sum_{j=1}^{k}\frac{\partial}{\partial s_j}H_k\left(-\lambda_{l_1},\, -\lambda_{l_2},\, \ldots,\, -\lambda_{l_k}\right)\right) \\
&=\sum_{k=1}^{M}\sum_{l_1=1}^{r}\cdots\sum_{l_k=1}^{r}\phi_{l_1,\, l_2,\, \ldots,\, l_k} \\
&\qquad\times\left(\sum_{j=1}^{k}\frac{\partial}{\partial s_j}H_{r_k}\left(-\lambda_{l_1},\, -\lambda_{l_2},\, \ldots,\, -\lambda_{l_k}\right)\right),
\end{aligned}$$

where $\phi_{l_1,\, l_2,\, \ldots,\, l_k}$ and $\lambda_{l_1},\ \lambda_{l_2},\ \ldots,\ \lambda_{l_k}$ are residues and poles of the transfer function H_{r_k} associated with ζ_r^M, respectively.

Algorithm 2 lists TBIRKA.

Both BIRKA and TBIRKA in turn require solving large sparse linear systems of equations. If we compare Algorithm 1 and 2, we realize that the number of linear solves at each step of the `While` loop in the former is 2 systems of size $nr \times nr$ and in the latter is $2M$ systems of size $nr \times nr$. This makes it seem that TBIRKA is more expensive than BIRKA. However, TBIRKA is implemented in such a way that the Kronecker products are avoided making it more efficient than BIRKA. For further details on this see chapter 4 in [21] and Section 5.3 in [22]. These implementation details do not affect our stability analysis, and hence, we use Algorithm 2 in the current form as our base.

As mentioned earlier, using iterative methods for solving such linear systems introduces approximation errors. We have done a detailed stability analysis of BIRKA with respect to the inexact linear solves in [31], and we briefly revisit this next. Generally, accuracy is the metric that tells about the correctness in the output of an inexact algorithm. Due to unavailability of the exact output, it is not possible to determine accuracy [26], [31]. A more easier metric is stability. Backward stability is one such notation, which says "A backward stable algorithm gives exactly the right output to nearly the right input" [26]. In our context, theoretically we obtain two reduced systems. One by applying an inexact MOR algorithm (with iterative linear solves) on the original full model, and other by applying the same MOR algorithm but exactly (with direct linear solves) on a perturbed full model (the perturbation is introduced in the original full model as part of stability analysis, and is an unknown quantity). If these two reduced systems are equal (*first condition*), with the difference between the original full model and the perturbed full model equal to the order of perturbation (*second condition*), then the MOR algorithm under consideration is called backward stable. The two theorems summarizing this stability analysis for BIRKA are listed below.

Algorithm 2 TBIRKA [21], [22]

```
1: Given an input bilinear dynamical system $A,\ N,\ b,\ c$.
2: Select an initial guess for the reduced system as $\breve{A},\ \breve{N},\ \breve{b},\ \breve{c}$. Also select the truncation index $M$ and stopping tolerance $tbtol$.
3: While $\left(\text{relative change in eigenvalues of } \breve{A} \geq tbtol\right)$
    a. $R\Lambda R^{-1} = \breve{A},\ \breve{\breve{b}} = \breve{b}^T R^{-T},\ \breve{\breve{c}} = \breve{c}R,$
       $\breve{\breve{N}} = R^T \breve{N} R^{-T}$.
    b. Compute
       $vec\,(\mathbf{V}_1) = (-\Lambda \otimes I_n - I_r \otimes A)^{-1}\left(\breve{\breve{b}}^T \otimes b\right),$
       $vec\,(\mathbf{W}_1) = \left(-\Lambda \otimes I_n - I_r \otimes A^T\right)^{-1}\left(\breve{\breve{c}}^T \otimes c^T\right).$
    c. For $j = 2, \ldots, M$, solve
       $vec\,(\mathbf{V}_j) = (-\Lambda \otimes I_n - I_r \otimes A)^{-1}\left(\breve{\breve{N}}^T \otimes N\right) vec\,(\mathbf{V}_{j-1}),$
       $vec\,(\mathbf{W}_j) = \left(-\Lambda \otimes I_n - I_r \otimes A^T\right)^{-1}\left(\breve{\breve{N}} \otimes N^T\right) vec\,(\mathbf{W}_{j-1}).$
    d. $\mathbf{V} = \sum_{j=1}^{M}\mathbf{V}_j,\ \mathbf{W} = \sum_{j=1}^{M}\mathbf{W}_j.$
    e. $\mathbf{V}_r = orth\,(\mathbf{V}),\ \mathbf{W}_r = orth\,(\mathbf{W}).$
    f. $\breve{A} = (\mathbf{W}_r^T\mathbf{V}_r)^{-1}\mathbf{W}_r^T A \mathbf{V}_r,$
       $\breve{N} = \left(\mathbf{W}_r^T\mathbf{V}_r\right)^{-1}\mathbf{W}_r^T N \mathbf{V}_r,$
       $\breve{b} = \left(\mathbf{W}_r^T\mathbf{V}_r\right)^{-1}\mathbf{W}_r^T b, \quad \breve{c} = c\mathbf{V}_r.$
4: $A_r = \breve{A}, \quad N_r = \breve{N}, \quad b_r = \breve{b}, \quad c_r = \breve{c}.$
```

Theorem 3 [31] : If the inexact linear solves in BIRKA (lines 3b. and 3c. of Algorithm 1) are solved using a Petrov-Galerkin framework, then BIRKA satisfies the first condition of backward stability with respect to these solves.

Theorem 4 [31] : Let $\widehat{Q} = \left(-\begin{bmatrix} A & 0 \\ 0 & A\end{bmatrix} \otimes \begin{bmatrix} I_n & 0 \\ 0 & I_n\end{bmatrix} - \begin{bmatrix} I_n & 0 \\ 0 & I_n\end{bmatrix} \otimes \begin{bmatrix} A & 0 \\ 0 & A\end{bmatrix} - \begin{bmatrix} N & 0 \\ 0 & N\end{bmatrix} \otimes \begin{bmatrix} N & 0 \\ 0 & N\end{bmatrix}\right)$*, where I_n is an identity matrix of size $n \times n$ and $\otimes$ denotes the standard Kronecker product. Also, let* $\widehat{\widehat{F}} = \left(I_{2n} \otimes \widehat{F} + \widehat{F} \otimes I_{2n}\right)$ *with* $\widehat{F} = \begin{bmatrix} 0 & 0 \\ 0 & F\end{bmatrix}$*, where F is the perturbation introduced in A matrix of the input dynamical system and I_{2n} is an identity matrix of size $2n \times 2n$. If $\widehat{Q}$ is invertible, $\left\|\widehat{Q}^{-1}\right\|_2 < 1$, and $\left\|\widehat{\widehat{F}}\right\|_2 < 1$, then BIRKA satisfies the second condition of backward stability with respect to the inexact linear solves.*

III. BACKWARD STABILITY OF TBIRKA

Here, the *first condition* is satisfied in a way similar to that of BIRKA except that some extra orthogonality conditions are imposed on the linear solver (discussed below).

