Generalized tangential interpolation for model reduction of discrete-time MIMO bilinear systems

Peter Benner   Tobias Breiten
Max Planck Institute for Dynamics of Complex Technical Systems
Sandtorstr. 1
D-39106 Magdeburg, Germany,
Phone: +49 391 6110 - 474,
Fax: +49 391 6110 - 453,
benner@mpi-magdeburg.mpg.de,
breiten@mpi-magdeburg.mpg.de

Tobias Damm
Fachbereich Mathematik
Technische Universität Kaiserslautern
Erwin Schrödinger Straße
D-67663 Kaiserslautern, Germany
Phone: +49 631 205 - 4489,
Fax: +49 631 205 - 4986,
damm@mathematik.uni-kl.de

Abstract

In this paper, we discuss a model order reduction method for multiple-input and multiple-output discrete-time bilinear control systems. Similar to the continuous-time case, we will show that a system can be characterized by a series of generalized transfer functions. This will be achieved by a multivariate Z-transform of kernels corresponding to an explicit solution formula for discrete-time systems. We will further address the problem of generalized tangential interpolation which naturally comes along with this approach. We will introduce a reasonable generalization of the linear $\mathcal{H}_2$-norm. Based on this concept, we discuss the choice of interpolation points. Furthermore, an efficient discretization of continuous-time systems is provided. The performance of the proposed method is evaluated in some numerical examples and compared with the method of balanced truncation for bilinear systems.

Keywords: model order reduction, bilinear systems, tangential interpolation, discrete systems

1   Introduction

Nowadays many technical and industrial processes require accurate and systematic analysis and simulation with the help of mathematical models. However, the accurate modelling frequently leads to very large-scale control systems which prevent efficient numerical treatment. Therefore, model order reduction is concerned with the construction of a system of much smaller state dimension that still faithfully reproduces the original dynamics or transfer behaviour. While for linear systems there are well-established techniques to construct reduced order models satisfying certain error bounds or interpolation properties, in the presence of nonlinearities much less is known.

Here, we want to consider discrete-time bilinear systems of the form

$$\Sigma: \quad \begin{cases} x(k+1) = Ax(k) + N\left(I_m \otimes x(k)\right)u(k) + Bu(k), \\ \quad\, y(k) = Cx(k), \quad x(0) = x_0, \end{cases} \tag{1}$$

where $A \in \mathbb{R}^{n \times n}$, $N = [N_1, \ldots, N_m] \in \mathbb{R}^{n \times nm}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $u(k) \in \mathbb{R}^m$, $y(k) \in \mathbb{R}^p$. These systems arise naturally as theoretical model for different real-life dynamics, see e.g. [14, 15, 16, 7]. Moreover, for systems with weak nonlinearities, the concept of Carleman bilinearization, see [19], yields satisfying approximations by enlarged bilinear systems. An interesting feature of this special class of nonlinear systems now is their close relationship to linear systems which allows generalizing some of the concepts from the linear case, see e.g. [1]. Formally, throughout this paper, our goal will be the construction of a reduced-order system

$$\widehat{\Sigma}: \quad \begin{cases} \hat{x}(k+1) = \hat{A}\hat{x}(k) + \hat{N}\left(I_m \otimes \hat{x}(k)\right)u(k) + \hat{B}u(k), \\ \quad\, \hat{y}(k) = \hat{C}\hat{x}(k), \quad \hat{x}(0) = \hat{x}_0, \end{cases} \tag{2}$$

with $\hat{A} \in \mathbb{R}^{\hat{n} \times \hat{n}}$, $\hat{N} = [\hat{N}_1, \ldots, \hat{N}_m] \in \mathbb{R}^{\hat{n} \times \hat{n}m}$, $\hat{B} \in \mathbb{R}^{\hat{n} \times m}$, $\hat{C} \in \mathbb{R}^{p \times \hat{n}}$, $u(k) \in \mathbb{R}^m$, $\hat{y}(k) \in \mathbb{R}^p$, $\hat{y} \approx y$ and $\hat{n} \ll n$. For simplicity, we impose a zero initial condition on the system, i.e. $x_0 = 0$. Generalizations to $x_0 \neq 0$ can be obtained analogously. Throughout the paper we will denote $\tilde{N} = [N_1^T, \ldots, N_m^T]$.

We proceed as follows. In Section 2, we derive an explicit solution formula for discrete-time bilinear systems. Based on this expression, we show that a multivariate Z-transform allows an input-output characterization via some generalized transfer functions known from the continuous-time case. In Section 3, we provide an interpolation-based model reduction method that tangentially approximates the transfer functions of the original system. Based on the definition of the $\mathcal{H}_2$-norm for bilinear systems, we will discuss the choice of optimal interpolation points together with tangential directions. A possible discretization of continuous-time bilinear systems follows in Section 5. We compare the performance of the new approach with the method of balanced truncation in Section 6 and conclude with a summary and possible topics of future studies in Section 7. Throughout the rest of the paper, we will use the following notations for a matrix A shifted by σI and the k-fold Kronecker product, respectively:

$$A_\sigma := \sigma I - A, \quad I_m^{\otimes^{k+1}} := \underbrace{I_m \otimes \cdots \otimes I_m}_{k+1}.$$

2 Discrete-Time Bilinear Systems

Recall that the solution of a continuous-time bilinear system is given by a Volterra series which can be obtained by the method of successive approximations. The following lemma, initially stated in [22], provides an analogue discrete-time solution formula. For the sake of completeness, we provide a proof in compact notation using the Kronecker product formalism.

Lemma 2.1. *The solution of a discrete-time bilinear control system Σ can be expressed as*

$$x(k) = \sum_{p=1}^{k} w_p(k),$$

with

$$w_p(k) = \sum_{i_p=p-1}^{k-1} \sum_{i_{p-1}=p-2}^{i_p-1} \cdots \sum_{i_1=0}^{i_2-1} A^{k-i_p-1} N\left(I_m \otimes A^{i_p-i_{p-1}-1}N\right) \cdots \left(I_m^{\otimes^{p-2}} \otimes A^{i_3-i_2-1}N\right)$$

$$\left(I_m^{\otimes^{p-1}} \otimes A^{i_2-i_1-1}B\right) u(i_p) \otimes \cdots \otimes u(i_1)$$

