Adv Comput Math (2020) 46: 83
https://doi.org/10.1007/s10444-020-09825-8

Sampling-free model reduction of systems with low-rank parameterization

Christopher Beattie[1] · **Serkan Gugercin**[1] · **Zoran Tomljanović**[2]

Received: 24 December 2019 / Accepted: 30 October 2020 /
Published online: 19 November 2020

Abstract

We consider the reduction of parametric families of linear dynamical systems having an affine parameter dependence that allow for low-rank variation in the state matrix. Usual approaches for parametric model reduction typically involve exploring the parameter space to identify representative parameter values and the associated models become the principal focus of model reduction methodology. These models are then combined in various ways in order to interpolate the response. The initial exploration of the parameter space can be a forbiddingly expensive task. A different approach is proposed here that requires neither parameter sampling nor parameter space exploration. Instead, we represent the system response function as a composition of four subsystem response functions that are *non-parametric* with a purely parameter-dependent function. One may apply any one of a number of standard (non-parametric) model reduction strategies to reduce the subsystems independently, and then conjoin these reduced models with the underlying parameterization to obtain the overall parameterized response. Our approach has elements in common with the parameter mapping approach of Baur et al. (PAMM **14**(1), 19–22 2014) but offers greater flexibility and potentially greater control over accuracy. In particular, a data-driven variation of our approach is described that exercises this flexibility through the use of limited frequency-sampling of the underlying non-parametric models. The parametric structure of our system representation allows for a priori guarantees of system stability in the resulting reduced models across the full range of parameter values. Incorporation of system theoretic error bounds allows us to determine appropriate approximation orders for the non-parametric systems sufficient to yield uniformly high accuracy across the parameter range. We illustrate our approach on a class of structural damping optimization problems and on a benchmark model of thermal conduction in a

Communicated by: Stefan Volkwein

✉ Zoran Tomljanović
ztomljan@mathos.hr

Extended author information available on the last page of the article.

semiconductor chip. The parametric structure of our reduced system representation lends itself very well to the development of optimization strategies making use of efficient cost function surrogates. We discuss this in some detail for damping parameter and location optimization for vibrating structures.

Keywords Sampling-free · Model reduction · Damping optimization

Mathematics subject classification (2010) 93C05 · 49J15 · 70Q05 · 70H33

1 Introduction

Consider a linear time invariant dynamical system, parameterized with a k-dimensional parameter vector $\mathsf{p} = \left[p_1, p_2, \ldots, p_k\right]^T \in \Omega \subseteq \mathbb{R}^k$ and represented in state-space form as:

$$\begin{aligned} E\dot{x}(t;\mathsf{p}) &= A(\mathsf{p})x(t;\mathsf{p}) + Bw(t), \\ y(t;\mathsf{p}) &= Cx(t;\mathsf{p}), \end{aligned} \tag{1}$$

where $E, A(\mathsf{p}) \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{\ell\times n}$ are constant (time-invariant) matrices. In (1), $x(t;\mathsf{p}) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, and $y(t;\mathsf{p}) \in \mathbb{R}^\ell$ denote the state vector, inputs, and outputs, respectively. We assume throughout that the matrix E is invertible.

1.1 Basic structure

The presumed structural feature of (1) that we will exploit extensively will be that the system matrix $A(\mathsf{p})$ has the parametric form:

$$A(\mathsf{p}) = A_0 - U\ \mathrm{diag}(p_1, p_2, \ldots, p_k)V^T = A_0 - \sum_{i=1}^{k} p_i u_i v_i^T, \tag{2}$$

where $U = \left[u_1, u_2, \ldots, u_k\right] \in \mathbb{R}^{n\times k}$ and $V = \left[v_1, v_2, \ldots, v_k\right] \in \mathbb{R}^{n\times k}$ are constant matrices with $u_i, v_i \in \mathbb{R}^k$ for $i = 1, \ldots, k$. The parameterization in (2) is a special case of a general affine parametrization $A(\mathsf{p}) = A_0 - \sum_i p_i A_i$ with an added rank constraint that $\mathsf{rank}(A_i) = 1$. Even for a general affine parametrization with $A_i \in \mathbb{R}^{n\times n}$ individually having unrestricted rank but having $\mathsf{rank}(\sum_i p_i A_i) = k \ll n$, in aggregate the form of (2) may be assumed without loss of generality (allowing for the possibility that the parameters are replaced by functions of $\{p_i\}_i$). The condition $k \ll n$ is a practical constraint leading to the prospect of computational efficiency, but there is no theoretical restriction on the size of k.

Taking the Laplace transform of (1), the full-order transfer function of the parametrized system is obtained as:

$$\mathcal{H}(s;\mathsf{p}) = C(sE - A(\mathsf{p}))^{-1}B = C\left(sE - \left(A_0 - U\ \mathrm{diag}(p_1, p_2, \ldots, p_k)V^T\right)\right)^{-1}B. \tag{3}$$

The goal of parametric model reduction, in this setting, is to find a reduced parametric system:

$$\begin{aligned} E_r \dot{x}_r(t;\mathsf{p}) &= A_r(\mathsf{p}) x_r(t;\mathsf{p}) + B_r w(t), \\ y_r(t;\mathsf{p}) &= C_r x_r(t;\mathsf{p}), \end{aligned} \tag{4}$$

where E_r, $A_r(\mathsf{p}) \in \mathbb{R}^{r\times r}$, $B_r \in \mathbb{R}^{r\times m}$, and $C_r \in \mathbb{R}^{\ell\times r}$ with $r \ll n$ such that the reduced transfer function:

$$\mathcal{H}_r(s;\mathsf{p}) = C_r(sE_r - A_r(\mathsf{p}))^{-1}B_r$$

approximates $\mathcal{H}(s;\mathsf{p})$ accurately for the parameter range of interest.

Several approaches to parametric model order reduction (pMOR) exist. One of the most common approaches involves state-space projection using globally defined bases: Choose a set of parameter samples $\mathsf{p}^1, \ldots, \mathsf{p}^{n_s}$. For every parameter sample p^i, the full-order model $\mathcal{H}(s;\mathsf{p}^k)$ becomes a non-parametric linear time-invariant system, for which a plethora of model reduction methods are available. Whatever choice is made, let Z_r^i and W_r^i denote *local* model reduction bases for the parameter sample p^i, for each $i = 1, \ldots, n_s$. Then concatenate these local bases to form the *global* model reduction bases Z_r and W_r: $Z_r = [Z_r^1, \ldots, Z_r^{n_s}]$ and $W_r = [W_r^1, \ldots, W_r^{n_s}]$. This concatenation step is usually followed by a rank-revealing QR or truncated SVD computation to compute and condense orthogonal bases. The parametric reduced model quantities in (4) are obtained via a Petrov-Galerkin projection, i.e.,

$$E_r = W_r^T E Z_r, \quad A_r(\mathsf{p}) = W_r^T A(\mathsf{p}) Z_r, \quad B_r = W_r^T B, \text{ and } C_r = C Z_r. \tag{5}$$

Reviews of methods that consider such a reduction framework can be found, e.g., in [3, 4, 6, 9, 11, 13, 14, 16, 29, 59]. These approaches are widely studied especially for systems inheriting structure tied with particular applications; see, e.g., [6, 13–15, 59, 62, 65, 66].

The global basis approach has been applied successfully in many circumstances requiring parametric model reduction and in some cases it may be the only viable approach. Nonetheless, it comes with some drawbacks, the main issue being the need to sample the parameter domain adequately in order to construct representative local bases. Except for special cases [9, 34], how one chooses optimal parameter sampling points with respect to a joint global frequency-parameter error measure has not been known until recently. In [43], Hund et al. tackle this joint-optimization problem by deriving optimality conditions and then constructing model reduction bases that enforce those conditions. The most widely used approaches for global basis construction in pMOR are greedy or optimization-based sampling strategies; see [14] for a survey. However, especially in the case of high-dimensional parameter domains, this off-line stage could prove prohibitively expensive since it requires a large-number of full-order function evaluations. One may try to avoid the overhead of these high-fidelity sampling techniques and pick parameter samples heuristically. Since the global bases directly depend on the initial sampling and this, in turn, influences the final accuracy of the parametric reduced model, if this initial sampling stage is not done properly the reduced model could not be expected to provide good approximations over a wide parameter range.

In this paper, we focus on systems having the special structure described in (2). We develop a novel parametric model order reduction approach that is sampling free (that is, there is no need for parameter sampling) yet it still offers uniformly high fidelity across the full parameter range. Significantly, the reduced model retains the parametric structure of the original full model.

1.2 A motivating example: damping optimization

Consider the vibrational system described by:

$$\begin{aligned} M\ddot{q}(t) + D\dot{q}(t) + Kq(t) &= B_2 w(t), \\ y(t) &= C_2 q(t), \end{aligned} \tag{6}$$

where M and K are real, symmetric positive definite matrices of size $n \times n$, denoting the mass and stiffness matrices, respectively. The state variables are described by the coordinate vector $q \in \mathbb{R}^n$ representing structure displacements. The time-dependent vector $w(t) \in \mathbb{R}^m$ is the primary excitation and typically represents an input disturbance. $B_2 \in \mathbb{R}^{n\times m}$ is the *primary excitation matrix*, i.e., the input-to-state mapping. Similarly, $y(t) \in \mathbb{R}^\ell$ is the performance output, representing a quantity of interest that is obtained from the state-vector q via a mapping by the state-output matrix $C_2 \in \mathbb{R}^{\ell\times n}$.

The damping matrix $D \in \mathbb{R}^{n\times n}$ is modeled as:

$$D = D_{int} + D_{ext},$$

where D_{ext} represents external damping and D_{int} represents internal damping. The internal damping D_{int} is usually taken to be a small multiple of the critical damping denoted by D_{crit} or a small multiple of proportional damping (see, e.g., [15, 19]):

$$D_{\text{int}} = \alpha_c D_{\text{crit}}, \quad \text{where} \quad D_{\text{crit}} = 2M^{1/2}\sqrt{M^{-1/2}KM^{-1/2}}M^{1/2}. \tag{7}$$

Other possibilities for modeling internal damping have been considered, e.g., in [46].

We are mainly interested in external damping taking the form:

$$D_{\text{ext}} = U_2 \,\text{diag}\,(p_1, p_2, \ldots, p_k) U_2^T,$$

where the non-negative entries p_i for $i = 1, \ldots, k$ represent the friction coefficients of the dampers, usually called gains or viscosities, and the matrix U_2 encodes the damper positions and geometry; for more details, see, e.g., [15, 18, 21, 49, 62, 64].

In damping optimization problems, the principal goal is to determine an *optimal* external damping matrix D_{ext} that will minimize the influence of the input w (viewed as a disturbance) on the output, y. One can consider different optimality measures in quantifying this influence. For input-state-output representations of dynamical systems, natural optimization criteria are usually based on system norms such as the $\mathcal{H}_2$ or the $\mathcal{H}_\infty$ system norm (see, e.g., [15, 21, 63]; a mixed performance measure was considered in [52]). Evidently, the specific choice of optimization criterion strongly depends on the application context at hand. The pMOR methods we develop in this paper allow for the use of different optimization criteria, thus enabling a broad range of applications.

By defining the state-vector as $x = [q^T \ \dot{q}^T]^T$, we obtain a first-order state-space representation of the vibrational system:

$$\begin{aligned} E\dot{x}(t) &= A(\mathsf{p})x(t) + Bw(t), \\ y(t) &= Cx(t), \end{aligned} \tag{8}$$

where

$$E = \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ B_2 \end{bmatrix}, \quad C = \begin{bmatrix} C_2 & 0 \end{bmatrix}, \tag{9}$$

$$\text{and} \quad A(\mathsf{p}) = \underbrace{\begin{bmatrix} 0 & I \\ -K & -D_{int} \end{bmatrix}}_{A_0} - \underbrace{\begin{bmatrix} 0 \\ U_2 \end{bmatrix}}_{U} \operatorname{diag}(p_1, p_2, \ldots, p_k) \underbrace{\begin{bmatrix} 0 & U_2^T \end{bmatrix}}_{V^T = U^T}, \tag{10}$$

with $\mathsf{p} = \begin{bmatrix} p_1, p_2, \ldots, p_k \end{bmatrix}^T$. Note that the model for damped vibration in (9)–(10) has the parametric structure described in (2).

The optimization of damper locations can be formulated effectively as optimization over a finite (but potentially large) number of configurations for the matrix B_2; this is a demanding combinatorial optimization problem and for each B_2-configuration, one must optimize over p, the parameter vector. pMOR approaches seeking to make this task cheaper have been considered previously: for optimization based on the $\mathcal{H}_2$ norm criterion, [15] used a global basis approach, as described above, where local bases were obtained via the dominant pole algorithm [60]. Using the same optimization criterion, [62] applied a global basis approach where local bases were obtained via the Iterative Rational Krylov Algorithm (IRKA) [37], an $\mathcal{H}_2$-optimal model reduction approach. Even though both approaches show great promise, success in each case depends on the initial parameter sampling used to construct the global basis, an issue that is faced in most pMOR cases. For the damping optimization problem, an efficient heuristic that can guide a parameter sampling strategy is not available, and the natural alternative, a preliminary offline greedy sampling stage, can be computationally very demanding and potentially negate the gains one would anticipate from model reduction.

In subsequent sections, we will propose two frameworks that remove the need for parametric sampling in problems structured as in (2), including in particular the damping optimization problem discussed above.

2 PMOR based on subsystem model reduction

We introduce first a sampling-free reduction method for (3), providing both error bounds and a discussion of the stability guarantees that can be made for the reduced models across the full parameter range.

