Interpolatory Model Reduction for Quadratic-Bilinear Systems using Error Estimators

Mian Ilyas Ahmad[1,*], Peter Benner[2,3], Lihong Feng[2]

[1]Research Center for Modeling and Simulation, National University of Science and Technology, Islamabad, Pakistan
[2]Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany
[3]Otto-von-Guericke Universität Magdeburg, Germany
[*]m.ilyas@rcms.nust.edu.pk

Abstract: We propose an interpolation based projection framework for model reduction of quadratic-bilinear descriptor systems. The approach constructs projection matrices from the bilinear part of the original quadratic-bilinear descriptor system and uses these matrices to project the original system. The projection matrices are constructed by viewing the bilinear system as a linear parametric system, where the input associated with the bilinear part is treated as a parameter. The advantage of this approach is that the projection matrices can be constructed reliably by using an a posteriori error bound for linear parametric systems. The use of the error bound allows us to select a good choice of interpolation points and parameter samples for the construction of the projection matrices by employing a greedy-type framework. The results are compared with the standard quadratic-bilinear projection methods and it is observed that the approximations through the proposed methods are comparable to the standard method, but at a lower computational cost ("offline time"). Furthermore, we extend a one-sided parametric moment matching method to a two-sided method that doubles the number of moments matched in the reduction method.

1. Introduction

We consider interpolatory model reduction techniques for a single-input single-output quadratic-bilinear descriptor system of the form:

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Nx(t)u(t) + Hx(t) \otimes x(t) + Bu(t), \\ y(t) &= Cx(t), \end{aligned} \tag{1}$$

where E, A, $N \in \mathbb{R}^{n \times n}$, $H \in \mathbb{R}^{n \times n^2}$, B, $C^T \in \mathbb{R}^n$ are the coefficient matrices and vectors. $x(t) \in \mathbb{R}^n$ is the state vector and $u(t), y(t) \in \mathbb{R}$ are the input and output of the system. A large class of nonlinear systems can be written in quadratic-bilinear form by using exact transformations [19], in addition to systems that naturally appear in the quadratic-bilinear form. This includes nonlinear electrical circuits, biochemical rate equations, reaction diffusion equations and fluid mechanics, see for example [19, 9]. Note that the notion "quadratic" is due to the term $Hx(t) \otimes x(t)$ and the "bilinear" is due to the term $Nx(t)u(t)$, which is linear in $x(t)$ and linear in $u(t)$ separately, but together it is bilinear. Simulation, control and optimization of such nonlinear systems are computationally inefficient for large n. A remedy to this issue is the use of model order reduction.

The goal of model order reduction is to construct a reduced system of dimension $r \ll n$:

$$E_r \dot{x}_r(t) = A_r x_r(t) + N_r x_r(t) u(t) + H_r x_r(t) \otimes x_r(t) + B_r u(t),$$
$$y_r(t) = C_r x_r(t), \tag{2}$$

with the output response $y_r(t)$ approximately equal to $y(t)$. Various techniques have been proposed in the literature to compute such reduced-order models, cf. [4, Chapters 7-12] and [10]. In this paper, we discuss projection-based interpolatory techniques [18, 5] that are well used in the linear case and are recently extended to quadratic-bilinear systems [19, 9, 2]. Projection involves approximating the state vector $x(t)$ in an r-dimensional subspace spanned by the column vectors of $V \in \mathbb{R}^{n \times r}$, so that the residual in the state equation is orthogonal to another r-dimensional subspace spanned by the column vectors of $W \in \mathbb{R}^{n \times r}$. That is, we approximate $x(t) \approx V x_r(t)$ such that the Petrov-Galerkin orthogonality condition holds:

$$W^T \Big(EV\dot{x}_r(t) - \big(AVx_r(t) + NVx_r(t)u(t) + HVx_r(t) \otimes Vx_r(t) + Bu(t)\big)\Big) = 0,$$
$$\hat{y}(t) = CVx_r(t).$$

If $W = V$, the projection is orthogonal and is often called one-sided projection, otherwise it is oblique and called two-sided projection. The oblique projection framework leads to a set of reduced system matrices of the form:

$$E_r = W^T EV, \quad A_r = W^T AV, \quad H_r = W^T H(V \otimes V), \quad N_r = W^T NV,$$
$$B_r = W^T B, \quad C_r = CV. \tag{3}$$

Analogous to the linear case, the basis matrices V and W can be selected such that the input-output representation in frequency domain is interpolated at some predefined interpolation points. However, unlike linear systems, quadratic-bilinear systems involve a series of multivariate transfer functions, each representing a subsystem of the original system. Thus, the problem is how to construct V and W such that the first K multivariate transfer functions associated with the reduced system are interpolating the corresponding original multivariate transfer functions, at multiple frequency shifts. To achieve this, orthogonal projections [19] as well as oblique projections [9] have been used in the literature with some simplifications. For example, the approach in [9] constructs V and W such that the reduced system ensures interpolation of the first two subsystems only. Also the choice of interpolation points for each frequency variable is assumed to be the same. We discuss these results further in Section 2.

Recently a new interpolation framework [3] for quadratic-bilinear systems has been proposed that relaxes the restrictions in [9]. The idea is to construct basis matrices V and W from the bilinear part of the quadratic-bilinear system and uses these matrices to project the quadratic-bilinear system. This means that if the projection approach interpolates the first K ($K \geq 2$) bilinear transfer functions, then a major part of each of the first K quadratic-bilinear transfer functions is exactly interpolated and the remaining parts are well approximated. It is observed in [3] that if the procedure of selecting the interpolation points is similar to the one used in [9], then the quality of the reduced model through this new approach is comparable to the direct method.

In this paper, we use this new projection approach for model order reduction (MOR) of quadratic-bilinear systems and identify a good choice of interpolation points for the bilinear part of the quadratic-bilinear system by utilizing a greedy type framework on the error bound expressions derived recently in [16]. The error bound expressions in [16] are derived for general linear parametric

systems by introducing primal and dual systems for identifying projection matrices and their error expressions. A similar primal and dual system formulation is also developed in [25], while greedy procedures to derive sampling points have a long tradition in reduced-basis methods, see e.g., [23] or [12, Chapters 2, 3]. In our proposed method, we first rewrite the bilinear part of the quadratic-bilinear system as an equivalent linear parametric system and then iteratively identify a set of interpolation points corresponding to the maximum value of the error bound, starting from an initial interpolation point. For each choice of interpolation points, we interpolate not only the original transfer function and its first derivate, but also higher derivatives, so that the reduced bilinear system is well approximated. The iteration stops when the approximation error is less than the prescribed tolerance level. Each of the iterations contributes to constructing a better set of basis matrices, that are then used for oblique projection of the original quadratic-bilinear system.

The remaining part of the paper is organized as follows. Section 2 shows projection techniques, including direct and indirect approaches, for quadratic-bilinear systems. Section 3 presents moment matching techniques for the bilinear part of the quadratic-bilinear system. In particular, we extend the one-sided projection method in [11] to a two-sided method. Section 4 shows the error bound expression and a greedy type algorithm for selecting the interpolation points. Finally in Section 5, numerical results are shown for some benchmark examples.

2. MOR for Quadratic-Bilinear Systems

In this section, we briefly review the concept of projection based interpolation for quadratic-bilinear systems. As discussed in the introduction, analogous to bilinear systems, quadratic-bilinear systems in frequency domain involve multivariate transfer functions, see [1] for the bilinear case and [19, 9] for the quadratic-bilinear case. The structure of the first two multivariate transfer functions in the so called symmetric form [24, Chapter 3] is given as:

$$\mathbf{H}_1(s_1) = C(s_1 E - A)^{-1} B =: CF(s_1)^{-1}B =: CG_1(s_1),$$

$$\mathbf{H}_2(s_1, s_2) = \frac{1}{2!}CF(s_1 + s_2)^{-1}\Big(N\big[G_1(s_1) + G_1(s_2)\big] +$$

$$H\big[G_1(s_1)\otimes G_1(s_2) + G_1(s_2)\otimes G_1(s_1)\big]\Big),$$

$$=: CG_2(s_1, s_2),$$

(4)

where $F(s) := sE - A$. For general k-variate transfer functions, the problem of interpolation can be written as:

$$\mathbf{H}_k(\sigma_{1i}, \ldots, \sigma_{ki}) = \hat{\mathbf{H}}_k(\sigma_{1i}, \ldots, \sigma_{ki}),$$

(5)

where $\sigma_{ki} \in \mathbb{C}$ are the interpolation points or shift frequencies for $k = 1, \ldots, K$ and $i = 1, \ldots, n_r$. Since the structure of the multivariate transfer functions is too complex for $K > 2$, often in the literature simplifications are used in the projection framework. The following are the two main assumptions often employed in the existing projection techniques:

1. The value of K is restricted to 2, so that only the first two transfer functions, as given in (4), are interpolated.

2. All the frequency variables s_k's are varying in the same way, meaning that $s_1 = \ldots = s_k = s$, so that we need a single set of interpolation points for all multivariate transfer functions.