2

Proof. First of all, note the identity

$$w_1(k+1) = \sum_{i_1=0}^{k} A^{k-i_1} Bu(i_1) = \sum_{i_1=0}^{k-1} A^{k-i_1} Bu(i_1) + Bu(k) = A \sum_{i_1=0}^{k-1} A^{k-i_1-1} Bu(i_1) + Bu(k)$$

$$= Aw_1(k) + Bu(k) .$$

For $1 < p < k+1$ we thus obtain

$$w_p(k+1) = \sum_{i_p=p-1}^{k} \sum_{i_{p-1}=p-2}^{i_p-1} \cdots \sum_{i_1=0}^{i_2-1} A^{k-i_p} N \left(I_m \otimes A^{i_p-i_{p-1}-1} N \right) \cdots \left(I_m^{\otimes p-2} \otimes A^{i_3-i_2-1} N \right)$$

$$\left(I_m^{\otimes p-1} \otimes A^{i_2-i_1-1} B \right) u(i_p) \otimes \cdots \otimes u(i_1)$$

$$= \sum_{i_p=p-1}^{k-1} \sum_{i_{p-1}=p-2}^{i_p-1} \cdots \sum_{i_1=0}^{i_2-1} A^{k-i_p} N \left(I_m \otimes A^{i_p-i_{p-1}-1} N \right) \cdots \left(I_m^{\otimes p-2} \otimes A^{i_3-i_2-1} N \right)$$

$$\left(I_m^{\otimes p-1} \otimes A^{i_2-i_1-1} B \right) u(i_p) \otimes \cdots \otimes u(i_1)$$

$$+ \sum_{i_{p-1}=p-2}^{k-1} \cdots \sum_{i_1=0}^{i_2-1} A^{k-k} N \left(I_m \otimes A^{k-i_{p-1}-1} N \right) \cdots \left(I_m^{\otimes p-2} \otimes A^{i_3-i_2-1} N \right)$$

$$\left(I_m^{\otimes p-1} \otimes A^{i_2-i_1-1} B \right) \left(I_m \otimes u(i_{p-1}) \otimes \cdots \otimes u(i_1) \right) u(k)$$

$$= Aw_p(k) + N \left(I_m \otimes w_{p-1}(k) \right) u(k),$$

$$w_{k+1}(k+1) = \sum_{i_{k+1}=k}^{k} \sum_{i_k=k-1}^{i_{k+1}-1} \cdots \sum_{i_1=0}^{i_2-1} A^{k+1-i_{k+1}-1} N \left(I_m \otimes A^{i_{k+1}-i_k-1} N \right) \cdots \left(I_m^{\otimes k-1} \otimes A^{i_3-i_2-1} N \right)$$

$$\left(I_m^{\otimes k} \otimes A^{i_2-i_1-1} B \right) u(i_{k+1}) \otimes u(i_k) \otimes \cdots \otimes u(i_1)$$

$$= N \left(I_m \otimes N \right) \cdots \left(I_m^{\otimes k-1} \otimes N \right) \left(I_m^{\otimes k} \otimes B \right) \left(I_m \otimes u(i_k) \otimes \cdots \otimes u(i_1) \right) u(i_{k+1})$$

$$= N \left(I_m \otimes w_k(k) \right) u(k)$$

We can now prove the statement by induction over k. For the first two cases, we end up with

$$x(1) = Ax(0) + N \left(I_m \otimes x(0) \right) u(0) + Bu(0) = Bu(0) = w_1(1)$$

$$x(2) = Ax(1) + N \left(I_m \otimes x(1) \right) u(1) + Bu(1) = Aw_1(1) + w_2(2) + Bu(1) = w_1(2) + w_2(2) .$$

Next, let us assume that we have

$$x(k) = \sum_{p=1}^{k} w_p(k).$$

Making use of the properties shown above, we obtain

$$
\begin{aligned}
x(k+1) &= Ax(k) + N\left(I_m \otimes x(k)\right)u(k) + Bu(k) \\
&= A\left(\sum_{p=1}^{k} w_p(k)\right) + N\left(I_m \otimes \sum_{p=1}^{k} w_p(k)\right)u(k) + Bu(k) \\
&= Aw_1(k) + Bu(k) + \sum_{p=2}^{k} Aw_p(k) + \sum_{p=1}^{k-1} N\left(I_m \otimes w_p(k)\right)u(k) + N\left(I_m \otimes w_k(k)\right)u(k) \\
&= \underbrace{Aw_1(k) + Bu(k)}_{w_1(k+1)} + \sum_{p=2}^{k} \underbrace{\left(Aw_p(k) + N\left(I_m \otimes w_{p-1}(k)\right)u(k)\right)}_{w_p(k+1)} + \underbrace{N\left(I_m \otimes w_k(k)\right)u(k)}_{w_{k+1}(k+1)} \\
&= \sum_{p=1}^{k+1} w_p(k+1).
\end{aligned}
$$

$\square$

Now that we have found an expression for the solution of a discrete-time bilinear system, we know that the corresponding output at time point k is given by

$$
y(k) = \sum_{p=1}^{k} \sum_{i_p=p-1}^{k-1} \sum_{i_{p-1}=p-2}^{i_p-1} \cdots \sum_{i_1=0}^{i_2-1} CA^{k-i_p-1}N\left(I_m \otimes A^{i_p-i_{p-1}-1}N\right)\cdots\left(I_m^{\otimes^{p-2}} \otimes A^{i_3-i_2-1}N\right)
$$
$$
\left(I_m^{\otimes^{p-1}} \otimes A^{i_2-i_1-1}B\right)u(i_p)\otimes\cdots\otimes u(i_1). \tag{3}
$$

Note the close connection to the theory of continuous-time systems. Here, the system can be analyzed by means of the Volterra series of a bilinear system. A multivariate Laplace transform of the nonlinear kernels corresponding to this series representation leads to generalized transfer functions that pave the way for Krylov-based reduction techniques. For a more detailed analysis of the continuous-time case, we refer to [2, 3, 6, 9, 18, 17]. Having said this, it seems reasonable to derive discrete-time transfer functions that characterize the input-output behavior in the frequency domain. For this, we perform the following change of variables:

$$
j_p := k - i_p, \ j_r = i_{r+1} - i_r, \ r < p.
$$

Thus, expression (3) can be rewritten as

$$
y(k) = \sum_{p=1}^{k} \sum_{j_p=1}^{k-p+1} \sum_{j_{p-1}=1}^{k-j_p-p+2} \cdots \sum_{j_1=1}^{k-j_p-\ldots-j_2}
$$
$$
\underbrace{CA^{j_p-1}N\left(I_m \otimes A^{j_{p-1}-1}N\right)\cdots\left(I_m^{\otimes^{p-2}} \otimes A^{j_2-1}N\right)\left(I_m^{\otimes^{p-1}} \otimes A^{j_1-1}B\right)}_{h_{reg(j_1,\ldots,j_p)}}
$$
$$
u(k-j_p)\otimes\cdots\otimes u(k-j_p-\ldots-j_1),
$$

where the term $h_{reg(j_1,\ldots,j_p)}$ is called *degree-p kernel*. Finally, a multivariate Z-transform of the *degree-p*

kernel can be interpreted as the *p-th transfer function* of the corresponding bilinear system:

$$
\begin{aligned}
H(z_1,\ldots,z_p) &= \sum_{j_p=1}^{\infty}\sum_{j_{p-1}=1}^{\infty}\cdots\sum_{j_1=1}^{\infty} CA^{j_p-1}N\left(I_m \otimes A^{j_{p-1}-1}N\right)\cdots\left(I_m^{\otimes p-2} \otimes A^{j_2-1}N\right) \\
&\qquad \left(I_m^{\otimes p-1} \otimes A^{j_1-1}B\right) z_1^{-j_1}\cdots z_p^{-j_p} \\
&= \sum_{j_p=0}^{\infty}\sum_{j_{p-1}=0}^{\infty}\cdots\sum_{j_1=0}^{\infty} Cz_p^{-1}(z_p^{-1}A)^{j_p}N\left(I_m \otimes z_{p-1}^{-1}(z_{p-1}^{-1}A)^{j_{p-1}}N\right)\cdots \\
&\qquad \left(I_m^{\otimes p-2} \otimes z_2^{-1}(z_2^{-1}A)^{j_2}N\right)\left(I_m^{\otimes p-1} \otimes z_1^{-1}(z_1^{-1}A)^{j_1}B\right) \\
&= Cz_p^{-1}(I - z_p^{-1}A)^{-1}N\left(I_m \otimes z_{p-1}^{-1}(I - z_{p-1}^{-1}A)^{-1}N\right)\cdots \\
&\qquad \left(I_m^{\otimes p-2} \otimes z_2^{-1}(I - z_2^{-1}A)^{-1}N\right)\left(I_m^{\otimes p-1} \otimes z_1^{-1}(I - z_1^{-1}A)^{-1}B\right) \\
&= C(z_pI - A)^{-1}N\left(I_m \otimes (z_{p-1}I - A)^{-1}N\right)\cdots\left(I_m^{\otimes p-2} \otimes (z_2I - A)^{-1}N\right) \\
&\qquad \left(I_m^{\otimes p-1} \otimes (z_1I - A)^{-1}B\right).
\end{aligned}
$$