2.1 Reformulation of the parametric transfer function

The crucial observation and starting point for our development is that we can rewrite the structured transfer function (3) in a form that separates the s and p dependence by making use of the Sherman-Morrison-Woodbury formula [32].

Proposition 1 *Consider the structured transfer function*

$$\mathcal{H}(s;\mathsf{p}) = C\left(sE - \left(A_0 - U\ \mathrm{diag}(p_1, p_2, \ldots, p_k)V^T\right)\right)^{-1} B, \tag{11}$$

where $p_i \geq 0$, for $i = 1, 2, \ldots, k$. Then,

$$\mathcal{H}(s;\mathsf{p}) = \mathcal{H}_1(s) - \mathcal{H}_2(s)D(\mathsf{p})\left[I + D(\mathsf{p})\mathcal{H}_3(s)D(\mathsf{p})\right]^{-1} D(\mathsf{p})\mathcal{H}_4(s), \tag{12}$$

where the diagonal matrix

$$D(\mathsf{p}) = \mathrm{diag}(\sqrt{p}_1, \sqrt{p}_2, \ldots, \sqrt{p}_k) \tag{13}$$

encodes the parameters, and $\mathcal{H}_1(s)$, $\mathcal{H}_2(s)$, $\mathcal{H}_3(s)$, and $\mathcal{H}_4(s)$ are transfer functions that are independent of the parameters, given by

$$\begin{aligned} \mathcal{H}_1(s) &= C(sE - A_0)^{-1}B, & \mathcal{H}_2(s) &= C(sE - A_0)^{-1}U, \\ \mathcal{H}_3(s) &= V^T(sE - A_0)^{-1}U, \text{ and} & \mathcal{H}_4(s) &= V^T(sE - A_0)^{-1}B. \end{aligned} \tag{14}$$

Proof Let $T \in \mathbb{C}^{n\times n}$ be invertible and assume for the moment that $X \in \mathbb{C}^{n\times k}$ and $Y \in \mathbb{C}^{n\times k}$ are such that $I_k + Y^T T^{-1} X$ is invertible. The Sherman-Morrison-Woodbury formula asserts that $T + XY^T$ is invertible and

$$(T + XY^T)^{-1} = T^{-1} - T^{-1}X(I_k + Y^T T^{-1}X)^{-1}Y^T T^{-1}.$$

Conversely, if $T + XY^T$ is invertible, then $I_k + Y^T T^{-1} X$ will be invertible as well. The conclusion follows from (11) and (13) with the assignments: $T = sE - A_0$, $X = UD(\mathsf{p})$, and $Y = VD(\mathsf{p})$. □

Remark 1 Notice that if we define the extended parameterized transfer function:

$$\widetilde{\mathcal{F}}(s;\mathsf{p}) = \begin{bmatrix} \mathcal{H}_1(s) & \mathcal{H}_2(s)D(\mathsf{p}) \\ D(\mathsf{p})\mathcal{H}_4(s) & I + D(\mathsf{p})\mathcal{H}_3(s)D(\mathsf{p}) \end{bmatrix},$$

then (12) is the Schur complement of $\widetilde{\mathcal{F}}(s;\mathsf{p})$ with respect to the (2,2) block:

$$\mathcal{H}(s;\mathsf{p}) = \left[\widetilde{\mathcal{F}}(s;\mathsf{p})/(I + D(\mathsf{p})\mathcal{H}_3(s)D(\mathsf{p}))\right].$$

Remark 2 Proposition 1 assumes that the parameter vector p has non-negative entries, leading to the form (12) where the diagonal matrix $D(p)$ appears symmetrically in the second term. This form was motivated in part by the damping optimization problem described in Section 1.2, but evidently neither this form nor

the non-negativity of the parameter range is a necessary assumption and the general case can be accommodated in a similar way by defining:

$$\widetilde{D}(\mathsf{p}) = \mathrm{diag}(p_1, p_2, \ldots, p_k)$$

and

$$\mathcal{H}(s;\mathsf{p}) = \mathcal{H}_1(s) - \mathcal{H}_2(s)\widetilde{D}(\mathsf{p})\left[I + \mathcal{H}_3(s)\widetilde{D}(\mathsf{p})\right]^{-1}\mathcal{H}_4(s), \tag{15}$$

where $\mathcal{H}_1(s)$, $\mathcal{H}_2(s)$, $\mathcal{H}_3(s)$, and $\mathcal{H}_4(s)$ are as defined in (14). We will use the symmetric formulation presented in Proposition 1, noting that all results to follow are easily generalized to non-positive parameter classes using the non-symmetric form (15).

2.2 Subsystem model reduction

Proposition 1 displays a decomposition of the full-order transfer function $\mathcal{H}(s;\mathsf{p})$ in terms of four *non-parametric* transfer functions with the parameter dependency entering as an interconnection coupling the four systems. Since $\mathcal{H}_i(s)$, for $i = 1, 2, 3, 4$, are non-parametric, they may be reduced without any need for sampling via well-established model reduction techniques such as balanced truncation (BT) [48, 50], Hankel norm approximation (HNA) [31], or iterative rational Krylov algorithm IRKA [37]; see [4, 6, 13] for further choices. Moreover, each system $\mathcal{H}_i(s)$, may be reduced independently of the others, potentially using different reduction orders and even different reduction methodologies.

Let the reduced model for $\mathcal{H}_i(s)$ be denoted by $\widehat{\mathcal{H}}_i(s)$, for $i = 1, \ldots, 4$. The resulting parametric reduced model for $\mathcal{H}(s;\mathsf{p})$ is $\widehat{\mathcal{H}}(s;\mathsf{p})$ given by:

$$\widehat{\mathcal{H}}(s;\mathsf{p}) = \widehat{\mathcal{H}}_1(s) - \widehat{\mathcal{H}}_2(s)D(\mathsf{p})(I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p}))^{-1}D(\mathsf{p})\widehat{\mathcal{H}}_4(s). \tag{16}$$

Online evaluation or simulation of $\widehat{\mathcal{H}}(s;\mathsf{p})$ for a given parameter value is trivial as it only involves reduced quantities and evaluation of the matrix $D(\mathsf{p})$. We have constructed a parameterized, easy-to-evaluate, reduced model without any need for parameter sampling. Algorithm 1 below gives a sketch of this process.

Algorithm 1 Parametric reduced order model based on reduction of subsystems.

1: Offline Stage: Calculate the four non-parametric reduced systems

$$\mathcal{H}_1(s) \to \widehat{\mathcal{H}}_1(s), \mathcal{H}_2(s) \to \widehat{\mathcal{H}}_2(s), \mathcal{H}_3(s) \to \widehat{\mathcal{H}}_3(s), \mathcal{H}_4(s) \to \widehat{\mathcal{H}}_4(s),$$

(Reductions can be performed via a variety of non-parametric reduction techniques)

2: Online Stage: For any given parameter $\mathsf{p} = \left[p_1, p_2, \ldots, p_k\right]^T$, obtain the parametric model by

$$\mathcal{H}(s;\mathsf{p}) \approx \widehat{\mathcal{H}}(s;\mathsf{p}) = \widehat{\mathcal{H}}_1(s) - \widehat{\mathcal{H}}_2(s)D(\mathsf{p})(I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p}))^{-1}D(\mathsf{p})\widehat{\mathcal{H}}_4(s).$$

Some observations are appropriate regarding step 1 in Algorithm 1. The model $\mathcal{H}_1(s)$ has the same input-output dimension as $\mathcal{H}(s;\mathsf{p})$. The model $\mathcal{H}_2(s)$ has the

same number of outputs (ℓ) as $\mathcal{H}(s;\mathsf{p})$ and k inputs. Analogously, the model $\mathcal{H}_4(s)$ has the same number of inputs (m) as $\mathcal{H}(s;\mathsf{p})$, and k outputs. Provided that the input/output dimension is modest, reducing $\mathcal{H}_1(s)$, $\mathcal{H}_2(s)$, and $\mathcal{H}_4(s)$ will not be expected to be strongly influenced by the size of k since in most cases the smaller of the input/output dimensions determines the difficulty in reducing dynamical systems. On the other hand, the model $\mathcal{H}_3(s)$ will have k-inputs and k-outputs. Therefore, if k is significantly larger than ℓ and m, this may be the most difficult subsystem of the four to reduce with high fidelity. Although the framework and theoretical analysis we develop here would also apply to a system with transfer function:

$$\mathcal{H}(s;\,p) = C(sE - (A_0 - pA_1))^{-1}B$$

even when A_1 has full rank. It is unlikely that any computational benefits would be seen with this strategy since $\mathcal{H}_3(s)$ will then be an n-dimension dynamical system with n-inputs and n-outputs. Such models are not generally amenable to model reduction.

We next consider error bounds for parameterized reduced models obtained from Algorithm 1.

Theorem 1 *Let the full order transfer function $\mathcal{H}(s;\mathsf{p})$, and the corresponding subsystems $\mathcal{H}_i(s)$, for $i = 1,\ldots,4$ be given as in (12) and (14). Assume that the non-parametric reduced models $\mathcal{H}_i(s)$ are reduced so that*

$$\|\mathcal{H}_i - \widehat{\mathcal{H}}_i\| \le \epsilon_i, \quad \textit{for} \quad i = 1,\ldots,4, \tag{17}$$

and that the corresponding parametric reduced model $\widehat{\mathcal{H}}(s;\mathsf{p})$ is constructed as in (16). Then,

$$\begin{aligned}\|\mathcal{H}(\cdot;\mathsf{p})) - \widehat{\mathcal{H}}(\cdot;\mathsf{p}))\| \le\ & \epsilon_1 \ +\ f_1(\mathsf{p},\widehat{\mathcal{H}}_3,\widehat{\mathcal{H}}_4)\epsilon_2 \\ & + f_1(\mathsf{p},\widehat{\mathcal{H}}_3,\widehat{\mathcal{H}}_4)\,f_2(\mathsf{p},\mathcal{H}_2,\mathcal{H}_3)\epsilon_3 \ +\ f_2(\mathsf{p},\mathcal{H}_2,\mathcal{H}_3)\epsilon_4,\end{aligned} \tag{18}$$

where

$$f_1(\mathsf{p},\mathcal{G}_1,\mathcal{G}_2) = \|D(\mathsf{p})\big(I + D(\mathsf{p})\mathcal{G}_1(\cdot)D(\mathsf{p})\big)^{-1}D(\mathsf{p})\mathcal{G}_2(\cdot)\| \ \textit{and} \tag{19}$$

$$f_2(\mathsf{p},\mathcal{G}_1,\mathcal{G}_2) = \|\mathcal{G}_1(\cdot)D(\mathsf{p})\big(I + D(\mathsf{p})\mathcal{G}_2(\cdot)D(\mathsf{p})\big)^{-1}D(\mathsf{p})\|. \tag{20}$$

Proof First, by using the formulae (12) and (16), we obtain:

$$\mathcal{H}(s;\mathsf{p}) - \widehat{\mathcal{H}}(s;\mathsf{p}) = \mathcal{H}_1(s) - \widehat{\mathcal{H}}_1(s) + \widehat{\mathcal{H}}_2(s)\widehat{\mathcal{E}}_1 - \mathcal{H}_2(s)\mathcal{E}_1, \tag{21}$$

where

$$\mathcal{E}_1 = D(\mathsf{p})(I + D(\mathsf{p})\mathcal{H}_3(s)D(\mathsf{p}))^{-1}D(\mathsf{p})\mathcal{H}_4(s)$$

and

$$\widehat{\mathcal{E}}_1 = D(\mathsf{p})(I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p}))^{-1}D(\mathsf{p})\widehat{\mathcal{H}}_4(s).$$

The last two terms in (21) can be manipulated together as:

$$\begin{aligned}\widehat{\mathcal{H}}_2(s)\widehat{\mathcal{E}}_1 - \mathcal{H}_2(s)\mathcal{E}_1 &= [\widehat{\mathcal{H}}_2(s) - \mathcal{H}_2(s)]\widehat{\mathcal{E}}_1 + \mathcal{H}_2(s)(\widehat{\mathcal{E}}_1 - \mathcal{E}_1) \\ &= [\widehat{\mathcal{H}}_2(s) - \mathcal{H}_2(s)]\widehat{\mathcal{E}}_1 + \mathcal{H}_2(s)[\widehat{\mathcal{E}}_2\widehat{\mathcal{H}}_4(s) - \mathcal{E}_2\mathcal{H}_4(s)],\end{aligned}$$

where

$$\mathcal{E}_2 = D(\mathsf{p})(I + D(\mathsf{p})\mathcal{H}_3(s)D(\mathsf{p}))^{-1}D(\mathsf{p})$$

and

$$\widehat{\mathcal{E}}_2 = D(\mathsf{p})(I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p}))^{-1}D(\mathsf{p}),$$

and then, the previous expression can be rewritten as:

$$\begin{aligned}\widehat{\mathcal{H}}_2(s)\widehat{\mathcal{E}}_1 - \mathcal{H}_2(s)\mathcal{E}_1 \;=\; & [\widehat{\mathcal{H}}_2(s) - \mathcal{H}_2(s)]\widehat{\mathcal{E}}_1 \\ & + \mathcal{H}_2(s)[(\widehat{\mathcal{E}}_2 - \mathcal{E}_2)\widehat{\mathcal{H}}_4(s) + \mathcal{E}_2(\widehat{\mathcal{H}}_4(s) - \mathcal{H}_4(s))].\end{aligned} \tag{22}$$