Theorem 5: Let the inexact linear solves in TBIRKA (lines 3b. and 3c. of Algorithm 2) are solved satisfying

$$\begin{bmatrix}\mathbf{V}_1^T\\ \mathbf{V}_2^T\\ \vdots\\ \mathbf{V}_M^T\end{bmatrix}\begin{bmatrix}R_{c_1} & R_{c_2} & \cdots & R_{c_M}\end{bmatrix}=0 \qquad \text{and}$$

$$\begin{bmatrix}\mathbf{W}_1^T\\ \mathbf{W}_2^T\\ \vdots\\ \mathbf{W}_M^T\end{bmatrix}\begin{bmatrix}R_{b_1} & R_{b_2} & \cdots & R_{b_M}\end{bmatrix}=0, \tag{7}$$

where $\mathbf{V}_1$ and $\mathbf{V}_j$ are given by the first equations of lines 3b. and 3c. of Algorithm 2, respectively; R_{c_1} and R_{c_j} are the residuals in the second equations of lines 3b. and 3c. of Algorithm 2, respectively; $\mathbf{W}_1$ and $\mathbf{W}_j$ are given by the second equations of lines 3b. and 3c. of Algorithm 2, respectively; R_{b_1} and R_{b_j} are the residuals in the first equations of lines 3b. and 3c. of Algorithm 2, respectively; and $j = 2, \ldots, M$. Then, TBIRKA satisfies the first condition of backward stability with respect to these solves.

Proof: Follows the same pattern as the proof for Theorem 3 in [31]. □

From the above theorem, we infer that the underlying iterative solver should *firstly* be based upon a Petrov-Galerkin framework to achieve

$$\mathbf{V}_j^T R_{c_j} = 0 \quad \text{and} \quad \mathbf{W}_j^T R_{b_j} = 0, \tag{8}$$

for $j = 1, \ldots, M$. Since BiConjugate Gradient (i.e., BiCG) is one such algorithm [23], we propose its use in TBIRKA. This is exactly same as for BIRKA. *Secondly,* this particular solver should also satisfy the remaining orthogonalities of (7).

These orthogonalities have a form similar to the orthogonalities required while reducing second order linear dynamical systems ((23) and (24) in [29]), and can be easily satisfied by using a recycling variant of the underlying iterative solver. In [29], the ideal iterative solver to be used is Conjugate Gradient (i.e., CG) [23] (due to the use of Galerkin projection). Hence, to satisfy the similar orthogonalities there, without any extra cost, the authors use Recycling Conjugate Gradient (i.e., RCG) [32]. Since here BiCG is the ideal iterative solver (as discussed above), we propose the use of Recycling BiConjugate Gradient (i.e., RBiCG) [33], [34], which would ensure that the remaining orthogonalities of (7) (besides (8)) are satisfied without any extra cost.

To satisfy the *second condition* of backward stability of TBIRKA, we need to show that

$$\left\|\zeta^M - \tilde{\zeta}^M\right\|_{H_2} = \mathcal{O}\left(\|F\|_2\right), \tag{9}$$

where ζ^M is given by (6) or

$$\zeta^M = \left\{H_1(s_1),\ H_2(s_1,\ s_2),\ H_3(s_1,\ s_2,\ s_3),\ \ldots, H_M(s_1,\ \ldots,\ s_M)\right\} \tag{10a}$$

with $H_k(s_1,\ \ldots,\ s_k)$ for $k \in \{1,\ \ldots,\ M\}$ given by (4) or

$$H_k(s_1,\ \ldots,\ s_k) = c\,(s_k I - A)^{-1} N\,(s_{k-1} I - A)^{-1} \ldots N\,(s_1 I - A)^{-1} b, \tag{10b}$$

$$\tilde{\zeta}^M = \left\{\tilde{H}_1(s_1),\ \tilde{H}_2(s_1, s_2),\ \tilde{H}_3(s_1,\ s_2,\ s_3),\ \ldots, \tilde{H}_M(s_1,\ \ldots,\ s_M)\right\}, \tag{11a}$$

with for $k \in \{1,\ \ldots,\ M\}$

$$\tilde{H}_k(s_1,\ \ldots,\ s_k) = c\,(s_k I - (A+F))^{-1} N\,(s_{k-1} I - (A+F))^{-1} \ldots N\,(s_1 I - (A+F))^{-1} b, \tag{11b}$$

and assuming perturbation F in A matrix of the input dynamical system (as for BIRKA stability; see Theorem 4 earlier).

One way to satisfy (9) is to use the definition of the H_2−norm of $\zeta^M - \tilde{\zeta}^M$, i.e., from Lemma 5.1 of [22]

$$\begin{aligned}
&\left\|\zeta^M - \tilde{\zeta}^M\right\|_{H_2}^2 \\
&= \left(\begin{bmatrix}c & -c\end{bmatrix} \otimes \begin{bmatrix}c & -c\end{bmatrix}\right) \\
&\times \sum_{j=0}^{M}\left[\left(-\begin{bmatrix}A & 0\\ 0 & A+F\end{bmatrix} \otimes \begin{bmatrix}I_n & 0\\ 0 & I_n\end{bmatrix} - \begin{bmatrix}I_n & 0\\ 0 & I_n\end{bmatrix}\right.\right. \\
&\left.\left.\otimes \begin{bmatrix}A & 0\\ 0 & A+F\end{bmatrix}\right)^{-1} \begin{bmatrix}N & 0\\ 0 & N\end{bmatrix} \otimes \begin{bmatrix}N & 0\\ 0 & N\end{bmatrix}\right]^j \\
&\times \left(-\begin{bmatrix}A & 0\\ 0 & A+F\end{bmatrix} \otimes \begin{bmatrix}I_n & 0\\ 0 & I_n\end{bmatrix} - \begin{bmatrix}I_n & 0\\ 0 & I_n\end{bmatrix}\right. \\
&\left.\otimes \begin{bmatrix}A & 0\\ 0 & A+F\end{bmatrix}\right)^{-1} \left(\begin{bmatrix}b\\ b\end{bmatrix} \otimes \begin{bmatrix}b\\ b\end{bmatrix}\right).
\end{aligned} \tag{12}$$

This approach is followed in satisfying the *second condition* of backward stability of BIRKA, but for TBIRKA it turns out to be more challenging. The reason for this is that the definition of the H_2−norm of $\zeta - \tilde{\zeta}$ used in BIRKA is different from (12), [4] i.e., from Corollary 4.1 of [15] or Theorem 4.5 of [21]

$$\begin{aligned}
&\|\zeta - \tilde{\zeta}\|_{H_2}^2 \\
&= \left(\begin{bmatrix}c & -c\end{bmatrix} \otimes \begin{bmatrix}c & -c\end{bmatrix}\right) \\
&\times \left(-\begin{bmatrix}A & 0\\ 0 & A+F\end{bmatrix} \otimes \begin{bmatrix}I_n & 0\\ 0 & I_n\end{bmatrix} - \begin{bmatrix}I_n & 0\\ 0 & I_n\end{bmatrix}\right. \\
&\left.\otimes \begin{bmatrix}A & 0\\ 0 & A+F\end{bmatrix} - \begin{bmatrix}N & 0\\ 0 & N\end{bmatrix} \otimes \begin{bmatrix}N & 0\\ 0 & N\end{bmatrix}\right)^{-1} \left(\begin{bmatrix}b\\ b\end{bmatrix} \otimes \begin{bmatrix}b\\ b\end{bmatrix}\right).
\end{aligned}$$

From (10) and (11), we know that both ζ^M and $\tilde{\zeta}^M$ are represented by a finite set of transfer functions, respectively. Hence, another way to satisfy (9) in case of TBIRKA, which

[4] Recall, in BIRKA we work with ζ rather than ζ^M.

turns out to be more easier and mathematically equivalent, is to show that the norm of the difference between the respective order transfer functions of (10) and (11) is equal to the norm of the perturbation. That is, instead of (9) we can show that

$$\begin{aligned} \left\| H_1(s_1) - \widetilde{H}_1(s_1) \right\|_{H_2} &\propto \mathcal{O}(\|F\|_2), \\ \left\| H_2(s_1, s_2) - \widetilde{H}_2(s_1, s_2) \right\|_{H_2} &\propto \mathcal{O}(\|F\|_2), \\ &\vdots \\ \left\| H_M(s_1, \ldots, s_M) - \widetilde{H}_M(s_1, \ldots, s_M) \right\|_{H_2} &\propto \mathcal{O}(\|F\|_2), \end{aligned} \tag{13}$$

where $H_k(s_1, \ldots, s_k)$ for $k \in \{1, \ldots, M\}$ is given by (4) and (10b), and $\widetilde{H}_k(s_1, \ldots, s_k)$ for $k \in \{1, \ldots, M\}$ is given by (11b). This way was not possible in BIRKA because there $M \rightarrow \infty$ (see (2)-(4)).