3 Generalized Tangential Interpolation

Based on the derivation in the previous section, we can now proceed similarly to the continuous-time case. As has been shown in e.g. [3, 6, 18], one might match so-called multimoments around a specified frequency, i.e. values and derivatives of the transfer functions at these points. More generally, we consider the concept of tangential interpolation of a fixed number r of bilinear transfer functions by a smaller system $\hat{\Sigma}$ at a set of prescribed complex points $S = \{\sigma_1,\ldots,\sigma_q\}$ together with left and right tangential directions $\{l_1,\ldots,l_q\}$ and $\{r_1,\ldots,r_q,\}$ respectively. This will be an essential tool to investigate the $\mathcal{H}_2$-model order reduction in the next section.

Formally, we are heading for

$$H_k(s_1,\ldots,s_{k-1},\sigma_j)r_j = \hat{H}_k(s_1,\ldots,s_{k-1},\sigma_j)r_j, \quad k = 1,\ldots,r, \quad s_i \in S, \tag{4}$$

$$l_j H_k(\sigma_j,s_2,\ldots,s_k) = l_j \hat{H}_k(\sigma_j,s_2,\ldots,s_k), \quad k = 1,\ldots,r, \quad s_i \in S, \tag{5}$$

which can be achieved by a series of specific rational Krylov subspaces as stated in the following theorem.

Theorem 3.1. *Given a bilinear system as in (1), let the ranges of V and W contain the union of the column spaces of*

$$
\begin{aligned}
V_1 &= \left[A_{\sigma_1}^{-1}Br_1,\ldots,A_{\sigma_q}^{-1}Br_q\right], \\
V_k &= \left[A_{\sigma_1}^{-1}N\left(I_m \otimes V_{k-1}\right),\ldots,A_{\sigma_q}^{-1}N\left(I_m \otimes V_{k-1}\right)\right], \quad k = 2,\ldots,r,
\end{aligned}
$$

and

$$
\begin{aligned}
W_1 &= \left[A_{\sigma_1}^{-T}C^T l_1,\ldots,A_{\sigma_q}^{-T}C^T l_q\right], \\
W_k &= \left[A_{\sigma_1}^{-T}\tilde{N}\left(I_m \otimes W_{k-1}\right),\ldots,A_{\sigma_q}^{-T}\tilde{N}\left(I_m \otimes W_{k-1}\right)\right], \quad k = 2,\ldots,r,
\end{aligned}
$$

respectively. Then the reduced system $\hat{\Sigma}$ arising from an oblique projection $P = VZ$, with $Z = (W^TV)^{-1}W^T$,

i.e., $\hat{A} = ZAV$, $\hat{N} = ZN\left(I_m \otimes V\right)$, $\hat{B} = ZB$, $C = CV$, *satisfies*

$$H_k(s_1, \ldots, s_{k-1}, \sigma_j)r_j = \hat{H}_k(s_1, \ldots, s_{k-1}, \sigma_j)r_j, \quad 1 \le k \le r, \tag{6}$$

$$l_j^* H_k(\sigma_j, s_2, \ldots, s_k) = l_j^* \hat{H}_k(\sigma_j, s_2, \ldots, s_k), \quad 1 \le k \le r, \tag{7}$$

$$l_i^* H_k(\sigma_i, s_2, \ldots, s_{k-1}, \sigma_j)r_j = l_i^* \hat{H}_k(\sigma_i, s_2, \ldots, s_{k-1}, \sigma_j)r_j, \quad r < k \le 2r, \tag{8}$$

$$l_i^* \frac{\partial}{\partial s_\ell} H_k(\sigma_i, s_2, \ldots, s_{k-1}, \sigma_j)r_j = l_i^* \frac{\partial}{\partial s_\ell} \hat{H}_k(\sigma_i, s_2, \ldots, s_{k-1}, \sigma_j)r_j, \quad 1 \le k \le r, \ 1 \le \ell \le k, \tag{9}$$

$$l_i^* \frac{\partial}{\partial s_\ell} H_k(\sigma_i, s_2, \ldots, s_{k-1}, \sigma_j)r_j = l_i^* \frac{\partial}{\partial s_\ell} \hat{H}_k(\sigma_i, s_2, \ldots, s_{k-1}, \sigma_j)r_j, \quad r < k \le 2r-1, \ k-r < \ell \le r, \tag{10}$$

for all $s_i \in S$.

Proof. W.l.o.g. let us consider the case $s_i = \sigma$ for all $2 \le i \le k-1$. First of all, note that we have

$$V\hat{A}_\sigma^{-1}\hat{N} \cdots \left(I_m^{\otimes^{k-1}} \otimes \hat{A}_{\sigma_j}^{-1}\hat{B}r_j\right) = A_\sigma^{-1}N \cdots \left(I_m^{\otimes^{k-1}} \otimes A_{\sigma_j}^{-1}Br_j\right),$$

for $k \le r$. This can be shown via induction with respect to k. For $k = 1$, the result is known from the linear case. Hence, let us assume equality for k. Recall the fact that VZ is a projector onto

$$V \supset A_\sigma^{-1}N \cdots \left(I_m^{\otimes^{k-1}} \otimes A_{\sigma_j}^{-1}Br_j\right)$$

and hence

$$VZA_\sigma^{-1}N \cdots \left(I_m^{\otimes^{k-1}} \otimes A_{\sigma_j}^{-1}Br_j\right) = A_\sigma^{-1}N \cdots \left(I_m^{\otimes^{k-1}} \otimes A_{\sigma_j}^{-1}Br_j\right),$$

