Peter Benner Serkan Gugercin Karen Willcox

A Survey of Model Reduction Methods for Parametric Systems

Max Planck Institute Magdeburg Preprints

MPIMD/13-14 **August 14, 2013**

Impressum:

Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg

Publisher:
Max Planck Institute for Dynamics of Complex Technical Systems

Address:
Max Planck Institute for Dynamics of Complex Technical Systems
Sandtorstr. 1
39106 Magdeburg

www.mpi-magdeburg.mpg.de/preprints

A Survey of Model Reduction Methods for Parametric Systems *

Peter Benner† Serkan Gugercin‡ Karen Willcox§

August 14, 2013

Abstract

Numerical simulation of large-scale dynamical systems plays a fundamental role in studying a wide range of complex physical phenomena; however, the inherent large-scale nature of the models leads to unmanageable demands on computational resources. Model reduction aims to reduce this computational burden by generating reduced models that are faster and cheaper to simulate, yet accurately represent the original large-scale system behavior. Model reduction of linear, non-parametric dynamical systems has reached a considerable level of maturity, as reflected by several survey papers and books. However, parametric model reduction has emerged only more recently as an important and vibrant research area, with several recent advances making a survey paper timely. Thus, this paper aims to provide a resource that draws together recent contributions in different communities to survey state-of-the-art in parametric model reduction methods.

Parametric model reduction targets the broad class of problems for which the equations governing the system behavior depend on a set of parameters. Examples include parameterized partial differential equations and large-scale systems of parameterized ordinary differential equations. The goal of parametric model reduction is to generate low cost but accurate models that characterize system response for different values of the parameters. This paper surveys state-of-the-art methods in parametric model reduction, describing the different approaches within each class of methods for handling parametric variation and providing a comparative discussion that lend insights to potential advantages and disadvantages in applying each of the methods. We highlight the important role played by parametric model reduction in design, control, optimization, and uncertainty quantification—settings that require repeated model evaluations over a potentially large range of parameter values.

Keywords. Dynamical systems, parameterized model reduction, interpolation, proper orthogonal decomposition, balanced truncation, greedy algorithm.

AMS subject classifications. 35B30, 37M99, 41A05, 65K99, 93A15, 93C05.

1 Introduction

Dynamical systems are the basic framework for modeling and control of an enormous variety of complex systems of scientific interest or industrial value. Examples include heat transfer, fluid

*The first author acknowledges support by DFG grant BE 2174/16-1 "Multivariate Interpolation Methods for Parametric Model Reduction". The second author acknowledges support of NSF Grant DMS-1217156 (Program Manager L.M. Jameson). The third author acknowledges support of AFOSR Computational Mathematics Grant FA9550-12-1-0420 (Program Manager F. Fahroo).

†Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany; benner@mpi-magdeburg.mpg.de

‡Department of Mathematics, Virginia Tech, Blacksburg, VA, 24061-0123, USA; gugercin@math.vt.edu

§Department of Aeronautics & Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA; kwillcox@mit.edu

dynamics, chemically reacting flows, biological systems, signal propagation and interference in electric circuits, wave propagation and vibration suppression in large structures, and design of micro-electro-mechanical systems (MEMS). Numerical simulation of the associated models has been one of the few available means for studying complex underlying physical phenomena. However, the ever increasing need for improved accuracy requires the inclusion of more detail in the modeling stage, leading inevitably to larger-scale, more complex models of dynamical systems. The ever increasing push towards improving system performance leads to a need to simulate many different possible realizations of the system. Performing multiple simulations in such large-scale settings often presents unmanageably large demands on computational resources. Alleviation of this computational burden is the main motivation for deriving *reduced models*—low-dimensional, efficient models that are fast to solve but that approximate well the underlying high-resolution simulations. The field of *model reduction* encompasses a broad set of mathematical methods to generate and evaluate these reduced models.

In this paper, we focus on the broad class of problems for which the equations representing the system dynamics depend on a set of parameters and the goal is to characterize system response for different values of the parameters. These parameters may enter the models in many ways, representing, for example, material properties, system geometry, system configuration, initial conditions, and boundary conditions. This parametric dependence presents a unique set of challenges for model reduction, since one cannot afford to create a new reduced model for every change in the parameter values. Hence, the desired approach is to generate a parametric reduced model, that approximates the original full-order dynamical system with high fidelity over a range of parameters. This is the goal of *parametric model reduction*. This survey provides an overview of state-of-the-art methods in parametric model reduction for large-scale dynamical systems. Recent years have seen considerable progress in this field, with several classes of methods emerging. This paper highlights the different approaches within each class of methods for handling parametric variation. We provide a comparative discussion that lend insights to potential advantages and disadvantages in applying each of the methods. We rarely elaborate on technical details and for proofs mostly give pointers to the relevant literature.

Why is parametric model reduction important and useful? It fills a critical need in design, control, optimization, and uncertainty quantification settings—settings that require repeated model evaluations over a potentially large range of parameter values. These are settings in which we are willing to forego a large up-front cost, the so-called *offline* cost, in order to obtain a reduced model that allows rapid yet accurate simulation over the range of parameters, the so-called *online* phase.

The *design* setting may require evaluation of system performance over a range of parameter values representing critical design constraints. For example, parametric reduced models capturing the coupled fluid dynamic and structural dynamic behavior of an aircraft configuration provide rapid evaluation of aeroelastic performance over a range of aircraft operating conditions [109, 108, 2]. This enables rapid characterization of the aircraft's flight envelope, calculations that would otherwise require many weeks of computation time. Parametric reduced models have also shown to be an important enabling technology in the synthesis and design of interconnect [48, 37] and MEMS [21, 63] as well as in electro-chemical [65] and electro-thermal applications [66].

In *control* design we desire to drive the system dynamics into a desired configuration, while accounting for parametrically varying dynamics. Examples include design process control in Rayleigh-Bénard convection with varying Rayleigh number [112], and control for fluid flow with varying Reynolds number and/or shape parameters [88, 115]. Instead of designing a new controller for every new parameter, which would be a large computational burden and intractable for online controller design, one could design either a single reduced-order controller that performs effectively over the full parameter space or one that parametrically adapts to the changes.

Both design and control can involve *optimization*, where the goal may be an optimal system configuration with respect to a certain performance objective (e.g., maximal throughput, minimal weight, minimal energy consumption, etc.) or an optimal controller. Most optimization algorithms require multiple evaluations of the forward model for varying parameter configurations; this is where parametric reduced models can play a role. Past work has used trust regions to manage the

reduced model as the optimization proceeds in an optimal control problem [11]. Another approach for optimal control incorporates the optimality conditions into derivation of the reduced model [100]. A combination of domain decomposition and model reduction has also been developed for optimal control and shape optimization problems [6, 7]. Another recent framework employed parametric reduced models that ensure exact matching of the objective function and gradient evaluations for a subset of parameter values [20, 22, 56, 160].

Uncertainty quantification is another area that demands repeated model evaluations–often many thousands of evaluations are needed to sample the uncertainty space (e.g., using Monte Carlo sampling). For example, using parametric reduced models, the forward propagation of uncertainty through complex systems, such as those typically modeled using large-scale computational fluid dynamic (CFD) models, can be achieved in turnaround times useful for design [42]. Parametric reduced basis models have also been combined with stochastic collocation to solve PDEs with random coefficients [59]. Another example is large-scale statistical inverse problems for which Markov chain Monte Carlo methods are computationally intractable, either because they require excessive amounts of CPU time or because the parameter space is too large to be explored effectively by state-of-the-art sampling methods. In these cases, parametric model reduction over both state and parameter spaces can make tractable the solution of large-scale inverse problems that otherwise cannot be solved [156, 69, 107, 56, 22].

The rest of the paper is organized as follows: In Section 2, we define the problem setup and introduce the general framework for projection-based model reduction of parameterized dynamical systems. We also discuss measures of reduced model error. Sections 3 and 4 discuss in detail the building blocks of parametric model reduction. In particular, Section 3 presents three different methods for deriving the reduced-order basis: rational interpolation methods, balanced truncation and proper orthogonal decomposition. We also discuss methods for sampling the parameter space. Section 4 covers methods for constructing the parameterized reduced model. These construction methods are divided into those that use global information over the parameter space, and those that use local information combined with an interpolation method. Section 5 provides a comparative discussion of the various approaches, along with their relative advantages and disadvantages. Finally, Section 6 concludes the paper with a discussion of open challenges and future outlook.

2 General problem setup

In this section we define the parameterized dynamical systems of interest. We present the general projection-based model reduction framework for systems that are linear in state but have general nonlinear parametric dependence. The projection framework also applies to systems that are nonlinear in state, but as discussed in Section 6, the model reduction theory in this case is much less developed. Throughout the discussion, we indicate which aspects of the methods carry over to the nonlinear-in-state case. This section concludes with a discussion of error measures to assess the quality of the reduced model.

2.1 Parameterized dynamical systems

We consider dynamical systems that are linear in state and parameterized with d parameters $\mathbf{p} = [p_1, ..., p_d]^T \in \Omega \subset \mathbb{R}^d$ (usually, Ω is a bounded domain) as:

$$\mathbf{E}(\mathbf{p})\,\dot{\mathbf{x}}(t;\mathbf{p}) = \mathbf{A}(\mathbf{p})\,\mathbf{x}(t;\mathbf{p}) + \mathbf{B}(\mathbf{p})\,\mathbf{u}(t), \quad \mathbf{y}(t;\mathbf{p}) = \mathbf{C}(\mathbf{p})\,\mathbf{x}(t;\mathbf{p}). \tag{1}$$

The state-vector is denoted by $\mathbf{x}(t;\mathbf{p}) \in \mathbb{R}^n$. $\mathbf{u}(t) \in \mathbb{R}^m$ and $\mathbf{y}(t;\mathbf{p}) \in \mathbb{R}^q$ denote, respectively, the inputs (excitations) and outputs (observations) of the underlying model. Hence, the model has m inputs and q outputs. The state-space matrices, then, have the dimensions $\mathbf{E}(\mathbf{p})$, $\mathbf{A}(\mathbf{p}) \in \mathbb{R}^{n\times n}$, $\mathbf{B}(\mathbf{p}) \in \mathbb{R}^{n\times m}$, and $\mathbf{C}(\mathbf{p}) \in \mathbb{R}^{q\times n}$. We focus on models that are linear in state, but we allow nonlinear parametric dependency in all system matrices. The length of the state-vector $\mathbf{x}(t)$ (i.e., n) is called the dimension of the parametric model (1). We will assume that for every $\mathbf{p} \in \Omega$, $\mathbf{E}(\mathbf{p})$ is nonsingular even though most of the discussion can be extended to the singular $\mathbf{E}(\mathbf{p})$ case. We

will further assume that the original model in (1) is asymptotically stable for every $\mathbf{p} \in \Omega$; i.e., the eigenvalues of the pencil $\lambda \mathbf{E}(\mathbf{p}) - \mathbf{A}(\mathbf{p})$ have negative real parts; however, most of the methods can be applied with small modifications to unstable systems, as discussed further in Section 6.

We are interested in cases where n is very large, typically exceeding hundreds of thousands. The goal is to replace the original large-scale model (1), sometimes called the "truth model," with a reduced model of the form

$$\mathbf{E}_r(\mathbf{p})\,\dot{\mathbf{x}}_r(t;\mathbf{p}) = \mathbf{A}_r(\mathbf{p})\,\mathbf{x}_r(t;\mathbf{p}) + \mathbf{B}_r(\mathbf{p})\,\mathbf{u}(t), \quad \mathbf{y}_r(t;\mathbf{p}) = \mathbf{C}_r(\mathbf{p})\,\mathbf{x}_r(t;\mathbf{p}), \tag{2}$$

such that the reduced output $\mathbf{y}_r(t;\mathbf{p}) \in \mathbb{R}^q$ is a good approximation of $\mathbf{y}(t;\mathbf{p})$ with respect to an appropriate error measure. Note that the reduced state-vector $\mathbf{x}_r(t;\mathbf{p})$ has length $r \ll n$, and the reduced state-space matrices have dimensions $\mathbf{E}_r(\mathbf{p})$, $\mathbf{A}_r(\mathbf{p}) \in \mathbb{R}^{r\times r}$, $\mathbf{B}_r(\mathbf{p}) \in \mathbb{R}^{r\times m}$, and $\mathbf{C}_r(\mathbf{p}) \in \mathbb{R}^{q\times r}$; hence the dimension is reduced from n down to $r \ll n$. This reduction process is illustrated pictorially in Figure 1 below:

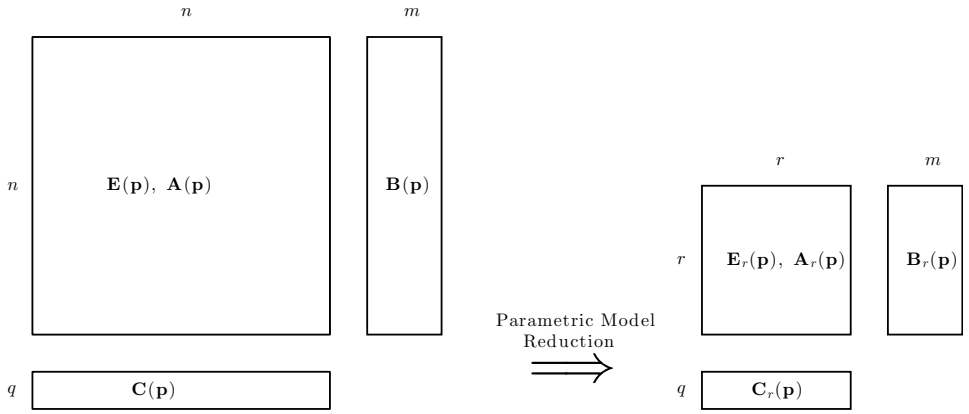

Figure 1: Parametric Model Reduction

2.2 Projection-based model reduction

Parametric model reduction can be approached from a variety of viewpoints. This paper focuses on projection-based approaches. To motivate the challenges associated with parameterized systems, first consider the general projection-based reduction approach for a system with no parametric dependence; i.e.,

$$\mathbf{E}\,\dot{\mathbf{x}}(t) = \mathbf{A}\,\mathbf{x}(t) + \mathbf{B}\,\mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{C}\,\mathbf{x}(t). \tag{3}$$

We approximate the unknown state quantities in a basis of reduced dimension and project the governing equations onto a suitably defined low-dimensional subspace. In particular, choose r-dimensional test and trial subspaces, denoted by $\mathcal{V}$ and $\mathcal{W}$, respectively, where $r \ll n$. We define the associated basis matrices $\mathbf{V} \in \mathbb{R}^{n\times r}$ and $\mathbf{W} \in \mathbb{R}^{n\times r}$, where $\mathcal{V} = \text{Range}(\mathbf{V})$ and $\mathcal{W} = \text{Range}(\mathbf{W})$. Using the approximation that the full state $\mathbf{x}(t)$ evolves in the r-dimensional subspace $\mathcal{V}$, we write $\mathbf{x}(t) \approx \mathbf{V}\mathbf{x}_r(t)$, where $\mathbf{x}_r(t) \in \mathbb{R}^r$. Using this approximation in (3) defines a residual $(\mathbf{E}\mathbf{V}\dot{\mathbf{x}}_r(t) - \mathbf{A}\mathbf{V}\mathbf{x}_r(t) - \mathbf{B}\,\mathbf{u}(t))$ and defines the reduced model output $\mathbf{y}_r(t) = \mathbf{C}\mathbf{V}\mathbf{x}_r(t)$. Then, enforcing the Petrov-Galerkin condition

$$\mathbf{W}^T\Big(\mathbf{E}\mathbf{V}\dot{\mathbf{x}}_r(t) - \mathbf{A}\mathbf{V}\mathbf{x}_r(t) - \mathbf{B}\,\mathbf{u}(t)\Big) = \mathbf{0},$$

leads to the reduced system

$$\mathbf{E}_r\;\dot{\mathbf{x}}_r(t) = \mathbf{A}_r\,\mathbf{x}_r(t) + \mathbf{B}_r\,\mathbf{u}(t), \quad \mathbf{y}_r(t) = \mathbf{C}_r\,\mathbf{x}_r(t),$$

where the reduced matrices are given by

$$\mathbf{E}_r = \mathbf{W}^T\mathbf{E}\mathbf{V}, \quad \mathbf{A}_r = \mathbf{W}^T\mathbf{A}\mathbf{V}, \quad \mathbf{B}_r = \mathbf{W}^T\mathbf{B}, \quad \text{and} \quad \mathbf{C}_r = \mathbf{C}\mathbf{V}. \tag{4}$$

2.3 Projection framework for parameterized systems

In the case of a system with no parametric dependence, the reduced quantities in (4) are precomputed constant matrices and the reduced model can be evaluated with no further reference to the full model. However, in the case of a parameterized system, the reduced model will also be parameter-dependent. Several challenges arise in achieving efficient construction and evaluation of a reduced model as the parameters vary.

The first challenge is how to introduce parametric dependence into the basis matrices $\mathbf{V}$ and $\mathbf{W}$. One option is to construct "global" basis matrices over the parameter space; that is, a single matrix $\mathbf{V}$ and a single matrix $\mathbf{W}$, each of which captures parametric dependence by embedding information regarding the entire parameter space (e.g., information collected by sampling multiple parameter values). A second option is to construct "local" basis matrices. That is, we consider K parameter sample points $\hat{\mathbf{p}}_1, \ldots, \hat{\mathbf{p}}_K$. For the realization of the dynamical system corresponding to $\hat{\mathbf{p}}_i$ (i.e., $\mathbf{E}(\hat{\mathbf{p}}_i)\,\dot{\mathbf{x}}(t) = \mathbf{A}(\hat{\mathbf{p}}_i)\,\mathbf{x}(t) + \mathbf{B}(\hat{\mathbf{p}}_i)\,\mathbf{u}(t),\ \mathbf{y}(t) = \mathbf{C}(\hat{\mathbf{p}}_i)\,\mathbf{x}(t)$), the state-space matrices are constant and we compute appropriate local basis matrices $\mathbf{V}_i$ and $\mathbf{W}_i$. There are several ways one could then use these local basis matrices to construct the parametric reduced model. For example, one might interpolate the local basis matrices over the parameter space and construct a single reduced model, or one might construct local reduced models and then interpolate the reduced models themselves. These global and local approaches are discussed in detail in Section 4.

A second challenge is achieving efficient evaluations of the parametric reduced model. For example, for a given $\mathbf{V}$ and $\mathbf{W}$ (either global or local) consider evaluating $\mathbf{A}_r(\mathbf{p}) = \mathbf{W}^T\mathbf{A}(\mathbf{p})\mathbf{V}$. For general parametric dependence, $\mathbf{A}_r$ cannot be precomputed; instead, evaluating the reduced model for a new parameter value $\mathbf{p}$ requires computing $\mathbf{A}(\mathbf{p})$ and subsequent pre- and post-multiplication by $\mathbf{W}^T$ and $\mathbf{V}$, respectively. These operations all depend on the (large) dimension n of the original problem. Fortunately, in many cases the structure of the problem admits an efficient strategy. For example, consider the case of affine parameter dependence

$$\mathbf{A}(\mathbf{p}) = \mathbf{A}_0 + \sum_{i=1}^{M} f_i(\mathbf{p})\mathbf{A}_i, \tag{5}$$

where the scalar functions f_i determine the parametric dependency, which can be nonlinear, and $\mathbf{A}_i \in \mathbb{R}^{n\times n}$ for $i = 0, \ldots, M$. Then, the reduced matrix is given by

$$\mathbf{A}_r(\mathbf{p}) = \mathbf{W}^T\mathbf{A}(\mathbf{p})\mathbf{V} = \mathbf{W}^T\mathbf{A}_0\mathbf{V} + \sum_{i=1}^{M} f_i(\mathbf{p})\mathbf{W}^T\mathbf{A}_i\mathbf{V}.$$

For affine parametric dependence in the other matrices $\mathbf{E}(\mathbf{p})$, $\mathbf{B}(\mathbf{p})$, and $\mathbf{C}(\mathbf{p})$, analogous expressions can be derived for the reduced-order counterparts. The two most important advantages of an affine parameterization are clear from the structure of the reduced-model: First, once the basis matrices $\mathbf{V}$ and $\mathbf{W}$ are chosen, the component reduced-order matrices, (e.g., $\mathbf{W}^T\mathbf{A}_i\mathbf{V}$, $i = 0, \ldots, M$) can be precomputed in the offline phase. Hence, the reduced model for a given $\mathbf{p}$ can be constructed without referring back to the original system, thus having a small online cost. Second, the reduced model has the same parametric structure as the original one, which may be appealing to designers who work with these models.