With these assumptions, the multivariate transfer functions in (4) become

$$\mathbf{H}_1(s) = C(sE - A)^{-1}B =: CF(s)^{-1}B,$$
$$\mathbf{H}_2(s, s) = C(2sE - A)^{-1}\Big(NG_1(s) + H[G_1(s) \otimes G_1(s)]\Big),$$

(6)

and the interpolation problem reduces to:

$$\hat{\mathbf{H}}_1(\sigma_j) = \mathbf{H}_1(\sigma_j), \quad \hat{\mathbf{H}}_2(\sigma_j, \sigma_j) = \mathbf{H}_2(\sigma_j, \sigma_j),$$

where $\sigma_j \in \mathbb{R}$ for $j = 1, \ldots, n_r$. Based on the results in [19], it is possible to construct the basis matrix V for a union of shifted Krylov subspaces that ensure the above interpolation conditions. Recently, these results were extended in [9] to an oblique projection framework so that the reduced model also matches the derivatives of the first two transfer functions. In the following, we briefly review the projection technique introduced in [9].

Notice that $\mathbf{H}(s_1)$ is a simple univariate transfer function and interpolating such a linear transfer function is well investigated in the literature. To ensure, in addition, interpolation of the second transfer function and its derivatives, the authors in [9] utilized tensor theory to introduce a special identity for the Hessian H. That is

$$w^T H(u \otimes v) = u^T H^{(2)}(v \otimes w),$$

with $u, v, w \in \mathbb{R}^n$ and where $H^{(2)}$ is the mode-2 matricization of a 3-dimensional tensor $\mathcal{H} \in \mathbb{R}^{n \times n \times n}$, see [21] for details on tensor theory. The mode-1 matricization $H = H^{(1)}$ of the tensor $\mathcal{H}$ also satisfies the commutativity rule: $H(u \otimes v) = H(v \otimes u)$. These observations are used to prove the main result of [9], that identifies a good choice of V and W:

Proposition 2.1. [9] Let $\sigma_i \in \mathbb{C}$ be the interpolation points and $\sigma_i \notin \{\Lambda(A, E), \Lambda(\hat{A}, \hat{E})\}$, where $\Lambda(A, E)$ represents the generalized eigenvalues of the matrix pencil $\lambda E - A$. Assume that $\hat{E} = W^T EV$ is nonsingular and $\hat{A}, \hat{H}, \hat{N}, \hat{B}, \hat{C}$ are as in (3) with full rank matrices $V, W \in \mathbb{R}^{n \times r}$ such that

$$\text{span}(V) = \text{span}_{i=1,\ldots,n_r} \{F(\sigma_i)^{-1}B, F(2\sigma_i)^{-1}[H(F(\sigma_i)^{-1}B \otimes F(\sigma_i)^{-1}B) + NF(\sigma_i)^{-1}B]\}$$

$$\text{span}(W) = \text{span}_{i=1,\ldots,n_r} \{F(2\sigma_i)^{-T}C^T, F(\sigma_i)^{-T}[H^{(2)}(F(\sigma_i)^{-1}B \otimes F(2\sigma_i)^{-T}C^T)$$

$$+ \frac{1}{2}N^T F(2\sigma_i)^{-T}C^T]\}.$$

Then the reduced quadratic-bilinear differential algebraic equation (QBDAE) satisfies the following interpolation conditions:

$$\mathbf{H}_1(\sigma_i) = \hat{\mathbf{H}}_1(\sigma_i), \qquad \mathbf{H}_1(2\sigma_i) = \hat{\mathbf{H}}_1(2\sigma_i),$$
$$\mathbf{H}_2(\sigma_i, \sigma_i) = \hat{\mathbf{H}}_2(\sigma_i, \sigma_i), \quad \frac{\partial}{\partial s_l}\mathbf{H}_2(\sigma_i, \sigma_i) = \frac{\partial}{\partial s_l}\hat{\mathbf{H}}_2(\sigma_i, \sigma_i),$$

(7)

where $l = 1, 2$ and $i = 1, \ldots, n_r$.

Proposition 2.1 suggests that oblique projection techniques can ensure interpolation as in (7), but with the two assumptions discussed before. It is also possible to match higher derivatives as

4

shown in [8], however in that case the size of the reduced model grows much faster than the number of derivatives matched.

An alternate oblique projection technique [3] for quadratic-bilinear systems is to interpolate the bilinear part of the multivariate transfer functions without the restrictions of interpolating only the first two subsystems with a single set of interpolation points. To see the structure of the bilinear part, note that each of the multivariate transfer functions associated with the quadratic-bilinear system can be divided into three parts: the bilinear part, the quadratic part and the quadratic-bilinear part,

$$\mathbf{H}_k(s_1, \ldots, s_k) = \mathbf{H}_{k_B}(s_1, \ldots, s_k) + \mathbf{H}_{k_Q}(s_1, \ldots, s_k) + \mathbf{H}_{k_{QB}}(s_1, \ldots, s_k). \tag{8}$$

The quadratic-bilinear part is nonzero for $k > 2$ implicating that the direct interpolation approach [9] completely ignores this part of the multivariate transfer function. The quadratic part is nonzero for $k > 1$. This means that for $k = 1$, the original transfer function is exactly equal to the bilinear part: $\mathbf{H}_1(s_1) = \mathbf{H}_{1_B}(s_1)$. For $k = 2$, we have the bilinear and the quadratic parts, $\mathbf{H}_2(s_1, s_2) = \mathbf{H}_{2_B}(s_1, s_2) + \mathbf{H}_{2_Q}(s_1, s_2)$ where

$$\mathbf{H}_{2_B}(s_1, s_2) = \frac{1}{2!}CF(s_1 + s_2)^{-1}N\big[G_1(s_1) + G_1(s_2)\big] =: CG_{2_B}(s_1, s_2),$$

$$\mathbf{H}_{2_Q}(s_1, s_2) = \frac{1}{2!}CF(s_1 + s_2)^{-1}Q\big[G_1(s_1) \otimes G_1(s_2) + G_1(s_2) \otimes G_1(s_1)\big]$$
$$=: CG_{2_Q}(s_1, s_2) \tag{9}$$

Similarly for $k = 3$, we have to consider all the three parts,

$$\mathbf{H}_3(s_1, s_2, s_3) = \mathbf{H}_{3_B}(s_1, s_2, s_3) + \mathbf{H}_{3_Q}(s_1, s_2, s_3) + \mathbf{H}_{3_{QB}}(s_1, s_2, s_3),$$

in which

$$\mathbf{H}_{3_B}(s_1, s_2, s_3) = \frac{1}{3!}CF(s_1 + s_2 + s_3)^{-1}N\big[G_{2_B}(s_1, s_2) + G_{2_B}(s_1, s_3) + G_{2_B}(s_2, s_3)\big]$$

$$\mathbf{H}_{3_Q}(s_1, s_2, s_3) = \frac{1}{3!}CF(s_1 + s_2 + s_3)^{-1}H\big[G_1(s_1) \otimes G_{2_Q}(s_2, s_3)$$
$$+ G_{2_Q}(s_2, s_3) \otimes G_1(s_1) + G_1(s_2) \otimes G_{2_Q}(s_1, s_3) + G_{2_Q}(s_1, s_3) \otimes G_1(s_2)$$
$$+ G_1(s_2) \otimes G_{2_Q}(s_1, s_3) + G_{2_Q}(s_1, s_3) \otimes G_1(s_2)\big]$$

$$\mathbf{H}_{3_{QB}}(s_1, s_2, s_3) = \frac{1}{3!}CF(s_1 + s_2 + s_3)^{-1}N\big[G_{2_Q}(s_1, s_2) + G_{2_Q}(s_1, s_3) + G_{2_Q}(s_2, s_3)$$
$$+ H\big[G_1(s_1) \otimes G_{2_B}(s_2, s_3) + G_{2_B}(s_2, s_3) \otimes G_1(s_1) + G_1(s_2) \otimes G_{2_B}(s_1, s_3)$$
$$+ G_{2_B}(s_1, s_3) \otimes G_1(s_2) + G_1(s_2) \otimes G_{2_B}(s_1, s_3) + G_{2_B}(s_1, s_3) \otimes G_1(s_2)\big]$$
$$\tag{10}$$

Thus the general form of the bilinear part for $k > 1$ has the following structure:

$$\mathbf{H}_{k_B}(s_1, \ldots, s_k) = \frac{1}{k!}C\big((\sum_{j=1}^{k} s_j)E - A\big)^{-1}\Big(N\big[\sum_{\pi(\cdot)} G_{k-1_B}(s_{\pi(1)}, \ldots, s_{\pi(k-1)})\big]\Big), \tag{11}$$

where the second summation with $\pi(\cdot)$ is over all $(k-1)!$ factorials of the integers 1 to $k-1$. It is easy to see that $\mathbf{H}_{k_B}(s_1, \ldots, s_k)$ corresponds to the symmetric form of the multivariate transfer

5

functions associated with a bilinear system as in (1) with H being a zero matrix. This means that the bilinear part corresponds to the following system

$$
\begin{aligned}
E\dot{x}_b(t) &= Ax_b(t) + Nx_b(t)u(t) + Bu(t), \\
y_b(t) &= Cx_b(t).
\end{aligned}
\tag{12}
$$