for $k \le r$. Hence, we then obtain

$$\begin{aligned}
&V\hat{A}_\sigma^{-1}\hat{N} \left(I_m \otimes \hat{A}_\sigma^{-1}\hat{N}\right) \cdots \left(I_m^{\otimes^k} \otimes \hat{A}_{\sigma_j}^{-1}\hat{B}r_j\right) \\
=&V\hat{A}_\sigma^{-1}\hat{N} \left\{I_m \otimes \left[\hat{A}_\sigma^{-1}\hat{N} \cdots \left(I_m^{\otimes^{k-1}} \otimes \hat{A}_{\sigma_j}^{-1}\hat{B}r_j\right)\right]\right\} \\
=&V\hat{A}_\sigma^{-1}ZN(I_m \otimes V) \left\{I_m \otimes \left[\hat{A}_\sigma^{-1}\hat{N} \cdots \left(I_m^{\otimes^{k-1}} \otimes \hat{A}_{\sigma_j}^{-1}\hat{B}r_j\right)\right]\right\} \\
=&V\hat{A}_\sigma^{-1}ZN \left\{I_m \otimes \left[V\hat{A}_\sigma^{-1}\hat{N} \cdots \left(I_m^{\otimes^{k-1}} \otimes \hat{A}_{\sigma_j}^{-1}\hat{B}r_j\right)\right]\right\} \\
=&V\hat{A}_\sigma^{-1}ZN \left\{I_m \otimes \left[A_\sigma^{-1}N \cdots \left(I_m^{\otimes^{k-1}} \otimes A_{\sigma_j}^{-1}Br_j\right)\right]\right\} \\
=&V\hat{A}_\sigma^{-1}ZN \left(I_m \otimes A_\sigma^{-1}N\right) \cdots \left(I_m^{\otimes^k} \otimes A_{\sigma_j}^{-1}Br_j\right) \\
=&V\left(ZA_\sigma V\right)^{-1} ZN \left(I_m \otimes A_\sigma^{-1}N\right) \cdots \left(I_m^{\otimes^k} \otimes A_{\sigma_j}^{-1}Br_j\right) \\
=&V\left(ZA_\sigma V\right)^{-1} ZA_\sigma A_\sigma^{-1}N \left(I_m \otimes A_\sigma^{-1}N\right) \cdots \left(I_m^{\otimes^k} \otimes A_{\sigma_j}^{-1}Br_j\right) \\
=&V\left(ZA_\sigma V\right)^{-1} ZA_\sigma VZA_\sigma^{-1}N \left(I_m \otimes A_\sigma^{-1}N\right) \cdots \left(I_m^{\otimes^k} \otimes A_{\sigma_j}^{-1}Br_j\right) \\
=&A_\sigma^{-1}N \left(I_m \otimes A_\sigma^{-1}N\right) \cdots \left(I_m^{\otimes^k} \otimes A_{\sigma_j}^{-1}Br_j\right).
\end{aligned}$$

Here, the last identity is due to the construction of V and the fact that $P = VZ$ is a projection. A similar argumentation leads to

$$l_j^* \hat{C}\hat{A}_{\sigma_j}^{-1}\hat{N} \cdots \left(I_m^{\otimes^{k-1}} \otimes \hat{A}_\sigma^{-1}Z\right) = l_j^* CA_{\sigma_j}^{-1}N \cdots \left(I_m^{\otimes^{k-1}} \otimes A_\sigma^{-1}\right),$$

for $k \le r$. Hence, we have shown assertions (6) and (7), respectively. For equation (8), note that it holds

$$H_{k+f}(\sigma, \ldots, \sigma) = CA_\sigma^{-1}N \cdots \left(I_m^{\otimes^{k-1}} \otimes A_\sigma^{-1}N\right) \cdot \left(I_m^{\otimes^k} \otimes \left(A_\sigma^{-1}N \cdots \left(I_m^{\otimes^{f-1}} \otimes A_\sigma^{-1}B\right)\right)\right).$$

Thus, a combination of the equalities shown before together with the fact that $ZV = I$ again yields the desired result. Finally, the first derivative of the k-th transfer function can be rewritten as follows:

$$
\begin{aligned}
&\frac{\partial}{\partial \ell_j} H_k(\sigma, \ldots, \sigma) \\
&= -CA_\sigma^{-1}N \cdots \left(I_m^{\otimes^{j-1}} \otimes A_\sigma^{-2}N\right) \cdots \left(I_m^{\otimes^{k-1}} \otimes A_\sigma^{-1}B\right) \\
&= -CA_\sigma^{-1}N \cdots \left(I_m^{\otimes^{j-1}} \otimes A_\sigma^{-1}\right) \cdot \left(I_m^{\otimes^{j-1}} \otimes A_\sigma^{-1}N\right) \cdots \left(I_m^{\otimes^{k-1}} \otimes A_\sigma^{-1}B\right) \\
&= -CA_\sigma^{-1}N \cdots \left(I_m^{\otimes^{j-1}} \otimes A_\sigma^{-1}\right) \cdot \left(I_m^{\otimes^{j-1}} \otimes \left(A_\sigma^{-1}N \cdots \left(I_m^{\otimes^{k-j}} \otimes A_\sigma^{-1}B\right)\right)\right).
\end{aligned}
$$

For $k \leq r$, we now again can make use of the identities shown in the beginning of the proof, readily justifying (9). Finally, for $k > r$, the combination that creates the term A_σ^{-2} can only be achieved when this term appears sufficiently centered, explaining the restriction for ℓ in (10). $\qquad\square$

4 $\mathcal{H}_2$-Norm for Bilinear Systems

Obviously, the choice of the interpolation points as well as the tangential directions will play a crucial role in the construction of a reduced order model. For linear systems, Gugercin et al. and Bunse-Gerstner et al. have shown in [12] and [8], respectively, that a reduced model has to interpolate the values and first derivatives of the original transfer function at the reciprocals of its own eigenvalues in order to solve an $\mathcal{H}_2$-model reduction problem. An iterative rational Krylov algorithm (IRKA/MIRIAm) has proven to be an efficient and easily implementable tool to construct $\hat{\Sigma}$. In order to provide a similar method for bilinear systems, we first have to define a reasonable generalization of the common $\mathcal{H}_2$-norm for linear systems. While this has been done for the continuous-time case in [21], we present the following adaption for discrete-time systems.

Definition 4.1. *Let Σ be a discrete-time bilinear systems and let H_j denote its generalized j-th transfer function. Then we define*

$$
\|\Sigma\|_{\mathcal{H}_2}^2 = \mathrm{tr}\left(\sum_{k=1}^\infty \underbrace{\int_0^{2\pi} \cdots \int_0^{2\pi} \frac{1}{(2\pi)^k} \overline{H_k(e^{i\theta_1}, \ldots, e^{i\theta_k})} \left(H_k(e^{i\theta_1}, \ldots, e^{i\theta_k})\right)^T d\theta_1 \ldots d\theta_k}_{J_k^2}\right).
$$

Under the assumption of convergence of the above series, we can make use of linearity as well as commutativity of the trace functional. Moreover, the fact that $\mathrm{tr}(A) = \mathrm{tr}(A^T)$ shows that each summand defines an inner product $\langle A, B \rangle = \mathrm{tr}(A^*B)$, showing the norm properties of our definition. Since for $N = 0$, the above definition coincides with the linear $\mathcal{H}_2$-norm, this may be a reasonable generalization.