For the more general case where the parametric dependence is non-affine, typically one must introduce an approximation strategy in order to avoid costly $\mathcal{O}(n)$ evaluations in forming the reduced matrices for each different parameter value–though some parametric model reduction methods do not require affine parameter dependence, see Section 4. In some cases, the low-order terms of a Taylor series expansion provide a suitable approximate affine decomposition of the system matrices [154, 75, 42]. A more general approach that has been used successfully for nonlinear model reduction is interpolation. Among this class of methods, the Missing Point Estimation [12] and Gauss Newton with approximated tensors (GNAT) [45] methods both build upon the gappy POD interpolation method [60]; the Empirical Interpolation Method (EIM) [17] and its discrete variant, the Discrete Empirical Interpolation Method (DEIM) [46], conduct interpolation on a

low-dimensional basis for the nonlinear term. The EIM has been recently extended to the case where $\mathbf{A}(\mathbf{p})$ represents a PDE operator (see [53], for example).

In our finite-dimensional state-space framework as in (1), the fast evaluation of $\mathbf{W}^T\mathbf{A}(\mathbf{p})\mathbf{V}$ can be explained by exploiting how DEIM is used in handling nonlinearities for model reduction of finite-dimensional nonlinear dynamical systems $\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t))$, where $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear mapping. Let $\mathbf{a}(\mathbf{p}) = \mathrm{vec}(\mathbf{A}(\mathbf{p})) \in \mathbb{R}^{n^2}$ denote the vector obtained by stacking the columns of $\mathbf{A}(\mathbf{p})$. Similarly, define

$$\mathbf{a}_r(\mathbf{p}) = \mathrm{vec}(\mathbf{A}_r(\mathbf{p})) = \mathrm{vec}(\mathbf{W}^T\mathbf{A}(\mathbf{p})\mathbf{V}) = (\mathbf{V}^T \otimes \mathbf{W}^T)\mathbf{a}(\mathbf{p}) \in \mathbb{R}^{r^2}.$$

The goal is to approximate $\mathbf{a}(\mathbf{p})$ as $\mathbf{a}(\mathbf{p}) \approx \tilde{\mathbf{a}}(\mathbf{p}) = \mathbf{\Phi}\boldsymbol{\alpha}(\mathbf{p})$, where $\mathbf{\Phi} \in \mathbb{R}^{n^2 \times M}$ is constant and $\boldsymbol{\alpha}(\mathbf{p}) \in \mathbb{R}^M$ so that $\mathbf{a}_r(\mathbf{p})$ can be approximately computed, independent of n, using

$$\mathbf{a}_r(\mathbf{p}) = (\mathbf{V}^T \otimes \mathbf{W}^T)\mathbf{a}(\mathbf{p}) \approx (\mathbf{V}^T \otimes \mathbf{W}^T)\tilde{\mathbf{a}}(\mathbf{p}) = (\mathbf{V}^T \otimes \mathbf{W}^T)\mathbf{\Phi}\boldsymbol{\alpha}(\mathbf{p}) = \tilde{\mathbf{a}}_r(\mathbf{p}). \tag{6}$$

In (6), $(\mathbf{V}^T \otimes \mathbf{W}^T)\mathbf{\Phi}$ can be precomputed and only M evaluations appearing in $\boldsymbol{\alpha}(\mathbf{p})$ need computing. One can interpret $\tilde{\mathbf{a}}(\mathbf{p})$ as $\tilde{\mathbf{a}}(\mathbf{p}) = \mathrm{vec}(\tilde{\mathbf{A}}(\mathbf{p}))$ such that $\tilde{\mathbf{A}}(\mathbf{p})$ is an approximation to $\mathbf{A}(\mathbf{p})$ that will allow faster on-line computation of $\mathbf{A}_r(\mathbf{p})$. The DEIM basis $\mathbf{\Phi}$ is constructed by sampling the coefficient matrix $\mathbf{A}(\mathbf{p})$ at parameter values $\mathbf{p}_1, \ldots, \mathbf{p}_M$. If we let $\boldsymbol{\phi}_i = \mathrm{vec}(\mathbf{A}(\mathbf{p}_i))$, then the basis $\mathbf{\Phi}$ is given by $\mathbf{\Phi} = [\boldsymbol{\phi}_1,\ \boldsymbol{\phi}_2 \ldots, \boldsymbol{\phi}_M] \in \mathbb{R}^{n^2 \times M}$. Using the DEIM algorithm, we choose $\boldsymbol{\alpha}(\mathbf{p})$ to enforce that selected entries of $\tilde{\mathbf{a}}(\mathbf{p}_i)$ interpolate the corresponding entries of $\mathbf{a}(\mathbf{p}_i)$. Once mapped back to $\mathbf{A}(\mathbf{p})$ (i.e., when the vec operation is reversed), this corresponds to a selected set of entries of $\tilde{\mathbf{A}}(\mathbf{p}_i)$ exactly matching the corresponding entries of $\mathbf{A}(\mathbf{p}_i)$.

Let $z_1, z_2, \ldots, z_M$ be the indices to be exactly matched. There are a variety of methods to select these interpolation indices [17, 123, 46]. Construct the permutation matrix $\mathbf{Z} = [\mathbf{e}_{z_1}, \mathbf{e}_{z_2}, \ldots, \mathbf{e}_{z_M}] \in \mathbb{R}^{n^2 \times M}$ where $\mathbf{e}_i$ is the i^{th} canonical vector in $\mathbb{R}^{n^2}$. Then, forcing interpolation at the selected rows implies

$$\mathbf{Z}^T\mathbf{a}(\mathbf{p}) = \mathbf{Z}^T\mathbf{\Phi}\boldsymbol{\alpha}(\mathbf{p}) \quad \Longrightarrow \quad \boldsymbol{\alpha}(\mathbf{p}) = (\mathbf{Z}^T\mathbf{\Phi})^{-1}\mathbf{Z}^T\mathbf{a}(\mathbf{p}). \tag{7}$$

Hence, the approximation is given by

$$\tilde{\mathbf{a}}(\mathbf{p}) = \mathbf{\Phi}(\mathbf{Z}^T\mathbf{\Phi})^{-1}\mathbf{Z}^T\mathbf{a}(\mathbf{p}). \tag{8}$$

Note that, as discussed in [46], $\mathbf{\Phi}(\mathbf{Z}^T\mathbf{\Phi})^{-1}\mathbf{Z}^T$ in an oblique projector onto the range of $\mathbf{\Phi}$ and

$$\mathbf{Z}^T\tilde{\mathbf{a}}(\mathbf{p}) = \mathbf{Z}^T\mathbf{a}(\mathbf{p}),$$

i.e., the interpolation conditions are met at the required entries. Then, using (8) in (6), we obtain

$$\tilde{\mathbf{a}}_r(\mathbf{p}) = \mathrm{vec}(\tilde{\mathbf{A}}_r(\mathbf{p})) = (\mathbf{V}^T \otimes \mathbf{W}^T)\mathbf{\Phi}(\mathbf{Z}^T\mathbf{\Phi})^{-1}\mathbf{Z}^T\mathbf{a}(\mathbf{p}).$$

After reversing the vec operation, we obtain the reduced parametric coefficient matrix

$$\tilde{\mathbf{A}}_r(\mathbf{p}) = \mathbf{W}^T\Big(\sum_{i=1}^{M} \boldsymbol{\alpha}_i(\mathbf{p})\mathbf{A}_i(\mathbf{p})\Big)\mathbf{V} = \sum_{i=1}^{M} \boldsymbol{\alpha}_i(\mathbf{p})\Big(\mathbf{W}^T\mathbf{A}(\mathbf{p}_i)\mathbf{V}\Big). \tag{9}$$

Note that in (9), the matrices $\mathbf{W}^T\mathbf{A}(\mathbf{p}_i)\mathbf{V}$ can be precomputed. Hence, for a new parameter value $\mathbf{p}$, the reduced-order matrix $\tilde{\mathbf{A}}_r(\mathbf{p})$ can be computed without any $\mathcal{O}(n)$ operations, by exploiting (9) and (7).

2.4 Error measures

In model reduction, parametric or non-parametric, one needs to quantify the error introduced by the underlying approximation scheme. Different communities have used different, but closely

related, error measures. In the POD and reduced-basis communities, one usually measures the error in time domain; for a given $\hat{\mathbf{p}}$, for example, mostly using either

$$\|\mathbf{y}(\cdot\,;\hat{\mathbf{p}}) - \mathbf{y}_r(\cdot\,;\hat{\mathbf{p}})\|_{L_2} = \left(\int_0^\infty \left\|\mathbf{y}(t;\hat{\mathbf{p}}) - \mathbf{y}_r(t;\hat{\mathbf{p}})\right\|_2^2 \mathrm{d}t\right)^{1/2}, \tag{10}$$

or

$$\|\mathbf{y}(\cdot\,;\hat{\mathbf{p}}) - \mathbf{y}_r(\cdot\,;\hat{\mathbf{p}})\|_{L_\infty} = \sup_{t\geq 0} \left\|\mathbf{y}(t;\hat{\mathbf{p}}) - \mathbf{y}_r(t;\hat{\mathbf{p}})\right\|_\infty. \tag{11}$$

In the systems and control theory community, on the other hand, the concept of *transfer function* is used to analyze the accuracy of the reduced model by measuring the error in the frequency domain. For simplicity, assume $\mathbf{x}(0;\mathbf{p}) = \mathbf{0}$. Let $\mathbf{Y}(s;\mathbf{p})$ and $\mathbf{U}(s)$ denote the Laplace transforms of $\mathbf{y}(t;\mathbf{p})$ and $\mathbf{u}(t)$, respectively, where $s \in \mathbb{C}$ is the Laplace variable. Then, one can take Laplace transforms of (1) and (2) to obtain

$$\mathbf{Y}(s;\mathbf{p}) = \mathbf{H}(s;\mathbf{p})\,\mathbf{U}(s) \quad \text{and} \quad \mathbf{Y}_r(s;\mathbf{p}) = \mathbf{H}_r(s;\mathbf{p})\,\mathbf{U}(s), \tag{12}$$

where $\mathbf{H}(s;\mathbf{p})$ and $\mathbf{H}_r(s;\mathbf{p})$ are, respectively, the (parameterized) full and reduced-order transfer functions defined by

$$\mathbf{H}(s;\mathbf{p}) = \mathbf{C}(\mathbf{p})\,(s\,\mathbf{E}(\mathbf{p}) \;-\; \mathbf{A}(\mathbf{p}))^{-1}\,\mathbf{B}(\mathbf{p}) \tag{13}$$

and

$$\mathbf{H}_r(s;\mathbf{p}) = \mathbf{C}_r(\mathbf{p})\,(s\,\mathbf{E}_r(\mathbf{p}) \;-\; \mathbf{A}_r(\mathbf{p}))^{-1}\,\mathbf{B}_r(\mathbf{p}). \tag{14}$$

From (12), one can see that the output error $\mathbf{Y}(s;\mathbf{p}) - \mathbf{Y}_r(s;\mathbf{p})$ in the frequency domain is directly related to how well the reduced-order transfer function $\mathbf{H}_r(s;\mathbf{p})$ approximates $\mathbf{H}(s;\mathbf{p})$. For a given $\hat{\mathbf{p}}$, the two most common error measures are the $\mathcal{H}_\infty$ error norm, defined as

$$\left\|\mathbf{H}(\cdot\,;\hat{\mathbf{p}}) - \mathbf{H}_r(\cdot\,;\hat{\mathbf{p}})\right\|_{\mathcal{H}_\infty} = \sup_{\omega\in\mathbb{R}} \left\|\mathbf{H}(\imath\omega;\hat{\mathbf{p}}) - \mathbf{H}_r(\imath\omega;\hat{\mathbf{p}})\right\|_2, \tag{15}$$

and the $\mathcal{H}_2$ error norm defined as

$$\left\|\mathbf{H}(\cdot\,;\hat{\mathbf{p}}) - \mathbf{H}_r(\cdot\,;\hat{\mathbf{p}})\right\|_{\mathcal{H}_2} = \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \left\|\;\mathbf{H}(\imath\omega;\hat{\mathbf{p}}) - \mathbf{H}_r(\imath\omega;\hat{\mathbf{p}})\right\|_F^2 \mathrm{d}\omega\right)^{1/2}, \tag{16}$$

where $\|\cdot\|_{\mathbf{F}}$ denotes the Frobenius form.

There is a strong connection between the frequency-domain error measures (15)–(16) and the aforementioned time-domain error measures (10)–(11); namely for a given input $\mathbf{u}(t)$ with bounded L_2 norm,

$$\|\mathbf{y}(\cdot\,;\hat{\mathbf{p}}) - \mathbf{y}_r(\cdot\,;\hat{\mathbf{p}})\|_{L_2} \leq \left\|\mathbf{H}(\cdot\,;\hat{\mathbf{p}}) - \mathbf{H}_r(\cdot\,;\hat{\mathbf{p}})\right\|_{\mathcal{H}_\infty} \|\mathbf{u}\|_{L_2} \tag{17}$$

and

$$\|\mathbf{y}(\cdot\,;\hat{\mathbf{p}}) - \mathbf{y}_r(\cdot\,;\hat{\mathbf{p}})\|_{L_\infty} \leq \left\|\mathbf{H}(\cdot\,;\hat{\mathbf{p}}) - \mathbf{H}_r(\cdot\,;\hat{\mathbf{p}})\right\|_{\mathcal{H}_2} \|\mathbf{u}\|_{L_2}. \tag{18}$$

Hence, there is an equivalency in measuring the error in the time domain and the frequency domain. Even though these types of error measures, which are pointwise in the parameter $\mathbf{p}$, prove effective in practice, the ultimate goal is to minimize a joint error measure defined on a composite frequency and parameter space. Baur *et al.* [20] introduced such a composite error measure, $\mathcal{L}_2$ error in the parameter space and $\mathcal{H}_2$ error in the frequency domain, leading to

$$\|\mathbf{H} - \mathbf{H}_r\|^2_{\mathcal{H}_2\otimes\mathcal{L}_2(\Omega)} \overset{\text{def}}{=} \frac{1}{2\pi}\int_{-\infty}^{+\infty} \int_{\Omega} \left\|\mathbf{H}(\imath\omega,\,\mathbf{p}) - \mathbf{H}_r(\imath\omega,\,\mathbf{p})\right\|_F^2 \mathrm{d}p_1 \ldots \mathrm{d}p_d\,\mathrm{d}\omega. \tag{19}$$

Then, for a specific parameter dependency, [20] has introduced a method to minimize this error measure, as explained in Section 3.4. One can similarly define a composite error measure in uniform norm in both frequency and parameter domain, namely

$$\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_\infty\otimes\mathcal{L}_\infty(\Omega)} \overset{\text{def}}{=} \sup_{\omega\in\mathbb{R},\;\mathbf{p}\in\Omega} \|\mathbf{H}(\imath\omega,\,\mathbf{p}) - \mathbf{H}_r(\imath\omega,\,\mathbf{p})\|_2\,. \tag{20}$$

The ultimate goal is to construct parametric reduced models minimizing these composite measures; however, except for special cases as considered in [20], this goal remains an open challenge.

3 Basis computation

This section presents three different methods for deriving the reduced basis matrices $\mathbf{V}$ and $\mathbf{W}$: rational interpolation methods, balanced truncation and proper orthogonal decomposition (POD). The section concludes with a discussion of strategies for sampling the parameter space.

3.1 Rational interpolation methods

Over the last decade, numerous parametric model reduction methods based on a rational interpolation framework have been suggested. In accordance with the historical development, we first present the idea of multivariate Padé approximation, or "multi-moment matching," and then discuss the more general tangential interpolation approach. The tangential interpolation setting proposed in [20] provides a unifying framework for interpolatory projection-based model reduction of parametric systems and also paves the way to $\mathcal{H}_2 \otimes \mathcal{L}_2$-type optimal (locally, at least) parametric reduced models.

Moment-Matching

Moment-matching and Padé(-type) approximation for linear non-parametric systems determine a reduced model that satisfies the Hermite interpolation conditions

$$\frac{d^k}{ds^k}\mathbf{H}(\hat{s}) = \frac{d^k}{ds^k}\mathbf{H}_r(\hat{s}) \quad \text{for} \quad k = 0, 1, \ldots, N,$$

up to a maximal order N for $\hat{s} \in \mathbb{C}$ not a pole of $\mathbf{H}$. This yields a reduced model whose transfer function $\mathbf{H}_r(s)$ coincides in as many coefficients of its Taylor expansion (also called "moments") about $\hat{s}$ as possible for a given order of the reduced model. See, e.g., [13, 68] for a review of this approach and its close connection to the (nonsymmetric) Lanczos process. The case $\hat{s} = 0$ is generally referred to as moment-matching, while for $\hat{s} \neq 0$ we obtain shifted moments, and $\hat{s} = \infty$ leads to matching of the Markov parameters of the full system. One can also match moments (i.e., Taylor series coefficients) around multiple expansion (interpolation) points $\hat{s}_1, \ldots \hat{s}_k$ as opposed to a single expansion point $\hat{s}$, leading to the concept of multi-point moment-matching, also called multi-point rational interpolation. Surveys on this class of model reduction methods can be found in [13, 68], see also [9, 8, 29, 141].

The moment-matching idea can easily be extended to parametric model reduction by using multivariate Taylor expansion of $\mathbf{H}(s, \mathbf{p})$ about $(\hat{s}, \hat{\mathbf{p}})$. This has been discussed in numerous publications in the past two decades, e.g., [64, 38, 48, 58, 61, 63, 66, 82, 81, 104, 106, 113, 121, 158]. Moment-matching/interpolation properties can be proved (see, e.g., [64, 48, 158, 82]) analogously as for standard moment-matching methods such as Padé-via-Lanczos [62, 70].