It follows that one can use bilinear interpolatory techniques [13] to identify a choice of projection matrices $V_B \in \mathbb{R}^{n \times r}$ and $W_B \in \mathbb{R}^{n \times r}$ such that

$$
\text{span}(V_B) = \underset{i=1,\ldots,n_r}{\text{span}} \{F(\sigma_i)^{-1}B, F(2\sigma_i)^{-1}NF(\sigma_i)^{-1}B\}
$$

$$
\text{span}(W_B) = \underset{i=1,\ldots,n_r}{\text{span}} \{F(2\sigma_i)^{-T}C^T, \frac{1}{2}F(\sigma_i)^{-T}N^TF(2\sigma_i)^{-T}C^T]\},
$$

satisfying

$$
\mathbf{H}_{k_B}(\sigma_{1i}, \ldots, \sigma_{ki}) = \hat{\mathbf{H}}_{k_B}(\sigma_{1i}, \ldots, \sigma_{ki}),
\tag{13}
$$

where $\sigma_{ji} \in \mathbb{C}$ are the interpolation points. Here, in the bilinear case we can avoid the restrictions used in the quadratic-bilinear interpolation. So, if we use V_B and W_B as the basis matrices to construct reduced system matrices:

$$
\hat{E} = W_B^T EV_B, \quad \hat{A} = W_B^T AV_B, \quad \hat{H} = W_B^T H(V_B \otimes V_B), \quad \hat{N} = W_B^T NV_B,
$$
$$
\hat{B} = W_B^T B, \quad \hat{C} = CV_B,
\tag{14}
$$

then we can not only ensure (13) but also hope to approximate the quadratic and the quadratic-bilinear parts:

$$
\begin{aligned}
\mathbf{H}_{k_Q}(\sigma_{1i}, \ldots, \sigma_{ki}) &\approx \hat{\mathbf{H}}_{k_Q}(\sigma_{1i}, \ldots, \sigma_{ki}), \\
\mathbf{H}_{k_{QB}}(\sigma_{1i}, \ldots, \sigma_{ki}) &\approx \hat{\mathbf{H}}_{k_{QB}}(\sigma_{1i}, \ldots, \sigma_{ki}).
\end{aligned}
\tag{15}
$$

Remarks 2.1. In contrast to the projection framework discussed in Proposition 2.1, here the basis matrices V_B and W_B are constructed without the $n \times n^2$ quadratic matrices H and $H^{(2)}$. This saves significant computational cost for each sparse linear system that is to be solved for a fixed interpolation point. Once V_B and W_B are constructed, only then the quadratic matrix H is used as in (14) to obtain the reduced model.

It is observed in [3] that a good choice of bilinear basis matrices V_B and W_B improves the quality of the reduced quadratic-bilinear system. This means that we require a good choice of interpolation points μ_{ki} with which we can identify a reduced model that in addition to ensuring (13), also matches some higher derivatives of $\mathbf{H}_B(s_1, \ldots, s_k)$. That is

$$
\frac{\partial^j}{\partial s_l^j}\mathbf{H}_{k_B}(\sigma_{1i}, \ldots, \sigma_{ki}) = \frac{\partial^j}{\partial s_l^j}\hat{\mathbf{H}}_{k_B}(\sigma_{1i}, \ldots, \sigma_{ki}),
\tag{16}
$$

where $l = 1, \ldots, k$ and j is the order of the partial derivatives.

One possibile approach would be to use the two-sided bilinear iterative rational Krylov algorithm (BIRKA) [7] that iteratively updates a set of interpolation points such that the reduced model satisfies the necessary conditions [26] for an $\mathcal{H}_2$-optimal reduced model. The BIRKA method iteratively converges to a reduced model that not only ensures (13) but also matches the first derivative of $\mathbf{H}_B(s_1, \ldots, s_k)$ with $j = 1$ in (16). However, the computational cost of BIRKA for large n and

r values is highly expensive and also it is not possible to match higher derivatives. Regarding the computational cost of BIRKA, some effort has been done in [17] to reduce the computational cost. In the next section, we propose to apply error bounds for adaptive selection of interpolation points σ_{ki} and identification of projection matrices V_B and W_B that also match some of the higher order derivatives $\mathbf{H}_B(s_1, \ldots, s_k)$, as shown in (16) with $j \geq 2$.

3. Moment matching techniques

For the purpose of model reduction, we can consider a linear parametric system instead of the bilinear system (12) by replacing the control in the bilinear system formally by a parameter.

$$
\begin{aligned}
E\dot{x}_p(t) &= Ax_p(t) + pNx_p(t) + Bu(t), \\
y_p(t) &= Cx_p(t),
\end{aligned}
\tag{17}
$$

The Laplace transform of the linear parametric system in the frequency domain yields the transfer function

$$
\mathbf{H}(s, p) = C(sE - A - pN)^{-1}B.
\tag{18}
$$

Since in the linear parametric form, we have a posteriori error bounds on model reduction of $H(s, p)$, we are interested in using a linear parametric model order reduction (PMOR) method for computing the projection matrices. Thus the problem is to identify a choice of projection matrices $V_p \in \mathbb{R}^{n \times r}$ and $W_p \in \mathbb{R}^{n \times r}$ such that

$$
\mathbf{H}(\sigma_i, p_i) = \hat{\mathbf{H}}(\sigma_i, p_i),
\tag{19}
$$

where $p_i \in \mathbb{C}$ are some predefined fixed parameters, and construct reduced quadratic-bilinear system matrices as in (14) using V_p and W_p as projection matrices. In this way, the bilinear and the quadratic state matrices are involved in the reduction process and the reduced system is in the quadratic-bilinear form.

In the following, we discuss multi-moment matching PMOR techniques for computation of V_B and W_B. Multi-moment matching PMOR methods for linear parametric systems are well used in the literature [15, 20, 11]. In this section, we review a numerically efficient method [11] for multi-moment matching of a linear parametric system. We extend this framework to oblique projection which theoretically allows us to double the number of multi-moments being matched. Before going to the derivation of multi-moment matching, it is important to clarify the following point.

Remarks 3.1. Since the multi-moments are related to the corresponding transfer function or their derivatives evaluated at the interpolation point, multi-moment matching properties of the reduced bilinear system correspond to the ability of the reduced model to ensure multivariate interpolation conditions as discussed before.

As discussed in [18], projection based approximation of $H(s, p)$ can be viewed as the approximation of two shifted linear systems given as:

$$
(sE - A - pN)x(s, p) = B,
\tag{20}
$$

$$
(sE - A - pN)^T y(s, p) = C^T,
\tag{21}
$$

since $H(s, p) = y(s, p)^T (sE - A - pN)x(s, p)$. The state $x(s, p)$ can be expanded into a Taylor series at an expansion point $\mu_0 = (\sigma_0, p_0)$ as

$$\begin{aligned} x(s, p) &= [I - (\delta_1 M_1 - \delta_2 M_2)]^{-1} B_M, \\ &= B_M + (\delta_1 M_1 - \delta_2 M_2) B_M + (\delta_1 M_1 - \delta_2 M_2)^2 B_M + \dots \end{aligned} \tag{22}$$

where, if $\tilde{E} = (\sigma_0 E - A - p_0 N)$, then $M_1 = \tilde{E}^{-1} E$, $M_2 = \tilde{E}^{-1} N$ and $B_M = \tilde{E}^{-1} B$. Also $\delta_1 = -(s - \sigma_0)$ and $\delta_2 = -(p - p_0)$. The coefficient matrices in the above series multiplied by C from the left are called the multi-moments of the parameterized system at μ_0. Now by defining some auxiliary variables as

$$\begin{aligned} x(s, p) &= x_0 + x_1 + x_2 + \dots + x_j + \dots, \quad \text{where } x_0 = B_M, \\ x_1 &= (\delta_1 M_1 - \delta_2 M_2)x_0, \ x_2 = (\delta_1 M_1 - \delta_2 M_2)x_1, \dots \\ x_j &= (\delta_1 M_1 - \delta_2 M_2)x_{j-1}, \dots, \end{aligned}$$

we can easily find a recursive method to compute the coefficient matrices implicitly. An orthonormal projection matrix $V_{p\mu_0}$ of the subspace spanned by the coefficient matrices can then be identified as below,

$$\begin{aligned} \text{range}\left(V_{p\mu_0}\right) &= \text{span}\{R_0, R_1, \dots, R_q\}_{\mu_0}, \\ \text{where } R_0 &= B_M, \ R_1 = [M_1 R_0, M_2 R_0], \ R_2 = [M_1 R_1, M_2 R_1], \dots, \\ R_q &= [M_1 R_{q-1}, M_2 R_{q-1}]. \end{aligned}$$

Likewise, projection matrices corresponding to other interpolation points μ_i, $i = 1, \dots, m$, can be computed. The final projection matrix V_p becomes

$$V_p = \text{orth}\{V_{p\mu_0}, \dots, V_{p\mu_m}\}, \tag{23}$$

where "orth" means orthogonalization of the columns in $V_{p\mu_0}, \dots, V_{p\mu_m}$. It is shown in [11] that the above choice of projection matrix V_B ensures that the multi-moments of the linear parametric system are implicitly matched by the reduced system. In the following, we extend these results by showing multi-moment matching properties in the oblique projection framework.