Like in the linear case, we can express this norm with the help of generalized Gramians. If the operators $P \mapsto APA^T - P + N(I_m \otimes P)N^T$ and $Q \mapsto A^TQA - Q + \tilde{N}(I_m \otimes Q)\tilde{N}^T$ are Hurwitz stable (i.e. all eigenvalues in $\mathbb{C}_-$), then the reachability Gramian $P \geq 0$ and the observability Gramian $Q \geq 0$ are given as solutions of the generalized Stein equations

$$
APA^T - P + N(I_m \otimes P)N^T + BB^T = 0, \tag{11}
$$

$$
A^TQA - Q + \tilde{N}(I_m \otimes Q)\tilde{N}^T + C^TC = 0. \tag{12}
$$

Each of them can be computed as the limit of a series of linear Stein equations

$$
AP_1A^T - P_1 + BB^T = 0,
$$
$$
A^TQ_1A - Q_1 + C^TC = 0,
$$
$$
AP_jA^T - P_j + N(I_m \otimes P_{j-1})N^T = 0,
$$
$$
A^TQ_jA - Q_j + \tilde{N}(I_m \otimes Q_{j-1})\tilde{N}^T = 0,
$$

with $P = \sum_{j=1}^\infty P_j$ and $Q = \sum_{j=1}^\infty Q_j$.

7

Lemma 4.1. *Let P and Q be the solutions of the generalized Stein equations (11) and (12), respectively. Then the $\mathcal{H}_2$-norm of Σ can be computed as*

$$\|\Sigma\|_{\mathcal{H}_2}^2 = \operatorname{tr}\left(CPC^T\right) = \operatorname{tr}\left(B^TQB\right).$$

Proof. To prove the assertion for the reachability Gramian P we can show that $J_k^2 = CP_kC^T$ for all k. In the case $k = 1$, this is known from linear system theory (e.g. [12]) and follows directly from the formula (see [11])

$$\frac{1}{2\pi}\int_0^{2\pi}\left(e^{-i\theta}I - A\right)^{-1}X\left(e^{i\theta}I - A^T\right)^{-1}d\theta = \sum_{j=0}^{\infty}A^jX(A^T)^j\,, \tag{13}$$

where for $X = BB^T$ the right hand side is equal to P_1. Using the same formula again, we obtain

$$P_2 = \sum_{j=0}^{\infty}A^j\left(P_2 - AP_2A\right)(A^T)^j = \sum_{j=0}^{\infty}A^j\left(-N(I_m \otimes P_1)N^T\right)(A^T)^j$$

$$= \frac{1}{2\pi}\int_0^{2\pi}\left(e^{-i\theta}I - A\right)^{-1}\left(-N(I_m \otimes P_1)N^T\right)\left(e^{i\theta}I - A^T\right)^{-1}d\theta$$

$$= \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi}\left(e^{-i\theta_2}I - A\right)^{-1}N\left(I_m \otimes \left(e^{-i\theta_1} - A\right)^{-1}BB^T\left(e^{i\theta_1} - A^T\right)^{-1}\right)N^T\left(e^{i\theta_2}I - A^T\right)^{-1}d\theta_1 d\theta_2.$$

Multiplication with C and C^T gives the desired formula. For higher order subsystems and for the observability Gramian the argument is analogous. $\square$

An alternative formula, which can be used to specify necessary conditions for $\mathcal{H}_2$-optimal approximation, is given in the following lemma.

Lemma 4.2. *Let Σ be a MIMO bilinear system and let $\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}$ denote the spectrum of the system matrix A. Then the $\mathcal{H}_2$-norm of Σ can be computed as follows.*

$$\|\Sigma\|_{\mathcal{H}_2}^2 = \operatorname{tr}\left(\sum_{k=1}^{\infty}\sum_{\ell_1=1}^{n}\cdots\sum_{\ell_k=1}^{n}\frac{\Phi_{\ell_1,\ldots,\ell_k}}{\lambda_{\ell_1}\cdots\lambda_{\ell_k}}\left(H_k\left(\frac{1}{\lambda_{\ell_1}},\ldots,\frac{1}{\lambda_{\ell_k}}\right)\right)^T\right),$$

where

$$\Phi_{\ell_1,\ldots,\ell_k} = \lim_{s_j\to\lambda_{\ell_j}}H_k\left(s_1,\ldots,s_k\right)\left(s_1 - \lambda_{\ell_1}\right)\cdots\left(s_k - \lambda_{\ell_k}\right)$$

denotes a generalized residue associated with the k-th transfer function.

Proof. First, note that

$$\|\Sigma\|_{\mathcal{H}_2}^2 = \operatorname{tr}\left(\sum_{k=1}^{\infty}\int_0^{2\pi}\cdots\int_0^{2\pi}\frac{1}{(2\pi)^k}\overline{H_k(e^{i\theta_1},\ldots,e^{i\theta_k})}\left(H_k(e^{i\theta_1},\ldots,e^{i\theta_k})\right)^T d\theta_1\ldots d\theta_k\right)$$

$$= \operatorname{tr}\left(\sum_{k=1}^{\infty}\int_0^{2\pi}\cdots\int_0^{2\pi}\frac{1}{(2\pi)^k}H_k(e^{-i\theta_1},\ldots,e^{-i\theta_k})\left(H_k(e^{i\theta_1},\ldots,e^{i\theta_k})\right)^T d\theta_1\ldots d\theta_k\right)$$

Introduce new variables

$$s_j = e^{i\theta_j} \Rightarrow d\theta_j = \frac{ds_j}{i\,s_j}.$$

We then have

$$\|\Sigma\|_{\mathcal{H}_2}^2 = \operatorname{tr}\left(\sum_{k=1}^{\infty}\int_0^{2\pi}\cdots\int_0^{2\pi}\frac{1}{(2\pi i)^k}\frac{1}{s_1\cdots s_k}H_k(\frac{1}{s_1},\ldots,\frac{1}{s_k})\left(H_k(s_1,\ldots,s_k)\right)^T ds_1\ldots ds_k\right).$$

The residue theorem applied to each of the above integrals together with the fact that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ and $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$ shows the assumption. $\square$

A straightforward consideration yields an error expression between the reduced and the original system.

Corollary 4.1. *Let Σ and $\hat{\Sigma}$ denote the original and a reduced bilinear system, respectively. Then the $\mathcal{H}_2$-norm of the error system $\Sigma - \hat{\Sigma}$ is given as*

$$\|\Sigma - \hat{\Sigma}\|_{\mathcal{H}_2}^2 = \operatorname{tr}\left(\sum_{k=1}^{\infty}\sum_{\ell_1=1}^{n}\cdots\sum_{\ell_k=1}^{n}\frac{\Phi_{\ell_1,\ldots,\ell_k}}{\lambda_{\ell_1}\cdots\lambda_{\ell_k}}\left(H_k(\frac{1}{\lambda_{\ell_1}},\ldots,\frac{1}{\lambda_{\ell_k}}) - \hat{H}_k(\frac{1}{\lambda_{\ell_1}},\ldots,\frac{1}{\lambda_{\ell_k}})\right)^T\right)$$
$$+ \operatorname{tr}\left(\sum_{k=1}^{\infty}\sum_{\hat{\ell}_1=1}^{\hat{n}}\cdots\sum_{\hat{\ell}_k=1}^{\hat{n}}\frac{\Phi_{\hat{\ell}_1,\ldots,\hat{\ell}_k}}{\hat{\lambda}_{\hat{\ell}_1}\cdots\hat{\lambda}_{\hat{\ell}_k}}\left(\hat{H}_k(\frac{1}{\hat{\lambda}_{\hat{\ell}_1}},\ldots,\frac{1}{\hat{\lambda}_{\hat{\ell}_k}}) - H_k(\frac{1}{\hat{\lambda}_{\hat{\ell}_1}},\ldots,\frac{1}{\hat{\lambda}_{\hat{\ell}_k}})\right)^T\right).$$

Besides the commonly used method of comparing outputs of the original and the reduced system in the time domain, the above results allow to judge the approximation quality in terms of the relative $\mathcal{H}_2$-error. Note that, analogue to linear system theory, the error is due to the mismatch of the transfer functions at the reciprocals of the original system poles and the reduced system poles. Although we have not obtained $\mathcal{H}_2$-optimality conditions for the bilinear case so far, we expect the reduced system poles to be connected to the optimal interpolation points. Unfortunately, an iterative construction based on Theorem 3.1 would increase the reduced system dimension in each step. For this reason, we propose to perform a truncated SVD of the obtained projection matrices V and W and to continue the iteration afterwards, see Algorithm 1.