Tangential Interpolation

In the tangential interpolation setting, the model reduction task is posed as follows: Given a frequency interpolation point $\hat{s} \in \mathbb{C}$, a parameter interpolation point $\hat{\mathbf{p}} \in \mathbb{R}^d$, and a nontrivial direction vector $\hat{\mathbf{r}} \in \mathbb{C}^m$, construct a reduced parametric model via projection as in (4) such that $\mathbf{H}_r(s, \mathbf{p})$ interpolates $\mathbf{H}(s, \mathbf{p})$ at $(s, \mathbf{p}) = (\hat{s}, \hat{\mathbf{p}})$ along the direction $\hat{\mathbf{r}}$, i.e., $\mathbf{H}_r(\hat{s}, \hat{\mathbf{p}})\hat{\mathbf{r}} = \mathbf{H}(\hat{s}, \hat{\mathbf{p}})\hat{\mathbf{r}}$. In this case, we say $\mathbf{H}_r(s, \mathbf{p})$ tangentially interpolates $\mathbf{H}(s, \mathbf{p})$ at $(s, \mathbf{p}) = (\hat{s}, \hat{\mathbf{p}})$ along the right direction vector $\hat{\mathbf{r}}$. Similarly, if $\hat{\boldsymbol{\ell}}^T \mathbf{H}_r(\hat{s}, \hat{\mathbf{p}}) = \hat{\boldsymbol{\ell}}^T \mathbf{H}(\hat{s}, \hat{\mathbf{p}})$ for a nontrivial vector $\hat{\boldsymbol{\ell}} \in \mathbb{C}^q$, we say $\mathbf{H}_r(s, \mathbf{p})$ tangentially interpolates $\mathbf{H}(s, \mathbf{p})$ along the left direction vector $\hat{\boldsymbol{\ell}}$. Finally, if $\hat{\boldsymbol{\ell}}^T \mathbf{H}_r'(\hat{s}, \hat{\mathbf{p}})\hat{\mathbf{r}} = \hat{\boldsymbol{\ell}}^T \mathbf{H}'(\hat{s}, \hat{\mathbf{p}})\hat{\mathbf{r}}$, where $'$ denotes differentiation with respect to the frequency variable s, we say $\mathbf{H}_r(s, \mathbf{p})$ is a bitangential Hermite interpolant to $\mathbf{H}(s, \mathbf{p})$. For a discussion of projection-based tangential interpolation for *nonparametric* systems, we refer the reader to [71]. Note that if $m > 1$ or $q > 1$, i.e. if the system is not SISO (single-input single-output), the tangential interpolation

framework is different to the standard interpolation in moment-matching where one enforces matrix interpolation, i.e., $\mathbf{H}(\hat{s},\hat{\mathbf{p}}) = \mathbf{H}_r(\hat{s},\hat{\mathbf{p}})$. This difference can prove crucial, especially for systems with a large number of inputs and outputs, as discussed in more detail below in Remark 1.

Computing the basis for rational interpolation methods

The recent work by Baur *et al.* [20] provided a unifying projection-based framework for structure-preserving, tangential interpolatory parametric model reduction, which also permitted a robust implementation. In what follows, we present the main results using the framework of [20]. As stated earlier, optimal control and optimization are two of the common applications for parametric model reduction. In these settings, when the objective functions are approximated, [1] shows that to establish convergence of the underlying optimization technique it is sufficient that the approximate models are first-order accurate, meaning that the gradient with respect to the optimization variable is matched. In our setting, this first-order accuracy corresponds to matching the gradient of the transfer function $\mathbf{H}(s,\mathbf{p})$. Therefore, we present the results only up to matching the first derivatives with respect to $\mathbf{p}$ (although higher derivatives can be matched in this framework).

The following theorem from [20] describes the conditions on the matrices $\mathbf{V}$ and $\mathbf{W}$ such that $\mathbf{H}_r(s,\mathbf{p}) = \mathbf{C}_r(\mathbf{p})\left(s\,\mathbf{E}_r(\mathbf{p}) - \mathbf{A}_r(\mathbf{p})\right)^{-1}\mathbf{B}_r(\mathbf{p})$, obtained by projection as in (4), is a rational tangential interpolant to $\mathbf{H}(s,\mathbf{p})$. $\hat{s}\in\mathbb{C}$ denotes an interpolation point in the frequency domain and $\hat{\mathbf{p}}$ is an interpolation point in the parameter domain.

Theorem 1 *Let $\hat{s}\in\mathbb{C}$ and $\hat{\mathbf{p}}\in\mathbb{C}^d$ be chosen such that both $\hat{s}\,\mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}})$ and $\hat{s}\,\mathbf{E}_r(\hat{\mathbf{p}}) - \mathbf{A}_r(\hat{\mathbf{p}})$ are invertible. Suppose $\hat{\mathbf{r}}\in\mathbb{C}^m$ and $\hat{\boldsymbol{\ell}}\in\mathbb{C}^q$ are two nontrivial vectors.*

(a) If $\left(\hat{s}\,\mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}})\right)^{-1}\mathbf{B}(\hat{\mathbf{p}})\hat{\mathbf{r}} \in \mathrm{Ran}(\mathbf{V})$, then

$$\mathbf{H}(\hat{s},\hat{\mathbf{p}})\hat{\mathbf{r}} = \mathbf{H}_r(\hat{s},\hat{\mathbf{p}})\hat{\mathbf{r}}. \tag{21}$$

(b) If $\left(\hat{\boldsymbol{\ell}}^T\mathbf{C}(\hat{\mathbf{p}})\left(\hat{s}\,\mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}})\right)^{-1}\right)^T \in \mathrm{Ran}(\mathbf{W})$, then

$$\hat{\boldsymbol{\ell}}^T\mathbf{H}(\hat{s},\hat{\mathbf{p}}) = \hat{\boldsymbol{\ell}}^T\mathbf{H}_r(\hat{s},\hat{\mathbf{p}}). \tag{22}$$

(c) Suppose, in addition, that $\mathbf{E}(\mathbf{p})$, $\mathbf{A}(\mathbf{p})$, $\mathbf{B}(\mathbf{p})$, and $\mathbf{C}(\mathbf{p})$ are continuously differentiable in a neighborhood of $\hat{\mathbf{p}}$. If $\left(\hat{s}\,\mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}})\right)^{-1}\mathbf{B}(\hat{\mathbf{p}})\hat{\mathbf{r}} \in \mathrm{Ran}(\mathbf{V})$ and $\left(\hat{\boldsymbol{\ell}}^T\mathbf{C}(\hat{\mathbf{p}})\left(\hat{s}\,\mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}})\right)^{-1}\right)^T \in \mathrm{Ran}(\mathbf{W})$, then

$$\nabla_{\mathbf{p}}\left(\hat{\boldsymbol{\ell}}^T\mathbf{H}(\hat{s},\hat{\mathbf{p}})\hat{\mathbf{r}}\right) = \nabla_{\mathbf{p}}\left(\hat{\boldsymbol{\ell}}^T\mathbf{H}_r(\hat{s},\hat{\mathbf{p}})\hat{\mathbf{r}}\right) \tag{23}$$

and

$$\hat{\boldsymbol{\ell}}^T\mathbf{H}'(\hat{s},\hat{\mathbf{p}})\hat{\mathbf{r}} = \hat{\boldsymbol{\ell}}^T\mathbf{H}_r'(\hat{s},\hat{\mathbf{p}})\hat{\mathbf{r}}, \tag{24}$$

where $'$ denotes differentiation with respect to s and $\nabla_{\mathbf{p}}$ denotes differentiation with respect to $\mathbf{p}$.

This theorem states that given $\hat{s}$, $\hat{\mathbf{p}}$ and a right direction vector $\hat{\mathbf{r}}$, adding one vector to the basis matrix $\mathbf{V}$ will satisfy the required right tangential interpolation condition (and analogously for the left direction vector $\hat{\boldsymbol{\ell}}$). The cost in each case is simply solving a (sparse) linear system, namely $\left(\hat{s}\,\mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}})\right)\mathbf{v} = \mathbf{B}(\hat{\mathbf{p}})\hat{\mathbf{r}}$ in the case of right tangential interpolation. This simplicity, both conceptually and with respect to implementation, is an advantage of the rational interpolation methods for parametric model reduction. As shown in Theorem 1, they also require basically no assumptions on system properties. The reduced model is generated quickly, usually requiring less time than generating a trajectory of the full-order model, making the "offline" phase cheap.

Theorem 1(c) reveals an important additional fact. By adding one vector each to the bases $\mathbf{V}$ and $\mathbf{W}$ to tangentially interpolate $\mathbf{H}(s,p)$, one would match additionally, *for free*, the gradient

of the transfer function $\mathbf{H}(s, \mathbf{p})$ with respect to $\mathbf{p}$ and the derivative with respect to s as shown in (23) and (24), respectively. Note that these quantities are matched without being computed; in other words no gradient with respect to $\mathbf{p}$ and no derivative with respect to s is computed in the construction of $\mathbf{V}$ or $\mathbf{W}$, yet they are still matched. This differs from the earlier work in interpolatory parametric model reduction methods, such as [81, 82, 48, 38, 37], where one-sided projection is employed, i.e., $\mathbf{W} = \mathbf{V}$. In one-sided approaches, to satisfy (23), gradients of $s\mathbf{E}(\mathbf{p}) - \mathbf{A}(\mathbf{p})$ with respect to $\mathbf{p}$ need to be computed and added to the basis $\mathbf{V}$. Of course, we emphasize that in applications where $\mathbf{E}(\mathbf{p})$ and $\mathbf{A}(\mathbf{p})$ are symmetric and symmetry needs to be preserved in the reduced model, one might be restricted to the one-sided projection due to the nature of the problem. However, for the general non-symmetric case, by taking advantage of the flexibility in choosing $\mathbf{W}$, one can obtain greater accuracy.

To illustrate Theorem 1 for multiple points, assume that the frequency interpolation points $\{s_i\}_{i=1}^{L} \in \mathbb{C}$, the parameter interpolation points $\{\mathbf{p}_j\}_{j=1}^{K} \in \mathbb{R}^d$ are given together with the right tangential directions $\{\mathbf{r}_{ij}\}_{i=1,j=1}^{K,L} \in \mathbb{C}^m$ and the left tangential directions $\{\boldsymbol{\ell}_{ij}\}_{i=1,j=1}^{K,L} \in \mathbb{C}^q$. For $i = 1, \ldots, K$ and $j = 1, \ldots, L$, define the vectors

$$\mathbf{v}_{ij} = \big(s_i\mathbf{E}(\mathbf{p}_j) - \mathbf{A}(\mathbf{p}_j)\big)^{-1}\mathbf{B}(\mathbf{p}_j)\mathbf{r}_{ij} \quad \text{and} \quad \mathbf{w}_{ij} = \big(s_i\mathbf{E}(\mathbf{p}_j) - \mathbf{A}(\mathbf{p}_j)\big)^{-T}\mathbf{C}(\mathbf{p}_j)^T\boldsymbol{\ell}_{ij},$$

and construct the model reduction bases $\mathbf{V}$ and $\mathbf{W}$ as follows:

$$\begin{aligned}\mathbf{V} &= [\mathbf{v}_{11}, \ldots, \mathbf{v}_{1L}, \mathbf{v}_{21}, \ldots, \mathbf{v}_{2L}, \ldots, \mathbf{v}_{K1}, \ldots, \mathbf{v}_{KL}] \in \mathbb{C}^{n\times(KL)},\\ \mathbf{W} &= [\mathbf{w}_{11}, \ldots, \mathbf{w}_{1L}, \mathbf{w}_{21}, \ldots, \mathbf{w}_{2L}, \ldots, \mathbf{w}_{K1}, \ldots, \mathbf{w}_{KL}] \in \mathbb{C}^{n\times(KL)}.\end{aligned}$$

Then, the resulting projection-based parametric reduced model satisfies the interpolation conditions of Theorem 1 for every pair $(\hat{s}, \hat{\mathbf{p}}) = (s_i, \mathbf{p}_j)$ where $i = 1, \ldots, L$ and $j = 1, \ldots, K$. We can go one step further and match the Hessian (curvature) information with respect to the parameters as well. This is done in a similar way by adding additional vectors to the reduction subspaces. For details, see [20].

Remark 1 *Given $\hat{s} \in \mathbb{C}$, $\hat{\mathbf{p}} \in \mathbb{C}^d$, and $\hat{\mathbf{r}} \in \mathbb{C}^m$, tangential interpolation requires $\big(\hat{s}\,\mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}})\big)^{-1}\mathbf{B}(\hat{\mathbf{p}})\hat{\mathbf{r}} \in \mathrm{Ran}(\mathbf{V})$; i.e., a single vector is added to the subspace. On the other hand, full matrix interpolation requires $\big(\hat{s}\,\mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}})\big)^{-1}\mathbf{B}(\hat{\mathbf{p}}) \in \mathrm{Ran}(\mathbf{V})$; i.e., m new vectors are added to the subspace. If we require interpolation at, for example, L frequency points and K parameter points, full matrix interpolation can lead to $\mathrm{Ran}(\mathbf{V})$ having dimension as large as mLK. For applications where the system input dimension m is large, this would lead to a rather large reduced model dimension. In comparison, tangential interpolation will at most lead to a reduced dimension of LK, thus making the reduced order independent of the input dimension. We note that in the full matrix interpolation case one can keep the dimension grow modest by truncating the linearly dependent components from the model reduction bases $\mathbf{V}$ and $\mathbf{W}$ while still obtaining accurate reduced models; see, for example, [39, 47, 22].*

Remark 2 *As noted above, if the system is not SISO, the tangential interpolation is different than the full-matrix interpolation, i.e., full moment matching, such as in [64, 38, 48, 58, 61, 82, 106, 121, 158]. Different methods proposed in these references differ in the way moments are computed (implicitly vs. explicitly) and in the number of (mixed) moments that are matched. Approaches based on explicitly computed moments suffer from the same numerical instabilities as analogous methods for model reduction of non-parametric systems. Implicit approaches appear to provide a robust resolution of these difficulties at least for low-dimensional parameter spaces.*

Optimal-$\mathcal{H}_2$ tangential interpolation for nonparametric systems

In rational interpolation methods, the choice of expansion methods and tangential directions determines the accuracy of the reduced model. Until recently, this choice was largely *ad hoc*. For systems without parametric dependency, recently Gugercin *et al.* [78] have introduced the Iterative

Rational Krylov Algorithm (IRKA) to compute the frequency interpolation points optimally (at least locally) to minimize the $\mathcal{H}_2$ error norm defined in (16) using the optimality conditions by Meier and Luenberger in [118]; see also, [44, 151], for related work. IRKA can be used for parametric systems as well to compute locally optimal reduced models for the given parameter sample. This local information can then be used to construct a parametric reduced model using the ideas of Section 4, as done in [20, 56].

Suppose $\mathbf{H}_r(s)$ is an $\mathcal{H}_2$ optimal approximation to $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}$. Let $\widetilde{\mathbf{H}}_r(s) = \sum_{i=1}^{r} \frac{1}{s-\widetilde{\lambda}_i}\tilde{\mathbf{c}}_i\tilde{\mathbf{b}}_i^T$ be a partial fraction expansion of $\mathbf{H}_r(s)$. Note that $\tilde{\mathbf{c}}_i\tilde{\mathbf{b}}_i^T$ is the residue of $\mathbf{H}_r(s)$ at $s = \widetilde{\lambda}_i$; $\tilde{\mathbf{c}}_i$ and $\tilde{\mathbf{b}}_i^T$ are called the residue directions. Then, the $\mathcal{H}_2$ optimality of $\mathbf{H}_r(s)$ means that, for $i = 1, \cdots, r$,

$$\begin{aligned} \mathbf{H}(-\widetilde{\lambda}_i)\tilde{\mathbf{b}}_i = \widetilde{\mathbf{H}}_r(-\widetilde{\lambda}_i)\tilde{\mathbf{b}}_i, \quad \tilde{\mathbf{c}}_i^T\mathbf{H}(-\widetilde{\lambda}_i) = \tilde{\mathbf{c}}_i^T\widetilde{\mathbf{H}}_r(-\widetilde{\lambda}_i), \\ \text{and} \quad \tilde{\mathbf{c}}_i^T\mathbf{H}'(-\widetilde{\lambda}_i)\tilde{\mathbf{b}}_i = \tilde{\mathbf{c}}_i^T\widetilde{\mathbf{H}}_r'(-\widetilde{\lambda}_i)\tilde{\mathbf{b}}_i. \end{aligned} \tag{25}$$

In other words, an $\mathcal{H}_2$-optimal reduced model $\mathbf{H}_r(s)$ is a bitangential Hermite interpolant to $\mathbf{H}(s)$. The optimal interpolation points are the mirror images of the poles of $\mathbf{H}_r(s)$ and the optimal tangential directions are the corresponding residue directions. Thus, the optimal points and associated tangent directions depend on the reduced model and are not known *a priori*. Then, starting with an initial selection of interpolation points and tangential directions, IRKA [78], using successive substitution, iteratively corrects the interpolation points and tangential directions until the optimality conditions in (25) are satisfied. For details of IRKA, we refer the reader to [78, 9].

3.2 Balanced truncation

In the systems and control theory community, balanced truncation (BT) [122, 120] is one of the most common techniques for approximating linear dynamical systems without parametric dependency. In the parametric setting, BT can be employed to construct local reduced models at given parameter values. These local models can be used in various ways to construct a parametric reduced model, as will be discussed in Section 4. Here we describe construction of a BT reduced model at a single parameter point, $\hat{\mathbf{p}}$. For ease of notation, for the remainder of this subsection we denote $\mathbf{E}(\hat{\mathbf{p}})$ by $\mathbf{E}$ and similarly for the other matrices, so that $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}$ is our dynamical system (1) evaluated at $\hat{\mathbf{p}}$.

The BT basis matrices $\mathbf{V}$ and $\mathbf{W}$ depend on the two system Gramians, which are defined by the integrals

$$\mathcal{P} = \int_0^\infty e^{\mathbf{E}^{-1}\mathbf{A}t}\,\mathbf{E}^{-1}\mathbf{B}\mathbf{B}^T\mathbf{E}^{-T}\,e^{\mathbf{A}^T\mathbf{E}^{-T}t}dt \tag{26}$$

and

$$\mathcal{Q} = \int_0^\infty \mathbf{E}^{-T}e^{\mathbf{A}^T\mathbf{E}^{-T}t}\,\mathbf{C}^T\mathbf{C}\,e^{\mathbf{E}^{-1}\mathbf{A}t}\mathbf{E}^{-1}dt. \tag{27}$$

Here, $\mathcal{P} \in \mathbb{R}^{n\times n}$ and $\mathcal{Q} \in \mathbb{R}^{n\times n}$ are called *the reachability Gramian* and *the observability Gramian*, respectively. The "reachability" of a state $\mathbf{x}$ is a measure of how easy it is to reach the state $\mathbf{x}$ from the zero state. On the other hand, the observability of a state $\mathbf{x}_0$ is a measure of how easy it is to distinguish the initial state $\mathbf{x}_0$ from the zero state by observing the output $\mathbf{y}(t)$ in the case of zero input.

To explain these concepts further, define

$$\mathcal{J}_r(\mathbf{x}) \;:=\; \min_{\mathbf{x}(-\infty)=\mathbf{0},\ \mathbf{x}(0)=\mathbf{x}} \|\mathbf{u}(t)\|^2, \quad t \le 0 \tag{28}$$

$$\mathcal{J}_o(\mathbf{x}_0) \;:=\; \|\mathbf{y}(t)\|^2, \quad \mathbf{x}(0) = \mathbf{x}_0,\ \mathbf{u}(t) = \mathbf{0},\ t \ge 0. \tag{29}$$

$\mathcal{J}_r(\mathbf{x})$ is the minimal energy required to drive the system from the zero state at $t = -\infty$ to the state $\mathbf{x}$ at $t = 0$. On the other hand, $\mathcal{J}_o(\mathbf{x}_0)$ is the energy obtained by observing the output due

to the initial state $\mathbf{x}_o$ with zero input. The smaller the reachability energy $\mathcal{J}_r(\mathbf{x})$, the easier to reach the state $\mathbf{x}$. The larger the observability energy $\mathcal{J}_o(\mathbf{x}_0)$, the easier to observe the state $\mathbf{x}_0$. These two energies are completely determined by the gramians $\mathcal{P}$ and $\mathcal{Q}$:

$$\mathcal{J}_r(\mathbf{x}) = \mathbf{x}^T \mathcal{P}^{-1} \mathbf{x}, \qquad \text{and} \qquad \mathcal{J}_r(\mathbf{x}_0) = \mathbf{x}_0^T \mathcal{Q}\, x_0.$$

Thus, $\mathcal{P}$ and $\mathcal{Q}$ explain how important a state $\mathbf{x}$ is for the input-to-state and the state-to-output mapping, respectively.