To see this, we now do a Taylor series expansion for $y(s, p)$ in (21) at an expansion point $\mu_0 = (\sigma_0, p_0)$:

$$\begin{aligned} y(s, p) &= [I - (\delta_1 L_1 - \delta_2 L_2)]^{-1} C_L, \\ &= C_L + (\delta_1 L_1 - \delta_2 L_2) C_L + (\delta_1 L_1 - \delta_2 L_2)^2 C_L + \dots, \end{aligned} \tag{24}$$

where, $L_1 = \tilde{E}^{-T} E^T$, $L_2 = \tilde{E}^{-T} N^T$ and $C_L = \tilde{E}^{-T} C^T$. Also $\delta_1 = -(s - \sigma_0)$ and $\delta_2 = -(p - p_0)$. This implies that analogous to the computation of V_p, we can compute the oblique projection matrix W_p as

$$W_p = \text{orth}\{W_{p\mu_0}, \dots, W_{p\mu_m}\} \tag{25}$$

where, $\text{range}\{W_{p\mu_i}\} = \text{span}\{T_0, T_1, \dots, T_q\}_{\mu_i}$ in which

$$\begin{aligned} T_0 &= C_L, \ T_1 = [L_1 T_0, L_2 T_0], \ T_2 = [L_1 T_1, L_2 T_1], \dots, \\ T_q &= [L_1 T_{q-1}, L_2 T_{q-1}]. \end{aligned}$$

Using V_p and W_p, we can construct a reduced-order model of the parametric system (17) as,

$$\hat{E}\dot{\hat{x}}(t) = \hat{A}\hat{x}(t) + p\hat{N}\hat{x}(t) + \hat{B}u(t),$$
$$\hat{y}(t) = \hat{C}\hat{x}(t), \tag{26}$$

with transfer function

$$\hat{\mathbf{H}} = \hat{C}(s\hat{E} - \hat{A} - p\hat{N})^{-1}\hat{B}. \tag{27}$$

To show that V_p and W_p ensure moment matching, i.e., $\hat{\mathbf{H}}(s, p)$ matches multi-moments of $\mathbf{H}(s, p)$, we consider the basis matrices corresponding to a fixed interpolation point μ_0. We first define coefficient matrices for the two reduced shifted linear systems:

$$\hat{R}_0 = \hat{B}_M, \ \hat{R}_1 = [\hat{M}_1\hat{R}_0, \hat{M}_2\hat{R}_0], \ \hat{R}_2 = [\hat{M}_1\hat{R}_1, \hat{M}_2\hat{R}_1], \ldots,$$
$$\hat{R}_q = [\hat{M}_1\hat{R}_{q-1}, \hat{M}_2\hat{R}_{q-1}], \ \hat{M}_1 = (W_p^T\tilde{E}V_p)^{-1}\hat{E}, \ \hat{M}_2 = (W_p^T\tilde{E}V_p)^{-1}\hat{N},$$
$$\hat{T}_0 = \hat{C}_L, \ \hat{T}_1 = [\hat{L}_1\hat{T}_0, \hat{L}_2\hat{T}_0], \ \hat{T}_2 = [\hat{L}_1\hat{T}_1, \hat{L}_2\hat{T}_1], \ldots, \tag{28}$$
$$\hat{T}_q = [\hat{L}_1\hat{T}_{q-1}, \hat{L}_2\hat{T}_{q-1}], \ \hat{L}_1 = (W_p^T\tilde{E}V_p)^{-T}\hat{E}^T, \ \hat{L}_2 = (W_p^T\tilde{E}V_p)^{-T}\hat{N}^T,$$

where $\hat{B}_M = (W_p^T\tilde{E}V_p)^{-1}\hat{B}$ and $\hat{C}_L = (W_p^T\tilde{E}V_p)^{-T}\hat{C}^T$. Now since

$$\text{range}\{V_p\} = \text{span}\{R_0, R_1, \ldots, R_q\} =: \mathcal{V},$$
$$\text{range}\{W_p\} = \text{span}\{T_0, T_1, \ldots, T_q\} =: \mathcal{W}, \tag{29}$$

we can use the following results to show the moment matching properties.

Lemma 3.1. Let V_p and W_p be defined as in (29), then for any $v \in \mathcal{V}$ and $w \in \mathcal{W}$,

$$v = V_p(W_p^T\tilde{E}V_p)^{-1}W_p^T\tilde{E}v, \ \ w = W_p(V_p^T\tilde{E}^TW_p)^{-1}V_p^T\tilde{E}^Tw. \tag{30}$$

Proof. See [18] for a proof. □

Lemma 3.2. Let all variables be defined as in (28) and (29), then the following holds:

$$V_p\hat{R}_i = R_i, \ \ W_p\hat{T}_i = T_i. \tag{31}$$

Proof. We begin with $i = 0$. As $\hat{R}_0 = (W_p^T\tilde{E}V_p)^{-1}W_p^TB$ and $\tilde{E}^{-1}B = R_0 \in \mathcal{V}$, we have

$$V_p\hat{R}_0 = V_p(W_p^T\tilde{E}V_p)^{-1}W_p^T\tilde{E}\tilde{E}^{-1}B$$
$$= \tilde{E}^{-1}B, \tag{32}$$

where the last equation follows from Lemma 3.1. Also, since $\hat{T}_0 = (V_p^T\tilde{E}^TW_p)^{-1}V_p^TC^T$ and $\tilde{E}^{-T}C^T = T_0 \in \mathcal{W}$, we have

$$W_p\hat{T}_0 = W_p(V_p^T\tilde{E}^TW_p)^{-1}V_p^T\tilde{E}^T\tilde{E}^{-T}C^T$$
$$= \tilde{E}^{-T}C^T \qquad \text{(follows from Lemma 3.1).} \tag{33}$$

9

Thus Lemma 3.2 is true for $i = 0$. Next, we assume that the Lemma holds for $i \leq j$, so that $V_p \hat{R}_j = R_j$ and $W_B \hat{T}_j = T_j$. We will prove that Lemma 3.2 also holds for $i = j + 1 \leq q$.

$$
\begin{aligned}
V_p \hat{R}_{j+1} &= V_p[(W_p^T \tilde{E} V_p)^{-1} W_p^T E V_p \hat{R}_j, (W_p^T \tilde{E} V_p)^{-1} W_p^T N V_p \hat{R}_j] \\
&= [V_p (W_p^T \tilde{E} V_p)^{-1} W_p^T \tilde{E} \tilde{E}^{-1} E R_j, V_p (W_p^T \tilde{E} V_p)^{-1} W_p^T \tilde{E} \tilde{E}^{-1} N R_j] \\
&= [V_p (W_p^T \tilde{E} V_p)^{-1} W_p^T \tilde{E} M_1 R_j, V_p (W_p^T \tilde{E} V_p)^{-1} W_p^T \tilde{E} M_2 R_j] \\
&= [M_1 R_j, M_2 R_j] \qquad \text{(follows from Lemma 3.1)} \\
&= R_{j+1}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
W_p \hat{T}_{j+1} &= W_p[(V_p^T \tilde{E}^T W_p)^{-1} V_p^T E^T W_p \hat{T}_j, (V_p^T \tilde{E}^T W_p)^{-1} V_p^T N^T W_p \hat{T}_j] \\
&= [W_p (V_p^T \tilde{E}^T W_p)^{-1} V_p^T \tilde{E}^T \tilde{E}^{-T} E^T T_j, W_p (V_p^T \tilde{E}^T W_p)^{-1} V_p^T \tilde{E}^T \tilde{E}^{-T} N^T T_j] \\
&= [W_p (V_p^T \tilde{E}^T W_p)^{-1} V_p^T \tilde{E}^T L_1 T_j, W_p (V_p^T \tilde{E}^T W_p)^{-1} V_p^T \tilde{E}^T L_2 T_j] \\
&= [L_1 T_j, L_2 T_j] \qquad \text{(follows from Lemma 3.1)} \\
&= T_{j+1}.
\end{aligned}
$$

$\square$

Theorem 3.1. If V_p and W_p satisfy (29), then the reduced system implicitly matches all the multi-moments of the original system that are of the form $\hat{C}^T \hat{R}_i = C^T R_i$ and $\hat{T}_i^T W_p^T \tilde{E} V_p \hat{R}_i = T_i^T \tilde{E} R_i$ for $i = 0, \ldots, q$.

Proof. It follows from Lemma 3.2 that by multiplying C from the left, we have $C V_p \hat{R}_i = C R_i$ and by multiplying B^T from the left, we have $B^T W_p \hat{T}_i = B^T T_i$ for $i = 0, \ldots, m$. Also, using $V_p \hat{R}_i = R_i$ and $W_p \hat{T}_i = T_i$, we have $\hat{T}_i^T W_p^T \tilde{E} V_p \hat{R}_i = T_i^T \tilde{E} R_i$, which proves that additional multi-moments are also matched. $\square$

Remarks 3.2. The above moment-matching property is the property of the reduced model in (26), which in turn matches the multi-moments of the transfer function H_B in (11), the bilinear part of the quadratic-bilinear system in (1). The link between the linear parametric system and the bilinear system is also discussed in [6]. The overall reduced model of the original system in (1) produced by the PMOR method should be in the form of (3), with W, V being replaced by W_p and V_p.