Algorithm 1 Bilinear Iterative Rational Krylov Algorithm (Bilinear-IRKA)

Require: A, N, B, C, r, q
Ensure: $\hat{A}$, $\hat{N}$, $\hat{B}$, $\hat{C}$
 1: Make an initial selection $\{\sigma_1,\ldots,\sigma_q\}$ with tangential directions $R = [r_1,\ldots,r_q]$ and $L = [l_1,\ldots,l_q]$.
 2: **while** (change in $\sigma_i > \epsilon$) **do**
 3:     Compute $V = [V_1,\ldots,V_r]$ and $W = [W_1,\ldots,W_r] \in \mathbb{R}^{n\times(q+\cdots+q^r)}$ as in Theorem 3.1.
 4:     Compute truncated SVD V_q and W_q of V and W.
 5:     $\hat{A} = (W_q^T V_q)^{-1}W_q^T A V_q$, $\hat{B} = (W_q^T V_q)^{-1}W_q^T B$, $\hat{C} = CV_q$.
 6:     Compute the eigenvalue decomposition $\hat{A} = Q \cdot \Lambda \cdot Q^{-1}$.
 7:     Set $\sigma_i \leftarrow \frac{1}{\lambda_i(\hat{A})}$, $L = \hat{C}Q$, $R = \hat{B}^* Q^{-*}$.
 8: **end while**
 9: $\hat{N} = (W_q^T V_q)^{-1}W_q^T N V_q$, $\hat{B} = (W_q^T V_q)^{-1}W_q^T B$, $\hat{C} = CV_q$.

---

5 Discretization of Bilinear Systems

Due to the lack of large-scale discrete-time test examples in the open literature, we will later on generate several artificial examples out of continuous-time bilinear systems. Hence, let us now briefly focus on a given continuous-time bilinear system which we want to transform into its discrete-time counterpart. For this, we consider the system

$$\Sigma_c : \quad \begin{cases} x(t) = A_c x(t) + N_c\left(I_m \otimes x(t)\right)u(t) + B_c u(t), \\ y(t) = C_c x(t), \quad x(0) = x_0, \end{cases}$$

where the dimensions of the system matrices are identical to the setting in (1). The subscript c will denote the continuous character of the equation. While in the linear case one can make use of the Tustin transform to create a discrete-time system which additionally preserves the stability properties of the continuous system, the situation becomes more complicated for bilinear models. For a more detailed overview on this topic, the reader is referred to [10, 20]. Since the main focus of this paper is directed to the problem of model order reduction, we will be content with a semi-implicit Euler discretization of the above system, i.e. the discrete matrices will be constructed according to

$$A = (I - hA_c)^{-1}, \; N = h(I - hA_c)^{-1}N_c, \; B = h(I - hA_c)^{-1}B_c, \; C = C_c,$$

where h denotes a sampling parameter. Since for large-scale matrices the explicit computation of the inverse of $(I - hA)$ might cause severe problems, we will briefly show how the discretization technique can be directly incorporated in the construction of the projection matrices V and W. Let us exemplarily concentrate on V_1. For each interpolation point σ_i in S, we have to compute

$$
\begin{aligned}
(\sigma_i I - A)^{-1} B b_i &= \left(\sigma_i I - (I - hA_c)^{-1}\right)^{-1} h(I - hA_c)^{-1} B_c b_i \\
&= h \left[(I - hA_c) \left(\sigma_i I - (I - hA_c)^{-1} \right) \right]^{-1} B_c b_i \\
&= h \left[\sigma_i I - h\sigma_i A_c - I \right]^{-1} B_c b_i \\
&= h \left[\sigma_i (I - hA_c) - I \right]^{-1} B_c b_i.
\end{aligned}
$$

It evidently follows that $(\sigma_i I - A)^{-1} N (I_m \otimes V_k) = h \left[\sigma_i (I - hA_c) - I \right]^{-1} N_c (I_m \otimes V_k)$, indicating that we only have to solve systems of linear equations specified by the original continuous-time matrices.

6 Numerical Examples

We implemented the rational interpolation approach for bilinear discrete-time systems as MATLAB® function, tested it for several examples and compared it with the method of balanced truncation for bilinear systems. In each case, we stuck to an approximation of the first two transfer functions, i.e. we chose $r = 2$ in Algorithm 1. Additionally, we compared our results with the randomly chosen interpolation points and tangential directions, respectively, used for the initialization of the algorithm. The following subsections show the results we obtained on an Intel® Core™i7 CPU 920, 8 MB cache, 12 GB RAM, openSUSE Linux 11.1 (x86_64).

6.1 Hinamoto and Maekawa

The first example we want to study was introduced by Hinamoto and Maekawa ([13]) and was used as a numerical test example in [22]. Even though the original model is only of dimension 5 and obviously far away from being large-scale, we will compare our results (denoted by "B-IRKA") with those presented in [22] obtained by balanced truncation (BT) since other detailed discussions on discrete-time bilinear systems do not exist to the authors' knowledge. The system matrices are as follows

$$
A = \begin{bmatrix} 0 & 0 & 0.024 & 0 & 0 \\ 1 & 0 & -0.26 & 0 & 0 \\ 0 & 1 & 0.9 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & -0.06 \\ 0 & 0 & 0.15 & 1 & 0.5 \end{bmatrix}, \quad B = \begin{bmatrix} 0.8 \\ 0.6 \\ 0.4 \\ 0.2 \\ 0.5 \end{bmatrix}, \quad C^T = \begin{bmatrix} 0.2 \\ 0.4 \\ 0.6 \\ 0.8 \\ 1.0 \end{bmatrix}, \quad N = \mathrm{diag}(0.1, 0.2, 0.3, 0.4, 0.5).
$$

We successively reduced the model to systems of dimension $1, 2, 3$ and 4 respectively. Figure 1 displays the particular response to a unit step input and the relative $\mathcal{H}_2$-error for each of the systems. As a matter of fact, the interpolation framework can compete with the results achieved by balanced truncation.