To prove this condition, we first abstract out the term containing the perturbation F from the normed difference between the two corresponding transfer functions (of the original system and the perturbed system) in Lemma 6. Next, in Lemma 7, for $k = 2$, we show that the norm of this term is order of the norm of F. Finally, we generalize the result of Lemma 7 in Lemma 8 (from $k = 2$ to any general k) by using induction.

Note, that in all our subsequent derivations, we assume that all inverses used exist. This is an acceptable assumption because the inverse of matrices arising here are of the form as in [28] and [31] (the papers that discuss stability of IRKA and BIRKA, respectively).

Lemma 6: Let the original bilinear dynamical system be defined as in (10) *and the perturbed bilinear dynamical system be defined as in (11). Then,*

$$\begin{aligned} &\left\| H_k(s_1, \ldots, s_k) - \widetilde{H}_k(s_1, \ldots, s_k) \right\|_{H_2}^2 \\ &\quad \leq \left\| c\mathcal{K}^{-1}(s_k) \right\|_{H_2}^2 \left\| \mathcal{K}^{-1}(s_{k-1}) \right\|_{H_2}^2 \ldots \left\| \mathcal{K}^{-1}(s_1) \right\|_{H_2}^2 \\ &\qquad \times \|U(s_1, \ldots, s_k)\|_{H_\infty}^2 \left\| \mathcal{K}^{-1}(s_1) b \right\|_{H_\infty}^2, \end{aligned}$$

where $\mathcal{K}(s_i) = (s_i I_n - A)$ *for* $i = 1, \ldots, k$, *and*

$$\begin{aligned} U(s_1, \ldots, s_k) = \mathcal{K}(s_1) \ldots \mathcal{K}(s_{k-1}) \\ \times \Bigg(N\mathcal{K}^{-1}(s_{k-1}) \ldots N\mathcal{K}^{-1}(s_2) N \\ - \left(I_n - F\mathcal{K}^{-1}(s_k) \right)^{-1} \\ \times N\mathcal{K}^{-1}(s_{k-1}) \left(I_n - F\mathcal{K}^{-1}(s_{k-1}) \right)^{-1} \\ \ldots N\mathcal{K}^{-1}(s_2) \left(I_n - F\mathcal{K}^{-1}(s_2) \right)^{-1} \\ \times N \left(I_n - \mathcal{K}^{-1}(s_1) F \right)^{-1} \Bigg). \end{aligned} \tag{14}$$

Proof: See Appendix A. □

Lemma 7: Let $\|F\|_2 < 1$, *where F is the perturbation introduced in the A matrix of the input dynamical system. Also, let* $\left\| \mathcal{K}^{-1}(s_i) \right\|_{H_\infty} < 1$ *for* $i = 1$ and 2, *where* $\mathcal{K}(s_i) = (s_i I_n - A)$ *with* I_n *being the identity matrix. Then,*

$$\|U_2\|_{H_\infty} \propto \mathcal{O}(\|F\|_2).$$

where $U_2 = U(s_1, s_2)$ *from* (14).

Proof: See Appendix B. □

Lemma 8: Let $\|F\|_2 < 1$, *where F is the perturbation introduced in the A matrix of the input dynamical system. Also, let* $\left\| \mathcal{K}^{-1}(s_i) \right\|_{H_\infty} < 1$ *for* $i = 1, 2, \ldots, k$, *where* $\mathcal{K}(s_i) = (s_i I_n - A)$ *with* I_n *being the identity matrix. Then,*

$$\|U_k\|_{H_\infty} \propto \mathcal{O}(\|F\|_2),$$

where $U_k = U(s_1, \ldots, s_k)$ *from* (14).

Proof: See Appendix C. □

Theorem 9: If hypotheses of Lemmas 6 and 8 holds, then

$$\left\| H_k(s_1, \ldots, s_k) - \widetilde{H}_k(s_1, \ldots, s_k) \right\|_{H_2}^2 = \mathcal{O}\left(\|F\|_2^2 \right)$$

or TBIRKA satisfies the second condition of backward stability with respect to inexact linear solves.

Proof: Directly follows from combining the results of Lemmas 6 and 8. □

IV. CONCLUSIONS & FUTURE WORK

In this paper, we apply iterative linear solvers during model order reduction (MOR) of bilinear dynamical systems. Since such solvers are inexact, the stability of the underlying MOR algorithm, with respect to these approximation errors, is important. Here, we extend the earlier stability analysis done for BIRKA in [31], to its cheaper variant TBIRKA. Proving that an algorithm is stable, typically requires satisfying two conditions. In TBIRKA, fulfilling the first condition for stability leads to constraints on the iterative linear solver, which are similar to those obtained during BIRKA's stability analysis. The second condition for a stable TBIRKA is satisfied using an approach different than the one used in BIRKA, and is more intuitive.

Our first future direction is to extend our analysis from SISO (Single Input Single Output) to MIMO (Multiple Input Multiple Output) systems. The stability analysis as done for BIRKA earlier and TBIRKA here, all give us sufficiency conditions for a stable underlying MOR algorithm. Hence, second, we plan to derive the necessary conditions for the same. In recent years, there have been a lot of efforts in performing data-driven MOR algorithm (specially using Leowner framework [19]). Our third future direction is to apply this stability analysis to such classes of algorithms as well. Finally, and fourth, our stability analysis can be extended to the cases when instead of a dynamical system, the underlying differential equation is studied [35]–[37].