When $\mathbf{H}(s)$ is asymptotically stable, both $\mathcal{P}$ and $\mathcal{Q}$ are positive semi-definite matrices. Square roots of the eigenvalues of the product $\mathcal{PQ}$ are called the Hankel singular values of $\mathbf{H}(s)$, denoted by $\eta_i(\mathbf{H})$, and they are are the singular values of the Hankel operator associated with $\mathbf{H}(s)$. The states corresponding to the smallest Hankel singular values are the least important states in terms of both the input-to-output map; that is, these are hard to reach and hard to observe states. Model reduction via BT corresponds to eliminating those states corresponding to small Hankel singular values.

In practice one does not need to evaluate the infinite integrals (26) and (27); instead one solves the corresponding Lyapunov equations that $\mathcal{P}$ and $\mathcal{Q}$ satisfy, namely

$$\mathbf{A}\mathcal{P}\mathbf{E}^T + \mathbf{E}\mathcal{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0} \quad \text{and} \quad \mathbf{A}^T\mathcal{Q}\mathbf{E} + \mathbf{E}^T\mathcal{Q}\mathbf{A} + \mathbf{C}^T\mathbf{C} = \mathbf{0}. \tag{30}$$

Instead of forming the full gramians $\mathcal{P}$ and $\mathcal{Q}$ explicitly, one computes $\mathcal{P} = \mathbf{U}\mathbf{U}^T$ and $\mathcal{Q} = \mathbf{L}\mathbf{L}^T$ in the factored form. See [132, 105, 79, 139, 155, 91, 95, 31, 54, 94, 142, 148] and the references therein for effective solution methods for Lyapunov equations.

Given the factorizations $\mathcal{P} = \mathbf{U}\mathbf{U}^T$ and $\mathcal{Q} = \mathbf{L}\mathbf{L}^T$, let $\mathbf{U}^T\mathbf{E}\mathbf{L} = \mathbf{Z}\mathbf{S}\mathbf{Y}^T$ be the singular value decomposition with $\mathbf{S} = \operatorname{diag}(\eta_1, \eta_2, \ldots, \eta_n)$. Let $\mathbf{S}_r = \operatorname{diag}(\eta_1, \eta_2, \ldots, \eta_r)$ with $\eta_{r+1} < \eta_r$ and $r < n$. BT chooses

$$\mathbf{V} = \mathbf{U}\mathbf{Z}_r\mathbf{S}_r^{-1/2} \qquad \text{and} \qquad \mathbf{W} = \mathbf{L}\mathbf{Y}_r\mathbf{S}_r^{-1/2}, \tag{31}$$

where $\mathbf{Z}_r$ and $\mathbf{Y}_r$ denote the leading r columns of left singular vectors, $\mathbf{Z}$, and right singular vectors, $\mathbf{Y}$, respectively. Then the reduced model is obtained by following the projection in (4). The reduced model $\mathbf{H}_r(s)$ obtained by BT is asymptotically stable and the $\mathcal{H}_\infty$ norm of the error system satisfies $\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_\infty} \leq 2(\eta_{r+1} + \cdots + \eta_n)$. For more details on BT, see [8].

3.3 Proper orthogonal decomposition

Due to its broad applicability to linear and nonlinear systems, the proper orthogonal decomposition (POD) has become widely used in many different application domains as a method for computing the reduced basis. While the rational interpolation methods of Section 3.1 formulate the basis computation task in the frequency domain, POD formulations typically use the time domain. In the case of LTI systems, duality between time and frequency domain formulations reveals the connections between POD and balanced truncation.

Time domain POD

POD was introduced for the analysis of turbulent flows by Lumley [111], and is closely related to methods used in other fields such as Karhunen-Loéve expansions in stochastic process modeling [110, 98], principal component analysis in statistical analysis [90], and empirical orthogonal eigenfunctions in atmospheric modeling [124]. POD basis vectors are computed empirically using sampled data collected over a range of relevant system dynamics, typically using the method of snapshots, introduced by Sirovich [143].

Consider a set of "snapshots", $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_s}$: state solutions computed at different instants in time and/or different parameter values, where $\mathbf{x}_j \in \mathbb{R}^n$ denotes the jth snapshot and we collect a total of n_s snapshots. Define the snapshot matrix $\mathbf{X} \in \mathbb{R}^{n \times n_s}$ whose jth column is the snapshot $\mathbf{x}_j$. The (thin) singular value decomposition of $\mathbf{X}$ is written

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{Y}^T, \tag{32}$$

where the columns of the matrices $\mathbf{U} \in \mathbb{R}^{n \times n_s}$ and $\mathbf{Y} \in \mathbb{R}^{n_s \times n_s}$ are the left and right singular vectors of $\mathbf{X}$, respectively. $\mathbf{\Sigma} \in \mathbb{R}^{n_s \times n_s} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{n_s})$, where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{n_s} \geq 0$, are the singular values of $\mathbf{X}$. The POD basis is chosen as the r left singular vectors of $\mathbf{X}$ that correspond to the r largest singular values, $\mathbf{V} = \mathbf{U}_{1:r}$.

The POD basis is "optimal" in the sense that, for a basis of size r, it minimizes the least squares error of snapshot reconstruction,

$$\min_{\mathbf{V} \in \mathbb{R}^{n \times r}} \sum_{i=1}^{n_s} ||\mathbf{x}_i - \mathbf{V}\mathbf{V}^T \mathbf{x}_i||_2^2 = \sum_{i=r+1}^{n_s} \sigma_i^2. \tag{33}$$

As can be seen from (33), the error in snapshot representation is given by the sum of the squares of the singular values corresponding to those left singular vectors not included in the POD basis. Thus, the singular values provide quantitative guidance for choosing the size of the POD basis. A typical approach is to choose r so that

$$\frac{\sum_{i=1}^{r} \sigma_i^2}{\sum_{i=1}^{n_s} \sigma_i^2} > \kappa, \tag{34}$$

where κ is a user-specified tolerance, often taken to be 99.9% or greater. The numerator of (34) is often referred to as the "energy" captured by the POD modes.

Since the POD basis is constructed from sampled solutions, it makes no assumptions about the form of the full model; it applies to both linear and nonlinear systems, as well as to parametrically-varying systems. It is important to note that the optimality of the POD basis applies only to error in reconstruction of the snapshots, not to the error in solution of the POD-based reduced model. Clearly, the choice of snapshots is critical to the quality of the reduced model, although the POD theory per se gives no guidance on how to select the snapshots. Strategies for snapshot selection are discussed in Section 3.4.

Frequency domain POD

For linear systems, one can derive the POD in the frequency domain. We present a brief discussion here because it highlights the connection between POD and balanced truncation. The POD basis vectors are the left singular vectors of the snapshot matrix $\mathbf{X}$, and thus are the eigenvectors of the matrix

$$\mathbf{K} = \mathbf{X}\mathbf{X}^T = \sum_{i=1}^{n_s} \mathbf{x}_i \mathbf{x}_i^T. \tag{35}$$

In the original formulation of the POD method of snapshots, $\mathbf{K}$ is referred to as the kernel [143]. In some applications of POD, the snapshots are centered to have zero mean, in which case $\mathbf{K}$ is a covariance matrix.

Kim [97] develops the frequency domain POD method by showing that through a simple application of Parseval's theorem, one can write the kernel as

$$\mathbf{K} = \frac{1}{2\pi} \sum_{i=1}^{n_s} \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^* \Delta\omega_i. \tag{36}$$

Here, $\bar{\mathbf{x}}_i$ is now the ith (complex) snapshot computed at sample frequency ω_i,

$$\bar{\mathbf{x}}_i = (j\omega_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{B}, \tag{37}$$

$\Delta\omega_i = \omega_i - \omega_{i-1}$, and $*$ denotes the complex conjugate transpose.

Balanced POD

The connection between POD and balanced truncation is described in [101]. As discussed in [159], the approximation introduced by the POD is further highlighted through the frequency domain formulation. Using (37) in (36), we write

$$\mathbf{K} = \frac{1}{2\pi} \sum_{i=1}^{n_s} (j\omega_i \mathbf{E} - \mathbf{A})^{-1} \mathbf{B}\mathbf{B}^* (-j\omega_i \mathbf{E}^* - \mathbf{A}^*)^{-1} \Delta\omega_i, \tag{38}$$

which shows that the frequency domain kernel POD approximates the integral defining the reachability gramian,

$$\mathcal{P} = \frac{1}{2\pi} \int_{-\infty}^{\infty} (j\omega \mathbf{E} - \mathbf{A})^{-1} \mathbf{B}\mathbf{B}^* (-j\omega \mathbf{E}^* - \mathbf{A}^*)^{-1} \, d\omega. \tag{39}$$

The POD basis vectors are the dominant eigenvectors of the POD kernel and thus approximate the most reachable modes in the system.

Based on this observation, [159] proposed an approximate balanced truncation approach using the POD method of snapshots. Computing a set of dual POD modes from snapshots of the dual (adjoint) system leads to an approximation of the most observable modes in the system. Appropriately combining the primal and dual snapshots leads to an approximate balanced truncation. A modified algorithm for computing this snapshot-based approximate balanced truncation was proposed in [136].

3.4 Parameter sampling

We conclude this section on basis computation with a discussion of parameter sampling. The choice of parameter sample points is a critical question that arises in all methods to compute the basis. In the rational interpolation methods, we must select parameter samples at which interpolation conditions are applied; for balanced truncation, we must select parameter samples to generate the local LTI systems at which balanced truncation is applied; and in the POD, we must select parameter samples at which snapshots are computed. For problems with a small number of parameters, a structured or random sampling method (e.g., grid-based sampling or Latin hypercube sampling) is the simplest approach and, with a sufficiently high number of samples, will generate a rich set of data that covers the parameter space. For a moderate number of parameters, a sparse grid sampling approach will likely be effective. However, when the dimension of the parameter space is large ($d > 10$), it becomes challenging to balance sampling cost with coverage of the parameter space. In these cases, we require more sophisticated sampling approaches, such as a problem-aware adaptive search of the parameter space. In this section, we review adaptive sampling via greedy search methods. We also discuss sparse grid sampling and optimal interpolation points for a special class of systems.

Adaptive parameter sampling via greedy search

Greedy sampling methods for model reduction approach the task of choosing parameter sample points one-by-one in an adaptive manner. The general steps in the greedy sampling approach are as follows. First, given a current reduced model,

$$\mathbf{E}_r(\mathbf{p})\, \dot{\mathbf{x}}_r(t;\mathbf{p}) = \mathbf{A}_r(\mathbf{p})\, \mathbf{x}_r(t;\mathbf{p}) + \mathbf{B}_r(\mathbf{p})\, \mathbf{u}(t), \mathbf{y}_r(t;\mathbf{p}) = \mathbf{C}_r(\mathbf{p})\, \mathbf{x}_r(t;\mathbf{p}),$$

find the parameter value for which the error between the reduced model and the full model is largest:

$$\hat{\mathbf{p}} = \arg\max_{\mathbf{p}} \|\mathbf{y}(\cdot\,;\mathbf{p}) - \mathbf{y}_r(\cdot\,;\mathbf{p})\|_{L_2}. \tag{40}$$

Second, sample $\hat{\mathbf{p}}$ (i.e., solve the full model at $\hat{\mathbf{p}}$) to generate new information with which to update the reduced model. Then with the updated reduced model, repeat these two steps until the error is acceptable. The greedy sampling method was first introduced in the context of reduced basis

methods in [134, 154, 153, 75, 74]. It has since been applied in conjunction with POD methods [41, 84] and rational interpolation methods [55]. The key advantage of the greedy approach is that the search over the parameter space embodies the structure of the problem, so that the underlying system dynamics guide the selection of appropriate parameter samples.

In the first step, the task of finding the worst-case parameter value can be approached in a variety of ways. In the general case, using the actual reduced model error as a metric leads to a computationally intractable algorithm, since it requires evaluating the full model solution at many parameter points. Instead, one can use *a posteriori* error estimators for parameterized elliptic and parabolic PDEs [134, 154, 153, 75, 74] or a residual-based error indicator [41]. The process of locating the parameter of maximal error indicator or error estimate can be conducted by a simple grid search if the parameter dimension is low. For problems with more than a handful of parameters, the greedy sampling approach can be formulated as a sequence of adaptive model-constrained optimization problems [41]. In the linear case, these optimization problems have explicit solutions in the form of an eigenvalue problem [18]. In [107], the optimization-based greedy sampling approach was extended to construct both a basis for the state and a basis for the parameter, leading to models that have both reduced state and reduced parameters.

Sparse grids

Another strategy for an effective and representative choice of parameter points in higher dimensional parameter spaces (for example, say, with $d = 10$) comes through the use of sparse grids. Based on Smolyak's [144] earlier construction, sparse grid methods have gained increasing popularity over the past two decades in various areas within computational science and engineering dealing with high-dimensional approximation. This popularity is due to the development of numerical algorithms to implement them, see, e.g., [43, 96, 76, 161]. Sparse grids are based on a hierarchical basis and a sparse tensor product construction. This way, a significant reduction of interpolation points is achieved while preserving the level of accuracy obtainable on a tensor grid with the same spatial resolution. Popular constructions on bounded domains are the Clenshaw-Curtis rules based on Gaussian quadrature. Since refining the corresponding mesh yields non-nested interpolation points, it is often more efficient to apply multivariate extensions of the nested Gauss-Kronrod rules (i.e., the Kronrod-Patterson or Gauss-Patterson rules). Although these methods do not yield the optimal convergence rate, due to the re-usability of computed snapshots of model evaluations from coarser grids, they are generally preferred from an efficiency point of view in parametric model reduction applications. They also provide easy-to-implement error estimators. The application of sparse grid methods in parametric model reduction was first discussed in [19]; applications of the resulting methods in the design of MEMS can be found in [21].

Optimal interpolation points for a special class of systems

Even though the parameter selection strategies outlined above have led to high quality reduced parametric models, they are not optimal in the error measures defined in Section 2.4. In this section, for a special class of parametric dependency we will outline an optimal parameter selection strategy introduced by Baur *et al.* [20]. Inspired by IRKA of [78] for the nonparametric systems, [20] proposed a method to minimize the system-theoretic composite error measure $\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2 \otimes \mathcal{L}_2(\Omega)}$ defined in (19) in order to find optimal sampling points in both frequency and parameter space.

To simplify the presentation, we will assume a single-input/single-output dynamical system; extension to the multi-input/multi-output dynamical system case is trivial. The special parametric dependency assumes affine parameterization parameters in $\mathbf{B}(\mathbf{p})$ and $\mathbf{C}(\mathbf{p})$ and no parameter dependency in either $\mathbf{A}(\mathbf{p})$ or $\mathbf{E}(\mathbf{p})$. Without loss of generality, assume only one parameter in $\mathbf{B}(\mathbf{p})$ and $\mathbf{C}(\mathbf{p})$. Let $\mathbf{p} = [p_1, p_2]^T$. Then, the parametric model we consider has the form

$$\mathbf{H}(s, \mathbf{p}) = (\mathbf{c}_0^T + p_1 \mathbf{c}_1^T)\,(s\,\mathbf{E} - \mathbf{A})^{-1}\,(\mathbf{b}_0 + p_2 \mathbf{b}_1), \tag{41}$$

where $\mathbf{c}_0, \mathbf{c}_1, \mathbf{b}_0, \mathbf{b}_1 \in \mathbb{R}^n$ and $[p_1, p_1] \in \Omega \in \mathbb{R}^2$. For $\mathbf{H}(s, \mathbf{p})$ as in (41), then, the goal is to find an *optimal* reduced model, $\mathbf{H}_r(s, \mathbf{p})$, such that the composite error $\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2 \otimes \mathcal{L}_2(\Omega)}$ defined in (19) is minimized. The crucial observation of [20] is that solving the parametric $\mathcal{H}_2 \otimes \mathcal{L}_2(\Omega)$ optimization problem is equivalent to solving a nonparametric $\mathcal{H}_2$ optimization problem. Then, [20] used IRKA to solve the equivalent nonparametric $\mathcal{H}_2$ optimal problem and constructed the $\mathcal{H}_2 \otimes \mathcal{L}_2(\Omega)$ *optimal* parametric reduced model $\mathbf{H}_r(s, \mathbf{p}) = (\mathbf{c}_{0,r} + p_1\, \mathbf{c}_{1,r})^T (s\mathbf{E}_r - \mathbf{A}_r)^{-1} (\mathbf{b}_{0,r} + p_2\, \mathbf{b}_{1,r})$. Let $\{\tilde{p}_i\}_{i=1}^r$ denote the resulting optimal interpolation points. Then, for $i = 1, \ldots, r$,

$$\begin{aligned} \mathbf{H}(-\widetilde{\lambda}_i, \tilde{p}_i) = \mathbf{H}_r(-\widetilde{\lambda}_i, \tilde{p}_i), \qquad \mathbf{H}'(-\widetilde{\lambda}_i, \tilde{p}_i) = \mathbf{H}_r'(-\widetilde{\lambda}_i, \tilde{p}_i), \\ \text{and} \quad \nabla_{\mathbf{p}} \mathbf{H}(-\widetilde{\lambda}_i, \tilde{p}_i) = \nabla_{\mathbf{p}} \widetilde{H}_r(-\widetilde{\lambda}_i, \tilde{p}_i), \end{aligned} \tag{42}$$

where $\{\widetilde{\lambda}_i\}_{i=1}^r$ are the eigenvalues of the reduced pencil $s\mathbf{E}_r - \mathbf{A}_r$. In other words, the optimal reduced model is an Hermite interpolant in both s and $\mathbf{p}$. Note that every $\widetilde{\lambda}_i$ is associated with a different $\tilde{p}_i$; in other words the frequency and the parameter interpolation points are selected *jointly* to minimize an associated system theoretic $\mathcal{H}_2 \otimes \mathcal{L}_2(\Omega)$ error measure.

4 Parameterized reduced model generation

In this section, we discuss different strategies for constructing the parameterized reduced model. Many of these construction options are broadly applicable in that they can be used with any of the basis computation methods discussed in Section 3. Section 4.1 discusses approaches that use a single global basis computed by sampling many parameter values, while Section 4.2 presents methods that use multiple local basis matrices each computed at different parameter values. In the following discussion, our parameter $\mathbf{p}$ belongs to a single domain $\Omega \subset \mathbb{R}^d$. Recent work has proposed approaches to split Ω into multiple subdomains and construct reduced models in each subdomain [83, 52, 57, 4, 157]. Any of the model generation strategies described in the following can be applied in a partitioning setting by replacing Ω with the corresponding subdomain.