Using the identity, $(I + \widehat{M})^{-1} - (I + M)^{-1} = (I + M)^{-1}(M - \widehat{M})(I + \widehat{M})^{-1}$, we find

$$\begin{aligned}\widehat{\mathcal{E}}_2 - \mathcal{E}_2 \;=\; & D(\mathsf{p})(I + D(\mathsf{p})\mathcal{H}_3(s)D(\mathsf{p}))^{-1}D(\mathsf{p}) \\ & \times[\mathcal{H}_3(s) - \widehat{\mathcal{H}}_3(s)]D(\mathsf{p})(I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p}))^{-1}D(\mathsf{p}).\end{aligned} \tag{23}$$

Substituting (23) into (22), which is then substituted into (21), yields:

$$\begin{aligned}\mathcal{H}(\cdot;\mathsf{p}) - \widehat{\mathcal{H}}(\cdot;\mathsf{p}) \;=\; & [\mathcal{H}_1(s) - \widehat{\mathcal{H}}_1(s)] \\ & + [\mathcal{H}_2(s) - \widehat{\mathcal{H}}_2(s)]D(\mathsf{p})\big(I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p})\big)^{-1}D(\mathsf{p})\widehat{\mathcal{H}}_4(s) \\ & + \mathcal{H}_2(s)D(\mathsf{p})\big(I + D(\mathsf{p})\mathcal{H}_3(s)D(\mathsf{p})\big)^{-1}D(\mathsf{p})[\widehat{\mathcal{H}}_4(s) - \mathcal{H}_4(s)] \\ & + \mathcal{H}_2(s)D(\mathsf{p})\big(I + D(\mathsf{p})\mathcal{H}_3(s)D(\mathsf{p})\big)^{-1}D(\mathsf{p})[\mathcal{H}_3(s) - \widehat{\mathcal{H}}_3(s)] \\ & \times D(\mathsf{p})\big(I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p})\big)^{-1}D(\mathsf{p})\widehat{\mathcal{H}}_4(s).\end{aligned}$$

The upper bound follows by taking norms on both sides. □

2.3 Uniform stability of the parameterized reduced model

In many applications, the parameters of interest are not only non-negative but are also restricted to some given subset Ω of $\mathbb{R}^k$. We will assume in all that follows that $\mathsf{p} \in \Omega$ for some fixed Ω given as a closed but not necessarily bounded subset of the non-negative orthant in $\mathbb{R}^k$. Likewise, it occurs in many applications that the full parameterized model $\mathcal{H}(s, \mathsf{p})$ defined in (11) is asymptotically stable for every $\mathsf{p} \in \Omega$, so that eigenvalues of the matrix pencil $\lambda E - A(\mathsf{p})$ (which are then poles of the transfer function $\mathcal{H}(s, \mathsf{p})$) have negative real parts for every $\mathsf{p} \in \Omega$. Following terminology introduced in [14, §5.4], we refer to this property as *uniform asymptotic stability*.[1] Evidently, it will be important that this property be maintained across the whole family of reduced models as well. For example, the damping optimization problem we considered in Section 1.2 is asymptotically stable for every parameter in the parameter domain of interest, and so having guarantees that the same is true for the parameterized reduced model would be both useful and reasonable. Except for some special cases when structural features guarantee asymptotic stability (e.g.,

[1] Note that this is a *pointwise* property with respect to $\mathsf{p} \in \Omega$, contrasting with other common uses of the qualifier "uniform."

when $E = E^T$ is positive define, $A(\mathsf{p}) = A^T(\mathsf{p})$ is negative definite for all p, and one chooses $W_r = Z_r$ in (5)), uniform asymptotic stability across the given parameter range may be difficult to enforce in parameterized reduced models. See [14, §5.4] for a brief discussion on this issue. Significantly, our present framework is able to provide guarantees of uniform asymptotic stability for a broad class of problems.

By examining the parametric structure of the decomposition of the original transfer function $\mathcal{H}(s, \mathsf{p})$ in (12), we can deduce conditions sufficient to guarantee asymptotic stability across the full parameter range. For the original parameterized system, the overall system poles for $\mathsf{p} = 0$ (i.e., for the unperturbed system) are eigenvalues of the matrix pencil $sE - A_0$ and hence, also poles for the subsystems $\mathcal{H}_1(s)$, $\mathcal{H}_2(s)$, $\mathcal{H}_3(s)$, and $\mathcal{H}_4(s)$. Thus, it is natural to adopt the assumption that all four original non-parametric subsystems are asymptotically stable. Secondly, by examining the corresponding parametric decomposition of the reduced transfer function, $\widehat{\mathcal{H}}(s, \mathsf{p})$, in (16), the poles of $\widehat{\mathcal{H}}(s, \mathsf{p})$ must be contained within the set of points made up by the poles of $\widehat{\mathcal{H}}_1(s)$, $\widehat{\mathcal{H}}_2(s)$, and $\widehat{\mathcal{H}}_4(s)$, together with the zeroes of $I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p})$ (i.e., values of s where $I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p})$ becomes singular). Not all values in this aggregate set are necessarily poles of $\widehat{\mathcal{H}}(s, \mathsf{p})$ since there could be pole-zero cancellation. In order to assure uniform asymptotic stability in $\widehat{\mathcal{H}}(s, \mathsf{p})$, we should require that the non-parametric reduced subsystems $\widehat{\mathcal{H}}_1(s)$, $\widehat{\mathcal{H}}_2(s)$, and $\widehat{\mathcal{H}}_4(s)$ should retain asymptotic stability. This can be done by employing any of a variety of reduction strategies that enforce asymptotic stability, for example, BT [48, 50], HNA [31], or IRKA with enforced stability [12, 36, 37].

We will also need to guarantee that all zeros of $I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p})$ occur in the left half plane for all values of $\mathsf{p} \in \Omega$, in order to assure uniform asymptotic stability in $\widehat{\mathcal{H}}(s, \mathsf{p})$. This translates into seeking conditions on $\widehat{\mathcal{H}}_3(s)$ such that whenever $\mathsf{Re}(s) \geq 0$ and $\mathsf{p} \in \Omega$ then $I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p})$ is invertible. We look to our damped vibration model, (8)–(10), for guidance. Due to the internal damping term, D_{int}, eigenvalues of the matrix pencil $sE - A_0$ have negative real parts; and thus all the subsystems, $\mathcal{H}_1(s)$, $\mathcal{H}_2(s)$, $\mathcal{H}_3(s)$, and $\mathcal{H}_4(s)$, are asymptotically stable. A key point to note is that $\mathcal{H}_3(s) = sU_2^T(s^2M + sD + K)^{-1}U_2$ is a transfer function describing a *passive* dynamical system; indeed, it is a *port-Hamiltonian system*. Port-Hamiltonian systems are a category of dynamical systems possessing representations that encode the manner in which energy passes through the system. In its simplest form, a port-Hamiltonian system has a transfer function that appears as $\mathcal{H}_{\mathsf{PH}}(s) = B^T Q(sI - (J - R)Q)^{-1}B$ where $B \in \mathbb{R}^{n\times k}$ and Q, J, and R are in $\mathbb{R}^{n\times n}$; Q is symmetric positive-definite; R is symmetric positive-semidefinite; and J is skew-symmetric. The subsystem transfer function, $\mathcal{H}_3(s)$, from our damped vibration problem has precisely this form as can be seen by assigning $J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$, $R = \begin{bmatrix} 0 & 0 \\ 0 & D \end{bmatrix}$, $Q = \begin{bmatrix} K & 0 \\ 0 & M^{-1} \end{bmatrix}$, and $B = \begin{bmatrix} 0 \\ U_2 \end{bmatrix}$.

The defining representation for port-Hamiltonian systems leads to a key property that we will exploit: the transfer function $\mathcal{H}_{\mathsf{PH}}(s)$ is *positive real*, which means that

$\mathcal{H}_{\mathsf{PH}}(s) + \overline{\mathcal{H}_{\mathsf{PH}}(s)}^T$ is positive semi-definite for all s with $\mathsf{Re}(s) \geq 0$. This can be easily seen from:

$$\begin{aligned}\mathcal{H}_{\mathsf{PH}}(s) + \overline{\mathcal{H}_{\mathsf{PH}}(s)}^T &= B^T \left((sQ^{-1} - (J-R))^{-1} + (\bar{s}Q^{-1} - (J-R)^T)^{-1} \right) B \\ &= F^\star \left((s+\bar{s})Q^{-1} + 2R \right) F \geq 0,\end{aligned}$$

for $s + \bar{s} = 2\mathsf{Re}(s) \geq 0$, with $F = (sQ^{-1} - (J-R))^{-1}B$.

Since it has port-Hamiltonian structure, the subsystem transfer function, $\mathcal{H}_3(s)$, from our damped vibration problem is positive real. When $\mathcal{H}_3$ is reduced in such a way that $\widehat{\mathcal{H}}_3$ is also positive real, we will see that the aggregate parameterized model $\widehat{\mathcal{H}}(s, \mathsf{p})$ will be uniformly asymptotically stable. Note that positive real model reduction can be accomplished in a variety of ways, e.g., using positive real balanced truncation [25, 53] or interpolatory port-Hamiltonian model reduction [28, 38, 58].

Theorem 2 *Suppose that the full parameterized model* $\mathcal{H}(s, \mathsf{p})$ *described in* (11) *has been decomposed as in* (12) *into subsystems* $\mathcal{H}_1(s)$, $\mathcal{H}_2(s)$, *and* $\mathcal{H}_4(s)$ *that are each asymptotically stable, and a subsystem* $\mathcal{H}_3(s)$ *that is positive real. If the corresponding reduced subsystems* $\widehat{\mathcal{H}}_1(s)$, $\widehat{\mathcal{H}}_2(s)$, *and* $\widehat{\mathcal{H}}_4(s)$ *retain asymptotic stability, and* $\widehat{\mathcal{H}}_3(s)$ *retains positive-realness, then the reduced parameterized model* $\widehat{\mathcal{H}}(s, \mathsf{p})$ *in* (16) *is uniformly asymptotically stable for* $\mathsf{p} \in \Omega$.

Proof Following on the discussion above, the reduced transfer function $\widehat{\mathcal{H}}(s, \mathsf{p})$ in (16) has all its poles drawn from a set composed of the aggregate poles of $\widehat{\mathcal{H}}_1(s)$, $\widehat{\mathcal{H}}_2(s)$, and $\widehat{\mathcal{H}}_4(s)$, together with the set of zeroes of $I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p})$. The poles of $\widehat{\mathcal{H}}_1(s)$, $\widehat{\mathcal{H}}_2(s)$, and $\widehat{\mathcal{H}}_4(s)$ are independent of p and all have negative real parts since these subsystems have been reduced so as to be asymptotically stable. If the reduced parameterized model $\widehat{\mathcal{H}}(s, \mathsf{p})$ were *not* uniformly asymptotically stable then it must be that at least one zero of $I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p})$ is *not* in the open left half-plane and, in particular, it must be true that $I + D(\mathsf{p}_0)\widehat{\mathcal{H}}_3(z_0)D(\mathsf{p}_0)$ is singular for some z_0 with $\mathsf{Re}(z_0) \geq 0$ and some $\mathsf{p}_0 \in \Omega$. In this case, there must then exist a non-trivial $x_0 \in \mathbb{C}^n$ such that $(I + D(\mathsf{p}_0)\widehat{\mathcal{H}}_3(z_0)D(\mathsf{p}_0))x_0 = 0$ and so,

$$\begin{aligned}0 \;>\; -\|x_0\|^2 &= x_0^\star D(\mathsf{p})\widehat{\mathcal{H}}_3(z_0)D(\mathsf{p})x_0 \\ &= \frac{1}{2}\,(D(\mathsf{p})x_0)^\star \left(\widehat{\mathcal{H}}_3(z_0) + \overline{\widehat{\mathcal{H}}_3(z_0)}^T \right) (D(\mathsf{p})x_0) \geq 0.\end{aligned}$$

The last inequality comes from the assumption that $\widehat{\mathcal{H}}_3(s)$ is positive real and leads us to a contradiction. Thus, all zeros of $I + D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p}))$ must necessarily occur in the open left half-plane, $\mathsf{Re}(s) < 0$ for all $\mathsf{p} \in \Omega$, and the reduced parameterized model $\widehat{\mathcal{H}}(s, \mathsf{p})$ is uniformly asymptotically stable. □

Another approach that guarantees uniform asymptotic stability under still weaker assumptions (but with sampling) was proposed by Baur and Benner [10]. $\mathcal{H}(s, \mathsf{p})$ is sampled at n_s parameter configurations, $\mathsf{p}^{(i)}$ for $i = 1, 2, \ldots, n_s$, and each of the resulting n_s systems is reduced independently using any of the stability-preserving reduction methods previously mentioned (see citations above for BT, HNA, and

IRKA with enforced stability). The final parametric reduced system is obtained by (multivariate) interpolation of these local reduced models in the p-domain. Since interpolation with respect to p will not shift pole locations with respect to s (in effect, the p-dependency is focussed completely in the **B**- and/or **C**-matrix), the resulting reduced parametric system is guaranteed to be uniformly asymptotically stable. Conversely, in our formulation, the poles vary with p, yet in such a way as to assure uniform asymptotic stability. Allowing for the movement of poles with variation of p gives both greater flexibility and efficiency, and supports our goal to reduce and possibly altogether avoid the need for parameter sampling.

Remark 3 Significantly, Theorem 2 guarantees for the damping optimization problem considered in Section 1.2 (as well as for other problems yielding $\mathcal{H}_3(s)$ subsystems with positive real or port-Hamiltonian structure), parameterized reduced models that are uniformly asymptotically stable and built without any parameter sampling. Uniform asymptotic stability is assured even in cases that the parameter domain Ω is an unbounded subset of the non-negative orthant in $\mathbb{R}^k$.