APPENDIX A

Proof of Lemma 6: Using the definition of H_2-norm (1), we get

$$
\begin{aligned}
&\left\| H_k\left(s_1, \ldots, s_k\right) - \widetilde{H}_k\left(s_1, \ldots, s_k\right) \right\|_{H_2}^2 \\
&= \left(\frac{1}{2\pi}\right)^k \underset{m\to\infty}{Lim} \int_{-m}^{m} \ldots \int_{-m}^{m} \Big\| c\mathcal{K}^{-1}\left(i\omega_k\right) N\mathcal{K}^{-1}\left(i\omega_{k-1}\right) \ldots N\mathcal{K}^{-1}\left(i\omega_1\right) b \\
&\quad - c\left(\mathcal{K}\left(i\omega_k\right) - F\right)^{-1} N\left(\mathcal{K}\left(i\omega_{k-1}\right) - F\right)^{-1} \ldots N\left(\mathcal{K}\left(i\omega_2\right) - F\right)^{-1} N\left(\mathcal{K}\left(i\omega_1\right) - F\right)^{-1} b \Big\|_F^2 d\omega_1 \ldots d\omega_k \\
&= \left(\frac{1}{2\pi}\right)^k \underset{m\to\infty}{Lim} \int_{-m}^{m} \ldots \int_{-m}^{m} \Big\| c\mathcal{K}^{-1}\left(i\omega_k\right) \Big(N\mathcal{K}^{-1}\left(i\omega_{k-1}\right) \ldots N\mathcal{K}^{-1}\left(i\omega_2\right) N \\
&\quad - \left(I_n - F\mathcal{K}^{-1}\left(i\omega_k\right)\right)^{-1} N\mathcal{K}^{-1}\left(i\omega_{k-1}\right) \left(I_n - F\mathcal{K}^{-1}\left(i\omega_{k-1}\right)\right)^{-1} \\
&\quad \ldots N\mathcal{K}^{-1}\left(i\omega_2\right) \left(I_n - F\mathcal{K}^{-1}\left(i\omega_2\right)\right)^{-1} N \left(I_n - \mathcal{K}^{-1}\left(i\omega_1\right) F\right)^{-1} \Big) \mathcal{K}^{-1}\left(i\omega_1\right) b \Big\|_F^2 d\omega_1 \ldots d\omega_k \\
&= \left(\frac{1}{2\pi}\right)^k \underset{m\to\infty}{Lim} \int_{-m}^{m} \ldots \int_{-m}^{m} \Big\| c\mathcal{K}^{-1}\left(i\omega_k\right) \mathcal{K}^{-1}\left(i\omega_{k-1}\right) \ldots \mathcal{K}^{-1}\left(i\omega_1\right) \\
&\quad \times \mathcal{K}\left(i\omega_1\right) \ldots \mathcal{K}\left(i\omega_{k-1}\right) \Big(N\mathcal{K}^{-1}\left(i\omega_{k-1}\right) \ldots N\mathcal{K}^{-1}\left(i\omega_2\right) N \\
&\quad - \left(I_n - F\mathcal{K}^{-1}\left(i\omega_k\right)\right)^{-1} N\mathcal{K}^{-1}\left(i\omega_{k-1}\right) \left(I_n - F\mathcal{K}^{-1}\left(i\omega_{k-1}\right)\right)^{-1} \\
&\quad \ldots N\mathcal{K}^{-1}\left(i\omega_2\right) \left(I_n - F\mathcal{K}^{-1}\left(i\omega_2\right)\right)^{-1} N \left(I_n - \mathcal{K}^{-1}\left(i\omega_1\right) F\right)^{-1} \Big) \mathcal{K}^{-1}\left(i\omega_1\right) b \Big\|_F^2 d\omega_1 \ldots d\omega_k.
\end{aligned}
$$

Using $U\left(s_1, \ldots, s_k\right)$ given by (14), $\|XYZ\|_F \leq \|X\|_F \|YZ\|_F$, $\|YZ\|_F \leq \|Y\|_F \|Z\|_2$, and comparison integral inequality[5] [38] for any matrices X, Y, and Z, in the above equation, we have

$$
\begin{aligned}
\left\| H_k\left(s_1, \ldots, s_k\right) - \widetilde{H}_k\left(s_1, \ldots, s_k\right) \right\|_{H_2}^2 &\leq \left(\frac{1}{2\pi}\right)^k \underset{m\to\infty}{Lim} \int_{-m}^{m} \ldots \int_{-m}^{m} \left\| c\mathcal{K}^{-1}\left(i\omega_k\right) \right\|_F^2 \left\| \mathcal{K}^{-1}\left(i\omega_{k-1}\right) \right\|_F^2 \\
&\quad \ldots \left\| \mathcal{K}^{-1}\left(i\omega_1\right) \right\|_F^2 \left\| U\left(i\omega_1, \ldots, i\omega_k\right) \right\|_2^2 \left\| \mathcal{K}^{-1}\left(i\omega_1\right) b \right\|_2^2 d\omega_1 \ldots d\omega_k. \qquad (15)
\end{aligned}
$$

From the mean value theorem of integration [38] we know

$$
\begin{aligned}
\int_{-m}^{m} \int_{-m}^{m} f\left(i\omega_2\right) g\left(i\omega_1, i\omega_2\right) h\left(i\omega_1\right) d\omega_1 d\omega_2 &= \int_{-m}^{m} f\left(i\omega_2\right) \left(\int_{-m}^{m} g\left(i\omega_1, i\omega_2\right) h\left(i\omega_1\right) d\omega_1 \right) d\omega_2 \\
&\leq \int_{-m}^{m} f\left(i\omega_2\right) \left(\underset{\mathrm{c}\in\mathbb{R}}{max}\left(g(\mathrm{ic}, i\omega_2)\right) \int_{-m}^{m} h\left(i\omega_1\right) d\omega_1 \right) d\omega_2 \\
&\leq \underset{\mathrm{c,d}\in\mathbb{R}}{max}\left(g(\mathrm{ic}, \mathrm{id})\right) \int_{-m}^{m} f\left(i\omega_2\right) d\omega_2 \int_{-m}^{m} h\left(i\omega_1\right) d\omega_1.
\end{aligned}
$$

Using this property in (15) we get[6]

$$
\begin{aligned}
\left\| H_k\left(s_1, \ldots, s_k\right) - \widetilde{H}_k\left(s_1, \ldots, s_k\right) \right\|_{H_2}^2 &\leq \left(\frac{1}{2\pi}\right)^k \underset{m\to\infty}{Lim} \int_{-m}^{m} \ldots \int_{-m}^{m} \left\| c\mathcal{K}^{-1}\left(i\omega_k\right) \right\|_F^2 \\
&\quad \times \left\| \mathcal{K}^{-1}\left(i\omega_{k-1}\right) \right\|_F^2 \ldots \left\| \mathcal{K}^{-1}\left(i\omega_1\right) \right\|_F^2 d\omega_1 \ldots d\omega_k \\
&\quad \times \underset{\omega_1, \ldots, \omega_k\in\mathbb{R}}{max} \left\| U\left(i\omega_1, \ldots, i\omega_k\right) \right\|_2^2 \underset{\omega_1\in\mathbb{R}}{max} \left\| \mathcal{K}^{-1}\left(i\omega_1\right) b \right\|_2^2 \\
&\leq \left\| c\mathcal{K}^{-1}\left(s_k\right) \right\|_{H_2}^2 \left\| \mathcal{K}^{-1}\left(s_{k-1}\right) \right\|_{H_2}^2 \ldots \left\| \mathcal{K}^{-1}\left(s_1\right) \right\|_{H_2}^2 \\
&\quad \times \left\| U\left(s_1, \ldots, s_k\right) \right\|_{H_\infty}^2 \left\| \mathcal{K}^{-1}\left(s_1\right) b \right\|_{H_\infty}^2.
\end{aligned}
$$

□

[5] This inequality says if $f(x)$ and $g(x)$ are integrable over $[a, b]$ and $f(x) \leq g(x)$, then $\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx$. Note that although we have improper integrals here, this inequality still holds because of the earlier assumption that such integrals give a finite value.

[6] As mentioned in Footnote 5, the improper integrals here do not affect application of this mean value theorem because all such integrals are assumed to give a finite value.