4.1 A global basis over the parameter space

A global basis is defined by a single pair of basis matrices $\mathbf{V}$ and $\mathbf{W}$, which are built by sampling information over a range of parameters. These basis matrices could be computed using any one of the methods described in Section 3. In this case, the parametric reduced model takes the form

$$\begin{aligned} \mathbf{W}^T \mathbf{E}(\mathbf{p}) \mathbf{V}\, \dot{\mathbf{x}}_r(t; \mathbf{p}) &= \mathbf{W}^T \mathbf{A}(\mathbf{p}) \mathbf{V}\, \mathbf{x}_r(t; \mathbf{p}) + \mathbf{W}^T \mathbf{B}(\mathbf{p})\, \mathbf{u}(t), \\ \mathbf{y}_r(t; \mathbf{p}) &= \mathbf{C}(\mathbf{p}) \mathbf{V}\, \mathbf{x}_r(t; \mathbf{p}). \end{aligned} \tag{43}$$

(43) shows that construction of the ROM requires evaluation of terms such as $\mathbf{A}_r(\mathbf{p}) = \mathbf{W}^T \mathbf{A}(\mathbf{p}) \mathbf{V}$, $\mathbf{E}_r(\mathbf{p}) = \mathbf{W}^T \mathbf{E}(\mathbf{p}) \mathbf{V}$, etc. As already discussed in Section 2, if we evaluate $\mathbf{A}_r$, $\mathbf{E}_r$, and the other reduced matrices in this way for every new parameter value $\mathbf{p}$, then our reduced model will be inefficient and computational savings over solving the original full system will be small. Section 2 showed that for system matrices that depend affinely on the parameter $\mathbf{p}$, computational efficiency can be recovered. In this case, the parameterized reduced model is decomposed into reduced-order matrices that do not depend on the parameter and hence can be precomputed. In the more general case, a method such as EIM or DEIM (also discussed in Section 2) is used to approximate $\mathbf{A}_r(\mathbf{p})$ and the other reduced-order matrices in a way that admits efficient evaluation over the parameter space.

Concatenation of the basis

One of the common approaches to obtain the global basis matrices $\mathbf{V}$ and $\mathbf{W}$ is to concatenate the local basis matrices obtained for several parameter samples $\mathbf{p}_1, \ldots, \mathbf{p}_K$. Suppose that $\mathbf{V}_1, \ldots, \mathbf{V}_K$

and $\mathbf{W}_1, \ldots, \mathbf{W}_K$ denote the local basis matrices corresponding to $\mathbf{p}_1, \ldots, \mathbf{p}_K$. Then, one can construct the global basis matrices $\mathbf{V}$ and $\mathbf{W}$ using

$$\mathbf{V} = [\mathbf{V}_1, \ \mathbf{V}_2, \ \ldots, \ \mathbf{V}_K] \ \text{ and } \ \mathbf{W} = [\mathbf{W}_1, \ \mathbf{W}_2, \ \ldots, \ \mathbf{W}_K].$$

However, it is quite possible that the local matrices have common components among each other; leading to potentially rank-deficient global basis matrices $\mathbf{V}$ and $\mathbf{W}$. To avoid this scenario, the concatenation step is usually followed by an SVD or a rank-revealing QR factorization to remove these rank deficient components from $\mathbf{V}$ and $\mathbf{W}$, leading to global basis matrices with orthonormal columns. It is important to note that even though theoretically it would not matter whether the local matrices $\mathbf{V}_i$ and $\mathbf{W}_i$ are orthogonalized prior to the concatenation step (since the reduced model is determined by the range, not by a specific basis), numerically it might.

The local basis matrices can be obtained using any one of the methods described in Section 3. However, the method of choice results in different properties in the reduced model. The concatenation approach is especially appropriate when the local basis matrices are obtained by rational interpolation methods of Section 3.1. Due to Theorem 1, even after concatenation, the final reduced parameterized model obtained by the global basis matrices will still interpolate the original model at every frequency and parameter interpolation point combination used in constructing every local basis matrix, assuming that the SVD performed on the concatenated matrices removed only the zero singular values. If this SVD step removes some small non-zero singular values below a numerical tolerance, then the interpolation will be approximate. This is in contrast to, for example, the balanced truncation (BT) approach. If the local basis matrices are obtained via BT (i.e., $\mathbf{V}_i$ and $\mathbf{W}_i$ are the BT basis matrices at the parameter values $\mathbf{p}_i$), then once the concatenation is performed the resulting reduced model is no longer guaranteed to be balanced even at the parameter value $\mathbf{p}_i$. Nevertheless, concatenation of local BT basis may yield a good reduced parametric model as balancing is not an important?property of the reduced model and is actually usually avoided in numerical implementations of BT even in the nonparametric case.

4.2 Local bases at multiple parameter points

As opposed to constructing fixed global basis matrices $\mathbf{V}$ and $\mathbf{W}$, one might construct several local ones by sampling the parameter space at points $\mathbf{p}_1, \ldots, \mathbf{p}_K$. As before, we denote these local bases by $\mathbf{V}_k$ and $\mathbf{W}_k$ for $k = 1, \ldots, K$.

Interpolating the local bases

Given a set of local basis matrices, $\{\mathbf{V}_k\}_{k=1}^K$ and $\{\mathbf{W}_k\}_{k=1}^K$, a basis matrix for a new parameter value can be obtained by interpolating the local reduced bases. Straightforward interpolation of the entries in the basis vectors can lead to an interpolated quantity that does not preserve desired properties (e.g., the interpolated vectors might no longer have desired orthogonality properties). A better method interpolates the local reduced bases $\mathbf{V}_k$ and $\mathbf{W}_k$ for $k = 1, \ldots, K$ on a tangent space to a manifold of these reduced bases [2]. The manifold is chosen so as to preserve desired properties. Below we briefly explain this approach for the particular case of constructing an orthonormal $\mathbf{V}$, where the $\mathbf{V}_k, k = 1, \ldots, K$, each represents an orthonormal basis, and the interpolation is done on the tangent space to the *Stiefel manifold*.

The Stiefel manifold, $\mathcal{S}_{n,r}$, is the set of all r-dimensional orthonormal bases of $\mathbb{R}^n$, for $1 \leq r \leq n$. The orthonormal basis $\mathbf{V}_k \in \mathbb{R}^{n \times r}$ is a point on the Stiefel manifold $\mathcal{S}_{n,r}$. The first step in the approach proposed by Amsallem et al. [2] is to choose a reference point. For simplicity, we will take $\mathbf{V}_1$ as the reference point. Let $\mathcal{T}_{\mathbf{V}_1}$ be the tangent space of $\mathcal{S}_{n,r}$ at $\mathbf{V}_1 \in \mathcal{S}_{n,r}$. The next step is to map all other local bases $\mathbf{V}_2, \mathbf{V}_3, \ldots, \mathbf{V}_K$ onto the tangent space defined by $\mathbf{V}_1$. A point $\mathbf{V}_k \in \mathcal{S}_{n,r}$ in a neighborhood of $\mathbf{V}_1$ can be mapped to $\mathcal{T}_{\mathbf{V}_1}$ by the logarithmic map:

$$\mathbf{T}_k = \mathrm{Log}_{\mathbf{V}_1}(\mathbf{V}_k) \quad \in \quad \mathcal{T}_{\mathbf{V}_1}. \tag{44}$$

To compute $\mathbf{T}_k$, one first computes a short SVD:

$$(\mathbf{I} - \mathbf{V}_1\mathbf{V}_1^T)\mathbf{V}_k(\mathbf{V}_1^T\mathbf{V}_k)^{-1} = \mathbf{U}_k\Sigma_k\mathbf{Z}_k^T.$$

Then,

$$\mathbf{T}_k = \mathbf{U}_k \arctan(\Sigma_k)\mathbf{Z}_k^T.$$

The map $\mathbf{T}_k$ defines a geodesic on the tangent manifold from $\mathbf{V}_1$ to $\mathbf{V}_k$.

Now, given a new parameter point $\hat{\mathbf{p}}$, the method interpolates the local bases in their mapped representations. That is, we interpolate $\{\mathbf{T}_k\}_{k=1}^K$ using the parameter interpolation points $\{\mathbf{p}_k\}_{k=1}^K$. For example,

$$\mathbf{T}(\hat{\mathbf{p}}) = \sum_{k=1}^{K} l_k(\hat{\mathbf{p}})\mathbf{T}_k,$$

where $l_k(\hat{\mathbf{p}})$ are the Lagrange basis functions. After $\mathbf{T}(\hat{\mathbf{p}}) \in \mathcal{T}_{\mathbf{V}_1}$ is computed, the exponential map

$$\mathbf{V} = \mathrm{Exp}_{\mathbf{V}_1}(\mathbf{T}) \quad \in \quad \mathcal{S}_{n,r}, \tag{45}$$

maps it back to the original manifold $\mathcal{S}_{n,r}$. The result is an interpolated basis $\mathbf{V}$ at the new parameter point ($\hat{\mathbf{p}}$) that preserves the desired properties (in this case, orthonormality). Numerically, the step mapping back to the original manifold is achieved by computing a short SVD,

$$\mathbf{T}(\hat{\mathbf{p}}) = \hat{\mathbf{U}}\,\hat{\Sigma}\,\hat{\mathbf{Z}}^T,$$

followed by

$$\mathbf{V}(\hat{\mathbf{p}}) = \mathbf{V}_1\hat{\mathbf{Z}}\cos(\hat{\Sigma}) + \hat{\mathbf{U}}\sin(\hat{\Sigma}).$$

Interpolating the local reduced model matrices

Interpolating the local basis matrices as outlined above has the disadvantage that when the new basis matrices are computed for a given $\hat{\mathbf{p}}$, the multiplications $\mathbf{W}^T\mathbf{E}(\hat{\mathbf{p}})\mathbf{V}$ and $\mathbf{W}^T\mathbf{A}(\hat{\mathbf{p}})\mathbf{V}$ need to be recomputed. In general, these multiplications depend on the original system dimension n and thus are expensive. Recent work has addressed this issue for affine parametric dependence (as in (5)) by precomputing those quantities that do not depend on the parameters [145]. An approach to overcome this problem in the general non-affine case is to interpolate reduced state-space quantities as opposed to the basis matrices themselves. This idea has been recently introduced in [130, 3, 50]. The methods proposed in [3, 130] first perform a *congruence transformation* of the local basis matrices in $\{\mathbf{V}_k\}_{k=1}^K$ (and similarly for $\{\mathbf{W}_k\}_{k=1}^K$), so that the reduced systems are expressed in the same generalized coordinate system. Then the reduced-order coefficient matrices constructed from these transformed projection matrices can be interpolated by using matrix manifold interpolation [3] and direct interpolation [130]. In [50], only one pair of basis matrices $\mathbf{V}_k = \overline{\mathbf{V}}$ and $\mathbf{W}_k = \overline{\mathbf{W}}$ is used for all $\mathbf{p}_k$ where $\overline{\mathbf{V}}$ and $\overline{\mathbf{W}}$ can be obtained, for example, by collecting all the local information into a pair of global basis matrices, and therefore there is no need to perform the congruence transformation on the bases before interpolating.

Suppose we have local reduced-order coefficient matrices $\mathbf{A}_r(\mathbf{p}_k) = \mathbf{W}_k^T\mathbf{A}(\mathbf{p}_k)\mathbf{V}_k$ corresponding to parameters $\{\mathbf{p}_k\}$ for $k = 1\ldots, K$. To perform the congruence transformation, we first select a reference reduced system. For simplicity, we will take $k = 1$ as the reference system. Next, we compute transformation matrices $\mathbf{Q}_k \in \mathbb{R}^{r\times r}$ and $\mathbf{P}_k \in \mathbb{R}^{r\times r}$, for all $k = 1, \ldots, K$, by solving

$$\mathbf{Q}_k = \arg\min_{\mathbf{Q}} \|\mathbf{V}_k\mathbf{Q} - \mathbf{V}_1\|_2 \quad \text{subject to} \quad \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_r \tag{46}$$

and

$$\mathbf{P}_k = \arg\min_{\mathbf{P}} \|\mathbf{W}_k\mathbf{P} - \mathbf{W}_1\|_2 \quad \text{subject to} \quad \mathbf{P}^T\mathbf{P} = \mathbf{I}_r. \tag{47}$$

The optimization problems (46) and (47) are solved using the SVD. Towards this goal, define the SVDs $\mathbf{V}_k^T\mathbf{V}_1 = \mathbf{U}_{\mathbf{v}_k}\Sigma_{\mathbf{v}_k}\mathbf{Z}_{\mathbf{v}_k}^T$ and $\mathbf{W}_k^T\mathbf{W}_1 = \mathbf{U}_{\mathbf{w}_k}\Sigma_{\mathbf{w}_k}\mathbf{Z}_{\mathbf{w}_k}^T$. Then the solutions to (46) and (47) are given by

$$\mathbf{Q}_k = \mathbf{U}_{\mathbf{v}_k}\mathbf{Z}_{\mathbf{v}_k}^T \quad \text{and} \quad \mathbf{P}_k = \mathbf{U}_{\mathbf{w}_k}\mathbf{Z}_{\mathbf{w}_k}^T, \quad \text{for} \quad k = 1, \ldots, K. \tag{48}$$

For $k = 1$, the reference system, $\mathbf{Q}_1 = \mathbf{P}_1 = \mathbf{I}_r$. For $k = 1, \ldots, K$, define the transformed local reduction matrices by $\widetilde{\mathbf{V}}_k := \mathbf{V}_k\mathbf{Q}_k$ and $\widetilde{\mathbf{W}}_k := \mathbf{W}_k\mathbf{P}_k$. Then the congruence-transformed local state-space matrices are given by

$$\widetilde{\mathbf{A}}_r(\mathbf{p}_k) := \mathbf{P}_k^T\mathbf{W}_k^T\mathbf{A}(\mathbf{p}_k)\mathbf{V}_k\mathbf{Q}_k = \mathbf{P}_k^T\mathbf{A}_r(\mathbf{p}_k)\mathbf{Q}_k, \quad \text{for} \quad k = 1, \ldots, K.$$

Note that the transformed basis matrices $\widetilde{\mathbf{V}}_k$ and $\widetilde{\mathbf{W}}_k$ are applied throughout the original system associated with $\mathbf{p}_k$ and therefore not only transform $\mathbf{A}_r(\mathbf{p}_k)$ to $\widetilde{\mathbf{A}}_r(\mathbf{p}_k)$, but also transform other coefficient matrices. More specifically, the k^{th} local reduced model

$$\mathbf{E}_r(\mathbf{p}_k)\dot{\mathbf{x}}_r = \mathbf{A}_r(\mathbf{p}_k)\mathbf{x}_r + \mathbf{B}_r(\mathbf{p}_k)\mathbf{u}, \qquad \mathbf{y}_r = \mathbf{C}_r(\mathbf{p}_k)\mathbf{x}_r, \tag{49}$$

is transformed to

$$\widetilde{\mathbf{E}}_r(\mathbf{p}_k)\dot{\widetilde{\mathbf{x}}}_r = \widetilde{\mathbf{A}}_r(\mathbf{p}_k)\widetilde{\mathbf{x}}_r + \widetilde{\mathbf{B}}_r(\mathbf{p}_k)\mathbf{u}, \qquad \widetilde{\mathbf{y}}_r = \widetilde{\mathbf{C}}_r(\mathbf{p}_k)\widetilde{\mathbf{x}}_r, \tag{50}$$

where $\widetilde{\mathbf{E}}_r(\mathbf{p}_k) = \mathbf{P}_k^T\mathbf{E}_r(\mathbf{p}_k)\mathbf{Q}_k$, $\widetilde{\mathbf{B}}_r(\mathbf{p}_k) = \mathbf{P}_k^T\mathbf{B}_r(\mathbf{p}_k)$, and $\widetilde{\mathbf{C}}_r(\mathbf{p}_k) = \widehat{\mathbf{C}}_r(\mathbf{p}_k)\mathbf{Q}_k$. In contrast to the local systems in (49), the transformed systems in (50) are expected to lie in the same generalized coordinate system.

Then, the question becomes how to construct the reduced order matrices $\mathbf{A}_r(\hat{\mathbf{p}})$, $\mathbf{E}_r(\hat{\mathbf{p}})$, $\mathbf{B}_r(\hat{\mathbf{p}})$ and $\mathbf{E}_r(\hat{\mathbf{p}})$ for a new parameter value $\hat{\mathbf{p}}$ using the local congruence-transformed matrices. $\mathbf{A}_r(\hat{\mathbf{p}})$ and $\mathbf{E}_r(\hat{\mathbf{p}})$ are obtained by applying (similar) manifold interpolation ideas explained in the previous section to the matrices $\{\widetilde{\mathbf{A}}_r(\mathbf{p}_k)\}_{k=1}^K$ and $\{\widetilde{\mathbf{E}}_r(\mathbf{p}_k)\}_{k=1}^K$, respectively, assuming nonsingular transformed matrices. On the other hand, when the transformed coefficient matrices are not square as in the case of $\widetilde{\mathbf{B}}_r(\mathbf{p}_k)$ and $\widetilde{\mathbf{C}}_r(\mathbf{p}_k)$, or are singular, the reduced-order matrices can be obtained for the new parameter $\hat{\mathbf{p}}$ by using direct interpolation from [3].

A similar approach based on matrix interpolation is also proposed in [130]. In this approach, the local reduced-order matrices are also first transformed and then interpolated; however, the transformation and the interpolation techniques are different from the ones used in [3]. Let $\mathbf{U}_r$ denote the first r dominant left singular vectors of $[\mathbf{V}_1, \ldots, \mathbf{V}_K]$. Define $\mathbf{M}_k := (\mathbf{W}_k^T\mathbf{U}_r)^{-1}$, and $\mathbf{T}_k := (\mathbf{V}_k^T\mathbf{U}_r)^{-1}$. Then, the reduced system (49) is transformed to

$$\bar{\mathbf{E}}_r(\mathbf{p}_k)\dot{\bar{\mathbf{x}}}_r = \bar{\mathbf{A}}_r(\mathbf{p}_k)\bar{\mathbf{x}}_r + \bar{\mathbf{B}}_r(\mathbf{p}_k)\mathbf{u}, \qquad \bar{\mathbf{y}}_r = \bar{\mathbf{C}}_r(\mathbf{p}_k)\bar{\mathbf{x}}_r, \tag{51}$$

where $\bar{\mathbf{E}}_r(\mathbf{p}_k) = \mathbf{M}_k\mathbf{E}_r(\mathbf{p}_k)\mathbf{T}_k$, $\bar{\mathbf{A}}_r(\mathbf{p}_k) = \mathbf{M}_k\mathbf{A}_r(\mathbf{p}_k)\mathbf{T}_k$, $\bar{\mathbf{B}}_r(\mathbf{p}_k) = \mathbf{M}_k\mathbf{B}_r(\mathbf{p}_k)$, and $\bar{\mathbf{C}}_r(\mathbf{p}_k) = \mathbf{C}_r(\mathbf{p}_k)\mathbf{T}_k$. After this transformation, for a given parameter $\hat{\mathbf{p}}$, the matrices $\{\bar{\mathbf{E}}_r(\mathbf{p}_k)\}_{k=1}^K$, $\{\bar{\mathbf{A}}_r(\mathbf{p}_k)\}_{k=1}^K$, $\{\bar{\mathbf{B}}_r(\mathbf{p}_k)\}_{k=1}^K$, and $\{\bar{\mathbf{C}}_r(\mathbf{p}_k)\}_{k=1}^K$ are directly interpolated using any appropriate interpolation method.

Interpolating the local transfer functions

Yet another option for interpolating local information is to interpolate the transfer functions of local reduced models. Given the sampled parameter points $\mathbf{p}_k, k = 1, \ldots, K$, the local reduced models have transfer functions

$$\mathbf{H}_r(s, \mathbf{p}_k) = \mathbf{C}_r(\mathbf{p}_k)(s\mathbf{E}_r(\mathbf{p}_k) - \mathbf{A}_r(\mathbf{p}_k))^{-1}\mathbf{B}_r(\mathbf{p}_k), \quad k = 1, \ldots, K, \tag{52}$$

where $\mathbf{H}_r(s, \mathbf{p}_k)$ denotes the transfer function of the reduced model constructed at $\mathbf{p} = \mathbf{p}_k$. We allow the order of each local reduced model to differ, and denote the order of reduced model $\mathbf{H}_r(s, \mathbf{p}_k)$ as r_k. These local reduced models can be constructed with any set of local bases.