In the above theorem, we used a single expansion point to show the moment matching property of the reduced model, but the same framework can also be used to obtain moment matching at multiple expansion points. The problem, however, is the selection of these expansion or interpolation points so that the reduced model accurately approximates the behaviour of the original system. In the next section, we present an algorithm for efficient selection of these interpolation points.

4. Selection of Interpolation Points using Error Bound

In this section, we address the problem of adaptively selecting the multiple expansion points using the error bound expression derived in [16] and show its use in the proposed two-sided projection

framework. The error bound estimates the error between $H(s,p)$ (18) and $\hat{H}(s,p)$ (27). To this end, we define the primal and dual systems as:

$$(sE - A - pN)x^{pr}(s,p) = B, \tag{34}$$

$$(sE - A - pN)^* x^{du}(s,p) = -C^T, \tag{35}$$

respectively, where * denotes the conjugate transpose. The primal system is the same as the shifted system in (20), whereas the dual system defined here has a minus sign on the right hand side, and the left-hand side is the complex conjugate of the one in (21). The error bound is constructed based on two residuals, which result from model reduction of the primal and dual systems, respectively. The primal system is reduced using the matrix pair W_p and V_p. Due to the dual relation between (34) and (35), we exchange W_p with V_p and vice versa to reduce the dual system. As a result, the reduced primal system is,

$$(s\hat{E} - \hat{A} - p\hat{N})\hat{x}^{pr}(s,p) = \hat{B}$$

where $\hat{E} = W_p^T E V_p$, $\hat{A} = W_p^T A V_p$, $\hat{N} = W_p^T N V_p$, $\hat{B} = W_p^T B$ and the reduced dual system is

$$(s\hat{E} - \hat{A} - p\hat{N})^* \hat{x}^{du}(s,p) = -\hat{C}^T,$$

where $\hat{E} = V_p^T E W_p$, $\hat{A} = V_p^T A W_p$, $\hat{N} = V_p^T N W_p$, $\hat{C} = V_p^T C^T$. The residuals associated with the reduction of the primal and dual systems can be written as

$$\begin{aligned} r^{pr}(s,p) &= B - (sE - A - pN)V_p\hat{x}^{pr}(s,p), \\ r^{du}(s,p) &= -C^T - (sE - A - pN)^* W_p\hat{x}^{du}(s,p). \end{aligned} \tag{36}$$

With these quantities, the following result provides an a posteriori upper bound on the approximation error, $|H(s,p) - \hat{H}(s,p)|$:

Theorem 4.1. [16] The upper bound on the approximation of $H(s,p) = C(sE - A - pN)^{-1}B$ can be written as $|H(s,p) - \hat{H}(s,p)| \leq \Delta(s,p)$, where

$$\Delta(s,p) := \frac{\|r^{du}(s,p)\|_2 \|r^{pr}(s,p)\|_2}{\beta(s,p)}, \tag{37}$$

in which $\beta(s,p) = \sigma_{\min}(G(s,p))$, where $\sigma_{\min}$ indicates the smallest singular value of $G(s,p) := sE - A - pN$.

The error bound can be used iteratively to identify a good choice of interpolation points in a predefined sample space, starting from an initial choice. The selection criteria is to choose the points in the sample space for which the error bound is maximized. The selected interpolation points are then used to construct and update the required basis matrices V_p and W_p, by using the multi-moment matching technique described before. This means that we are seeking a model that satisfies the bound pointwise, but globally in the whole considered (s,p) domain. Such a greedy framework for selection of interpolation points is shown in Algorithm 1.

Remarks 4.1. Notice that the proposed framework can also deal with multi-input multi-output (MIMO) systems, as the ik-th entry of the transfer matrix associated with the MIMO system corresponds to a single-input single-output transfer function. This is demonstrated by an example in the numerical results.

Algorithm 1 Greedy framework for selection of interpolation points
***
Inputs: σ_0, p_0, E, A, N, B, C and S_{sample}: a set of the samples of $\mu := (s, p)$, which covers the frequency and the parameter domain.

Outputs: μ, V_p, W_p, V_p^{du} and W_p^{du}

Initialization: $V_p = [\ ]$; $W_p = [\ ]$; $\epsilon = 1$; $i = -1$; $\epsilon_{tol} < 1$, $\mu^0 = (\sigma_0, p_0)$.

WHILE $\epsilon > \epsilon_{tol}$

   $i = i + 1$;

   compute $V_{p,i}(\mu^i)$ and $W_{p,i}(\mu^i)$ following (29)

   $V_p = [V_p, V_{p,i}]$; $W_p = [W_p, W_{p,i}]$; $W_p^{du} = V_p$; $V_p^{du} = W_p$;

   $\mu^i = \arg \max_{\mu \in S_{\text{sample}}} \Delta(\mu)$;

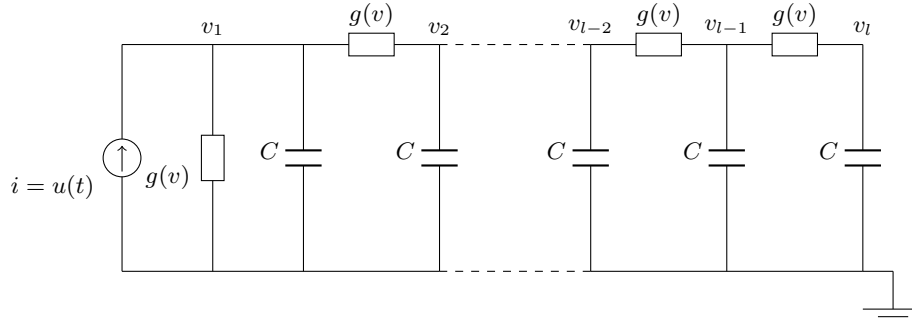 $\epsilon = \Delta(\mu^i)$;

END WHILE.
***

Fig. 1. *RC Circuit Diagram.*

5. Numerical Results

We use three benchmark examples to test the proposed approach. The results are compared with the direct interpolatory methods [19, 9] reviewed in Section 2 both, in the one-sided and two-sided projection frameworks. For reference, we use 1s-qbmor and 2s-qbmor to refer to the one-sided and two-sided direct methods. Similarly, 1s-pmor and 2s-pmor denote the indirect parametric model order reduction (PMOR) method without error estimation and 1s-pmor-ee and 2s-pmor-ee the proposed PMOR method with error estimation. A version of our MATLAB implementation is published at https://doi.org/10.5281/zenodo.1249930. All time-domian simulations of the full- and reduced-order systems were performed with the MATLAB function ode15s.

5.1. Nonlinear RC circuit

The nonlinear RC circuit (as shown in Figure 1) was first considered in [14] and since then it has been used as benchmark in many papers for nonlinear model reduction. The nonlinearity is due to the nonlinear resistor with I-V characteristics given as: $g(v) = e^{40v} + v - 1$, where $g(v)$ is the current function and v is the voltage across the resistor. All the capacitances are scaled to $C = 1$.

It is shown in [19] that the nonlinearity in the RC circuit example can be written in the quadratic-bilinear form as in (1) by introducing some auxiliary variables. The transformation is exact, but the dimension of the system increases to $n = 2 \cdot l$, where l is the number of nodes in Figure 1, and is also the dimension of the original nonlinear system.

12

For our results, we fixed the number of nodes to $l = 500$, so that the size of the quadratic-bilinear model (full order model (FOM)) is $n = 1000$. We randomly choose 9 candidate expansion points for s:

$$\sigma_i = \{0.1, 0.5, 1, 2, 5, 10, 25, 50, 100\}. \tag{38}$$

and 7 candidate expansion points of p:

$$p_j = \{0.15, 1, 1.2, 1.8, 4, 7, 10\}, \tag{39}$$

to constitute S_{sample} in Algorithm 1.

5.1.1. One-sided projection: We first consider the one sided projection method. Here, the direct interpolatory method (1s-qbmor) uses all the 9 candidates σ_i as interpolation points and constructs a reduced quadratic-bilinear system of size 18. The other two methods, PMOR with error estimation (1s-pmor-ee) and without error estimation (1s-pmor), select the interpolation points and parameter values from σ_i and p_j, such that the size of the reduced model is fixed to 18. In 1s-pmor, the value of q in (29) is fixed to 2 and the combination of the interpolation points and the parameter values are fixed to the following 8 possibilities:

$$S_\mu = \{(50, 4); (1, 1); (2, 1.2); (5, 10); (25, 4); (10, 1.8); (100, 7); (0.5, 0.15)\}$$

The 1s-pmor-ee method uses $q = 6$ in (29) and the initial expansion point for Algorithm 1 is $\mu^0 = (0.001, \ 0.001)$ with S_{sample} being defined as a set including all possible combinations of σ_i and p_j in (38) and (39), respectively.

Table 1 shows the iterative results of Algorithm 1, where 3 sets of expansion points are adaptively selected from the candidates in S_{sample}. The error bound $\Delta_{\max}$ accurately reflects the true error $\epsilon_{\max}$.