2.3.1 Uniform asymptotic stability in the general case

Theorem 2 guarantees uniform asymptotic stability when the decomposition described in (12) yields a subsystem $\mathcal{H}_3(s)$ that has positive real or port-Hamiltonian structure that is retained in $\widehat{\mathcal{H}}_3(s)$, and then uniform asymptotic stability can be assured even in cases that Ω is unbounded. Unfortunately, the hypothesis that is required on $\mathcal{H}_3(s)$ may either be difficult to determine in practice or it may be that the ensuing structure-preserving reduction may be awkward to implement. Thus, it may be useful to seek more broadly applicable strategies.

We find that if the full parametric model, written in terms of subsystems as in (11), has subsystems that are assumed, as before, to be asymptotically stable, and if model reduction techniques are applied that produce reduced subsystems that retain this asymptotic stability, then uniform asymptotic stability can still be assured but generally only on a subset of the original parameter range, Ω.

Corollary 1 *Suppose that the full parameterized model $\mathcal{H}(s, \mathsf{p})$ described in (11) has been decomposed as in (12) into subsystems $\mathcal{H}_1(s)$, $\mathcal{H}_2(s)$, $\mathcal{H}_3(s)$ and $\mathcal{H}_4(s)$ that are each asymptotically stable. Suppose further that the reduced subsystems $\widehat{\mathcal{H}}_1(s)$, $\widehat{\mathcal{H}}_2(s)$, $\widehat{\mathcal{H}}_3(s)$, and $\widehat{\mathcal{H}}_4(s)$ retain asymptotic stability. Let $M > 0$ be chosen so that $\|\widehat{\mathcal{H}}_3\|_{\mathcal{H}_\infty} < M$ and define:*

$$\Omega_M = \left\{ \mathsf{p} \in \Omega \;\middle|\; \|\mathsf{p}\|_\infty \leq \frac{1}{M} \right\} \subset \Omega.$$

Then the reduced parameterized model $\widehat{\mathcal{H}}(s, \mathsf{p})$ in (16) is uniformly asymptotically stable for $\mathsf{p} \in \Omega_M$.

Proof Following on to the proof of Theorem 2, since the poles of $\widehat{\mathcal{H}}_1(s)$, $\widehat{\mathcal{H}}_2(s)$, and $\widehat{\mathcal{H}}_4(s)$ are independent of p and all have negative real parts, in order to prove uniform

asymptotic stability of $\widehat{\mathcal{H}}(s,\mathsf{p})$ for $\mathsf{p}\in\Omega_M$, we need only assure that the zeroes of $I+D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p})$ lie in the left half plane for $\mathsf{p}\in\Omega_M$. Arguing for a contradiction, note that if the reduced parameterized model $\widehat{\mathcal{H}}(s,\mathsf{p})$ were *not* uniformly asymptotically stable then it must be that at least one zero of $I+D(\mathsf{p}_0)\widehat{\mathcal{H}}_3(s)D(\mathsf{p}_0)$ is *not* in the open left half-plane for some $\mathsf{p}_0\in\Omega_M$, and $I+D(\mathsf{p}_0)\widehat{\mathcal{H}}_3(z_0)D(\mathsf{p}_0)$ will be singular for some z_0 with $\mathsf{Re}(z_0)\geq 0$. In this case, there exists a non-trivial $x_0\in\mathbb{C}^n$ such that $(I+D(\mathsf{p}_0)\widehat{\mathcal{H}}_3(z_0)D(\mathsf{p}_0))x_0=0$, and so, $\|x_0\|=\|D(\mathsf{p}_0)\widehat{\mathcal{H}}_3(z_0)D(\mathsf{p}_0)x_0\|$ and $1\leq\|D(\mathsf{p}_0)\widehat{\mathcal{H}}_3(z_0)D(\mathsf{p}_0)\|$. Now,

$$1\leq\|D(\mathsf{p}_0)\widehat{\mathcal{H}}_3(z_0)D(\mathsf{p}_0)\|_{\mathcal{H}_\infty}\leq\|D(\mathsf{p}_0)\|_2^2\,\|\widehat{\mathcal{H}}_3\|_{\mathcal{H}_\infty}=\|\mathsf{p}_0\|_\infty\|\widehat{\mathcal{H}}_3\|_{\mathcal{H}_\infty}<1,$$

leading to a contradiction. Thus, all zeros of $I+D(\mathsf{p})\widehat{\mathcal{H}}_3(s)D(\mathsf{p}))$ must occur in the open left half-plane, $\mathsf{Re}(s)<0$ for all $\mathsf{p}\in\Omega_M$, and the reduced parametric model $\widehat{\mathcal{H}}(s,\mathsf{p})$ is uniformly asymptotically stable for $\mathsf{p}\in\Omega_M$. □

Remark 4 This corollary is evidently a relaxation of Theorem 2. The flexibility that we gain by allowing arbitrary model reduction approaches to be deployed on the subsystems (so long as asymptotic stability is retained) weakens the conclusions insofar as uniform asymptotic stability for all $\mathsf{p}\in\Omega$ is no longer a priori guaranteed but must be constrained by an a posteriori quantity, $\|\widehat{\mathcal{H}}_3\|_{\mathcal{H}_\infty}$. A priori guarantees of asymptotic stability could be recovered if we have an upper bound to $\|\mathcal{H}_3\|_{\mathcal{H}_\infty}$ and are able to construct $\widehat{\mathcal{H}}_3(s)$ in such a way as to retain this bound. Suppose that $\pi_{\max}$ denotes $\pi_{\max}=\sup_{\mathsf{p}\in\Omega}\|\mathsf{p}\|_\infty$ and assume that it happens that $\mathcal{H}_3(s)$ satisfies $\|\mathcal{H}_3\|_{\mathcal{H}_\infty}<\frac{1}{\pi_{\max}}$. We construct a reduced subsystem for $\widehat{\mathcal{H}}_3(s)$ in such a way that it retains the same bound: $\|\widehat{\mathcal{H}}_3\|_{\mathcal{H}_\infty}<\frac{1}{\pi_{\max}}$. This can be done, e.g., through a norm-preserving reduction technique such as bounded real balancing [53, 55]. We take $M=\frac{1}{\pi_{\max}}$ in Corollary 1 so that $\Omega_M\equiv\Omega$ and then the parameterized reduced model, $\widehat{\mathcal{H}}(s,\mathsf{p})$, in (16) is guaranteed to be uniformly asymptotically stable for all $\mathsf{p}\in\Omega$.

2.4 The parameter mapping approach of Baur et al. [8]

Parameterization given above in (2) appears as a (relatively) low-rank change from the base dynamic matrix, A_0, and it is this structural feature that we have exploited here. Another strategy that exploits this parametric structure has been proposed in [8]. How do the methods compare?

In [8], one augments and modifies the system by introducing a set of k additional synthetic inputs, $\omega(t)$, and outputs, $\eta(t)$, in such a way that the internal system parameterization is mapped to a feedthrough term. The original system response is recovered by constraining the synthetic inputs so as to null the synthetic outputs. The modified system has internal dynamics that is independent of parameters and so can be reduced independently of the parameterization. The final parameterized reduced model is recovered by imposing a constraint on the synthetic inputs analogous to the

original construction; they are chosen so as to null the (reduced) synthetic outputs. To illustrate, define:

$$\begin{aligned} E\dot{x}(t) &= A_0 x + [B \;\; U\, D(\mathsf{p})] \begin{bmatrix} w(t) \\ \omega(t) \end{bmatrix}, \\ \begin{bmatrix} \hat{y}(t) \\ \eta(t) \end{bmatrix} &= \begin{bmatrix} C \\ D(\mathsf{p}) V^T \end{bmatrix} x(t) + \begin{bmatrix} 0 & \\ & I \end{bmatrix} \begin{bmatrix} w(t) \\ \omega(t) \end{bmatrix}. \end{aligned} \tag{24}$$

Evidently, this system has $m + k$ inputs, $\ell + k$ outputs, and the parameterization now acts on a k-dimensional subpace that is common to both input and output spaces. Notice in particular that the parameterization no longer acts directly on the state vector. What relation does the response of (24) have with that of the original system (2)? If $\omega(t)$ is chosen so that $\eta(t) = 0$ (for example, if $\omega(t)$ is assigned by state feedback as $\omega(t) = -D(\mathsf{p}) V^T x(t)$), then the remaining output $\hat{y}(t)$ matches the output of the parameterized system described in (2): $\hat{y}(t) = y(t; \mathsf{p})$. Indeed, with this added constraint imposed on the synthetic inputs, the transfer function for the resulting system is identical to what has been defined by (11).

The dynamical system described in (24) may be reduced using any strategy appropriate for linear time-invariant MIMO (multiple input/multiple output) systems. Since the parameterization has been mapped to the synthetic input/output spaces and is now external to system dynamics, model reduction strategies can be pursued without the need of any parameter sampling. The approach described in [8] proposes a projective reduced model derived, say, as in (4)–(5) using projection bases defined by Z_r and W_r:

$$\begin{aligned} E_r \dot{x}_r(t) &= A_{0r} x_r + [B_r \;\; W_r^T U\, D(\mathsf{p})] \begin{bmatrix} w(t) \\ \hat{\omega}(t) \end{bmatrix}, \\ \begin{bmatrix} \hat{y}_r(t) \\ \eta_r(t) \end{bmatrix} &= \begin{bmatrix} C_r \\ D(\mathsf{p}) V^T Z_r \end{bmatrix} x_r(t) + \begin{bmatrix} 0 & \\ & I \end{bmatrix} \begin{bmatrix} w(t) \\ \hat{\omega}(t) \end{bmatrix}. \end{aligned} \tag{25}$$

Following [8], we set up a state feedback constraint to null the reduced synthetic outputs, similar to what has gone before. If $\hat{\omega}(t)$ is assigned via reduced state feedback as $\hat{\omega}(t) = -D(\mathsf{p}) V^T Z_r x_r(t)$, then $\eta_r(t) = 0$, and the reduced output $\hat{y}_r(t)$ will define the output of a reduced parametric system (2): $y_r(t; \mathsf{p}) = \hat{y}_r(t)$. The state-space representation of the resulting reduced model is given by:

$$W_r^T E Z_r \dot{x}_r(t) = W_r^T A_0 Z_r x_r - W_r^T U D^2(\mathsf{p}) V^T Z_r + W_r^T B w(t), \tag{26}$$

$$\hat{y}_r(t) = C Z_r x_r(t). \tag{27}$$

This dynamical system is identical to (4) and similarly we may rewrite it as in (16):

$$\widehat{\mathcal{H}}(s; \mathsf{p}) = \widehat{\mathcal{H}}_1(s) - \widehat{\mathcal{H}}_2(s) D(\mathsf{p}) (I + D(\mathsf{p}) \widehat{\mathcal{H}}_3(s) D(\mathsf{p}))^{-1} D(\mathsf{p}) \widehat{\mathcal{H}}_4(s).$$

However, in this case, each system, $\mathcal{H}_i(s)$, has been reduced with a single pair of projection bases, Z_r and W_r, for $i = 1, 2, 3, 4$; the *same projection bases* are used for all four systems. So, the framework developed here, summarized in Algorithm 1, contains that of [8] as a special case.

Although the approach we take here is evidently closely aligned with the approach of [8], an important distinction is that we are able here to reduce the individual subsystems, $\mathcal{H}_i(s)$, independently of one another. This allows us to better control the

fidelity of the final model, and as we described in Section 2.3, we are able to guarantee asymptotic stability of $\widehat{\mathcal{H}}(s;\mathsf{p})$ uniformly in p, so long as each reduced model $\widehat{\mathcal{H}}_i(s)$, $i = 1,2,3,4$ is asymptotically stable and the single reduced $\widehat{\mathcal{H}}_3(s)$ is also positive real. Naturally, these assertions may also be made with the approach of [8] or its recent formulation for second-order systems [56], but it can be substantially more difficult to guarantee these properties using a single choice of projecting bases, Z_r and W_r. We are able to exploit the flexibility of reducing the four subsystems independently of one another and do not suffer under these constraints.

It is also worth pointing out here connections to the work [66] where the authors performed sampling-free model reduction of a second-order dynamical system as in (6) where parametrization enters into the matrix D, in the frequency domain, as $D(s,\mathsf{p}) = \hat{U}Z(\mathsf{p},s)\hat{U}^T$ where $\hat{U} \in \mathbb{C}^{n\times k}$ and $Z(s,\mathsf{p}) \in \mathbb{C}^{k\times k}$. Then, [66] performs the model reduction in the second-order form (6) via a block second-order Arnoldi process using interpolatory model reduction (moment matching) around $s = 0$. As in [8], this is achieved by extending the input matrix by the low-rank factor $\hat{U}$ and applying model reduction to the new system. Therefore, similar to [8], one can consider [66] as a method in which all the subsystems are reduced by the same model reduction subspace. Moreover, [66] performs only one-sided model reduction and the output equation does not enter into the reduction step.

2.5 Numerical examples

We illustrate the performance of Algorithm 1 on two numerical examples. BT is used for subsystem reduction in both examples. Since the subsystems have the same E- and A-terms, only a single Schur decomposition (a significantly expensive preliminary step in solving Lyapunov equations for BT) is needed in the offline stage, which occupies step 1 of Algorithm 1. Thus, step 1 is significantly cheaper than applying BT independently to the four subsystems.