6.2 A Heat Transfer Model

The second application is a boundary controlled heat transfer system which has already been used for bilinear model reduction purposes several times, see e.g. [4, 6]. Here, the bilinear structure is due to mixed Dirichlet and Robin boundary conditions imposed on a single side of a square plate. In [5], the authors have shown that these conditions are met by e.g. spraying intensities of a cooling fluid which can be regulated. Since the process initially is modeled by a partial differential equation, the problem is spatially discretized by finite differences on a $k \times k$ grid, followed by the semi-implicit Euler method ($h = 0.005$) discussed in Section 5. Figure 2 shows the relative $\mathcal{H}_2$-error for $k = 15$ (hence, $n = 225$) and varying dimensions of the reduced order model, $r = 1, \ldots, 13$. Here, we have used boundary control for the left and the lower boundary. Even though the error decreases faster in case of balanced truncation, the performance of B-IRKA is surprisingly
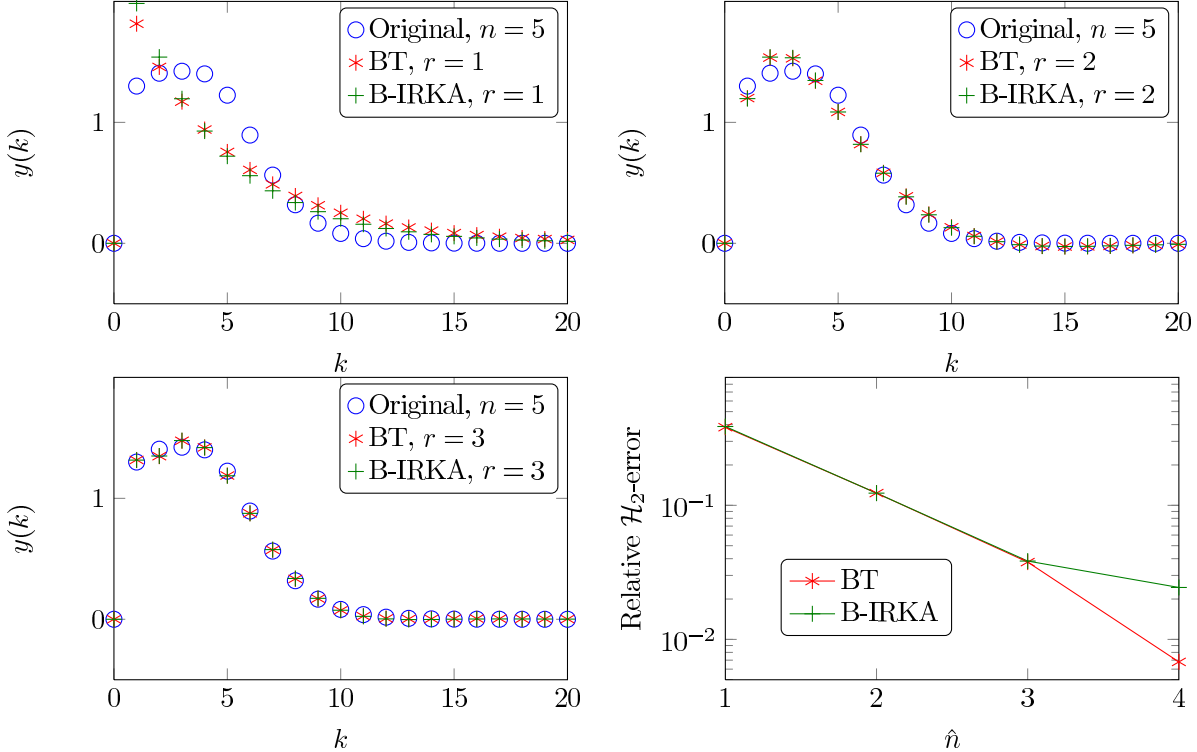
Figure 1: Transient response of Hinamoto model to a unit step input and relative $\mathcal{H}_2$-error for balanced truncation and B-IRKA.

well and certainly outperforms the random initial data. Moreover, the transient response of the system, which was chosen to be the average temperature, is faithfully approximated by the reduced system. Note that the only moderate size of the original system is due to the high complexity of the method of balanced truncation. In Figure 3 we refined the grid, using $k = 500$ (hence, $n = 250500$), making it impracticable to reduce the system by balanced truncation. However, it should be mentioned that we had some problems with the convergence of the algorithm. This explains also the missing values for some system dimensions in Figure 2.

6.3   A Nonlinear RC Circuit

The last example is an electrical circuit equipped with nonlinear resistors which is approximated by an augmented bilinear system by help of Carleman bilinearization. In the context of bilinear reduction techniques, this has been proven to be the most common test example and a more detailed explanation of this procedure can be found e.g. in [3]. A discrete-time system again was designed by the semi-implicit Euler method ($h = 0.01$). As shown in Figure 4, we only compared the new algorithm with its random initial data. Here, we have used a total of 100 resistors, leading to a large-scale bilinearized system of dimension $n = 10100$, which could not be reduced by balanced truncation.

7   Conclusions

In this paper, we have discussed an interpolation-based reduction approach for discrete-time bilinear control systems. Based on an explicit solution formula for multiple-input and multiple-output systems, we generalized the concepts of bilinear transfer functions known from the continuous-time case. This was achieved by multivariate Z-transforms and opened up the possibility of a generalized tangential interpolation method. Moreover, we have generalized the $\mathcal{H}_2$-norm for bilinear systems. Although we did not compute optimality
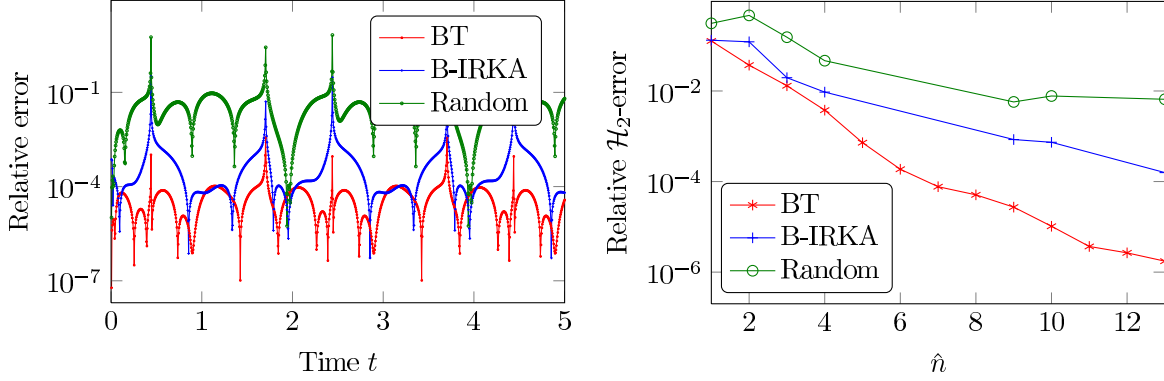
Figure 2: Relative error of heat transfer model ($n = 225$) to an input $u(t) = (\cos(\pi t), \cos(2\pi t))^T$ and relative $\mathcal{H}_2$-error for balanced truncation, B-IRKA and randomly chosen initial points.
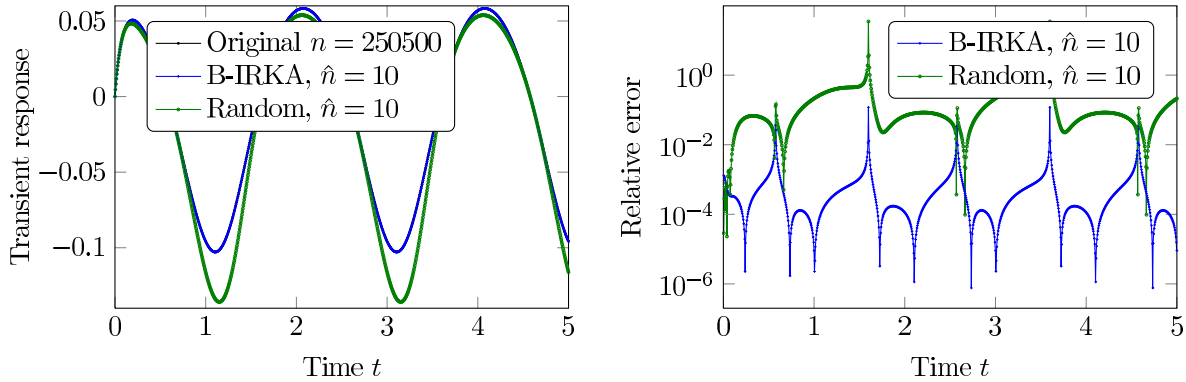