The reduced-order transfer function at a new parameter point $\hat{\mathbf{p}}$ can then be obtained by interpolating the $\mathbf{H}_r(s, \mathbf{p}_k)$:

$$\widehat{\mathbf{H}}_r(s, \hat{\mathbf{p}}) = \sum_{k=1}^{K} L_k(\hat{\mathbf{p}})\mathbf{H}_r(s, \mathbf{p}_k), \tag{53}$$

where $L_k(\mathbf{p}_j) = \delta_{kj}$, $k, j = 1, \ldots, K$. This yields the "functional" interpolation condition

$$\widehat{\mathbf{H}}_r(s, \mathbf{p}_k) = \mathbf{H}_r(s, \mathbf{p}_k) \text{ for } k = 1, \ldots, K.$$

This strategy was studied in detail in [19, 21]. Generally speaking, any multivariate interpolation technique could be used to select the functions L_k. Considering first the case of a scalar parameter, $d = 1$, one choice is the Lagrange polynomials, which were used in [19]. A computationally more efficient and reliable form is obtained if we use the barycentric form of Lagrange interpolation [35]. Other choices are rational interpolation (which is used in [21] in barycentric form based on [34]), variants of Hermite interpolation, Sinc or spline interpolation, etc. These interpolation techniques can all be generalized to $d > 1$, using, e.g., tensorization or more sophisticated approaches.

We will discuss one particular case to highlight the advantages and disadvantages of interpolating the reduced-order transfer functions. Consider the case of a scalar parameter, where balanced truncation with a specified error tolerance τ_{BT} is used for generating all the local reduced models, as in [19]. Noting that the representation (53) separates the variables s and $\mathbf{p}$, we are now free to put the parametric dependence in any one of the matrices realizing the reduced model. A convenient choice results from the following representation:

$$\sum_{k=1}^{K} L_k(\mathbf{p}) \mathbf{H}_r(s, \mathbf{p}_k) = \widehat{\mathbf{C}}_r(\mathbf{p})(s\widehat{\mathbf{E}}_r - \widehat{\mathbf{A}}_r)^{-1}\widehat{\mathbf{B}}_r = \widehat{\mathbf{H}}_r(s, \mathbf{p}),$$

where

$$\begin{aligned}
\widehat{\mathbf{A}}_r &= \operatorname{diag}\big(\mathbf{A}_r(\mathbf{p}_1), \ldots, \mathbf{A}_r(\mathbf{p}_K)\big) \in \mathbb{R}^{r \times r}, \quad \widehat{\mathbf{E}}_r = \operatorname{diag}\big(\mathbf{E}_r(\mathbf{p}_1), \ldots, \mathbf{E}_r(\mathbf{p}_K)\big) \in \mathbb{R}^{r \times r}, \\
\widehat{\mathbf{B}}_r &= \begin{bmatrix} \mathbf{B}_r(\mathbf{p}_1) \\ \vdots \\ \mathbf{B}_r(\mathbf{p}_K) \end{bmatrix} \in \mathbb{R}^{r \times m}, \qquad \widehat{\mathbf{C}}_r(\mathbf{p}) = [L_1(\mathbf{p})\mathbf{C}_r(\mathbf{p}_1), \ldots, L_K(\mathbf{p})\mathbf{C}_r(\mathbf{p}_K)] \in \mathbb{R}^{q \times r},
\end{aligned}$$

with $r = r_1 + r_2 + \cdots + r_K$.

The method is flexible in the sense that it applies directly to parametric systems with non-affine parameter dependence since only evaluations of $L_k(\mathbf{p})$ are necessary. In addition, the structure of the approximate transfer function permits using local reduced models of different orders. This is important since balanced truncation with adaptive order will in general yield models with varying r_k. Another advantage is that the main features of balanced truncation are inherited. In particular, if the full model is uniformly stable in the parameter domain Ω, then this property is preserved in the reduced model: for a uniformly stable system, $(\mathbf{E}(\mathbf{p}), \mathbf{A}(\mathbf{p}))$ is an asymptotically stable matrix pencil for all $\mathbf{p} \in \Omega$, hence $(\mathbf{E}(\mathbf{p}_k), \mathbf{A}(\mathbf{p}_k))$ is asymptotically stable for all k if the parameter samples are taken from Ω. By stability preservation of balanced truncation, then also $(\mathbf{E}_r(\mathbf{p}_k), \mathbf{A}_r(\mathbf{p}_k))$ is asymptotically stable for all k and due to the block-diagonal structure, $(\widehat{\mathbf{E}}_r, \widehat{\mathbf{A}}_r)$ inherits this property. Also, instead of balanced truncation, other balancing-based strategies can be employed (e.g., LQG balancing, which is also applicable in the unstable case). See [24, 126] for overviews of possible choices and corresponding properties preserved in the reduced models.

Using balanced truncation for the local models induces an error bound which is obtained by splitting the interpolation error in parameter space from the balanced truncation error. Since the expressions for the error bound depend strongly on the chosen interpolation method, we only provide a version for (polynomial) Lagrange interpolation, but writing the bounds down for other interpolation techniques is straightforward. Assume $\mathbf{H}$ is uniformly stable in Ω, at least K times differentiable with respect to $\mathbf{p}$, and let Λ_{K-1} denote the Lebesgue constant of our node set $\{\mathbf{p}_1, \ldots, \mathbf{p}_K\}$. Then, we obtain

$$\left\| \mathbf{H}(s, \hat{\mathbf{p}}) - \widehat{\mathbf{H}}_r(s, \hat{\mathbf{p}}) \right\|_2 \leq \rho(\mathbf{H}, \mathbf{p}, s) + \tau_{\mathrm{BT}} \Lambda_{K-1}, \tag{54}$$

where $\rho(\mathbf{H}, \hat{\mathbf{p}}, s)$ is the interpolation error, which for polynomial interpolation is given by

$$\rho(\mathbf{H}, \hat{\mathbf{p}}, s) = \frac{1}{K!} \left\| \frac{\partial^K}{\partial \mathbf{p}^K} \mathbf{H}(s, \xi(\mathbf{p})) \right\|_2 \cdot \prod_{k=1}^{K} |\hat{\mathbf{p}} - \mathbf{p}_k|,$$

with $\xi(\mathbf{p}) \in (\min_k \mathbf{p}_k, \max_k \mathbf{p}_k)$ (to be understood componentwise). The presence of the Lebesgue constant in the error bound suggests using a node set that produces a small Λ_{K-1}; hence, a uniform distribution of the $\mathbf{p}_k$ should be avoided. A reasonable choice is the Chebyshev nodes, particularly in their second form. As usual, the interpolation error will in general not be determined explicitly, but can be estimated, using, e.g., the reduced model instead of $\mathbf{H}$.

A clear disadvantage of the transfer function interpolation approach is that the state-space obtained from the realization $\widehat{\mathbf{A}}_r, \widehat{\mathbf{B}}_r, \widehat{\mathbf{E}}_r, \widehat{\mathbf{C}}_r(\mathbf{p})$ may grow quickly with the number of parameter samples. However, we note that for *evaluating* the reduced model online, it is not necessary to explicitly form this realization since the only computation required is the evaluation of the reduced-order transfer functions at the given value s and the evaluation of the (scalar) parametric functions L_k. This amounts to interpolating the outputs computed by the reduced-order models using the given L_k's. A similar idea is used in [112], where interpolation is used to create a parametric POD model for a convection problem where the mathematical model is unknown.

Another disadvantage is that by interpolating the reduced-order transfer functions as in (53), the poles of $\widehat{\mathbf{H}}_r$ are fixed and do not vary with $\mathbf{p}$ as the poles of $\mathbf{H}$ most likely do. In addition, some poles of the original model may be duplicated in several of the local reduced models, leading to spurious poles of the reduced-order transfer function. These observations are reported in [130], where it is suggested to interpolate the realizations of the local reduced models instead, as already discussed in the previous subsection.

Finally, we comment on extension of the transfer function interpolation to the multivariate case. For regular grids, tensorization of scalar interpolation methods is fairly easy, resulting, e.g., in bi-/tri-linear/-cubic (Hermite, spline) interpolation. Due to the curse of dimensionality, the resulting reduced models quickly become complex even for small d. In [19], the use of sparse grid interpolation [43, 76, 161] is suggested. The coupling of balanced truncation with piecewise polynomial interpolation using sparse grid points is reported in [19]. Promising results are obtained for this approach for examples with $d < 10$, see [21]. Another possibility to extend the idea of interpolating local transfer functions would be to use irregular grid interpolation approaches such as radial basis functions. Results reported in [150] are also promising and certainly deserve further investigation.

4.3 Comparing different reduced model generation approaches

This section has discussed several approaches for parametric reduced model generation, either using a global basis as in Section 4.1, or interpolating local information as in Section 4.2. The specific approaches presented are:

- using a single global basis;
- interpolating among local bases;
- interpolating among local reduced-order state-space matrices; and
- interpolating among local reduced-order transfer functions.

The following discussion provides some insight into the class of problems for which the different approaches may be most appropriate.

Parameter Sampling

All the presented reduced model generation approaches require an effective sampling of the parameter space, whether samples are used to construct global or local bases. For high-dimensional parameter spaces, the global basis approach is particularly amenable to the greedy adaptive parameter sampling approach discussed in Section 3.4, as shown in [41] for building a global POD basis. Similarly a greedy sampling approach can be easily combined with balanced truncation or rational interpolation methods as well to construct a global basis. In theory, it would be possible to wrap a greedy sampling approach around the local interpolation methods, although in practice

formulation of the problem would be more complicated (and perhaps expensive to solve). Adaptive selection of parameter sample points in conjunction with these local interpolation methods has not been explored in the literature. The parameter selection strategies that are optimal in the composite system-theoretic error measures defined in (19) and (20) remains an open challenge except for the very special parameter dependence as in (41).

Reduced model construction

In terms of basis construction, any of the methods for computing the basis matrices can be used with any of the reduced model generation approaches. However, the global basis approach is a more compelling choice to use with model reduction by rational (tangential) interpolation since the final parametric reduced model then retains the interpolation property in both the frequency and parameter domains. The global basis and interpolation among local bases approaches both require construction of a new reduced model (via projection) at every parameter value to be solved; hence it is important that the original system has an affine parametrization or is approximated by an affine decomposition so that the repeated reduced model construction is not too costly. However, for the other two methods of interpolating the local reduced matrices and interpolating local transfer functions, reduced models are only constructed at the parameter sample points and not at other parameter values to be solved. Thus, these latter two methods may be more appropriate in the case of general parameter dependence where a DEIM or other affine approximation is inaccurate.

Reduced model size

The different methods vary in how the overall size of the reduced model behaves, as well in their flexibility to adapt order in different regions of the parameter space. Interpolating local transfer functions allows for local models with different orders, whereas interpolating the local bases and local reduced matrices both require the same order of local models. The flexibility of having different local orders is important in problems where the system dynamics are significantly more or less rich (i.e., significantly harder or easier to approximate) at different parameter values. In the global basis approach there is only a single order of the basis, but adaptivity can be achieved by uneven sampling of the parameter space (e.g., using the greedy adaptive sampling approach). An advantage of interpolating the local bases or local reduced matrices is that the order of the reduced model does not grow with the number of parameter samples (it is fixed at the local reduced order). This is in contrast to the transfer function interpolation approach, where the order of the parametric reduced model can grow quickly due to the block diagonal structure of construction. For the global basis approach, the order will also grow if the global basis is obtained by concatenating the local bases. However, if the concatenation is followed by a rank revealing QR or SVD, growth in the overall reduced model order can be managed appropriately.

Error estimates

An advantage of interpolating local reduced-order transfer functions is that it separates the interpolation error in the parameter space from model reduction error (in the frequency domain) and allows an error bound as shown in (54), assuming that local reduced models are obtained with an error tolerance. Thus, balanced truncation, which has an *a priori* error bound, is appropriate to employ in this approach. For other methods, *a posteriori* error estimates may be possible for certain classes of problems and could be similarly combined with the interpolation error estimate.

Dynamical system poles

For the global basis, interpolating local bases, and interpolating local reduced matrices approaches, the poles of the reduced models vary with the parameter $\mathbf{p}$, due to the parametric dependency in the reduced $\mathbf{E}_r(\mathbf{p})$ and $\mathbf{A}_r(\mathbf{p})$ matrices. This is desirable, since the poles of the original model also vary with the parameter $\mathbf{p}$. This is not the case for the reduced models obtained by interpolating

the local transfer functions, for which the poles are fixed at the poles of the local models, which can lead to spurious poles in the reduced model.

5 Discussion

The preceding sections have presented the different elements for projection-based model reduction of parameterized systems: computing the basis, sampling the parameter space, constructing the reduced model, and representing parametric dependence in the reduced model. Within each element we have highlighted a number of different methods that can be employed. While in general there is no definitive "recipe" on how to combine elements to achieve the most effective reduction strategy for a given problem, we highlight here some of the relative advantages and disadvantages of the various methods and how that relates to their effectiveness for various classes of problems. We also cite examples from the literature that demonstrate the effectiveness of different reduction strategies for different applications.

Commonalities among model reduction methods

We begin with the observation that there are significant commonalities among the different methods for computing the reduction bases $\mathbf{V}$ and $\mathbf{W}$. While the specific strategies used in rational interpolation methods, balanced truncation, POD and the Reduced Basis Method are at surface level quite different, their commonalities are perhaps stronger than the usual presentation of the methods suggests. Rational interpolation methods, POD and the Reduced Basis Method are all snapshot-based methods. In classical rational interpolation methods, the snapshots correspond to sampled solutions over the complex frequency domain, with extensions to sampling over the parameter domain for the parametric model reduction variants. In the Reduced Basis Method, snapshots correspond to sampled solutions over the parameter domain, while for POD, the snapshots typically correspond to sampled solutions over both time and parameter domains. Different communities seem to prefer different strategies for sampling the parameter domain, but as discussed in Section 3.4, there is a great deal of flexibility in combining sampling strategies with basis computation strategies. For example, the combination of POD with a greedy parameter sampling strategy as in [84, 41] results in an overall reduction approach that is essentially the same as the Reduced Basis Method. The reduced basis community largely focuses on formulation of the model reduction problem in the continuous (PDE) domain; however, the resulting numerical algorithms build a projection basis as the span of a set of discrete snapshots over the parameter domain, just as in the POD. As discussed in Section 3.3, duality between time and frequency domain formulations for linear systems also reveals the connections between POD and balanced truncation, and between POD and rational interpolation methods. Nonetheless, some important differences remain among the methods, most notably the error bounds associated with balanced truncation (for a fixed parameter sample) and the applicability of POD to general nonlinear systems.

Applicability of the basis computation methods

POD is the most generally applicable of the methods for computing the basis, since it relies only on snapshots of the underlying simulation code. As a result, the POD basis can be computed easily, even when the simulation is a black-box code (although note that in the black-box case, the projection step to determine the reduced model remains a challenge). The POD can also be applied to general nonlinear problems, since computation of the POD basis does not rely on a specific problem structure. POD can also be used with any of the sampling strategies discussed. POD has shown to be effective in many application domains including fluid dynamics, structural dynamics, thermal modeling, atmospheric modeling and many more. The early applications of POD include unsteady flows and turbulence modeling [143, 49, 72], and unsteady fluid-structure interaction [51]. Perhaps the earliest example of POD snapshots being computed over the parameter domain is [112], which considered Rayleigh-Bénard convection with varying Rayleigh number.

In contrast, balanced truncation and rational interpolation methods are based on system-theoretic quantities such as Gramians and transfer functions, and exploit the specific dynamical system structure. This paper has focused on systems that are linear in state. Balanced truncation has been extended to bilinear systems [28]. Rational interpolation methods have been extended to bilinear [133, 15, 27, 40, 25, 26, 67] and quadratic-in-state systems [77, 26]. Even though these methods have proven effective for a wide range of problem settings, they are most widely used in circuit theory, such as [62, 127, 14, 135, 29], e.g., to analyze and predict signal propagation and interference in electric circuits; in structural mechanics, such as [149, 73, 119, 36, 138], to study, e.g., vibration suppression in large structures or behavior of micro-electro-mechanical systems; and in (optimal) control and controller reduction, such as [128, 32, 87, 5, 152, 162], e.g., in LQR/LQG control design.

Capturing parametric dependence

Section 4.3 discussed the different attributes of the approaches for generating the parametric reduced model, including both global and local approaches. Clearly there is no one "best" approach; rather, the different approaches will work well for different problems depending on the model's underlying parametric structure. As already discussed, the literature shows several examples of successful applications of these approaches; however, it is lacking a rigorous analysis that connects the performance of the methods with the structure of the parametric dependence in the problem. Since most of the methods rely on interpolation in one form or another, it might be expected that the large body of mathematical literature on analysis of interpolation methods is a fruitful avenue of investigation. The manifold interpolation methods are particularly novel and interesting; again, more work is needed to understand for what class of problems they are best.

Preservation of properties

When choosing a model reduction method, it is also important to consider what mathematical properties are essential to preserve in the reduced model. Rational interpolation methods have the advantage that the transfer function of the parametric reduced model exactly interpolates the transfer function of the full model at sampled frequency and parameter points (as long as the required vectors are included in the global basis). Stability of the parametric reduced model remains an important open question in most cases. Other than for a few exceptions related to specific problem structure, the approach of interpolating the local transfer functions is the only way to guarantee that the reduced model is stable for the full parameter range (provided each local reduced model is stable). Even for linear problems, the other approaches do not guarantee in general that an interpolation of stable models yields a globally stable parametric reduced model. Exceptions include the case of negative definite pencils $(\mathbf{E}(\mathbf{p}), \mathbf{A}(\mathbf{p}))$, where one-sided projection, $\mathbf{V} = \mathbf{W}$, combined with either a global basis approach or with positive interpolation weights on local reduced-order matrices will also guarantee stability. However, the general stability preservation property by transfer function interpolation comes at a cost—it results from the fact that the approach does not let the reduced model poles vary as the parameters change. Preservation of other properties such as passivity and contractivity may be important in some applications, but have yet received little attention in parametric model reduction methods.

Error bounds and error estimates

Guarantees on the quality of the parametric reduced models remains an important question. The reduced basis community in particular has promoted a strong emphasis on the derivation of error estimates for parametric model reduction [131, 85, 137, 154, 153, 75]. This work has created new methods that "certify" the reduced model through error estimates that can be computed without recourse to the full model [131]. An important observation is that these error estimates are related to the structure of the system (e.g., properties of the underlying partial differential equations) being approximated but are not specific to a particular basis computation method. In other

words, the basis behind the model reduction step can come from rational interpolation, balanced truncation or POD. This can be seen by analyzing these error estimates in state-space form as recently presented in [85]. Recall that projection-based model reduction as in (43) corresponds to approximating $\mathbf{x}(t;\mathbf{p})$ by $\mathbf{V}\mathbf{x}_r(t;\mathbf{p})$. As in [85], we take $\mathbf{E} = \mathbf{I}$ and let $\mathbf{e}(0;\hat{\mathbf{p}})$ denote the error in the state $\mathbf{x}(t;\mathbf{p})$ at time $t = 0$ and at parameter $\mathbf{p} = \hat{\mathbf{p}}$. Then, $\mathbf{e}(0;\hat{\mathbf{p}}) = (\mathbf{I} - \mathbf{V}\mathbf{W}^T)\mathbf{x}(0;\hat{\mathbf{p}})$. Similarly define the residual in the state equation

$$\mathbf{R}(t;\hat{\mathbf{p}}) = \mathbf{A}(\hat{\mathbf{p}})\mathbf{V}\mathbf{x}_r(t;\hat{\mathbf{p}}) + \mathbf{B}(\hat{\mathbf{p}})\mathbf{u}(t) - \mathbf{V}\dot{\mathbf{x}}_r(t;\hat{\mathbf{p}}).$$

Let $\xi(\hat{\mathbf{p}}) = \sup_t \left\| e^{\mathbf{A}(\hat{\mathbf{p}})t} \right\|_2$. Then, the error in the state variable at time t and at parameter $\hat{\mathbf{p}}$ is bounded by

$$\|\mathbf{x}(t;\hat{\mathbf{p}}) - \mathbf{V}\mathbf{x}(t;\hat{\mathbf{p}})\|_2 \leq \xi(\hat{\mathbf{p}}) \left(\|\mathbf{e}(0;\hat{\mathbf{p}})\|_2 + \int_0^t \|\mathbf{R}(\tau;\hat{\mathbf{p}})\|_2 d\tau \right). \tag{55}$$

The output error at time t and the parameter $\hat{\mathbf{p}}$ can be bounded similarly using

$$\|\mathbf{y}(t;\hat{\mathbf{p}}) - \mathbf{y}_r(t;\hat{\mathbf{p}})\|_2 \leq \|\mathbf{C}(\hat{\mathbf{p}})\| \, \|\mathbf{x}(t;\hat{\mathbf{p}}) - \mathbf{V}\mathbf{x}(t;\hat{\mathbf{p}})\|_2. \tag{56}$$

Although the error estimates apply to a general basis, the key remaining question is the computability of the constant $\xi(\hat{\mathbf{p}})$. This computation must be done in an efficient way, without recourse to the full model, otherwise the error estimate is of limited practical value. Again, the reduced basis community has derived efficient offline/online decomposition approaches to address this issue [131].