Table 1 Nonlinear RC-circuit, $q = 6$, $\epsilon_{tol} = 10^{-3}$, $n = 1000$, $r = 18$

iteration	μ	$\epsilon_{\max}$	$\Delta_{\max}$
1	(0.001, 0.001)	18.03	1.76×10^3
2	(10, 1.2)	0.26	4.38
3	(0.5, 0.05)	1.08×10^{-4}	1.14×10^{-4}

The results for all the three methods are shown in Figure 2. It is easy to see that the 1s-pmor-ee method produces much better results as compared to 1s-pmor and 1s-qbmor. We believe that a better set of selection samples S_{sample} can further improve the quality of the reduced-order model. The results also show that, in addition to the use of error estimation, matching multi-moments corresponding to higher derivatives is also crucial for better reduced-order models.

5.1.2. Two-sided projection: Next, we show our results for two-sided projection of the nonlinear RC circuit. Here we used 2s-qbmor, 2s-pmor and 2s-pmor-ee with starting parameters similar to those used in the one-sided projection case. As shown in Figure 3, the quality of the reduced model is much better in the case of two-sided projection.

Note that the relative error for the proposed 2s-pmor-ee method is smaller as compared to the other approaches but not at all time instants. We believe that the results will further improve with a good choice of initial interpolation points and parameter values, the sample set S_{sample} and the number of elements in S_{sample}. To show the savings in the computational time, it is observed that
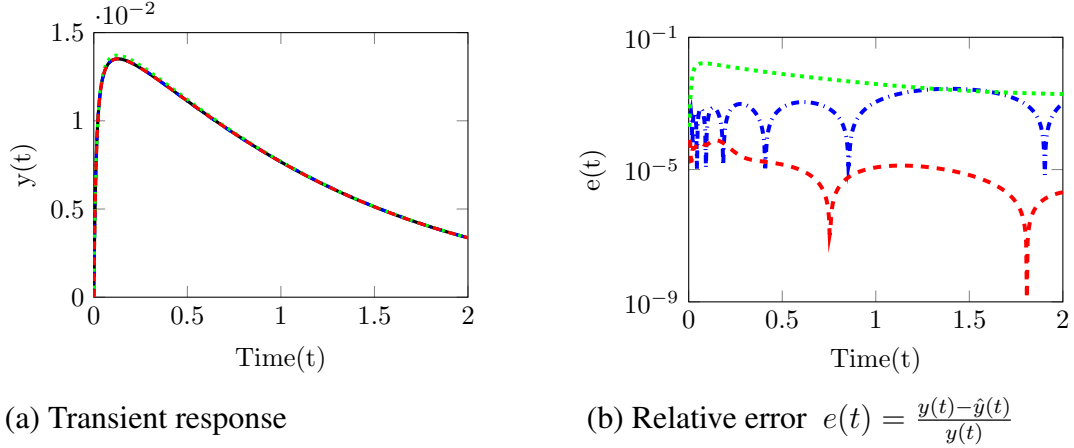
(a) Transient response

(b) Relative error $e(t) = \frac{y(t)-\hat{y}(t)}{y(t)}$

Fig. 2. _Nonlinear RC-circuit with_ $u(t) = e^{-t}$_; FOM of size_ $n = 1000$ _(——) ; ROM via 1s-qbmor of size_ $r = 18$_(–·–·) ; ROM via 1s-pmor of size_ $r = 18$ _(······) and ROM via 1s-pmor-ee with_ $r = 18$ _(– – –)_

the simulation time of the full order nonlinear RC circuit is 13.4287s while the time for simulation as well as computation of the reduced model through the three techniques 2s-qbmor, 2s-pmor and 2s-pmor-ee are 1.4804s, 0.9862s and 0.5732s, respectively.

5.2. One-Dimensional Burgers' Equation

As a second example, we consider the one-dimensional Burgers' equation, which is again well used for testing nonlinear model reduction techniques [22, 9]. The Burgers' equation with domain,
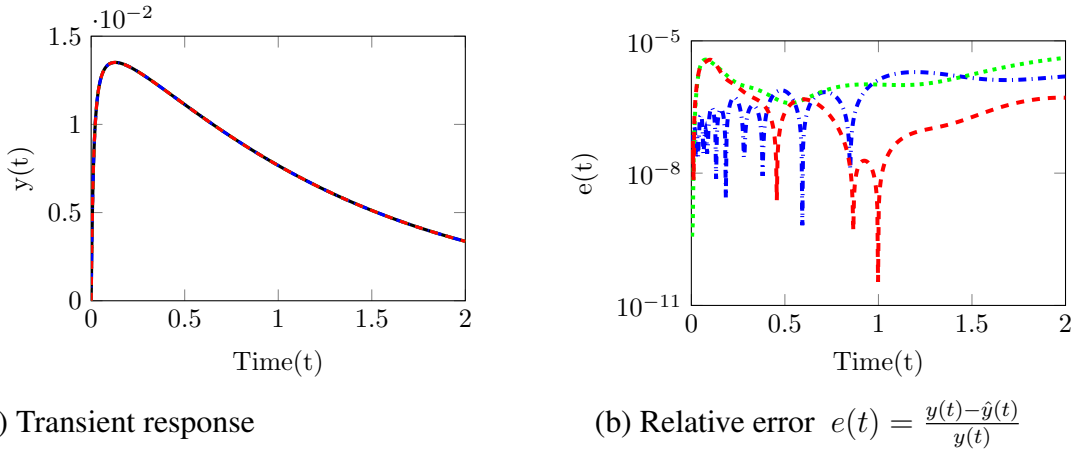

(a) Transient response

(b) Relative error $e(t) = \frac{y(t)-\hat{y}(t)}{y(t)}$

Fig. 3. _Nonlinear RC-circuit with_ $u(t) = e^{-t}$_; FOM of size_ $n = 1000$ _(——) ; ROM via 2s-qbmor of size_ $r = 18$_(–·–·) ; ROM via 2s-pmor of size_ $r = 18$ _(······) and ROM via 2s-pmor-ee with_ $r = 18$ _(– – –)_

14

$\Gamma = (0,1) \times (0,T)$ along with boundary conditions can be written as:

$$v_t + vv_x = \nu \cdot v_{xx}, \qquad\qquad\qquad \text{in } \Gamma,$$
$$\alpha v(0,t) + \beta x(0,t) = u(t), \quad v_x(1,t) = 0, \qquad\qquad t \in (0,T),$$
$$v(x,0) = v_0(x), \quad v_0(x) = 0, \qquad\qquad x \in (0,1),$$

where ν is the vicosity and $v_0(x)$ is the initial condition. Semi-discretization of such partial differential equations naturally leads to a quadratic-bilinear system as given in (1). For our results, we set the state dimensions to $n = 1000$ and the viscosity to $\nu = 0.05$. The size of the reduced model in the proposed indirect projection method is based on convergence of the error bound and number of higher moments being matched. To compare our results with the direct method, we vary the number of higher moments and sample of interpolation points such that the size of the reduced model is the same for each method. This however results in different sizes of the reduced model for one sided and two sided projections.

5.2.1. One-Sided Projection: Here we use the first 7 elements of σ_i given in (38) as interpolation points. The direct interpolatory method uses these interpolation points to construct a reduced quadratic-bilinear system of size 14. For the proposed error estimation approach, we use a sample space that includes all the combinations of the σ_i and p_j (given in (38) and (39)) along with some other random choices (mainly to ensure that the resulting reduced system is also of size 14 for error tolerance $tol = 10^{-6}$). With $q = 4$ and starting from the first combination of S_{sample}, Algorithm 1 iteratively converges to the actual error, as shown in the following table:

Table 2 Burgers' equation, $q = 4$, $\epsilon_{tol} = 10^{-6}$, $n = 1000$, $r = 14$

iteration	μ	ϵ_{max}	Δ_{max}
1	(50, 4)	4.22×10^{-2}	7.66×10^2
2	(0.1, 0.05)	4.19×10^{-2}	3.43×10^{-1}
3	(5, 0.05)	1.57×10^{-6}	1.96×10^{-6}
4	(1, 10)	1.20×10^{-7}	1.20×10^{-7}
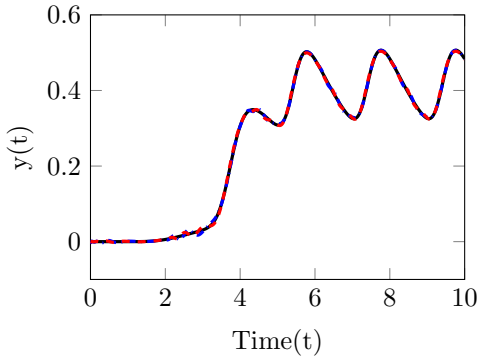
(a) Transient response
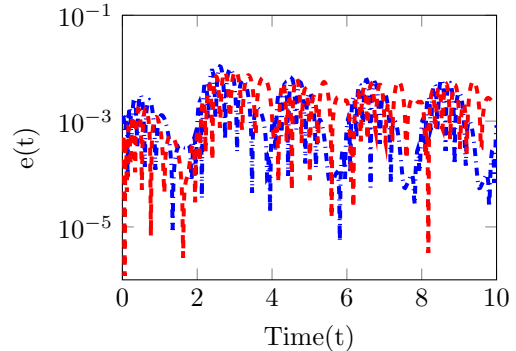
(b) Absolute error $e(t) = |y(t) - \hat{y}(t)|$

Fig. 4. *Burgers' equation with $u(t) = cos(\pi t)$; FOM of size $n = 1000$ (——) ; ROM via 1s-qbmor of size $r = 14$(⋅–⋅⋅) and ROM via 1s-pmor-ee with $r = 14$ (- - -)*

The reduced system obtained from Algorithm 1 is compared with that obtained from the direct projection technique in Figure 4. The performance of the proposed approach is better or comparable to the direct method with the advantage that the computations associated with the quadratic matrix H can be avoided.