APPENDIX B

Proof of Lemma 7: Substituting $k = 2$ in (14), we get

$$U_2 = \mathcal{K}(s_1)\left(N - \left(I_n - F\mathcal{K}^{-1}(s_2)\right)^{-1} N \left(I_n - \mathcal{K}^{-1}(s_1)F\right)^{-1}\right).$$

If $\left\|F\mathcal{K}^{-1}(s_2)\right\|_{H_\infty} < 1$ and $\left\|\mathcal{K}^{-1}(s_1)F\right\|_{H_\infty} < 1$, then by the Neumann series, we get[7]

$$\begin{aligned}
U_2 &= \mathcal{K}(s_1)\left(N - \left(I_n + F\mathcal{K}^{-1}(s_2) + \left(F\mathcal{K}^{-1}(s_2)\right)^2 + \cdots\right) N \left(I_n + \mathcal{K}^{-1}(s_1)F + \left(\mathcal{K}^{-1}(s_1)F\right)^2 + \cdots\right)\right) \\
&= \mathcal{K}(s_1)\Big(N - N - N\mathcal{K}^{-1}(s_1)F\left(I_n + \mathcal{K}^{-1}(s_1)F + \cdots\right) - F\mathcal{K}^{-1}(s_2)\left(I_n + F\mathcal{K}^{-1}(s_2) + \cdots\right)N \\
&\quad \times \left(I_n + \mathcal{K}^{-1}(s_1)F + \left(\mathcal{K}^{-1}(s_1)F\right)^2 + \cdots\right)\Big) \\
&= \mathcal{K}(s_1)\left(-N\mathcal{K}^{-1}(s_1)F\left(I_n - \mathcal{K}^{-1}(s_1)F\right)^{-1} - F\mathcal{K}^{-1}(s_2)\left(I_n - F\mathcal{K}^{-1}(s_2)\right)^{-1} N\left(I_n - \mathcal{K}^{-1}(s_1)F\right)^{-1}\right) \\
&= \mathcal{K}(s_1)\left(-N\mathcal{K}^{-1}(s_1)F - F\mathcal{K}^{-1}(s_2)\left(I_n - F\mathcal{K}^{-1}(s_2)\right)^{-1} N\right)\left(I_n - \mathcal{K}^{-1}(s_1)F\right)^{-1}.
\end{aligned}$$

Taking H_∞−norm on both sides, and using $\|XY\|_2 \leq \|X\|_2\|Y\|_2$ and $\|X+Y\|_2 \leq \|X\|_2 + \|Y\|_2$, for any two matrices X and Y, we get

$$\begin{aligned}
\|U_2\|_{H_\infty} &\leq \max_{\omega_1,\omega_2 \in \mathbb{R}}\Bigg(\|\mathcal{K}(i\omega_1)\|_2\left(\|N\|_2\left\|\mathcal{K}^{-1}(i\omega_1)\right\|_2\|F\|_2 + \|F\|_2\left\|\mathcal{K}^{-1}(i\omega_2)\right\|_2\left\|\left(I_n - F\mathcal{K}^{-1}(i\omega_2)\right)^{-1}\right\|_2\|N\|_2\right) \\
&\quad \times \left\|\left(I_n - \mathcal{K}^{-1}(i\omega_1)F\right)^{-1}\right\|_2\Bigg) \\
&\leq \|\mathcal{K}(s_1)\|_{H_\infty}\|N\|_2\|F\|_2\left(\left\|\mathcal{K}^{-1}(s_1)\right\|_{H_\infty} + \left\|\mathcal{K}^{-1}(s_2)\right\|_{H_\infty}\max_{\omega_2 \in \mathbb{R}}\left\|\left(I_n - F\mathcal{K}^{-1}(i\omega_2)\right)^{-1}\right\|_2\right) \\
&\quad \times \max_{\omega_1 \in \mathbb{R}}\left\|\left(I_n - \mathcal{K}^{-1}(i\omega_1)F\right)^{-1}\right\|_2.
\end{aligned} \tag{16}$$

Technically by definition of the H_∞−norm and how $\mathcal{K}(s)$ is defined in our hypotheses, $\|\mathcal{K}(s_1)\|_{H_\infty} = \|\mathcal{K}(s_2)\|_{H_\infty} = \|\mathcal{K}(s)\|_{H_\infty}$, however, for sake of exposition, we keep them separate. Similarly for the H_∞−norm of inverses of $\mathcal{K}(s_1)$ and $\mathcal{K}(s_2)$.

To abstract $\|F\|_2$ out from the above inequality, let us look at $\max_{\omega_2 \in \mathbb{R}}\left\|\left(I_n - F\mathcal{K}^{-1}(i\omega_2)\right)^{-1}\right\|_2$ separately. Recall, while applying Neumann series we assumed that $\left\|F\mathcal{K}^{-1}(s_2)\right\|_{H_\infty} < 1$ or $\max_{\omega_2 \in \mathbb{R}}\left\|F\mathcal{K}^{-1}(i\omega_2)\right\|_2 < 1$. Since the maximum of such a norm is less than one, we have for all $\omega_2 \in \mathbb{R}$, $\left\|F\mathcal{K}^{-1}(i\omega_2)\right\|_2 < 1$. Using this along with Lemma 2.3.3 from [40][8] in the above expression, we get

$$\begin{aligned}
\max_{\omega_2 \in \mathbb{R}}\left\|\left(I_n - F\mathcal{K}^{-1}(i\omega_2)\right)^{-1}\right\|_2 &\leq \max_{\omega_2 \in \mathbb{R}}\frac{1}{1 - \left\|F\mathcal{K}^{-1}(i\omega_2)\right\|_2} \\
&\leq \frac{1}{1 - \max_{\omega_2 \in \mathbb{R}}\left\|F\mathcal{K}^{-1}(i\omega_2)\right\|_2} \\
&\leq \frac{1}{1 - \left\|F\mathcal{K}^{-1}(s_2)\right\|_{H_\infty}}.
\end{aligned} \tag{17}$$

If we assume $\|F\|_2 < 1$ and $\left\|\mathcal{K}^{-1}(s_2)\right\|_{H_\infty} < 1$ (as in our hypotheses), then using earlier used matrix norm properties, we get

$$\begin{aligned}
\left\|F\mathcal{K}^{-1}(s_2)\right\|_{H_\infty} = \max_{\omega_2 \in \mathbb{R}}\left\|F\mathcal{K}^{-1}(i\omega_2)\right\|_2 &\leq \|F\|_2 \max_{\omega_2 \in \mathbb{R}}\left\|\mathcal{K}^{-1}(i\omega_2)\right\|_2 \\
&\leq \|F\|_2\left\|\mathcal{K}^{-1}(s_2)\right\|_{H_\infty} \\
&\leq 1,
\end{aligned}$$

[7] From [39, page 527], we know $(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$ when $\|A\| < 1$ for any matrix norm. Here, for the first inequality we have $\left\|F\mathcal{K}^{-1}(s_2)\right\|_{H_\infty} < 1$ or $\max_{\omega_2 \in \mathbb{R}}\left\|F\mathcal{K}^{-1}(i\omega_2)\right\|_2 < 1$, and hence, the applicable matrix norm is 2−norm. Similarly for the second inequality.