Extensions to more general system structure

The extension of the parametric reduction methods to general nonlinear systems was discussed above. Another assumption on system structure in this paper is that the $\mathbf{E}(\mathbf{p})$ matrix is nonsingular for any parameter selection $\mathbf{p}$. However, in several important applications (e.g., incompressible flows, circuit design, etc.) one obtains a system of the form (1) where $\mathbf{E}(\mathbf{p})$ could be singular for some or all selections of $\mathbf{p}$. Such systems with a singular $\mathbf{E}(\mathbf{p})$ matrix are called systems of differential algebraic equations (DAEs). Projection-based model reduction of DAEs has been studied extensively; see, for example, [146, 30, 147, 135, 117, 33, 89, 80]. The theoretical discussions of this paper directly extend to this setting. While in many cases numerical methods can be implemented as effectively as for the ODE case, there exist scenarios in which reducing a system of DAEs might prove more costly due to the need for computing projectors on certain deflating subspaces; for details, we refer the reader to the above references.

We also assume that the full-order model $\mathbf{H}(s,\mathbf{p})$ is asymptotically stable for every $\mathbf{p} \in \Omega$; indeed the discussion as presented only requires that $\mathbf{H}(s,\mathbf{p}_i)$ be stable where $\mathbf{p}_i$, for $i = 1,\ldots,K$, are the sampled parameter points. However, even this is unnecessary since most of the model reduction methods presented have been already extended to reducing unstable systems. From a numerical implementation perspective, the rational interpolation methods do not require $\mathbf{H}(s,\mathbf{p}_i)$ to be stable. The only requirement is that the frequency interpolation points $\hat{s}$ should not be a pole of $\mathbf{H}(s,\mathbf{p}_i)$. In the asymptotically stable case, this is easily guaranteed by choosing $\hat{s} \in \mathbb{C}_+$. In the unstable case, however, one needs to be more careful since some of the poles lie in $\mathbb{C}_+$ and they need to be avoided. Even the optimal rational interpolation-based model reduction methods have been extended to unstable systems; see, e.g., [114, 99]. Balanced truncation has been also generalized to reducing unstable systems, by either appropriately re-defining the system Gramians [163, 16, 164] or by using different balancing techniques, such as LQG balancing [128]. For the POD, the frequency domain formulation will be the appropriate choice for unstable systems since the time-domain snapshots will grow exponentially in this case.

Non-projection based model reduction

In this paper, we have focused on projection-based model reduction techniques; i.e., we have assumed the availability of a state-space description of the original model (1) and an explicit

projection was applied to state-space dimension. However, in settings where equations describing the evolution of a system are not explicitly specified (i.e., $\mathbf{E}(\mathbf{p}), \mathbf{A}(\mathbf{p})$ etc. are not available) and the only access to dynamics is via input/output measurements. This might be in the form of system responses such as measurements of $\mathbf{H}(s, \mathbf{p})$, or simulation outputs of a black-box code. In these cases, a reduced parametric model as in (2) is obtained directly from measurements or from simulation outputs without access to internal dynamics. One commonly used approach is to assume a specified functional form (e.g., a polynomial response surface) of the surrogate model, which is fit using output data. Other ways to specify the form of the approximation in this setting include radial basis functions and Kriging. For the case of dynamical systems with no parametric dependence, the Loewner-based data-driven framework [116, 102, 103] has been applied with great success to construct reduced models using only transfer function samples. Recently, this approach has been used to produce even locally optimal reduced models, see [23]. The data-driven Loewner framework has been recently extended to the parameter dependent systems as well [10, 93, 92] where the reduced parametric model is a rational function not only in the frequency s but also in every parameter p_k, $k = 1, \ldots, d$. This allows choosing the order of approximation in s and $\mathbf{p}$ independently. Even though the theoretical discussion extends to the multiple parameter case directly, the numerical computation and construction of the reduced model might present a challenge and the order of the reduced model might grow undesirably as d, the number of parameters, grows.

6 Outlook

We close this review paper with a brief discussion of promising current research directions and some open challenges. As already mentioned, recent development of the EIM and DEIM methods has led to significant progress in model reduction for nonlinear systems using POD and the Reduced Basis Method. It remains an open question whether the rational interpolation approaches and balanced truncation can be extended beyond bilinear- and quadratic-in-state systems and to handle general nonlinear dependence in the state variable $\mathbf{x}(t)$. A balancing method for nonlinear systems was proposed in [140], but a scalable algorithmic implementation remains an open challenge. Even for the linear-in-state systems we consider in this paper, despite the existence of effective sampling strategies discussed above, the optimal parameter selection strategies for the composite system-theoretic error measures defined in (19) and (20) remains an open challenge for the general parametric dependence. Further open challenges include parametric model reduction for systems with time-dependent and/or stochastic parameters. Extending parametric model reduction methods to these broader classes of systems will require new approaches for defining the reduced subspaces, as well as new methods for constructing and solving the reduced model. Handling high-dimensional parameter spaces remains another challenging problem. While some progress has been made in this area, in particular with greedy sampling approaches, further work is needed to develop methods that exploit system structure to avoid the curse of dimensionality. Promising recent efforts towards this goal use tensor techniques [125, 129]. The combination of tensor calculus [86] and parametric model reduction techniques for time-dependent problems is still in its infancy, but offers a promising research direction.

References

[1] N. Alexandrov, J.E. Dennis, Jr., R.M. Lewis, and V. Torczon, *A trust region framework for managing the use of approximation models in optimization*, Structural Optimization, 15 (1998), pp. 16–23.

[2] D. Amsallem and C. Farhat, *Interpolation method for the adaptation of reduced-order models to parameter changes and its application to aeroelasticity*, AIAA Journal, 46 (2008), pp. 1803–1813.

[3] D. Amsallem and C. Farhat, *An online method for interpolating linear parametric reduced-order models*, SIAM Journal on Scientific Computing, 33 (2011), pp. 2169–2198.

[4] D. Amsallem, M. J. Zahr, and C. Farhat, *Nonlinear model order reduction based on local reduced-order bases*, International Journal for Numerical Methods in Engineering, 92 (2012), pp. 891–916.

[5] B.D.O. Anderson and Y. Liu, *Controller reduction: concepts and approaches*, IEEE Transactions on Automatic Control, 34 (2002), pp. 802–812.

[6] H. Antil, M. Heinkenschloss, and R. H. W. Hoppe, *Domain decomposition and balanced truncation model reduction for shape optimization of the Stokes system*, Optimization Methods and Software, 26 (2011), pp. 643–669.

[7] H. Antil, M. Heinkenschloss, R. H. W. Hoppe, C. Linsenmann, and A. Wixforth, *Reduced order modeling based shape optimization of surface acoustic wave driven microfluidic biochips*, Mathematics and Computers in Simulation, 82 (2012), pp. 1986–2003.

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

[9] A.C. Antoulas, C.A. Beattie, and S. Gugercin, *Interpolatory model reduction of large-scale dynamical systems*, in Efficient Modeling and Control of Large-Scale Systems, J. Mohammadpour and K. Grigoriadis, eds., Springer-Verlag, 2010, pp. 2–58.

[10] A.C. Antoulas, A.C. Ionita, and S. Lefteriu, *On two-variable rational interpolation*, Linear Algebra and Its Applications, 436 (2012), pp. 2889–2915.

[11] E. Arian, M. Fahl, and E. Sachs, *Trust-region proper orthogonal decomposition models by optimization methods*, in Proceedings of the 41st IEEE Conference on Decision and Control, Las Vegas, NV, 2002, pp. 3300–3305.

[12] P. Astrid, S. Weiland, K. Willcox, and T. Backx, *Missing point estimation in models described by proper orthogonal decomposition*, IEEE Transactions on Automatic Control, (2008), pp. 2237–2251.

[13] Z. Bai, *Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems*, Applied Numerical Mathematics, 43 (2002), pp. 9–44.

[14] Z. Bai, P. Feldmann, and R.W Freund, *How to make theoretically passive reduced-order models passive in practice*, in Custom Integrated Circuits Conference, Proceedings of the IEEE, 1998, pp. 207–210.

[15] Z. Bai and D. Skoogh, *A projection method for model reduction of bilinear dynamical systems*, Linear algebra and its applications, 415 (2006), pp. 406–425.

[16] M. Barahona, A.C. Doherty, M. Sznaier, H. Mabuchi, and J.C. Doyle, *Finite horizon model reduction and the appearance of dissipation in Hamiltonian systems*, in 41st IEEE Conference on Decision and Control, vol. 4, 2002, pp. 4563–4568.

[17] M. Barrault, Y. Maday, N.C. Nguyen, and A.T. Patera, *An "empirical interpolation" method: Application to efficient reduced-basis discretization of partial differential equations*, Comptes Rendus Mathématique. Académie des Sciences. Paris, I (2004), pp. 339–667.

[18] O. Bashir, K. Willcox, O. Ghattas, B. van Bloemen Waanders, and J. Hill, *Hessian-based model reduction for large-scale systems with initial condition inputs*, International Journal for Numerical Methods in Engineering, 73 (2008), pp. 844–868.

[19] U. Baur and P. Benner, *Model reduction for parametric systems using balanced truncation and interpolation*, at–Automatisierungstechnik, 57 (2009), pp. 411–420.

[20] U. Baur, P. Benner, C.A. Beattie, and S. Gugercin, *Interpolatory projection methods for parameterized model reduction*, SIAM Journal on Scientific Computing, 33 (2011), pp. 2489–2518.

[21] U. Baur, P. Benner, A. Greiner, J.G. Korvink, J.Lienemann, and C. Moosmann, *Parameter preserving model order reduction for MEMS applications*, Mathematical and Computer Modelling of Dynamical Systems, 17 (2011), pp. 297–317.

[22] C.A. Beattie, S. Chaturantabut, E. de Sturler, S. Gugercin, and M. Kilmer, *Interpolatory model reduction strategies for nonlinear parametric inversion*, in SIAM Conference on Applied Linear Algebra, Valencia, Spain, June 2012.

[23] C. Beattie and S. Gugercin, *Realization-independent $\mathcal{H}_2$-approximation*, in 51st IEEE Conference on Decision and Control, 2012, pp. 4953–4958.

[24] P. Benner, *System-theoretic methods for model reduction of large-scale systems: Simulation, control, and inverse problems*, in Proceedings of MathMod, Vienna, I. Troch and F. Breitenecker, eds., vol. 35 of ARGESIM Report, February 11–13, 2009, pp. 126–145.

[25] P. Benner and T. Breiten, *Interpolation-based $\mathcal{H}_2$-model reduction of bilinear control systems*, SIAM Journal on Matrix Analysis and Applications, 33 (2012), pp. 859–885.

[26] ——, *Two-sided moment matching methods for nonlinear model reduction*, Max Planck Institute Magdeburg Preprint MPIMD/12-12, (June, 2012).

[27] P. Benner, T. Breiten, and T. Damm, *Generalised tangential interpolation for model reduction of discrete-time MIMO bilinear systems*, International Journal of Control, 84 (2011), pp. 1398–1407.

[28] P. Benner and T. Damm, *Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems*, SIAM Journal on Control and Optimization, 49 (2011), pp. 686–711.

[29] P. Benner, M. Hinze, and E.J.W. ter Maten, eds., *Model Reduction for Circuit Simulation*, vol. 74 of Lecture Notes in Electrical Engineering, Springer-Verlag, Dordrecht, NL, 2011.

[30] P. Benner, M.-S. Hossain, and T. Stykel, *Model reduction of periodic descriptor systems using balanced truncation*, in Benner et al. [29], pp. 193–206.

[31] P. Benner, J.R. Li, and T. Penzl, *Numerical solution of large-scale Lyapunov equations, riccati equations, and linear-quadratic optimal control problems*, Numerical Linear Algebra with Applications, 15 (2008), pp. 755–777.

[32] P. Benner and J. Saak, *A semi-discretized heat transfer model for optimal cooling of steel profiles*, in Dimension Reduction of Large-Scale Systems, P. Benner, V. Mehrmann, and D. Sorensen, eds., Lecture Notes in Computational and Applied Mathematics, Springer, 2005, pp. 353–356.

[33] P. Benner and V.I. Sokolov, *Partial realization of descriptor systems*, Systems & Control Letters, 55 (2006), pp. 929–938.

[34] J.-P. Berrut, R. Baltensperger, and H.D. Mittelmann, *Recent developments in barycentric rational interpolation*, in Trends and applications in constructive approximation, vol. 151 of International Series of Numerical Mathematics, Birkhäuser, Basel, 2005, pp. 27–51.

[35] J.-P. Berrut and L.N. Trefethen, *Barycentric Lagrange interpolation*, SIAM Review, 46 (2004), pp. 501–517.

[36] D.S. Bindel, Z. Bai, and J.W. Demmel, *Model reduction for RF MEMS simulation*, in Applied Parallel Computing. State of the Art in Scientific Computing, Springer, 2006, pp. 286–295.

[37] B.N. Bond and L. Daniel, *A piecewise-linear moment-matching approach to parameterized model-order reduction for highly nonlinear systems*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 26 (2007), pp. 2116–2129.

[38] B. N. Bond and L. Daniel, *Parameterized model order reduction of nonlinear dynamical systems*, in IEEE/ACM International Conference on Computer-Aided Design, 2005, pp. 487–494.

[39] T. Bonin, H. Fessbender, A. Soppa, and M. Zaeh, *A global Arnoldi method for the model reduction of second-order structural dynamical systems*. Preprint, TU Braunschweig, Institut Computational Mathematics, 2010.

[40] T. Breiten and T. Damm, *Krylov subspace methods for model order reduction of bilinear control systems*, Systems & Control Letters, 59 (2010), pp. 443–450.

[41] T. Bui-Thanh, K. Willcox, and O. Ghattas, *Model reduction for large-scale systems with high-dimensional parametric input space*, SIAM Journal on Scientific Computing, 30 (2008), pp. 3270–3288.

[42] ——, *Parametric reduced-order models for probabilistic analysis of unsteady aerodynamic applications*, AIAA Journal, 46 (2008), pp. 2520–2529.

[43] H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numerica, 13 (2004), pp. 147–269.

[44] A. Bunse-Gerstner, D. Kubalinska, G. Vossen, and D. Wilczek, *H_2-norm optimal model reduction for large scale discrete dynamical MIMO systems*, Journal of Computational and Applied Mathematics, 233 (2010), pp. 1202–1216.

[45] K. Carlberg, C. Farhat, J. Cortial, and D. Amsallem, *The GNAT method for nonlinear model reduction: Effective implementation and application to computational fluid dynamics and turbulent flows*, Journal of Computational Physics, 242 (2013), pp. 623–647.

[46] S. Chaturantabut and D. Sorensen, *Nonlinear model reduction via discrete empirical interpolation*, SIAM Journal on Scientific Computing, 32 (2010), pp. 2737–2764.

[47] C. Chia-Chi, L. Ming-Hong, and F. Wu-Shiung, *MIMO interconnects order reductions by using the multiple point adaptive-order rational global Arnoldi algorithm*, IEICE transactions on electronics, 89 (2006), pp. 792–802.

[48] L. Daniel, O.C. Siong, S.C. Low, K.H. Lee, and J. White, *A multiparameter moment matching model reduction approach for generating geometrically parameterized interconnect performance models*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 23 (2004), pp. 678–693.

[49] A.E. Deane, I.G. Kevrekidis, G. Em Karniadakis, and S.A. Orszag, *Low-dimensional models for complex geometry flows: Application to grooved channels and circular cylinders*, Physics of Fluids A: Fluid Dynamics, 3 (1991), p. 2337.

[50] J. Degroote, J. Vierendeels, and K. Willcox, *Interpolation among reduced-order matrices to obtain parameterized models for design, optimization and probabilistic analysis*, International Journal for Numerical Methods in Fluids, 63 (2010), pp. 207–230.

[51] E.H. Dowell and K.C. Hall, *Modeling of fluid-structure interaction*, Annual Review of Fluid Mechanics, 33 (2001), pp. 445–490.

[52] M. Drohmann, B. Haasdonk, and M. Ohlberger, *Adaptive reduced basis methods for nonlinear convection-diffusion equations*, in Finite Volumes for Complex Applications VI Problems & Perspectives, J. Fot, J. Fürst, J. Halama, R. Herbin, and F. Hubert, eds., vol. 4 of Springer Proc. in Mathematics, Springer, 2011, pp. 369–377.

[53] ——, *Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation*, SIAM Journal on Scientific Computing, 34 (2012), pp. 937–969.

[54] V. Druskin, L. Knizhnerman, and V. Simoncini, *Analysis of the rational Krylov subspace and ADI methods for solving the Lyapunov equation*, SIAM Journal on Numerical Analysis, 49 (2011), pp. 1875–1898.

[55] V. Druskin, C. Lieberman, and M. Zaslavsky, *On adaptive choice of shifts in rational Krylov subspace reduction of evolutionary problems*, SIAM Journal on Scientific Computing, 32 (2010), pp. 2485–2496.

[56] V. Druskin, V. Simoncini, and M. Zaslavsky, *Solution of the time-domain inverse resistivity problem in the model reduction framework, Part I. one-dimensional problem with SISO data*, SIAM Journal on Scientific Computing, (2011). submitted; see http://www.dm.unibo.it/ simoncin/list.html.

[57] J. L. Eftang and B. Stamm, *Parameter multi-domain hp empirical interpolation*, International Journal for Numerical Methods in Engineering, 90 (2012), pp. 412–428.

[58] R. Eid, B. Salimbahrami, B. Lohmann, E.B. Rudnyi, and J.G. Korvink, *Parametric order reduction of proportionally damped second-order systems*, in Sensors and Materials, Tokyo, vol. 19, 2007, pp. 149–164.

[59] H. Elman and Q. Liao, *Reduced basis collocation methods for partial differential equations with random coefficients*, SIAM/ASA Journal on Uncertainty Quantification, (2013). To appear.

[60] R. Everson and L. Sirovich, *The Karhunen-Loeve Procedure for Gappy Data*, Journal of the Optical Society of America, 12 (1995), pp. 1657–1664.