5.2.2. Two-sided Projection: Here the interpolation points are predefined as $\sigma_i = \{1, 2, 3, 4\}$. The direct two-sided projection technique constructs a reduced system of size 8 and the proposed approach selects some interpolation points from a sample set S_{sample}, which includes the σ_i's and some other random points along with random values of the parameter samples. Algorithm 1 selects a set of interpolation points and parameter values as shown in Table 3.

Table 3 Burgers' equation, $q = 2$, $\epsilon_{tol} = 10^{-6}$, $n = 1000$, $r = 8$

iteration	μ	ϵ_{max}	Δ_{max}
1	(1, 640)	8.27×10^2	8.02×10^5
2	(0.0001, 80)	7.67×10^{-1}	9.58×10^3
3	(5, 80)	6.31×10^{-4}	5.39×10^{-2}
4	(2, 80)	2.41×10^{-7}	1.23×10^{-6}
5	(0.4, 80)	2.02×10^{-7}	2.39×10^{-7}

(a) Transient response

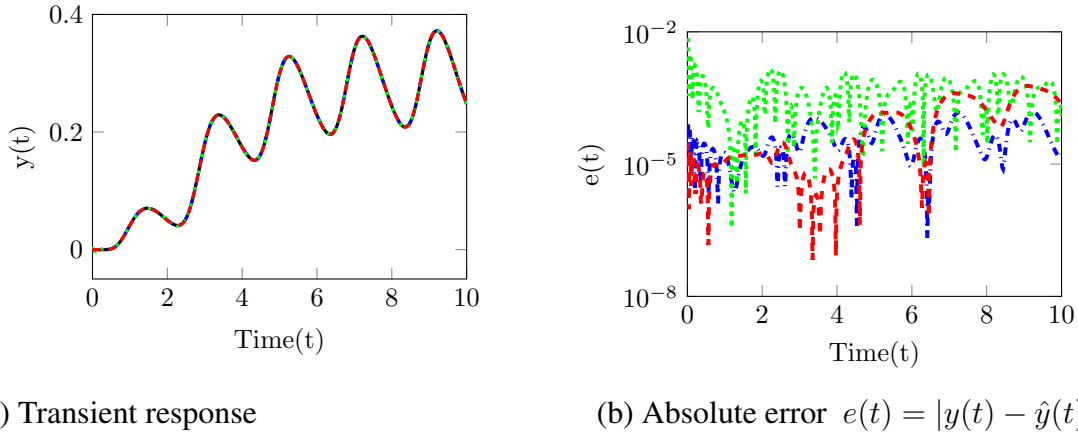(b) Absolute error $e(t) = |y(t) - \hat{y}(t)|$

Fig. 5. *Burger's equation with $u(t) = cos(\pi t)$; FOM of size $n = 1000$ (——) ; ROM via 2s-pmor with $r = 8$ (······) ; ROM via 2s-qbmor of size $r = 8$(-·-·) and ROM via 2s-pmor-ee with $r = 8$ (- - -)*

We compare the results with the direct two-sided projection approach as shown in Figure 5. The proposed method is, in particular, accurately approximating the transient response of the system. We believe that the results will further improve with the choice of the sample set S_{sample} and with the initial interpolation point and parameter values for Algorithm 1. Also the error estimation procedure sometimes computes projection matrices V and W of different sizes. In such a case, we simply truncate the extra columns in the V or W matrix. Though this destroys the theoretic interpolation properties, usually the approximation quality of the reduced-order model is not harmed by this. Regarding savings in the computational time, it is observed that the simulation of full order Burgers' equation takes 31.4055s and the time for simulation as well as computation of the reduced model via the three techniques 2s-qbmor, 2s-pmor and 2s-pmor-ee are 1.9689s, 1.4117s

16

and 1.1487s, respectively.

To obtain a different reduced order model from the proposed 2s-qbmor-ee, we set the tolerance level to $\epsilon_{tol} = 6.5 \times 10^{-2}$. Table 3 shows that this will stop the iterations after selecting the first three interpolation points. With $q = 2$ and the selected interpolation points, the size of the reduced system will change to 6. The resulting reduced system is compared with the direct approach (2s-qbmor) and the parametric method (2s-pmor) of the same size as shown in Figure 6. Clearly the performance of the proposed method is better than the direct method, especially in the transient behaviour, with low computational cost.
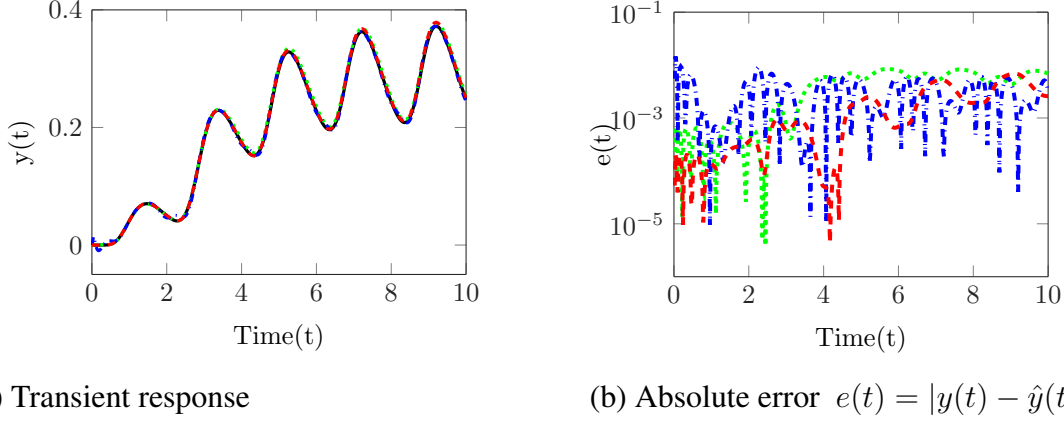

(a) Transient response (b) Absolute error $e(t) = |y(t) - \hat{y}(t)|$

Fig. 6. *Burger's equation with $u(t) = cos(\pi t)$; FOM of size $n = 1000$ (——) ; ROM via 2s-pmor with $r = 6$ (⋯⋯) ; ROM via 2s-qbmor of size $r = 6$ (-·-·-) and ROM via 2s-pmor-ee with $r = 6$ (- - -)*
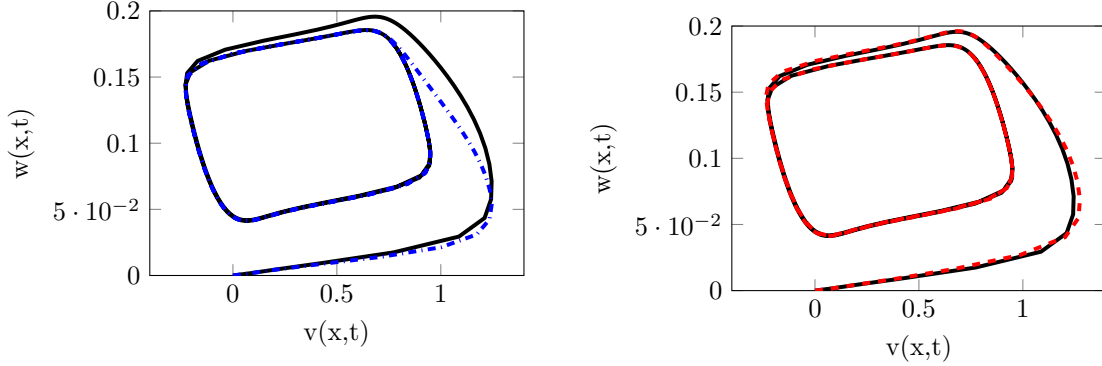
5.3. FitzHugh-Nagumo System

Our last example is the FitzHugh Nagumo system where the activation and deactivation dynamics of a spiking neuron are modeled by coupled nonlinear PDEs:

$$
\begin{aligned}
\epsilon v_t(x,t) &= \epsilon^2 v_{xx}(x,t) + f(v(x,t)) - w(x,t) + g, \\
w_t(x,t) &= hv(x,t) - \gamma w(x,t) + g,
\end{aligned}
\tag{40}
$$

with $f(v) = v(v - 0.1)(1 - v)$ and boundary conditions

$$
\begin{aligned}
v(x,0) &= 0, & w(x,0) &= 0, & x &\in [0,1], \\
v_x(0,t) &= -i_0(t), & v_x(1,t) &= 0, & t &\geq 0,
\end{aligned}
$$

where $\epsilon = 0.015$, $h = 0.5$, $\gamma = 2$, $g = 0.05$ and $i_0(t) = 5 \times 10^4 t^3 \exp(-15t)$. Applying the standard finite difference method results in a system of ODEs with cubic nonlinearities. The cubic terms can be represented in quadratic form by introducing a new dynamical variable $z_i = v_i^2$; $\dot{z}_i = 2v_i\dot{v}_i$. This means that the system in the QBDAE form will involve three dynamical variables v_i, w_i and z_i. So if the coupled PDE in (40) is discretized using $\bar{n}$ points, the system of QBDAEs will be of size $n = 3\bar{n}$. Note that the QBDAE system has to have two inputs and two outputs in order to incorporate the effect of the variable g and the initial stimulus $i_0(t)$ in the model. This means that 1s-qbmor (and 2s-qbmor) for computing the basis matrices V and W has to be repeated for each column of the input (and output) matrix. For our results, we set $\bar{n} = 750$, so that $n = 2250$.