Figure 3: Transient response and relative error of heat transfer model ($n = 250500$) to an input $u(t) = (\cos(\pi t), \cos(2\pi t))^T$ for B-IRKA and randomly chosen initial points.

conditions, we provided a new error expression explaining the norm of the error system. This lead to the idea of generalizing the iterative rational Krylov algorithm (IRKA/MIRIAm) for bilinear systems. As has been shown in our examples, we could improve the approximation quality for the approach of generalized tangential interpolation. Moreover, the fact that this method can be implemented efficiently, resulted in the possibility of reducing very large-scale bilinear systems which could not be handled by the method of balanced truncation. However, deriving first order $\mathcal{H}_2$-optimality conditions should be further studied in order to improve results.

References

[1] S. Al-Baiyat. Model reduction of bilinear systems described by input-output difference equation. *Internat. J. Syst. Sci.*, 35(9):503–510, 2004.

[2] Z. Bai. Krylov subspace techniques for reduced-order modeling of nonlinear dynamical systems. *Appl. Numer. Math.*, 43:9–44, 2002.

[3] Z. Bai and D. Skoogh. A projection method for model reduction of bilinear dynamical systems. *Linear Algebra Appl.*, 415(2–3):406–425, 2006.

[4] P. Benner and T. Damm. Lyapunov equations, energy functionals and model order reduction of bilinear and stochastic systems. *SIAM J. Cont. Optim.*, 2011. to appear.
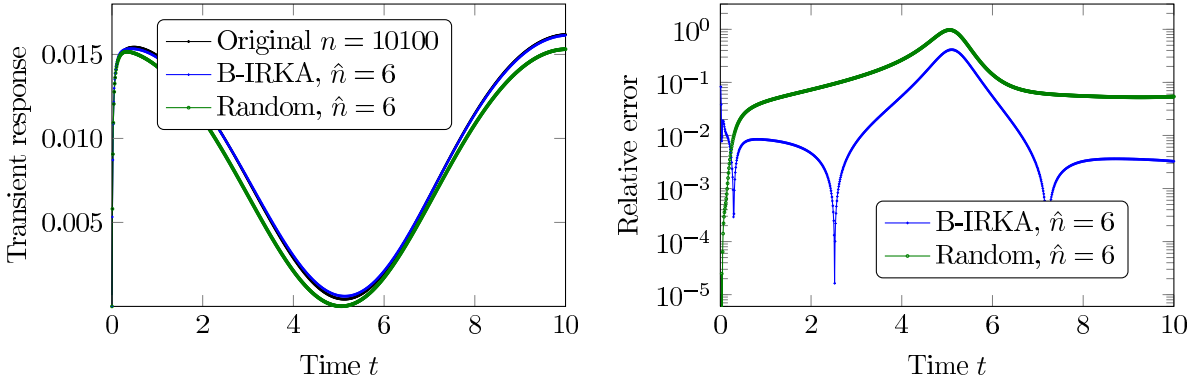
Figure 4: Transient response and relative error of bilinearized RC circuit for an input $u(t) = \frac{1}{2}\left(\cos\left(\frac{\pi t}{5}\right) + 1\right)$ B-IRKA and randomly chosen initial points.

[5] P. Benner and J. Saak. Linear-quadratic regulator design for optimal cooling of steel profiles. Technical Report SFB393/05-05, Sonderforschungsbereich 393 *Parallele Numerische Simulation f??r Physik und Kontinuumsmechanik*, TU Chemnitz, 09107 Chemnitz, FRG, 2005. Available from `http://www.tu-chemnitz.de/sfb393`.

[6] T. Breiten and T. Damm. Krylov subspace methods for model order reduction of bilinear control systems. *Sys. Control Lett.*, 59(8):443–450, 2010.

[7] C. Bruni, G. DiPillo, and G. Koch. On the mathematical models of bilinear systems. *Automatica*, 2:11–26, 1971.

[8] A. Bunse-Gerstner, D. Kubalinska, G. Vossen, and D. Wilczek. h2-norm optimal model reduction for large scale discrete dynamical MIMO systems. *J. Comput. Appl. Math.*, 233(5):1202–1216, 2010.

[9] M. Condon and R. Ivanov. Krylov subspaces from bilinear representations of nonlinear systems. *COMPEL*, 26:11–26, 2007.

[10] A. Dunoyer, L. Balmer, K.J. Burnham, and D.J.G. James. On the discretization of single-input single-output bilinear systems. *Int. J. Control*, 68(2):361–372, 1997.

[11] S.K. Godunov. *Modern aspects of linear algebra.* Amer. Mathematical Society, 1998.

[12] S. Gugercin, A. C. Antoulas, and S. Beattie. $\mathcal{H}_2$ Model Reduction for large-scale dynamical systems. *SIAM J. Matrix Anal. Appl.*, 30(2):609–638, 2008.

[13] T. Hinamoto and S. Maekawa. Approximation of polynomial state-affine discrete-time systems. *IEEE Trans. Circuits and Systems*, 31:713–721, 1984.

[14] R.R. Mohler. *Bilinear Control Processes.* New York Academic Press, 1973.

[15] R.R. Mohler. *Nonlinear systems (vol. 2): applications to bilinear control.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1991.

[16] R.R. Mohler. Natural bilinear control processes. *IEEE Transactions on Systems Science and Cybernetics*, 6(3):192–197, 2007.

[17] J. R. Phillips. Projection frameworks for model reduction of weakly nonlinear systems. In *Proceedings of DAC 2000*, pages 184–189, 2000.

[18] J. R. Phillips. Projection-based approaches for model reduction of weakly nonlinear, time-varying systems. *IEEE Trans. Circuits and Systems*, 22(2):171–187, 2003.

[19] W.J. Rugh. *Nonlinear System Theory*. The John Hopkins University Press, 1982.

[20] H. Schwarz. Stability of discrete-time equivalent homogeneous bilinear systems. *Contr. Theor. Adv. Tech.*, 3:263–269, 1987.

[21] L. Zhang and J. Lam. On $\mathcal{H}_2$ model reduction of bilinear systems. *Automatica*, 38(2):205–216, 2002.

[22] L. Zhang, J. Lam, B. Huang, and G. Yang. On gramians and balanced truncation of discrete-time bilinear systems. *Int. J. Control*, 76(4):414–427, 2003.