[8] If $F \in \mathbb{R}^{n \times n}$ and $\|F\|_p < 1$, then $I - F$ is nonsingular and $(I - F)^{-1} = \sum_{k=0}^{\infty} F^k$ with $\left\|(I - F)^{-1}\right\|_p \leq \frac{1}{1 - \|F\|_p}$.

as assumed for applying Neumann series earlier as well as Lemma 2.3.3 from [40] above. Thus, no extra assumptions beyond those in hypotheses are needed. Further, we also get

$$1 - \|F\|_2 \left\|\mathcal{K}^{-1}(s_2)\right\|_{H_\infty} \leq 1 - \left\|F\mathcal{K}^{-1}(s_2)\right\|_{H_\infty} \quad \text{or} \quad \frac{1}{1 - \left\|F\mathcal{K}^{-1}(s_2)\right\|_{H_\infty}} \leq \frac{1}{1 - \|F\|_2 \left\|\mathcal{K}^{-1}(s_2)\right\|_{H_\infty}}. \tag{18}$$

Similarly, by assuming $\|F\|_2 < 1$ and $\left\|\mathcal{K}^{-1}(s_1)\right\|_{H_\infty} < 1$ (as in our hypotheses), we can bound the last term of (16) as follows:

$$\max_{\omega_1 \in \mathbb{R}} \left\| \left(I_n - \mathcal{K}^{-1}(i\omega_1) F\right)^{-1} \right\|_2 \leq \frac{1}{1 - \left\|\mathcal{K}^{-1}(s_1) F\right\|_{H_\infty}} \quad \text{and} \tag{19}$$

$$\frac{1}{1 - \left\|\mathcal{K}^{-1}(s_1) F\right\|_{H_\infty}} \leq \frac{1}{1 - \left\|\mathcal{K}^{-1}(s_1)\right\|_{H_\infty} \|F\|_2}. \tag{20}$$

Substituting (17)-(18) and (19)-(20) in (16), we get

$$\|U_2\|_{H_\infty} \leq \|\mathcal{K}(s_1)\|_{H_\infty} \|N\|_2 \|F\|_2 \left[\left\|\mathcal{K}^{-1}(s_1)\right\|_{H_\infty} + \frac{\left\|\mathcal{K}^{-1}(s_2)\right\|_{H_\infty}}{1 - \|F\|_2 \left\|\mathcal{K}^{-1}(s_2)\right\|_{H_\infty}} \right] \left(\frac{1}{1 - \left\|\mathcal{K}^{-1}(s_1)\right\|_{H_\infty} \|F\|_2} \right).$$

From the above inequality it is clear that if $\|F\|_2 \left\|\mathcal{K}^{-1}(s_2)\right\|_{H_\infty} < 1$ and $\left\|\mathcal{K}^{-1}(s_1)\right\|_{H_\infty} \|F\|_2 < 1$, which are true from our hypotheses, then

$$\|U_2\|_{H_\infty} = \mathcal{O}(\|F\|_2).$$

□

APPENDIX C

Proof of Lemma 8: We prove this by mathematical induction.

Base Case :

$k = 1$ is the linear system case already proved in [28] (see Theorem 4.3 of [28]). $k = 2$ has been proved above (Lemma 7).

Induction Hypothesis :

From (14), we know for $k = L$

$$\begin{aligned} U_L = \mathcal{K}(s_1) \ldots \mathcal{K}(s_{L-1}) \Big(& N\mathcal{K}^{-1}(s_{L-1}) \ldots N\mathcal{K}^{-1}(s_2) N - \left(I_n - F\mathcal{K}^{-1}(s_L)\right)^{-1} \\ & \times N\mathcal{K}^{-1}(s_{L-1}) \left(I_n - F\mathcal{K}^{-1}(s_{L-1})\right)^{-1} \ldots N\mathcal{K}^{-1}(s_2) \left(I_n - F\mathcal{K}^{-1}(s_2)\right)^{-1} N \left(I_n - \mathcal{K}^{-1}(s_1) F\right)^{-1} \Big). \end{aligned} \tag{21}$$

Let $\|U_L\|_{H_\infty} = \mathcal{O}(\|F\|_2)$.

Induction Step :

We show the above for $k = L + 1$. Again, from (14), we know

$$\begin{aligned} U_{L+1} = \mathcal{K}(s_1) \ldots \mathcal{K}(s_L) \Big(& N\mathcal{K}^{-1}(s_L) \ldots N\mathcal{K}^{-1}(s_2) N - \left(I_n - F\mathcal{K}^{-1}(s_{L+1})\right)^{-1} \\ & \times N\mathcal{K}^{-1}(s_L) \left(I_n - F\mathcal{K}^{-1}(s_L)\right)^{-1} \ldots N\mathcal{K}^{-1}(s_2) \left(I_n - F\mathcal{K}^{-1}(s_2)\right)^{-1} N \left(I_n - \mathcal{K}^{-1}(s_1) F\right)^{-1} \Big). \end{aligned}$$

We first write U_{L+1} in terms of U_L. Using our hypotheses, we have $\left\|F\mathcal{K}^{-1}(s_{L+1})\right\|_{H_\infty} < \|F\|_2 \left\|\mathcal{K}^{-1}(s_{L+1})\right\|_{H_\infty} < 1$, and hence, applying Neumann series above, we get

$$\begin{aligned} U_{L+1} = \mathcal{K}(s_1) \ldots \mathcal{K}(s_L) \Big(& N\mathcal{K}^{-1}(s_L) \ldots N\mathcal{K}^{-1}(s_2) N - \left(I_n + F\mathcal{K}^{-1}(s_{L+1}) + \left(F\mathcal{K}^{-1}(s_{L+1})\right)^2 + \cdots\right) \\ & \times N\mathcal{K}^{-1}(s_L) \left(I_n - F\mathcal{K}^{-1}(s_L)\right)^{-1} \ldots N\mathcal{K}^{-1}(s_2) \left(I_n - F\mathcal{K}^{-1}(s_2)\right)^{-1} N \left(I_n - \mathcal{K}^{-1}(s_1) F\right)^{-1} \Big) \\ = \mathcal{K}(s_1) \ldots \mathcal{K}(s_L) \Big(& N\mathcal{K}^{-1}(s_L) \ldots N\mathcal{K}^{-1}(s_2) N \\ & - N\mathcal{K}^{-1}(s_L) \left(I_n - F\mathcal{K}^{-1}(s_L)\right)^{-1} \ldots N\mathcal{K}^{-1}(s_2) \left(I_n - F\mathcal{K}^{-1}(s_2)\right)^{-1} N \left(I_n - \mathcal{K}^{-1}(s_1) F\right)^{-1} \end{aligned}$$

$$-F\mathcal{K}^{-1}\left(s_{L+1}\right)\left(I_n-F\mathcal{K}^{-1}\left(s_{L+1}\right)\right)^{-1}N\mathcal{K}^{-1}\left(s_L\right)\left(I_n-F\mathcal{K}^{-1}\left(s_L\right)\right)^{-1}\ldots N\mathcal{K}^{-1}\left(s_2\right)\left(I_n-F\mathcal{K}^{-1}\left(s_2\right)\right)^{-1}$$
$$\times N\left(I_n-\mathcal{K}^{-1}\left(s_1\right)F\right)^{-1}\Big).$$