Example 1 This is a parametric version of the Penzl model [44, 57]. The full model transfer function $\mathcal{H}(s,\mathsf{p}) = C(sE - A(\mathsf{p}))^{-1}B$ is defined as:

$$E = I \quad \text{and} \quad A = \operatorname{diag}(A(p_1), A(p_2), A(p_3), -1, -2, \ldots, -M),$$
$$\text{where} \quad A(p_i) = \begin{bmatrix} -1 & p_i \\ -p_i & -1 \end{bmatrix}, \quad \text{for} \quad i = 1,\ldots,3;$$
$$C = [c_1\ c_2\ \ldots\ c_{M+6}] \in \mathbb{R}^{1\times(M+6)} \quad \text{where} \quad c_i = \begin{cases} 10, & i = 1,\ldots,6 \\ 1, & i = 7,\ldots,M{+}6; \end{cases}$$

and $B = C^T$. The parameters p_1, p_2, p_3 represent magnitudes of the imaginary parts of the conjugate-paired eigenvalues of $A(p_i)$, for $i = 1,2,3$, and so they will control the location of peaks in the frequency response.

This system can be represented in the structured form of (2) with:

$$A(\mathsf{p}) = A_0 - U\ \operatorname{diag}(p_1, p_1, p_2, p_2, p_3, p_3)V^T,$$

where $A_0 = \text{diag}(-I_6, -1, -2, \ldots, -M)$;

$$u_i = \begin{cases} -e_{i+1}, & i = 1, 3, 5, \\ e_{i-1}, & i = 2, 4, 6. \end{cases} \qquad \text{and} \qquad v_i = e_i \quad \text{for} \quad i = 1, \ldots, 6.$$

e_i denotes here the ith canonical vector. We choose $M = 100$, so $n = 106$. Based on the Hankel singular values of each subsystem, $\mathcal{H}_i(s)$, computed in step 1 of Algorithm 1, we decide on appropriate reduction orders, r_i, for subsystem $\widehat{\mathcal{H}}_i(s)$ and choose $r_1 = 10$, $r_2 = 1$, $r_3 = 6$, and $r_4 = 1$. We are exploiting an advantage of our framework in choosing reduction orders for the subsystems independently of one another.

In Fig. 1, we compare $\mathcal{H}(s, \mathsf{p}_\star)$ and $\widehat{\mathcal{H}}(s, \mathsf{p}_\star)$ for s on the imaginary axis at an arbitrarily chosen parameter configuration:

$$\mathsf{p}_\star = (p_1, p_2, p_3) = (10, 100, 5000).$$

Note that the reduced model has been obtained without any parameter sampling. The top plot in Fig. 1 shows $|\ \mathcal{H}(\imath\omega, \mathsf{p}_\star)\ |$ and $|\ \widehat{\mathcal{H}}(\imath\omega, \mathsf{p}_\star)\ |$ for $10^{-2} \le \omega \le 10^4$—visually, this is nearly an exact match. The bottom plot of Fig. 1 shows the relative error in the approximation at $\mathsf{p}_\star$, $\dfrac{|\ \mathcal{H}(\imath\omega, \mathsf{p}_\star) - \widehat{\mathcal{H}}(\imath\omega, \mathsf{p}_\star)\ |}{\max_\omega |\ \mathcal{H}(\imath\omega, \mathsf{p}_\star)\ |}$, across the same frequency range. The quality of the approximation at $\mathsf{p}_\star$ is evidently quite good, with relative errors uniformly smaller than 2×10^{-6}.

In order to illustrate the quality of approximation as the parameter configurations change as well, we present in Fig. 2a surface plot that shows the magnitude of $|$

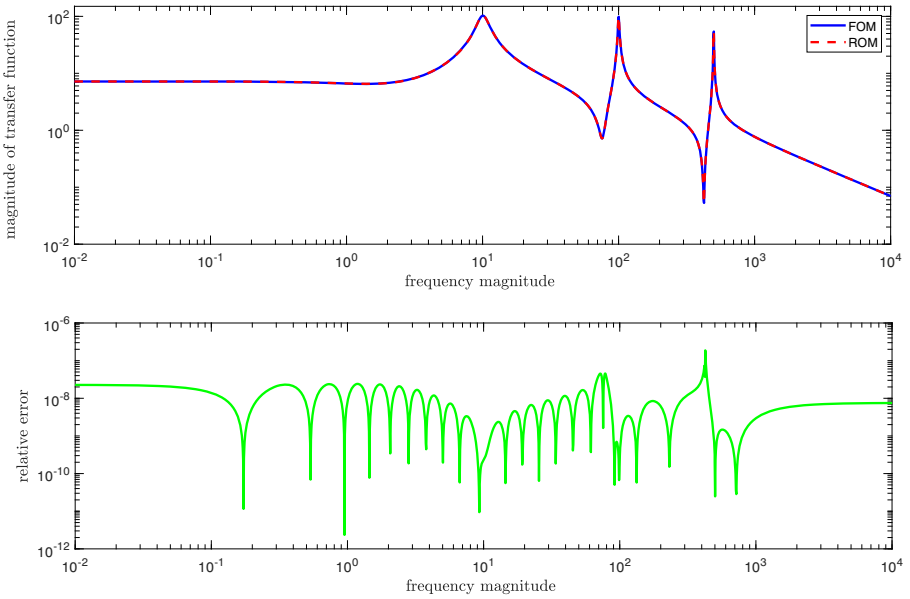


Fig. 1 Transfer function plot and relative error for $(p_1, p_2, p_3) = (10, 100, 5000)$

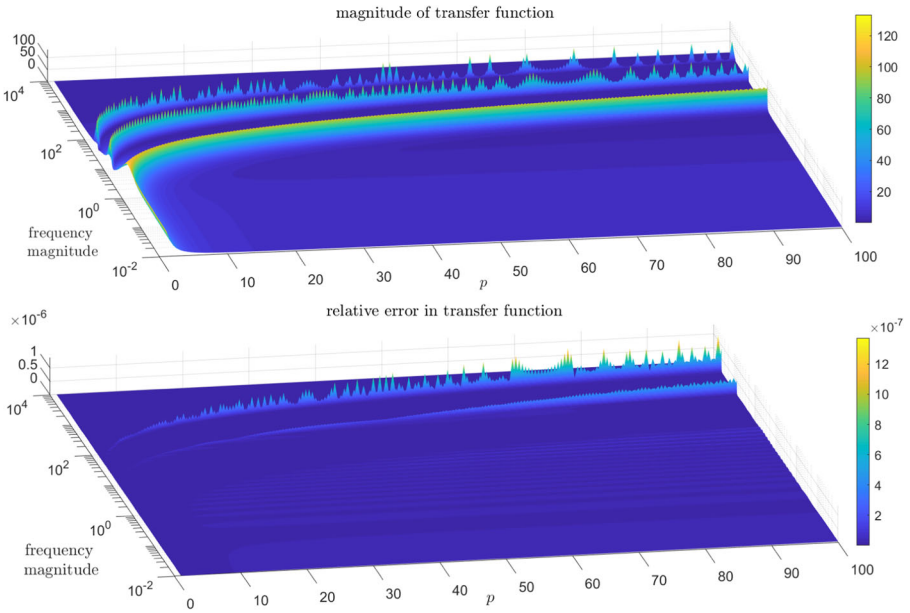


Fig. 2 Magnitude of transfer function and relative error for parameters given by (28)

$\mathcal{H}(\imath\omega, \mathsf{p})$ | (top) and relative error (bottom) as in Fig. 1 but now allowing variation of p_1, p_2, p_3 along a ray in parameter space:

$$\mathsf{p} = (p_1, p_2, p_3) = (p, 10\,p, 50\,p), \quad \text{with} \quad p \in [1, 100]. \tag{28}$$

As mentioned earlier, the parameters control the imaginary part of the complex poles and different parameter selections will move these peaks in the frequency domain. The top plot in Fig. 2 shows how the peaks of $|\ \mathcal{H}(\imath\omega, \mathsf{p})\ |$ are moving with the parameter p. The lower subplot shows magnitude of relative errors on the same range of frequency and parameter variation and demonstrates that the reduced model is accurate across the parameter domain (28), with the largest relative error remaining less than 10^{-6}. Note this accuracy was achieved with four non-parametric model reductions without any parameter sampling and the reduced model $\widehat{\mathcal{H}}(s, \mathsf{p})$ is asymptotically stable for every parameter sample.

Example 2 This is a model drawn from the Oberwolfach Benchmark Collection representing thermal conduction in a semiconductor chip [54]. The full model is described by:

$$\begin{aligned} E\dot{x} &= (A - p_t A_t - p_b A_b - p_s A_s)x + Bu, \\ y &= Cx, \end{aligned}$$

where $E,\ A \in \mathbb{R}^{4257\times 4257}$ represent respectively the distributed heat capacity and conductivity in the chip; the matrices A_t, A_b, $A_s \in \mathbb{R}^{4257\times 4257}$ are diagonal matrices resulting from the discretization of convection boundary conditions and have ranks 111, 99, and 31, respectively. The matrix $B \in \mathbb{R}^{1\times 4257}$ is a load profile vector and

$C \in \mathbb{R}^{7\times 4257}$ is an output matrix. The parameters p_t, p_b, p_s represent film interface coefficients and can be viewed as design parameters. For further details on this model, refer to [30, 54].

We fix $p_t = 1000$, assign the remaining parameters to $\mathsf{p} = [p_b \ p_s]$, allowing p_b and p_s to vary between 1 and 10^9. This model is rewritten in accord with our parameterized format (2) taking $A_0 = A - p_t A_t$, with $p_t = 1000$, while U and V are matrices with $\text{rank}(U) = \text{rank}(V) = \text{rank}(A_b) + \text{rank}(A_s) = 130$.

Based on Hankel singular values obtained in step 1 of Algorithm 1, we reduce each subsystem $\mathcal{H}_i(s)$ to $\widehat{\mathcal{H}}_i(s)$ via BT to a reduction order of r_i, choosing $r_1 = 46, r_2 = 66, r_3 = 200$, and $r_4 = 16$. Note that the reduction order r_3 is significantly bigger than the others; this may be anticipated since $\mathcal{H}_3(s)$ in this case represents a dynamical system with a relatively large input/output port dimension: 130 inputs and 130 outputs. In Fig. 3, we illustrate the quality of the approximation obtained by Algorithm 1 over the full parameter domain using the $\mathcal{H}_2$ system norm, $\|H(\cdot,\mathsf{p})\|_{\mathcal{H}_2} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} \|\mathcal{H}(\imath\omega,\mathsf{p})\|_F^2 \, d\omega}$ (system norms are discussed in Section 4). In Fig. 3, the horizontal axes represent parameter variation in p_s and p_t, while the vertical axis represents the relative error $\frac{\|\mathcal{H}(\cdot;\mathsf{p})-\widehat{\mathcal{H}}(\cdot;\mathsf{p})\|_{\mathcal{H}_2}}{\|\widehat{\mathcal{H}}(\cdot;\mathsf{p})\|_{\mathcal{H}_2}}$ for the reduced system $\widehat{\mathcal{H}}(s,\mathsf{p})$ as calculated by Algorithm 1. It is clear from the figure that $\widehat{\mathcal{H}}(s,\mathsf{p})$ accurately represents $\mathcal{H}(s,\mathsf{p})$ across the full parameter domain, with relative error uniformly smaller than 2.4×10^{-6} for all values of $(p_s, p_b) \in [1, 10^9]\times[1, 10^9]$. This high-fidelity approximation is obtained at the cost of four non-parametric subsystem model reduction *without any parameter sampling*. As in the previous example, $\widehat{\mathcal{H}}(s,\mathsf{p})$ is asymptotically stable throughout the parameter range.

3 Data-driven PMOR with subsystem frequency sampling

In Section 2, we proposed a sampling-free parametric model reduction approach that involved the preliminary reduction of four non-parametric models. In this section, we present a second approach that depends on the same decomposition into four subsystems yet uses instead a data-driven framework based on transfer function (frequency-domain) sampling to construct parametric reduced models in the offline stage. We first briefly review data-driven modeling frameworks and then present our approach to data-driven PMOR.

3.1 Data-driven modeling from transfer function samples

Let $\mathcal{H}(s)$ denote the transfer function of a (non-parametric) linear dynamical system. $\mathcal{H}(s)$ need not be a rational function of s and can contain, for example, internal delays or other non-rational dependence in s. Assume that we have access to samples of this transfer function, i.e., we have $\mathcal{H}(\xi_1), \mathcal{H}(\xi_2), \ldots, \mathcal{H}(\xi_N)$ where $\xi_i \in \mathbb{C}$ for $i = 1, 2, \ldots, N$ are the sampling points. When obtained experimentally, these sampling points are chosen on the imaginary axis. If an analytical evaluation of $\mathcal{H}(s)$ is possible, they can be chosen arbitrarily as long as they do not coincide with the poles

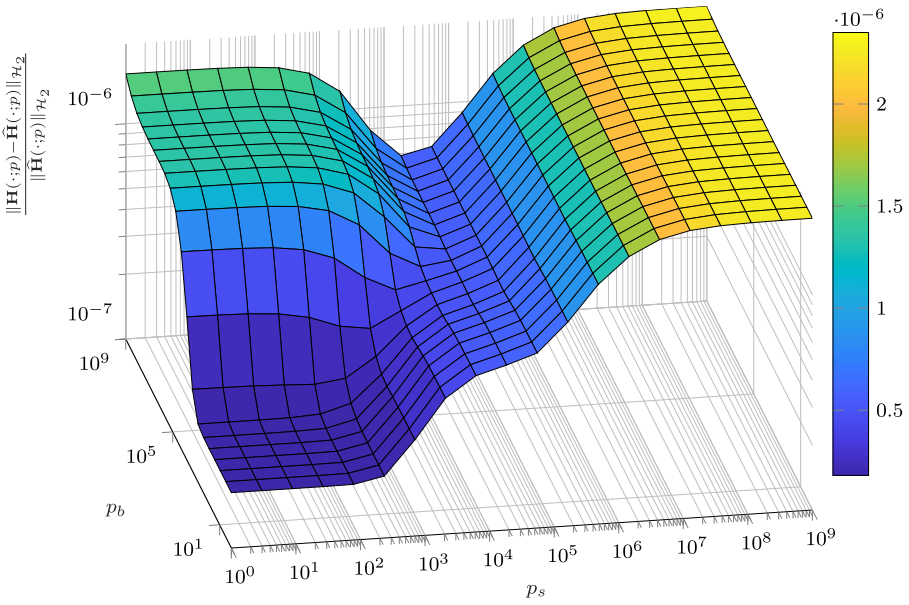


Fig. 3 Relative error for different parameters for thermal model

of $\mathcal{H}(s)$. In our numerical experiments, we will work with samples on the imaginary axis but the theoretical discussion applies to the general case.