17

(a) Transient response

(b) Transient response

Fig. 7. *FitzHugh-Nagumo system with $u(t) = 5 \times 10^4 t^3 e^{-15t}$; FOM of size $n = 1000$ (——) ; ROM via 1s-qbmor of size $r = 16$ (–·–·) (Left) and ROM via 1s-pmor-ee with $r = 30$ (- - -) (Right)*

5.3.1. One-sided Projection:
We used linear IRKA to identify 10 interpolation points of s and appended them with the elements of σ_i given in (38) as the candidates of the expansion points of s in S_{sample} of Algorithm 1. The samples of p in S_{sample} are all randomly selected. For the proposed error estimation approach, we use the complete sample space that includes all the combinations of the samples of s and p. With $q = 6$ and starting from the first combination, Algorithm 1 iteratively converges to the actual error, as shown in the following table:

Table 4 Fitzhugh Nagumo System, $q = 6$, $\epsilon_{tol} = 10^{-8}$, $n = 2250$, $r = 30$

No.	μ	ϵ_{max}	Δ_{max}
1	$(1.75 \; 10^4, \; 0.1210)$	$5.0123 \; 10^{-6}$	1.2016×10^{-2}
2	$(1.4276, \; 0.8213)$	7.6401×10^{-7}	8.1028×10^{-7}
3	$(4.3261 - i5.2842, \; 0.0631)$	2.1685×10^{-8}	2.1698×10^{-8}
4	$(50, \; 0.9841)$	5.45850×10^{-11}	5.45855×10^{-11}

At each iteration of Algorithm 1, the interpolation point $\mu := (\sigma_i, p_i)$ that corresponds to the maximized estimated error is used for construction of the basis matrices and therefore for reduction of the quadratic-bilinear system. The result is a reduced system of size 30 with transient response as shown in Figure 7b. To compare these results with the direct interpolatory method, we used 10 shift frequencies obtained through the linear IRKA to compute the reduced system using the direct method. The direct method leads to an unstable reduced system. However when we used only the first 4 interpolation points obtained through the linear IRKA, the result is a stable reduced system of size 16. The transient response of the computed reduced model is shown in Figure 7a. While the sizes of the two reduced models are different, the comparison clearly motivates the use of error estimation for the selection of interpolation points.

A 3-D plot for the limit cycle behavior of the original and the reduced system is presented in Figure 8. Note that in this case we used the MIMO version of the error estimator $\Delta(\mu)$ as discussed in [16]. In the MIMO case, the error estimation process is implemented on each element of the transfer matrix and corresponds to the maximized estimated error among all the elements, according to which, the sample points are selected, see [16] for details.
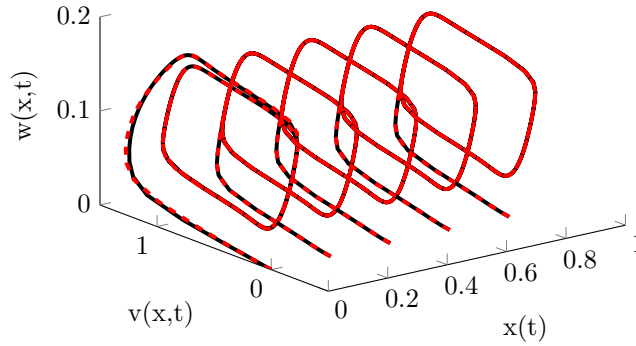
Fig. 8. *FitzHugh-Nagumo system with $u(t) = 5 \times 10^4 t^3 e^{-15t}$; FOM of size $n = 1000$ (——) and ROM via 1s-pmor-ee with $r = 30$ (- - -)*

6. Conclusions

A projection based model reduction technique for quadratic-bilinear systems is proposed, where the projection matrices are identified from an associated linear parametric system and interpolation points are selected such that an a posteriori error bound for the linear parametric system is minimized. Two-sided multi-moment matching properties are also derived for this framework. A better choice of the sample set S_{sample} used for the error bound is an important task, since S_{sample} varies with the original system to be reduced, and requires further research.

References

[1] M.I. Ahmad, U. Baur, and P. Benner. Implicit Volterra series interpolation for model reduction of bilinear systems. *J. Comput. Appl. Math.*, 316:15 – 18, 2017.

[2] M.I. Ahmad, P. Benner, and I. Jaimoukha. Krylov subspace methods for model reduction of quadratic-bilinear systems. *IET Control Theory Appl.*, 10:2010–2018(8), 2016.

[3] M.I. Ahmad, L. Feng, and P. Benner. A new interpolatory model reduction for quadratic bilinear descriptor systems. *Proc. Appl. Math. Mech.*, 15(1):589 – 590, 2015.

[4] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*. SIAM Publications, Philadelphia, PA, 2005.

[5] A. C. Antoulas, D. C. Sorensen, and S. Gugercin. A survey of model reduction methods for large-scale systems. *Contemp. Math.*, 280:193–219, 2001.

[6] P. Benner and T. Breiten. On $\mathcal{H}_2$-model reduction of linear parameter-varying systems. *Proc. Appl. Math. Mech.*, 11:805–806, 2011.

[7] P. Benner and T. Breiten. Interpolation-based $\mathcal{H}_2$-model reduction of bilinear control systems. *SIAM J. Matrix Anal. Appl.*, 33(3):859–885, 2012.

[8] P. Benner and T. Breiten. Two-sided moment matching methods for nonlinear model reduction. Technical report, Max Planck Institute Magdeburg Preprints, 2012.

[9] P. Benner and T. Breiten. Two-sided projection methods for nonlinear model order reduction. *SIAM J. Sci. Comput.*, 37(2):B239–B260, 2015.

[10] P. Benner, A. Cohen, M. Ohlberger, and K. Willcox (Eds.). *Model Reduction and Approximation: Theory and Algorithms*. SIAM Publications, Philadelphia, PA, 2017.

[11] P. Benner and L. Feng. A robust algorithm for parametric model order reduction based on implicit moment matching. In A. Quarteroni and G. Rozza, editors, *Reduced order methods for modeling and computational reduction*, pages 159–185. MS&A - Modeling, Simulation and Applications, Springer, 2014.

[12] P. Benner, M. Ohlberger, A. Cohen, and K. Willcox, editors. *Model Reduction and Approximation*. SIAM Publications, Philadelphia, PA, 2017.

[13] T. Breiten and T. Damm. Krylov subspace methods for model order reduction of bilinear control systems. *Sys. Control Lett.*, 59(10):443–450, 2010.

[14] Y. Chen. Model reduction for nonlinear systems. Master's thesis, Massachusetts Institute of Technology, 1999.

[15] L. Daniel, C.S. Ong, S.C. Low, K.H. Lee, and J. White. A multiparameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, 23(5):678–693, 2004.

[16] L. Feng, A. C. Antoulas, and P. Benner. Some a posteriori error bounds for reduced order modelling of (non-)parametrized linear systems. *M2AN Math. Model. Numer. Anal.*, 51:2127–2158, 2017.

[17] G. Flagg and S. Gugercin. Multipoint Volterra series interpolation and $\mathcal{H}_2$ optimal model reduction of bilinear systems. *SIAM J. Numer. Anal.*, 36(2):549–579, 2015.

[18] E. J. Grimme. *Krylov projection methods for model reduction*. Phd thesis, Univ. of Illinois at Urbana-Champaign, USA, 1997.

[19] C. Gu. QLMOR: a projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, 30(9):1307–1320, 2011.

[20] P.K. Gunupudi, R. Khazaka, M.S. Nakhla, T. Smy, and D. Celo. Passive parameterized time-domain macromodels for high-speed transmission-line networks. *IEEE Trans. Microw. Theory Techn.*, 51(12):2347–2354, 2003.

[21] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–475, 2009.

[22] K. Kunisch and S. Volkwein. Proper orthogonal decomposition for optimality systems. *ESAIM Math. Model. Numer. Anal.*, 42(1):1–23, 2008.

[23] A. T. Patera and G. Rozza. *Reduced Basis Approximation and a Posteriori Error Estimation for Parametrized Partial Differential Equations*. MIT, 2007. To appear in (tentative rubric) MIT Pappalardo Graduate Monographs in Mechanical Engineering.

[24] R. J. Rugh. *Nonlinear System Theory*. Johns Hopkins University Press Baltimore, MD, 1981.

[25] O. Zahm, M. Billaud-Friess, and A. Nouy. Projection-based model order reduction methods for the estimation of vector-valued variables of interest. *SIAM J. Sci. Comput.*, 39(4):A1647–A1674, 2017.

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