[61] O. Farle, V. Hill, P. Ingelström, and R. Dyczij-Edlinger, *Multi-parameter polynomial order reduction of linear finite element models*, Mathematical and Computer Modelling of Dynamical Systems, 14 (2008), pp. 421–434.

[62] P. Feldmann and R.W. Freund, *Efficient linear circuit analysis by Padé approximation via the Lanczos process*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 14 (1995), pp. 639–649.

[63] L. Feng, *Parameter independent model order reduction*, Mathematics and Computers in Simulation, 68 (2005), pp. 221–234.

[64] L. Feng and P. Benner, *A robust algorithm for parametric model order reduction based on implicit moment matching*, Proceedings in Applied Mathematics and Mechanics, 7 (2008), pp. 10215.01–10215.02.

[65] L. Feng, D. Koziol, E.B. Rudnyi, and J.G. Korvink, *Parametric Model Reduction for Fast Simulation of Cyclic Voltammograms*, Sensor Letters, 4 (2006), pp. 165–173.

[66] L. Feng, E.B. Rudnyi, and J.G. Korvink, *Preserving the film coefficient as a parameter in the compact thermal model for fast electrothermal simulation*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 24 (2005), pp. 1838–1847.

[67] G.M. Flagg, *Interpolation Methods for Model Reduction of Bilinear Systems*, PhD thesis, Virginia Tech, Department of Mathematics, May 2012.

[68] R. Freund, *Model reduction methods based on Krylov subspaces*, Acta Numerica, 12 (2003), pp. 267–319.

[69] D. Galbally, K. Fidkowski, K. Willcox, and O. Ghattas, *Nonlinear model reduction for uncertainty quantification in large-scale inverse problems*, International Journal for Numerical Methods in Engineering, 81 (2010), pp. 1581–1608.

[70] K. Gallivan, E.J. Grimme, and P. Van Dooren, *Asymptotic waveform evaluation via a Lanczos method*, Applied Mathematics Letters, 7 (1994), pp. 75–80.

[71] K. Gallivan, A. Vandendorpe, and P. Van Dooren, *Model reduction of MIMO systems via tangential interpolation*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 328–349.

[72] T.B. Gatski and M.N. Glauser, *Proper orthogonal decomposition based turbulence modeling*, in Instability, Transition, and Turbulence, Springer, 1992, pp. 498–510.

[73] W. Gawronski and J. Juang, *Model reduction for flexible structures*, Control and dynamic systems, 36 (1990), pp. 143–222.

[74] M. Grepl, *Reduced-Basis Approximation and A Posteriori Error Estimation for Parabolic Partial Differential Equations*, PhD thesis, MIT, Cambridge, MA, June 2005.

[75] M. Grepl and A. Patera, A posteriori *error bounds for reduced-basis approximations of parametrized parabolic partial differential equations*, ESAIM-Mathematical Modelling and Numerical Analysis (M2AN), 39 (2005), pp. 157–181.

[76] M. Griebel, *Sparse grids and related approximation schemes for higher dimensional problems*, in Foundations of Computational Mathematics (FoCM05), Santander, L. Pardo, A. Pinkus, E. Suli, and M.J. Todd, eds., Cambridge University Press, 2006, pp. 106–161.

[77] C. Gu, *QLMOR: A projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 30 (2011), pp. 1307–1320.

[78] S. Gugercin, A.C. Antoulas, and C.A. Beattie, $\mathcal{H}_2$ *model reduction for large-scale linear dynamical systems*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 609–638.

[79] S. Gugercin, D.C. Sorensen, and A.C. Antoulas, *A modified low-rank Smith method for large-scale Lyapunov equations*, Numerical Algorithms, 32 (2003), pp. 27–55.

[80] S. Gugercin, T. Stykel, and S. Wyatt, *Model reduction of descriptor systems by interpolatory projection methods*, SIAM Journal on Scientific Computing, (2013, in press.).

[81] P.K. Gunupudi, R. Khazaka, and M.S. Nakhla, *Analysis of transmission line circuits using multidimensional model reduction techniques*, IEEE Transactions on Advanced Packaging, 25 (2002), pp. 174–180.

[82] 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 Transactions on Microwave Theory and Techniques, 51 (2003), pp. 2347–2354.

[83] B. Haasdonk, M. Dihlmann, and M. Ohlberger, *A training set and multiple bases generation approach for parameterized model reduction based on adaptive grids in parameter space*, Mathematical and Computer Modelling of Dynamical Systems, 17 (2011), pp. 423–442.

[84] B. Haasdonk and M. Ohlberger, *Reduced basis method for finite volume approximations of parametrized linear evolution equations*, ESAIM: Mathematical Modelling and Numerical Analysis, 42 (2008), pp. 277–302.

[85] ——, *Efficient reduced models and a posteriori error estimation for parametrized dynamical systems by offline/online decomposition*, Mathematical and Computer Modelling of Dynamical Systems, 17 (2011), pp. 145–161.

[86] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, Springer, Berlin, 2012.

[87] Y. Halevi, D.S. Bernstein, and W.M. Haddad, *On stable full-order and reduced-order LQG controllers*, Optimal Control Applications and Methods, 12 (1991), pp. 163–172.

[88] A. Har, J.T. Borggaard, and D. Pelletier, *Local improvements to reduced-order models using sensitivity analysis of the proper orthogonal decomposition*, Journal of Fluid Mechanics, 629 (2009), pp. 41–72.

[89] M. Heinkenschloss, D.C. Sorensen, and K. Sun, *Balanced truncation model reduction for a class of descriptor systems with application to the Oseen equations*, SIAM Journal on Scientific Computing, 30 (2008), pp. 1038–1063.

[90] H. Hotelling, *Analysis of a complex of statistical variables with principal components*, Journal of Educational Psychology, 24 (1933), pp. 417–441,498–520.

[91] D.Y. Hu and L. Reichel, *Krylov-subspace methods for the Sylvester equation*, Linear Algebra and its Applications, 172 (1992), pp. 283–313.

[92] A.C. Ionita and A.C Antoulas, *Data-driven parametrized model reduction in the Loewner framework*. submitted, 2013.

[93] A.C. Ionita and A.C. Antoulas, *Model order reduction of parametrized systems*, in 17$^{\text{th}}$ International Linear Algebra Society Conference, TU Braunschweig, Germany, August 2011.

[94] I.M. Jaimoukha and E.M. Kasenally, *Krylov subspace methods for solving large Lyapunov equations*, SIAM Journal on Numerical Analysis, (1994), pp. 227–251.

[95] K. Jbilou, *Low rank approximate solutions to large Sylvester matrix equations*, Applied Mathematics and Computation, 177 (2006), pp. 365–376.

[96] J.Garcke and M. Griebel, eds., *Sparse Grids and Applications*, vol. 88 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, Heidelberg/Berlin, 2013.

[97] T. Kim, *Frequency-domain Karhunen-Loeve method and its application to linear dynamic systems*, AIAA Journal, 36 (1998), pp. 2117–2123.

[98] D.D. Kosambi, *Statistics in function space*, Journal of Indian Mathematical Society, 7 (1943), pp. 76–88.

[99] D. Kubalińska, *Optimal interpolation-based model reduction*, PhD thesis, Bremen, University Dissertation, 2008.

[100] K. Kunisch and S. Volkwein, *Proper orthogonal decomposition for optimality systems*, ESAIM: Mathematical Modelling and Numerical Analysis, 42 (2008), pp. 1–23.

[101] S. Lall, J.E. Marsden, and S. Glavaski, *A subspace approach to balanced truncation for model reduction of nonlinear control systems*, International Journal on Robust and Nonlinear Control, 12 (2002), pp. 519–535.

[102] S. Lefteriu and A.C. Antoulas, *A new approach to modeling multiport systems from frequency-domain data*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 29 (2010), pp. 14–27.

[103] S. Lefteriu and A. C. Antoulas, *A new adaptive approach to modeling measured multiport scattering parameters*, in Scientific Computing in Electrical Engineering, SCEE 2008, J. Roos and L.R.J Costa, eds., vol. 14 of Mathematics in Industry, Springer-Verlag, 2010, pp. 21–28.

[104] A.T.-M. Leung and R. Khazaka, *Parametric model order reduction technique for design optimization*, in IEEE International Symposium on Circuits and Systems (ISCAS), May 2005, pp. 1290–1293.

[105] J.R. Li and J. White, *Low-rank solution of Lyapunov equations*, SIAM Review, (2004), pp. 693–713.

[106] Y.T. Li, Z. Bai, and Y. Su, *A two-directional Arnoldi process and its application to parametric model order reduction*, Journal of Computational and Applied Mathematics, 226 (2009), pp. 10–21.

[107] C. Lieberman, K. Willcox, and O. Ghattas, *Parameter and state model reduction for large-scale statistical inverse problems*, SIAM Journal on Scientific Computing, 32 (2010), pp. 2523–2542.

[108] T. Lieu and C. Farhat, *Adaptation of aeroelastic reduced-order models and application to an F-16 configuration*, AIAA Journal, 45 (2007), pp. 1244–1257.

[109] T. Lieu, C. Farhat, and M. Lesoinne, *Reduced-order fluid/structure modeling of a complete aircraft configuration*, Computer Methods in Applied Mechanics and Engineering, 195 (2006), pp. 5730–5742.

[110] M. Loéve, *Probability Theory*, D. Van Nostrand Company Inc., New York, 1955.

[111] J.L. Lumley, *The Structures of Inhomogeneous Turbulent Flow*, Atmospheric Turbulence and Radio Wave Propagation, (1967), pp. 166–178.

[112] H.V. Ly and H.T. Tran, *Modeling and control of physical processes using proper orthogonal decomposition*, Journal of Mathematical and Computer Modeling, 33 (2001), pp. 223–236.

[113] M. Ma, A.T.-M. Leung, and R. Khazaka, *Sparse macromodels for parametric networks*, in IEEE International Symposium on Circuits and Systems (ISCAS), May 2006, pp. 2545–2548.

[114] C. Magruder, C.A. Beattie, and S. Gugercin, *Rational Krylov methods for optimal $\mathcal{L}_2$ model reduction*, in 49th IEEE Conference on Decision and Control, 2010, pp. 6797–6802.

[115] L. Mathelin and O. Le Maitre, *Robust control of uncertain cylinder wake flows based on robust reduced order models*, Computers and Fluids, 38 (2009), pp. 1168–1182.

[116] A.J. Mayo and A.C. Antoulas, *A framework for the solution of the generalized realization problem*, Linear Algebra and its Applications, 425 (2007), pp. 634–662.

[117] V. Mehrmann and T. Stykel, *Balanced truncation model reduction for large-scale systems in descriptor form*, in Dimension Reduction of Large-Scale Systems, P. Benner, V. Mehrmann, and D. Sorensen, eds., Lecture Notes in Computational and Applied Mathematics, 2005, pp. 83–115.

[118] L. Meier III and D. Luenberger, *Approximation of linear constant systems*, IEEE Transactions on Automatic Control, 12 (1967), pp. 585–588.

[119] D.G. Meyer and S. Srinivasan, *Balancing and model reduction for second-order form linear systems*, IEEE Transactions on Automatic Control, 41 (1996), pp. 1632–1644.

[120] B. Moore, *Principal component analysis in linear systems: Controllability, observability, and model reduction*, IEEE Transactions on Automatic Control, 26 (1981), pp. 17–32.

[121] K. Moosmann and J.G. Korvink, *Automatic parametric MOR for MEMS design*, in Tagungsband GMA-FA 1.30 "Modellbildung, Identifikation und Simulation in der Automatisierungstechnik", Workshop am Bostalsee, 27.–29.9.2006, B. Lohmann and A. Kugi, eds., 2006, pp. 89–99.

[122] C. Mullis and R. Roberts, *Synthesis of minimum roundoff noise fixed point digital filters*, IEEE Transactions on Circuits and Systems, 23 (1976), pp. 551–562.

[123] N.C. Nguyen, A.T. Patera, and J. Peraire, *A 'best points' interpolation method for efficient approximation of parmetrized functions*, International Journal for Numerical Methods in Engineering, 73 (2008), pp. 521–543.

[124] G.R. North, T.L. Bell, R.F. Cahalan, and F.J. Moeng, *Sampling errors in the estimation of empirical orthogonal functions*, Monthly Weather Review, 110 (1982), pp. 699–706.

[125] A. Nouy, *A priori model reduction through proper generalized decomposition for solving time-dependent partial differential equations*, Computer Methods in Applied Mechanics and Engineering, 199 (2010), pp. 1603–1626.

[126] G. Obinata and B.D.O. Anderson, *Model Reduction for Control System Design*, Communications and Control Engineering Series, Springer-Verlag, London, UK, 2001.

[127] A. Odabasioglu, M. Celik, and L.T Pileggi, *PRIMA: passive reduced-order interconnect macromodeling algorithm*, in Proceedings of the 1997 IEEE/ACM international conference on Computer-aided design, 1997, pp. 58–65.

[128] P. Opdenacker and E.A. Jonckheere, *LQG balancing and reduced LQG compensation of symmetric passive systems*, International Journal of Control, 41 (1985), pp. 73–109.

[129] I.V. Oseledets, *Tensor-train decomposition*, SIAM Journal on Scientific Computing, 33 (2011), pp. 2295–2317.

[130] H. Panzer, J. Mohring, R. Eid, and B. Lohmann, *Parametric model order reduction by matrix interpolation*, at–Automatisierungstechnik, 58 (2010), pp. 475–484.

[131] A.T. Patera and G. Rozza, *Reduced basis approximation and* a posteriori *error estimation for parametrized partial differential equations.* Version 1.0, Copyright MIT 2006.

[132] T. Penzl, *A cyclic low rank Smith method for large sparse Lyapunov equations*, SIAM Journal on Scientific Computing, 21 (2000), pp. 1401–1418.

[133] J.R. Phillips, *Projection-based approaches for model reduction of weakly nonlinear, time-varying systems*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 22 (2003), pp. 171–187.

[134] C. Prud'homme, D. Rovas, K. Veroy, Y. Maday, A.T. Patera, and G. Turinici, *Reliable real-time solution of parameterized partial differential equations: Reduced-basis output bound methods*, Journal of Fluids Engineering, 124 (2002), pp. 70–80.

[135] T. Reis and T. Stykel, *Positive real and bounded real balancing for model reduction of descriptor systems*, International Journal of Control, 83 (2010), pp. 74–88.

[136] C.W. Rowley, *Model reduction for fluids, using balanced proper orthogonal decomposition*, International Journal of Bifurcation and Chaos, 15 (2005), pp. 997–1013.

[137] G. Rozza, D.B.P. Huynh, and A.T. Patera, *Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: application to transport and continuum mechanics*, Archives of Computational Methods in Engineering, 15 (2008), pp. 229–275.

[138] E.B. Rudnyi and J.G. Korvink, *Review: Automatic model reduction for transient simulation of MEMS-based devices*, Sensors Update, 11 (2002), pp. 3–33.

[139] J. Sabino, *Solution of large-scale Lyapunov equations via the block modified Smith method*, PhD thesis, Rice University, 2007.

[140] J.M.A Scherpen, *Balancing for nonlinear systems*, Systems & Control Letters, 21 (1993), pp. 143–153.

[141] W.H.A. Schilders, H.A. van der Vorst, and J. Rommes, eds., *Model Order Reduction: Theory, Research Aspects and Applications*, vol. 13 of Mathematics in Industry, Springer-Verlag, Berlin, 2008.

[142] V. Simoncini, *A new iterative method for solving large-scale Lyapunov matrix equations*, SIAM Journal on Scientific Computing, 29 (2008), pp. 1268–1288.

[143] L. Sirovich, *Turbulence and the dynamics of coherent structures. Part 1: Coherent structures*, Quarterly of Applied Mathematics, 45 (1987), pp. 561–571.

[144] S.A. Smolyak, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, Dokl. Akad. Nauk SSSR, 148 (1963), pp. 1042–1043.

[145] N.T. Son, *A real time procedure for affinely dependent parametric model order reduction using interpolation on grassmann manifolds*, International Journal for Numerical Methods in Engineering, 93 (2013), pp. 818–833.

[146] T. Stykel, *Gramian-based model reduction for descriptor systems*, Mathematics of Control, Signals, and Systems, 16 (2004), pp. 297–319.

[147] ——, *Low-rank iterative methods for projected generalized Lyapunov equations*, Electronic Transactions on Numerical Analysis, 30 (2008), pp. 187–202.

[148] T. Stykel and V. Simoncini, *Krylov subspace methods for projected Lyapunov equations*, Applied Numerical Mathematics, 62 (2012), pp. 35–50.

[149] T.-J. Su and R.R. Craig Jr., *Model reduction and control of flexible structures using Krylov vectors*, Journal of Guidance, Control, and Dynamics, 14 (1991), pp. 260–267.

[150] G. van Bühren, N. Hornung, T. Clees, and L. Nikitinaon, *Aspects of adaptive hierarchical RBF metamodels for optimization*, Journal of Computational Methods in Science and Engineering, 12 (2012), pp. 5–23.

[151] P. Van Dooren, K.A. Gallivan, and P.-A. Absil, $\mathcal{H}_2$*-optimal model reduction of MIMO systems*, Applied Mathematics Letters, 21 (2008), pp. 1267–1273.

[152] A. Varga and B.D.O. Anderson, *Accuracy-enhancing methods for balancing-related frequency-weighted model and controller reduction*, Automatica, 39 (2003), pp. 919–927.

[153] K. Veroy and A. Patera, *Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations: Rigorous reduced-basis* a posteriori *error bounds*, International Journal for Numerical Methods in Fluids, 47 (2005), pp. 773–788.

[154] K. Veroy, C. Prud'homme, D.V. Rovas, and A.T. Patera, *A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations*, in Proceedings of the 16th AIAA Computational Fluid Dynamics Conference, Orlando, FL, 2003. AIAA Paper 2003-3847.

[155] E.L. Wachspress, *The ADI minimax problem for complex spectra*, Applied Mathematics Letters, 1 (1988), pp. 311–314.

[156] J. Wang and N. Zabaras, *Using Bayesian statistics in the estimation of heat source in radiation*, International Journal of Heat and Mass Transfer, 48 (2005), pp. 15–29.

[157] K. Washabaugh, D. Amsallem, M. Zahr, and C. Farhat, *Nonlinear model reduction for CFD problems using local reduced-order bases*, in 42nd AIAA Fluid Dynamics Conference and Exhibit, Fluid Dynamics and Co-located Conferences, AIAA Paper 2012-2686, June 2012.

[158] D.S. Weile, E. Michielssen, E.J. Grimme, and K.A. Gallivan, *A method for generating rational interpolant reduced order models of two-parameter linear systems*, Applied Mathematics Letters, 12 (1999), pp. 93–102.

[159] K. Willcox and J. Peraire, *Balanced model reduction via the proper orthogonal decomposition*, AIAA Journal, 40 (2002), pp. 2323–2330.

[160] Y. Yue and K. Meerbergen, *Accelerating optimization of parametric linear systems by model order reduction*, SIAM Journal on Optimization, 23 (2013), pp. 1344–1370.

[161] C. Zenger, *Sparse grids*, in Parallel algorithms for partial differential equations (Kiel, 1990), vol. 31 of Notes on Numerical Fluid Mechanics, Vieweg, Braunschweig, 1991, pp. 241–251.

[162] K. Zhou, J.C. Doyle, and K. Glover, *Robust and Optimal Control*, vol. 40, Prentice Hall New Jersey, 1996.

[163] K. Zhou, G. Salomon, and E. Wu, *Balanced realization and model reduction for unstable systems*, International Journal of Robust and Nonlinear Control, 9 (1999), pp. 183–198.

[164] A Zilouchian, *Balanced structures and model reduction of unstable systems*, in IEEE Proceedings of Southeastcon'91., 1991, pp. 1198–1201.