Data-driven modeling in this case amounts to the following question: Given the samples $\{\mathcal{H}(\xi_i)\}_{i=1}^N$ (but without access to internal dynamics of $\mathcal{H}(s)$, i.e., without access to a state-space realization), construct a rational approximation of degree-r, $\widehat{\mathcal{H}}(s)$, that fits the data in an appropriate sense. There are various ways to fit this frequency domain data. One can require that $\widehat{\mathcal{H}}(s)$ interpolates the data at every sampling point using the Loewner framework [5, 47]; one can construct $\widehat{\mathcal{H}}(s)$ so as to fit the data in a least-squares (LS) sense [20, 26, 40]; or finally, one can force $\widehat{\mathcal{H}}(s)$ to interpolate some of the data while minimizing the LS fit in the rest [23, 51]. In this work, we focus solely on fitting data in a LS sense.

Thus, given samples $\{\mathcal{H}(\xi_i)\}_{i=1}^N$, our goal is to construct a degree-r rational function $\widehat{\mathcal{H}}(s)$, i.e., a reduced transfer function, that minimizes the LS error $\sum_{i=1}^{N} \|\mathcal{H}(\xi_i) - \widehat{\mathcal{H}}(\xi_i)\|_F^2$. Note that due to the non-linear dependence on the poles of $\widehat{\mathcal{H}}(s)$, this is a non-linear LS problem. There are various approaches for solving this problem, see, e.g., [20, 26, 33, 40, 42, 51, 61]. Our approach employs the Vector Fitting (VF) framework of [40], though one could easily adapt any of the other LS methods. We view VF as a known tool to be employed (as we did with BT and IRKA in the subsystem reduction discussion of Section 2) and therefore we do not provide detailed background information. We do note that VF uses a barycentric-form for $\widehat{\mathcal{H}}(s)$ (as opposed to a state-space realization) and converts the non-linear LS problem into a sequence of weighted linear LS problems each of which could be solved

easily by well-established numerical linear algebra tools at every step. The variables determined in each step are coefficients of the barycentric form. Once the iteration is terminated, a state-space form is recovered. For details, we refer the reader to [24, 27, 39, 40], [35, Chap 7] and the references therein.

Regarding computational cost, VF performs $m \cdot \ell$ QR factorizations of size $N \times (2r)$ in every step [27]. When m and ℓ are modest, say $m, \ell < 10$, the computational effort is not significant. The cost increases as m and ℓ grow; however, there are various ways to speed up the process such as performing the $m \cdot \ell$ QR factorizations in parallel [24, 39] as they are independent of each other. We have not needed sophisticated refinements such as these; in our studies, a basic implementation proved to be relatively efficient. Assume that the underlying system is a rational function itself, i.e., $\mathcal{H}(s) = C(sE - A)^{-1}B$. Then, obtaining the samples, $\{\mathcal{H}(\xi_i)\}_{i=1}^N$, requires solving N linear systems of size $n \times n$ with multiple right-hand size. This is a much larger cost relative to all the remaining steps of VF. Indeed, the main cost of VF is the initial sampling step itself.

3.2 pMOR from offline samples

We consider now how to integrate the subsystem structure developed in Proposition 1 into VF for parameterized problems. As we noted above, the main cost in VF generally comes from computing the transfer function samples at selected frequencies. We want to avoid re-sampling $\mathcal{H}(s, \mathsf{p})$ from scratch for every new p considered.

Recall (12), repeated here:

$$\mathcal{H}(s; \mathsf{p}) = \mathcal{H}_1(s) - \mathcal{H}_2(s)D(\mathsf{p})\,[I + D(\mathsf{p})\mathcal{H}_3(s)D(\mathsf{p})]^{-1}\,D(\mathsf{p})\mathcal{H}_4(s).$$

Given predetermined points $\{\xi_1, \ldots, \xi_N\} \subset \mathbb{C}$, compute subsystem samples:

$$\mathcal{H}_1(\xi_i),\ \mathcal{H}_2(\xi_i),\ \mathcal{H}_3(\xi_i),\ \mathcal{H}_4(\xi_i) \quad \text{for} \quad i = 1, \ldots, N,$$

and note that these computed values do not depend on any p values. In particular, we need only perform this sampling once in a preliminary Offline Stage. Moreover, all four subsystem transfer functions share the same resolvent $(sE - A_0)^{-1}$ and so the evaluation of $\mathcal{H}_j(\xi_i)$ for $j = 1, \ldots, n$, can exploit this fact, significantly reducing the overall numerical cost of the step.

Subsequently, for any parameter, p, we can efficiently calculate the values $\mathcal{H}(\xi_i; \mathsf{p})$ using (12) as

$$\mathcal{H}(\xi_i; \mathsf{p}) = \mathcal{H}_1(\xi_i) - \mathcal{H}_2(\xi_i)D(\mathsf{p})(I + D(\mathsf{p})\mathcal{H}_3(\xi_i)D(\mathsf{p}))^{-1}D(\mathsf{p})\mathcal{H}_4(\xi_i), \quad (29)$$

for $i = 1, \ldots, N$. This step comes essentially at no cost and we can resample $\mathcal{H}(\xi_i; \mathsf{p})$ for any p with almost no effort. This allows us to employ a data-driven approach such as VF to construct cheaply a reduced model at any desired parameter value. This is summarized as Algorithm 2.

To determine the quality of approximations from Algorithm 2, we consider the discrete LS error examined through a relative error measure:

$$e(\mathcal{H}(\cdot; \mathsf{p})), \widehat{\mathcal{H}}(\cdot; \mathsf{p}))) = \sum_{i=1}^{N} \left\| \mathcal{H}(\xi_i; \mathsf{p}) - \widehat{\mathcal{H}}(\xi_i; \mathsf{p}) \right\|_F^2 \Big/ \sum_{i=1}^{N} \|\mathcal{H}(\xi_i; \mathsf{p})\|_F^2. \quad (30)$$

Algorithm 2 Parametric reduced order model based on VF.

1: Offline Stage: For predetermined points in the complex plane, $\{\xi_1, \ldots, \xi_N\}$, calculate

$$\mathcal{H}_1(\xi_i),\ \mathcal{H}_2(\xi_i),\ \mathcal{H}_3(\xi_i),\ \mathcal{H}_4(\xi_i) \quad \text{for} \quad i = 1, \ldots, N, \quad \text{using (14).}$$

2: Online Stage:

For any given parameter p calculate $\mathcal{H}(\xi_i; \mathsf{p})$ for $i = 1, \ldots, N$ using formula (29).

3: Based on $\mathcal{H}(\xi_1; \mathsf{p}), \ldots, \mathcal{H}(\xi_N; \mathsf{p})$ obtain reduced system $\widehat{\mathcal{H}}(s; \mathsf{p})$ using VF.

As was the case for Algorithm 1, our sampling-based Algorithm 2 is well suited for computationally efficient parameter optimization and the study of associated system properties. In Algorithm 2, step 1 is executed only once in the Offline Stage. Then, each time the parameter p is varied, (in the Online Stage), steps 2–3 can be executed efficiently. In the next section, we discuss how one can use Algorithms 1 and 2, informed by error estimates given by (18) and (30), in order to ensure robust and accurate parameter optimization.

Remark 5 State-space representation of $\widehat{\mathcal{H}}(s; \mathsf{p})$. Recall that parameterized reduced models obtained from Algorithm 1 will have the factored form given in (16) involving the individually reduced subsystems $\widehat{\mathcal{H}}_i(s)$ for $i = 1, \ldots, 4$. While an explicit system realization will be available for each of these reduced subsystems, an explicit system realization for the aggregate system, $\widehat{\mathcal{H}}(s; \mathsf{p})$, as described in (16) is not immediate and in some contexts, it may be useful to have such a realization. Using basic manipulations for the addition, multiplication, and inversion of transfer functions (see for example §3.6 of [67]), one may arrive at such a realization for $\widehat{\mathcal{H}}(s; \mathsf{p})$. These expressions rapidly become quite complicated and uninformative so we do not include them here, though such realizations are eminently feasible.

In contrast to Algorithm 1, Algorithm 2 will have available only frequency domain samples of the subsystems $\mathcal{H}_i(s)$ for $i = 1, \ldots, 4$ and so an explicit *parametrized* state-space realization will not generally be feasible. However, for any particular $\mathsf{p}_\star$ selected in the Online Stage, the output of VF in step 3 of Algorithm 2 will be directly a state-space realization of $\widehat{\mathcal{H}}(s; \mathsf{p}_\star)$.

4 Parameter optimization for systems with low-rank parameterization

In this section, we consider algorithms for parameter optimization that involve dynamic constraints reflecting low-rank structure as in (2). We will incorporate the proposed sampling-free parametric model reduction techniques of Sections 2 and 3 to allow for efficient surrogate optimization.

Parameter optimization plays a vital role in many applications. In the case of damping optimization, this is a computationally demanding problem even for moderate dimensions, principally due to the combinatorial character of the task of

optimizing damping parameters (viscosities) together with their positions within the structure. Optimization of damping parameters using criteria based on system norms was studied, e.g., in [15, 21, 52, 62]. We present here effective algorithms for parameter optimization in structured systems and then apply them in numerical experiments that illustrate efficient optimization of damping parameters. The use of model reduction for optimization in more general settings is discussed in, e.g., [1, 2, 7, 17, 22, 41, 45, 65] and the references therein.

4.1 The choice of cost function for damping optimization

For a fixed but otherwise arbitrary $\delta > 0$, consider the ODE-constrained optimization problem:

$$\begin{aligned} \mathsf{p}^\star &= \arg\min_{\mathsf{p}\in\Omega} \|y(\cdot,\mathsf{p})\| \\ \text{subject to} \quad E\dot{x}(t;\mathsf{p}) &= A(\mathsf{p})x(t;\mathsf{p}) + Bw(t), \\ y(t;\mathsf{p}) &= Cx(t;\mathsf{p}), \quad x(0;\mathsf{p}) = 0, \\ \text{and} \quad \|\, w\|_{L_2} &\le \delta. \end{aligned} \tag{31}$$

There are many viable choices for selecting the norm for $\|y(\cdot,\mathsf{p})\|$ and the algorithms we describe below will apply to various scenarios. For the damping optimization problem of Section 1.2, there will be a natural choice of norm that is discussed below.

Recall that in the damping optimization setting, the input w represents an input disturbance with energy bounded by δ and the goal is to minimize the influence of w on the output y. Therefore, one might choose to minimize $\|y(\cdot,\mathsf{p})\|_{L_\infty} := \sup_{t\ge 0}\|y(\cdot,\mathsf{p})\|_\infty$ or $\|y(\cdot,\mathsf{p})\|_{L_2} := \sqrt{\int_0^\infty \|y(\cdot,\mathsf{p})\|_2^2 dt}$. These norms can be equivalently represented using the transfer function $\mathcal{H}(s,\mathsf{p})$. The corresponding frequency-domain norms are the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms:

$$\|\mathcal{H}(\cdot,\mathsf{p})\|_{\mathcal{H}_2} := \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} \|\mathcal{H}(\imath\omega,\mathsf{p})\|_F^2 \, d\omega} \quad \text{and} \quad \|\mathcal{H}(\cdot,\mathsf{p})\|_{\mathcal{H}_\infty} := \sup_{\omega\in\mathbb{R}} \|\mathcal{H}(\imath\omega,\mathsf{p})\|_2\,, \tag{32}$$

where $\imath^2 = -1$ and $\|\cdot\|_F$ denotes the Frobenius norm. For a stable linear (parametric) dynamical systems with an input $w(t)$ having $\|w\|_{L_2} \le \infty$ and the corresponding output $y(t;\mathsf{p})$, it holds:

$$\|y(\cdot,\mathsf{p})\|_{L_\infty} \le \|\mathcal{H}(\cdot,\mathsf{p})\|_{\mathcal{H}_2}\|w\|_{L_2} \quad \text{and} \quad \|y(\cdot,\mathsf{p})\|_{L_2} \le \|\mathcal{H}(\cdot,\mathsf{p})\|_{\mathcal{H}_\infty}\|w\|_{L_2}.$$

If we choose to minimize $\|y(\cdot,\mathsf{p})\|_{L_\infty}$ in (31), then the resulting computational problem can be relaxed to:

$$\mathsf{p}^\star = \arg\min_{\mathsf{p}\in\Omega} \|\mathcal{H}(\cdot,\mathsf{p})\|_{\mathcal{H}_2} \quad \text{where} \quad \mathcal{H}(s,\mathsf{p}) = C\left(sE - \left(A_0 - U\,D^2(\mathsf{p})V^T\right)\right)^{-1} B. \tag{33}$$

For the special cases of a scalar-valued input or a scalar-valued output, this reformulation is exact since the the $\mathcal{H}_2$ norm in these cases is precisely the $L_2 - L_\infty$ induced

norm of the underlying convolution operator. In the discussion below, we present the analysis and algorithms for the parameter optimization problem (33) in the general multi-input/multi-output setting.