In the above equation, taking $N\mathcal{K}^{-1}\left(s_L\right)$ common from the first two terms of the bigger bracket, we have

$$\begin{aligned}&=\mathcal{K}\left(s_1\right)\ldots\mathcal{K}\left(s_L\right)\Bigg(N\mathcal{K}^{-1}\left(s_L\right)\Bigg(N\mathcal{K}^{-1}\left(s_{L-1}\right)\ldots N\mathcal{K}^{-1}\left(s_2\right)N-\left(I_n-F\mathcal{K}^{-1}\left(s_L\right)\right)^{-1}\\&\quad\times N\mathcal{K}^{-1}\left(s_{L-1}\right)\left(I_n-F\mathcal{K}^{-1}\left(s_{L-1}\right)\right)^{-1}\ldots N\mathcal{K}^{-1}\left(s_2\right)\left(I_n-F\mathcal{K}^{-1}\left(s_2\right)\right)^{-1}N\left(I_n-\mathcal{K}^{-1}\left(s_1\right)F\right)^{-1}\Bigg)\\&\quad-F\mathcal{K}^{-1}\left(s_{L+1}\right)\left(I_n-F\mathcal{K}^{-1}\left(s_{L+1}\right)\right)^{-1}N\mathcal{K}^{-1}\left(s_L\right)\left(I_n-F\mathcal{K}^{-1}\left(s_L\right)\right)^{-1}\ldots N\mathcal{K}^{-1}\left(s_2\right)\left(I_n-F\mathcal{K}^{-1}\left(s_2\right)\right)^{-1}\\&\quad\times N\left(I_n-\mathcal{K}^{-1}\left(s_1\right)F\right)^{-1}\Bigg).\end{aligned}\tag{22}$$

Now we look at expression of U_L. Multiplying $\mathcal{K}^{-1}\left(s_{L-1}\right)\ldots\mathcal{K}^{-1}\left(s_1\right)$ on both the sides of (21) from left, we get

$$\begin{aligned}\mathcal{K}^{-1}\left(s_{L-1}\right)\ldots\mathcal{K}^{-1}\left(s_1\right)U_L&=\Bigg(N\mathcal{K}^{-1}\left(s_{L-1}\right)\ldots N\mathcal{K}^{-1}\left(s_2\right)N-\left(I_n-F\mathcal{K}^{-1}\left(s_L\right)\right)^{-1}\\&\quad\times N\mathcal{K}^{-1}\left(s_{L-1}\right)\left(I_n-F\mathcal{K}^{-1}\left(s_{L-1}\right)\right)^{-1}\ldots N\mathcal{K}^{-1}\left(s_2\right)\left(I_n-F\mathcal{K}^{-1}\left(s_2\right)\right)^{-1}N\left(I_n-\mathcal{K}^{-1}\left(s_1\right)F\right)^{-1}\Bigg).\end{aligned}\tag{23}$$

Substituting (23) in (22), we get

$$\begin{aligned}U_{L+1}&=\mathcal{K}\left(s_1\right)\ldots\mathcal{K}\left(s_L\right)\Bigg(N\mathcal{K}^{-1}\left(s_L\right)\left(\mathcal{K}^{-1}\left(s_{L-1}\right)\ldots\mathcal{K}^{-1}\left(s_1\right)U_L\right)-F\mathcal{K}^{-1}\left(s_{L+1}\right)\left(I_n-F\mathcal{K}^{-1}\left(s_{L+1}\right)\right)^{-1}\\&\quad\times N\mathcal{K}^{-1}\left(s_L\right)\left(I_n-F\mathcal{K}^{-1}\left(s_L\right)\right)^{-1}\ldots N\mathcal{K}^{-1}\left(s_2\right)\left(I_n-F\mathcal{K}^{-1}\left(s_2\right)\right)^{-1}N\left(I_n-\mathcal{K}^{-1}\left(s_1\right)F\right)^{-1}\Bigg).\end{aligned}$$

Taking $H_\infty-$norm on both sides, and as earlier, using the norm inequality properties in the above equation, we get

$$\begin{aligned}\|U_{L+1}\|_{H_\infty}\leq\max_{\omega_1,\ \ldots,\ \omega_{L+1}\in\mathbb{R}}\Bigg[&\|\mathcal{K}\left(i\omega_1\right)\|_2\ldots\|\mathcal{K}\left(i\omega_L\right)\|_2\Bigg(\|N\|_2\left\|\mathcal{K}^{-1}\left(i\omega_L\right)\right\|_2\ldots\left\|\mathcal{K}^{-1}\left(i\omega_1\right)\right\|_2\\&\times\|U\left(i\omega_1,\ \ldots,\ i\omega_L\right)\|_2+\|F\|_2\left\|\mathcal{K}^{-1}\left(i\omega_{L+1}\right)\right\|_2\left\|\left(I_n-F\mathcal{K}^{-1}\left(i\omega_{L+1}\right)\right)^{-1}\right\|_2\\&\times\|N\|_2\left\|\mathcal{K}^{-1}\left(i\omega_L\right)\right\|_2\left\|\left(I_n-F\mathcal{K}^{-1}\left(i\omega_L\right)\right)^{-1}\right\|_2\ldots\|N\|_2\left\|\mathcal{K}^{-1}\left(i\omega_2\right)\right\|_2\left\|\left(I_n-F\mathcal{K}^{-1}\left(i\omega_2\right)\right)^{-1}\right\|_2\\&\times\|N\|_2\left\|\left(I_n-\mathcal{K}^{-1}\left(i\omega_1\right)F\right)^{-1}\right\|_2\Bigg)\Bigg].\end{aligned}$$

Similar to (17) and (18), here also, using Lemma 2.3.3 from [40] we get

$$\begin{aligned}\|U_{L+1}\|_{H_\infty}&\leq\|\mathcal{K}\left(s_1\right)\|_{H_\infty}\ldots\|\mathcal{K}\left(s_L\right)\|_{H_\infty}\|N\|_2\left\|\mathcal{K}^{-1}\left(s_L\right)\right\|_{H_\infty}\ldots\left\|\mathcal{K}^{-1}\left(s_2\right)\right\|_{H_\infty}\Bigg[\left\|\mathcal{K}^{-1}\left(s_1\right)\right\|_{H_\infty}\|U_L\|_{H_\infty}\\&\quad+\frac{\|N\|_2^{L-1}\left\|\mathcal{K}^{-1}\left(s_{L+1}\right)\right\|_{H_\infty}}{\left(1-\|F\|_2\left\|\mathcal{K}^{-1}\left(s_{L+1}\right)\right\|_{H_\infty}\right)\ldots\left(1-\|F\|_2\left\|\mathcal{K}^{-1}\left(s_2\right)\right\|_{H_\infty}\right)}\cdot\frac{\|F\|_2}{1-\left\|\mathcal{K}^{-1}\left(s_1\right)\right\|_{H_\infty}\|F\|_2}\Bigg].\end{aligned}$$

From induction hypothesis we know $\|U_L\|_{H_\infty}\propto\mathcal{O}\left(\|F\|_2\right)$. Using this we get

$$\|U_{L+1}\|_{H_\infty}\propto\mathcal{O}\left(\|F\|_2\right).$$

□

REFERENCES

[1] H. Kaper and H. Engler, *Mathematics and Climate*. Philadelphia, PA, USA: SIAM, 2013.

[2] J. D. Meiss, *Differential Dynamical Systems, Revised Edition*. Philadelphia, PA, USA: SIAM, 2017.

[3] J. T. Stuart, "Taylor-vortex flow: A dynamical system," *SIAM Review*, vol. 28, no. 3, pp. 315–342, Sep. 1986.

[4] P. D'Alessandro, A. Isidori, and A. Ruberti, "Realization and structure theory of bilinear dynamical systems," *SIAM J. Control Opt.*, vol. 12, no. 3, pp. 517–535, Aug. 1974.

[5] R. J. Wilson, *Nonlinear System Theory: The Volterra / Wiener Approach*. Baltimore, MD, USA: JHU Press, 1981.

[6] F. Carravetta, "Global exact quadratization of continuous-time nonlinear control systems," *SIAM J. Control Opt.*, vol. 53, no. 1, pp. 235–261, Jan. 2015.