4.2 Surrogate optimization with reduced parametric models

A major cost in solving (33) is the repeated evaluation of the $\mathcal{H}_2$ norm. For $\mathcal{H}(s;\mathsf{p}) = C(sE - A(p))^{-1}B$, the $\mathcal{H}_2$ norm is computed by solving a Lyapunov equation:

$$\|\mathcal{H}(\cdot,\mathsf{p})\|_{\mathcal{H}_2} = \sqrt{\mathsf{trace}(CPC^T)}$$
$$\text{where } P \text{ solves } \; A(\mathsf{p})PE^T + EPA(\mathsf{p})^T + BB^T = 0.$$

Solving a large-scale Lyapunov equation is computationally demanding and in this optimization setting, one has to repeat this task for many different p values. We will use the parametric reduced models from Algorithms 1 and 2 to relieve this computational burden, so that instead of (33), we solve a surrogate optimization problem:

$$\hat{\mathsf{p}}^\star = \arg\min_{\mathsf{p}\in\Omega} \left\|\widehat{\mathcal{H}}(\cdot,\mathsf{p})\right\|_{\mathcal{H}_2}, \tag{34}$$

where the reduced parametric transfer function $\widehat{\mathcal{H}}(\cdot,\mathsf{p})$ will be constructed using either Algorithms 1 or 2, without need for parameter sampling.

Assume $\mathsf{p}^\star$ is the minimizer of (33) and note that:

$$\left\|\mathcal{H}(\cdot,\mathsf{p}^\star)\right\|_{\mathcal{H}_2} \le \left\|\mathcal{H}(\cdot,\mathsf{p}^\star) - \widehat{\mathcal{H}}(\cdot,\mathsf{p}^\star)\right\|_{\mathcal{H}_2} + \left\|\widehat{\mathcal{H}}(\cdot,\mathsf{p}^\star)\right\|_{\mathcal{H}_2}. \tag{35}$$

The surrogate optimization problem (34) will minimize the second term in (35). Evidently, we will want the reduced model $\widehat{\mathcal{H}}(s,\mathsf{p})$ to be a sufficiently accurate approximation at the minimizer $\hat{\mathsf{p}}^\star$ so that the first term in (35) is comparatively insignificant.

4.2.1 Surrogate optimization with reduced model via Algorithm 1

To guarantee that $\widehat{\mathcal{H}}(s,\mathsf{p})$ is accurate enough at the optimizer $\hat{\mathsf{p}}^\star$, we need to evaluate the term $\left\|\mathcal{H}(\cdot,\hat{\mathsf{p}}^\star) - \widehat{\mathcal{H}}(\cdot,\hat{\mathsf{p}}^\star)\right\|_{\mathcal{H}_2}$. Therefore, an efficient evaluation (estimation) of this term during optimization is crucial for a numerically effective implementation.

When Algorithm 1 is employed to construct the reduced model $\widehat{\mathcal{H}}(s,p)$, Theorem 1 shows how $\left\|\mathcal{H}(\cdot,\hat{\mathsf{p}}^\star) - \widehat{\mathcal{H}}(\cdot,\hat{\mathsf{p}}^\star)\right\|_{\mathcal{H}_2}$ can be bounded using the subsystem errors $\epsilon_i = \|\mathcal{H}_i(\cdot) - \widehat{\mathcal{H}}_i(\cdot)\|$, for $i = 1,2,3,4$. Unfortunately, two of the terms in (18) depend on the full order quantities $\mathcal{H}_1(s)$ and $\mathcal{H}_2(s)$. Therefore, in our surrogate optimization routine, we will use an approximation to this upper bound. Assuming

$\widehat{\mathcal{H}}_2(s)$ and $\widehat{\mathcal{H}}_3(s)$ are accurate approximations to $\mathcal{H}_2(s)$ and $\mathcal{H}_3(s)$, i.e., $\mathcal{H}_2 \approx \widehat{\mathcal{H}}_2$, and $\mathcal{H}_3 \approx \widehat{\mathcal{H}}_3$(note that we can control this accuracy in the model reduction stage), we will approximate the upper bound using Theorem 1 as:

$$\begin{aligned} \|\mathcal{H}(\cdot;\mathsf{p})) - \widehat{\mathcal{H}}(\cdot;\mathsf{p}))\| \lesssim\ & \epsilon_1 + \epsilon_2\, f_1(\mathsf{p}, \widehat{\mathcal{H}}_3, \widehat{\mathcal{H}}_4) \\ & + \epsilon_3\, f_1(\mathsf{p}, \widehat{\mathcal{H}}_3, \widehat{\mathcal{H}}_4)\, f_2(\mathsf{p}, \widehat{\mathcal{H}}_2, \widehat{\mathcal{H}}_3) \\ & + \epsilon_4\, f_2(\mathsf{p}, \widehat{\mathcal{H}}_2, \widehat{\mathcal{H}}_3) \stackrel{\text{def}}{=} f(\mathsf{p}), \end{aligned} \tag{36}$$

where $\epsilon_i = \|\mathcal{H}_i(\cdot) - \widehat{\mathcal{H}}_i(\cdot)\|$ for $i = 1, \ldots, 4$ and f_1 and f_2 are given by (19) and (20), respectively. To simplify notation, we have denoted this upper bound estimate, (i.e., the right hand-side of (36)) as $f(\mathsf{p})$.

Now, using $f(\mathsf{p})$, we can efficiently estimate the accuracy of the reduced model at a given parameter value p. In Algorithm 3, we give an outline of a surrogate optimization method using this estimate. Starting with initial reduced subsystems in step 1 (constructed for a given accuracy), Algorithm 3 solves the surrogate optimization problem in step 2. Then, step 3 checks whether the reduced model is accurate enough at the current optimizer using the estimate $f(p)$ for the upper bound. If it is, the algorithm terminates. Otherwise, step 5 adaptively increases the reduced dimension and step 6 resolves the surrogate optimization for the updated reduced model. This procedure is repeated until a desired tolerance is met.

There are various algorithmic details that help speed up the computations. Instead of wading into those details, we will highlight some points. For example, assume that subsystem model reduction in Algorithm 1 is performed using BT. Then, to increase the reduced dimensions in step 5, one does not need to apply model reduction from scratch. In the case of BT, one will only need to add more vectors to the BT-based model reduction bases from already computed quantities. If IRKA is employed in Algorithm 1, then the current reduced-model poles, appended with some others, will be an effective initialization strategy for IRKA, yielding faster convergence.

4.2.2 Surrogate optimization with reduced model via Algorithm 2

We now focus on solving the optimization problem (33) using reduced models from the data-driven Algorithm 2. One major difference from Algorithm 1 is that in this case, there are no reduced subsystems. Instead, for any given p, we have a numerically efficient way to find an accurate approximation $\widehat{\mathcal{H}}(s, \mathsf{p})$ to $\mathcal{H}(s, \mathsf{p})$. Therefore, the subsystem-based error estimate in (36) does not apply here. However, we have the sample-based (discretized) version $e(\mathsf{p})$ defined in (30). For example, if VF is employed in Algorithm 2, the error $e(\mathsf{p})$ will be automatically calculated during the construction of $\widehat{\mathcal{H}}(s, \mathsf{p})$ and thus no additional effort is needed to compute $e(\mathsf{p})$. Therefore, following similar arguments to those found in Section 4.2.1, when solving the surrogate optimization problem (34), we need to ensure that the reduced model $\widehat{\mathcal{H}}(s, \mathsf{p})$ is an accurate approximation to $\mathcal{H}(s, \mathsf{p})$ at the optimizer $\mathsf{p} = \hat{\mathsf{p}}^\star$ where accuracy is now measured using $e(p)$.

Algorithm 3 Surrogate parameter optimization using reduced models via Algorithm 1.

Require: System matrices $A(\mathsf{p}), E, B, C$ defining (1);
initial point p^0 for optimization routine
tolerance $0 < \tau \ll 1$ for error bound
tolerance $0 < \nu \ll 1$ for optimization routine
starting reduced dimensions r_1, r_2, r_3, r_4 and the corresponding subsystems $\widehat{\mathcal{H}}_1$, $\widehat{\mathcal{H}}_2$, $\widehat{\mathcal{H}}_3$, and $\widehat{\mathcal{H}}_4$.

Ensure: approximation of optimal parameters

1: Choose the reduced orders r_1, r_2, r_3, r_4 (and thus the reduced subsystems $\widehat{\mathcal{H}}_1(s)$, $\widehat{\mathcal{H}}_2(s)$, $\widehat{\mathcal{H}}_3(s)$, and $\widehat{\mathcal{H}}_4(s)$ via Algorithm 1) so that $f(\mathsf{p}^0) < \tau \, \|\widehat{\mathcal{H}}(\cdot, \mathsf{p}^0)\|_{\mathcal{H}_2}$.
2: Solve the surrogate optimization problem

$$\hat{\mathsf{p}}^\star = \arg\min_{\mathsf{p}} \left\| \widehat{\mathcal{H}}(\cdot, \mathsf{p}) \right\|_{\mathcal{H}_2}$$

with the initial guess p^0 and tolerance ν.
3: **while** minimizer $\mathsf{p}^\star$ such that $f(\mathsf{p}^\star) > \tau \, \|\widehat{\mathcal{H}}(\cdot, \mathsf{p}^0)\|_{\mathcal{H}_2}$ **do**
4: $\mathsf{p}^0 = \hat{\mathsf{p}}^\star$
5: Increase the reduced orders r_1, r_2, r_3, r_4 (and thus the reduced subsystems via Algorithm 1) so that $f(\hat{\mathsf{p}}^\star) < \tau \, \|\widehat{\mathcal{H}}(\cdot, \mathsf{p}^0)\|_{\mathcal{H}_2}$.
6: Determine the new minimizer by solving the surrogate optimization problem

$$\hat{\mathsf{p}}^\star = \arg\min_{\mathsf{p}} \left\| \widehat{\mathcal{H}}(\cdot, \mathsf{p}) \right\|_{\mathcal{H}_2}$$

using the updated $\widehat{\mathcal{H}}$, the initial guess p^0, and tolerance ν.
7: **end while**

The resulting approach is briefly discussed in Algorithm 4. The fundamental structure is almost identical to that of Algorithm 3. We test whether $e(\hat{\mathsf{p}}^\star)$ is below a prespecified tolerance. If not, we increase the order of the reduced model in Algorithm 2 until we reach the desired accuracy.

We note that during the optimization procedure, for every new p, Algorithm 4 resamples $\widehat{\mathcal{H}}(s, \mathsf{p})$ (at practically no cost) and constructs a new reduced model based on these new samples. In analogy to Algorithm 3, it is as if the model reduction subspaces are changing with every p. Therefore, even though we cannot theoretically guarantee it, we expect that Algorithm 4 may produce smaller errors than Algorithm 3. This will be studied in Section 4.3 via numerical experiments.

As is the case with Algorithm 3, there are various numerical aspects that one could exploit to make the online computations faster. For example, one main factor determining the convergence speed of VF is the initial selection of poles. In damping optimization, the poles from the critical damping case are perfect candidates. If in step 5 of Algorithm 4, one must increase the order of the model, the already-converged poles from the previous optimization step (appended with a small number of additional ones) can be expected to speed up convergence. We will elaborate on these points in the numerical example below.

Algorithm 4 Surrogate parameter optimization using reduced models via Algorithm 2.

Require: System matrices $A(\mathsf{p}), E, B, C$ defining (1);
initial point p^0 for optimization routine
tolerance $0 < \tau \ll 1$ for error bound
tolerance $0 < \nu \ll 1$ for optimization routine
Samples $\{\mathcal{H}_i(\xi_i)\}_{i=1}^N$ at predetermined points in the complex plane $\xi_1, \ldots, \xi_N$.
Ensure: approximation of optimal parameters

1: Choose the reduced order in Algorithm 2 so that $e(\mathsf{p}^0) < \tau$.
2: Solve the surrogate optimization problem

$$\hat{\mathsf{p}}^\star = \arg\min_{\mathsf{p}} \left\| \widehat{\mathcal{H}}(\cdot, \mathsf{p}) \right\|_{\mathcal{H}_2}$$

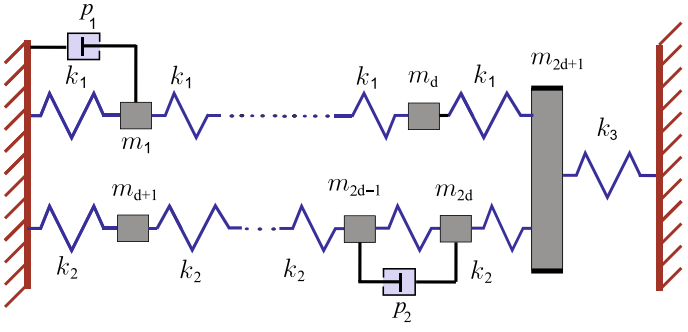
with the initial guess p^0 and tolerance ν with $\widehat{\mathcal{H}}_{\mathsf{p}}$ computed via Algorithm 2 using samples $\{\mathcal{H}_i(\xi_i)\}_{i=1}^N$.
3: **while** minimizer $\mathsf{p}^\star$ such that $e(\mathsf{p}^\star) > \tau$ **do**
4: $\mathsf{p}^0 = \hat{\mathsf{p}}^\star$
5: Increase the reduced order used in Algorithm 2 so that $e(\mathsf{p}^\star) < \tau$.
6: Determine the new minimizer by solving the surrogate optimization problem

$$\hat{\mathsf{p}}^\star = \arg\min_{\mathsf{p}} \left\| \widehat{\mathcal{H}}(\cdot, \mathsf{p}) \right\|_{\mathcal{H}_2}$$

using the updated $\widehat{\mathcal{H}}$, the initial guess p^0, and tolerance ν.
7: **end while**

4.3 Numerical example

We revisit the damping optimization problem described in Section 1.2 and focus on optimizing the viscosities for different sets of damping positions. We consider an n-mass oscillator with $2d + 1$ masses and $2d + 3$ springs as shown in the figure in Section 4.3. This oscillator has two rows of d masses connected with springs. The leading masses in each row on the left edge are connected to a fixed boundary while on the opposite (right) edge the masses (m_d and m_{2d}) are connected to a single mass m_{2d+1}, which, in turn, is connected to a fixed boundary. See [15, 62] for further details.