[7] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*. Philadelphia, PA, USA: SIAM, 2005.

[8] P. Benner, D. C. Sorensen, and V. Mehrmann, *Dimension Reduction of Large-Scale Systems* (Lecture Notes in Computational Science and Engineering). Berlin, Germany: Springer, 2005.

[9] E. J. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, Dept. Elect. Eng., Univ. Illinois at Urbana-Champaign, Urbana, IL, USA, 1997.

[10] Z. Bai, "Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems," *Appl. Numer. Math.*, vol. 43, nos. 1–2, pp. 9–44, Oct. 2002.

[11] K. Willcox and J. Peraire, "Balanced model reduction via the proper orthogonal decomposition," *AIAA J.*, vol. 40, no. 11, pp. 2323–2330, 2002.

[12] S. Gugercin, A. C. Antoulas, and C. A. Beattie, "$\mathcal{H}_2$ model reduction for large-scale linear dynamical systems," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 2, pp. 609–638, 2008.

[13] A. Bunse-Gerstner, D. Kubalińska, G. Vossen, and D. Wilczek, "h_2-norm optimal model reduction for large scale discrete dynamical MIMO systems," *J. Comput. Appl. Math.*, vol. 233, no. 5, pp. 1202–1216, Jan. 2010.

[14] Z. Bai and D. Skoogh, "A projection method for model reduction of bilinear dynamical systems," *Linear Algebra Appl.*, vol. 415, nos. 2–3, pp. 406–425, Jun. 2006.

[15] P. Benner and T. Breiten, "Interpolation-based H_2-model reduction of bilinear control systems," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 3, pp. 859–885, Aug. 2012.

[16] T. Bonin, H. Faßbender, A. Soppa, and M. Zaeh, "A fully adaptive rational global Arnoldi method for the model-order reduction of second-order MIMO systems with proportional damping," *Math. Comput. Simul.*, vol. 122, pp. 1–19, Apr. 2016.

[17] M. I. Ahmad, U. Baur, and P. Benner, "Implicit Volterra series interpolation for model reduction of bilinear systems," *J. Comput. Appl. Math.*, vol. 316, pp. 15–28, May 2017.

[18] P. Benner and T. Damm, "Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems," *SIAM J. Control Optim.*, vol. 49, no. 2, pp. 686–711, 2011.

[19] A. C. Antoulas, I. V. Gosea, and A. C. Ionita, "Model reduction of bilinear systems in the Loewner framework," *SIAM J. Sci. Comput.*, vol. 38, no. 5, pp. B889–B916, Oct. 2016.

[20] K.-L. Xu, Y.-L. Jiang, and Z.-X. Yang, "H_2 optimal model order reduction by two-sided technique on Grassmann manifold via the cross-gramian of bilinear systems," *Int. J. Control*, vol. 90, no. 3, pp. 616–626, Mar. 2017.

[21] G. M. Flagg, "Interpolation methods for the model reduction of bilinear systems," Ph.D. dissertation, Dept. Math., Virginia Polytechnic Inst. State Univ., Blacksburg, VA, USA, 2012.

[22] G. M. Flagg and S. Gugercin, "Multipoint Volterra series interpolation and H_2 optimal model reduction of bilinear systems," *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 2, pp. 549–579, May 2015.

[23] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Philadelphia, PA, USA: SIAM, 2003.

[24] S. Catalán, J. R. Herrero, E. S. Quintana-Ortí, R. Rodríguez-Sánchez, and R. Van De Geijn, "A case for malleable thread-level linear algebra libraries: The LU factorization with partial pivoting," *IEEE Access*, vol. 7, pp. 17617–17633, 2019.

[25] H.-L. Shen, S.-Y. Li, and X.-H. Shao, "The NMHSS iterative method for the standard Lyapunov equation," *IEEE Access*, vol. 7, pp. 13200–13205, 2019.

[26] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*. Philadelphia, PA, USA: SIAM, 1997.

[27] J. W. Demmel, *Applied Numerical Linear Algebra*. Philadelphia, PA, USA: SIAM, 1997.

[28] C. Beattie, S. Gugercin, and S. Wyatt, "Inexact solves in interpolatory model reduction," *Linear Algebra Appl.*, vol. 436, no. 8, pp. 2916–2943, Apr. 2012.

[29] N. P. Singh and K. Ahuja, "Stability analysis of inexact solves in moment matching based model reduction," 2018, *arXiv:1803.09283*. [Online]. Available: https://arxiv.org/abs/1803.09283

[30] D. Lu, Y. Su, and Z. Bai, "Stability analysis of the two-level orthogonal Arnoldi procedure," *SIAM J. Matrix Anal. Appl.*, vol. 37, no. 1, pp. 195–214, Feb. 2016.

[31] R. Choudhary and K. Ahuja, "Stability analysis of bilinear iterative rational Krylov algorithm," *Linear Algebra Appl.*, vol. 538, pp. 56–88, Feb. 2018.

[32] M. L. Parks, E. de Sturler, G. Mackey, D. D. Johnson, and S. Maiti, "Recycling Krylov subspaces for sequences of linear systems," *SIAM J. Sci. Comput.*, vol. 28, no. 5, pp. 1651–1674, Oct. 2006.

[33] K. Ahuja, "Recycling Krylov subspaces and preconditioners," Ph.D. dissertation, Dept. Math., Virginia Polytechnic Institute State Univ., Blacksburg, VA, USA, 2011.

[34] K. Ahuja, E. de Sturler, S. Gugercin, and E. R. Chang, "Recycling BiCG with an application to model reduction," *SIAM J. Sci. Comput.*, vol. 34, no. 4, pp. A1925–A1949, Jul. 2012.

[35] T. Han and Y. Han, "Numerical solution for super large scale systems," *IEEE Access*, vol. 1, pp. 537–544, 2013.

[36] M. Saqib, S. Hasnain, and D. S. Mashat, "Highly efficient computational methods for two dimensional coupled nonlinear unsteady convection-diffusion problems," *IEEE Access*, vol. 5, pp. 7139–7148, 2017.

[37] Z. Ren, Z. Zhao, Z. Wu, and T. Chen, "Dynamic optimal control of a one-dimensional magnetohydrodynamic system with bilinear actuation," *IEEE Access*, vol. 6, pp. 24464–24474, 2018.

[38] A. Jeffrey, *Handbook of Mathematical Formulas and Integrals*, 3rd ed. Cambridge, MA, USA: Academic, 2003.

[39] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA: SIAM, 2000.

[40] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: JHU Press, 2012.

RAJENDRA CHOUDHARY received the M.Tech. degree in robotics from the Indian Institute of Information Technology, Allahabad, India. He is currently pursuing the Ph.D. degree with IIT Indore.

His thesis is focused on stability analysis of inexact linear solves applied to different model reduction algorithms. His research interests are in the intersection of computer science and mathematics, especially numerical linear algebra, theory of computation, algorithms, optimization, dynamical systems, and group theory.

KAPIL AHUJA received the bachelor's degree from IIT (BHU), India, and the double master's degrees and the Ph.D. degree from Virginia Tech, Blacksburg, VA, USA.

He was also a Postdoctoral Research Fellow with the Max Planck Institute, Germany. He is currently an Associate Professor in computer science and engineering with IIT Indore. He focuses on applying mathematics and computation to solve science and engineering problems. He has a varied background in computer science, mathematics, and mechanical engineering. Specifically, his research interests include numerical linear algebra, numerical analysis, optimization, computational intelligence, big data, and social cloud.