The state-space model is given by (6) where the stiffness matrix is given by

$$K = \begin{bmatrix} K_{11} & & -\kappa_1 \\ & K_{22} & -\kappa_2 \\ -\kappa_1^T & -\kappa_2^T & k_1+k_2+k_3 \end{bmatrix}, \quad K_{ii} = k_i \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix},$$

with $\kappa_i = \begin{bmatrix} 0 & \dots & 0 & k_i \end{bmatrix}$ for $i = 1$ and $i = 2$, and the mass matrix is $M = \operatorname{diag}(m_1, m_2, \dots, m_{2d+1})$.

We pick 1801 masses ($d = 900$) with the values:

$$m_i = \begin{cases} 1000 - \frac{i}{2}, & i = 1, \dots, 450, \\ i + 325, & i = 451, \dots, 900, \\ 1300 - \frac{i}{4}, & i = 901, \dots, 2d+1. \end{cases}$$

The stiffness values are chosen as $k_1 = 500$, $k_2 = 200$, and $k_3 = 300$. The parameter α_c that determines the influence of the internal damping defined by (7) is set to 0.02. The primary excitation corresponds to the four masses closest to the left ground reference anchor (with larger influence on the mass that is closer to the anchor point) and the one mass closest to right ground reference anchor, i.e., $B_2 \in \mathbb{R}^{(2d+1)\times 5}$ with:

$$\begin{aligned} B_2(1:2, 1:2) &= \operatorname{diag}(20, 10), \\ B_2(901:902, 3:4) &= \operatorname{diag}(20, 10), \\ B_2(1801, 5) &= 30; \end{aligned}$$

and all other entries being zero. We are interested in the two displacements, yielding the output:

$$y(t; p) = \begin{bmatrix} q_{400}(t; p) & q_{1300}(t; p) \end{bmatrix}^T.$$

In this example, we consider optimization over four dampers with gains p_1, p_2, p_3 and p_4 with their positions encoded in:

$$U_2 = \begin{bmatrix} e_{j_1} - e_{j_1+10}, & e_{j_2}, & e_{j_3}, & e_{j_3} - e_{j_3+100} \end{bmatrix},$$

where e_i is the ith canonical vector and the indices j_1, j_2, j_3 determine the damping positions. To illustrate the performance of our surrogate optimization framework for different damping configurations, the following indices are considered: $j_1 \in \{100, 300, 500, 700\}$, $j_2 \in \{150, 350, 550, 750\}$, and $j_3 \in \{1400, 1700\}$. This results in 32 different damping configurations for which the $\mathcal{H}_2$ system norm is optimized (i.e., solve (33) and the surrogate problem (34)).

The full optimization problem (33) and the surrogate problem (34) were solved using MATLAB's built-in `fminsearch` with a transformation that allows constrained optimization. The starting point was $\mathsf{p}^0 = (100, 100, 100, 100)$ and for each parameter, the range was $[0, 5000]$. The stopping tolerance for optimization was set to $\nu = 5 \cdot 10^{-4}$. In solving the surrogate optimization problem, we employed both Algorithms 3 and 4. Our implementation takes advantage of the fact that the first subsystem $\mathcal{H}_1(s)$ is independent of not only the parameter p but also the damping positions. Therefore, for the 32 damping configurations considered, reducing $\mathcal{H}_1(s)$ in Algorithm 3 (or sampling of $\mathcal{H}_1(s)$ in Algorithm 4) needs to be done only once.

In Algorithm 3, for each damping configuration, we used:

$$\text{termination tolerance for error bound:} \quad \tau = 10^{-2};$$
$$\text{initial reduction dimensions:} \quad (r_1, r_2, r_3, r_4) = (280, 300, 480, 430).$$

The reduced orders r_i were chosen based on the Hankel singular values of each subsystem. Reduced subsystem updates were performed such that each time an update was needed, r_i was increased by 15%. Similarly, in Algorithm 4 for each damping configuration, we used:

$$\text{termination tolerance for error bound: } \tau = 10^{-4};$$
$$\text{initial reduction order set to 130;}$$
$$\text{number of predetermined sampling points } \{\xi_1, \ldots, \xi_N\} \text{ set to } N = 500.$$

The sampling points $\xi_1, \ldots, \xi_N$ were chosen to be logarithmically spaced between the smallest and largest (by magnitude) undamped eigenfrequencies. We initialized VF using dominant poles [15]. During the optimization process each time that $e(\mathcal{H}(\cdot; \mathsf{p})), \widehat{\mathcal{H}}(\cdot; \mathsf{p}))) > \tau$, the order was increased by 10%.

Across all damping configurations, Algorithm 3 entered the inner while loop only in 15% of the cases and Algorithm 4 in 34% of the cases. For Algorithm 3, this means that model reduction was performed only once for most configurations. For Algorithm 4, note that it employs Algorithm 2, which resamples $\mathcal{H}(s, \mathsf{p})$ almost at no cost, and then applies VF. Repeated application of VF constitutes only a modest cost increment since the required frequency sampling is obtained already at an earlier step. Remarkably, we have also observed that in nearly all cases that we consider, VF converges quickly. In particular, across all 32 damping configurations, application of VF for the initial surrogate optimization step took on average 13 iterations to converge. But in the vast majority of subsequent VF application, convergence occurred often after a single iteration (93.5% of total VF applications) or two iterations (3.3% of total VF applications).

Figure 4 depicts the relative errors in the optimal gains for different damping configurations calculated by Algorithm 3 (denoted by blue squares) and Algorithm 4 (denoted by black triangles). The relative errors in the optimal gain is calculated by $\|\mathsf{p}^\star - \hat{\mathsf{p}}^\star\|/\|\mathsf{p}^\star\|$, where $\mathsf{p}^\star$ and $\hat{\mathsf{p}}^\star$ denote the optimal gains calculated with, respectively, the full model (i.e., solving (33)) and the reduced model (i.e., solving (34)). The figure shows in most cases Algorithm 4 gave more accurate results.

However, since the cost function, the $\mathcal{H}_2$ norm, can be flat with respect to some damping parameters, the quality of the surrogate optimization is better illustrated in Fig. 5 where we show the relative errors in the cost function. For the optimal gain $\mathsf{p}^\star$ obtained by solving the full problem (33) and the optimal gain $\hat{\mathsf{p}}^\star$ obtained by solving the surrogate problem (34), the relative error is computed by $\frac{|\|\mathcal{H}(\cdot;\mathsf{p}^\star)\|_{\mathcal{H}_2} - \|\widehat{\mathcal{H}}(\cdot;\mathsf{p}^\star)\|_{\mathcal{H}_2}|}{\|\widehat{\mathcal{H}}(\cdot;\mathsf{p}^\star)\|_{\mathcal{H}_2}}$. These results are illustrated in Fig. 5. Even though both algorithms yield accurate results, the surrogate optimization with Algorithm 4 is consistently better with the largest relative error in the order of 10^{-4}.

In Algorithm 3, the relative termination tolerance for the error bound was set to 10^{-2}. But, Fig. 5 reveals that relative errors are usually a few order of magnitudes

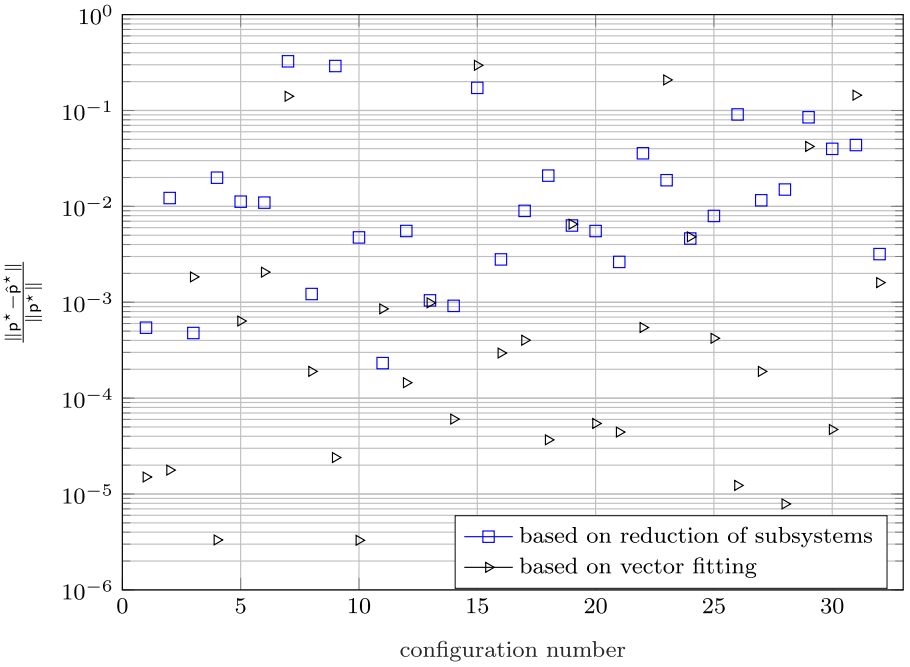


Fig. 4 Relative errors in the optimal gains for Algorithms 3 and 4

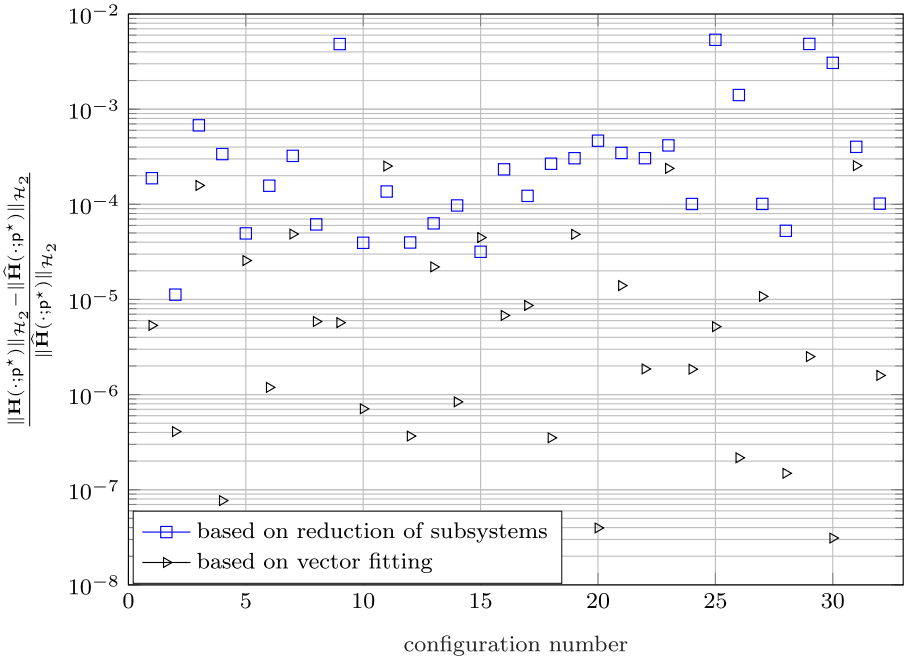


Fig. 5 Relative errors for $\mathcal{H}_2$ norm at optimal gain for Algorithm 3 and Algorithm 4

Table 1 Acceleration factors using surrogate optimization

	Algorithm 3	Algorithm 4
Acceleration factor	7.8	60

smaller. Thus, we can conclude that, (at least in this example) the approximate error bound (36) is not sharp in general and therefore in some cases, Algorithm 3 might lead to a larger reduced order than necessary.

Another important quantity to measure is the speed-up compared to the full problem. Table 1 shows the average speed-ups for the optimization process obtained by both algorithms.

Both methods produce optimized parameters with satisfactory relative errors and with considerable acceleration of the optimization process. For this damping optimization problem, Algorithm 4 not only produced more accurate results but also yielded bigger speed-up than Algorithm 3. The significant difference in timing seen in Table 1 is mainly due to the fact that in the VF the most expensive step is data-collection, which is essentially free here. Additional speed up is obtained because convergence within VF was usually extremely fast. Therefore, for this problem, Algorithm 4 was more efficient. However, we also note that Algorithm 3, being based on subsystem reduction that includes estimation of the true error $f(\mathsf{p})$ as opposed to the sampling-based error $e(\mathsf{p})$, might improve the robustness of the optimization process.

5 Conclusions

We have introduced a framework for producing reduced order models of dynamical systems having an affine, low-rank parameter structure. The new framework does not require any sampling in the parameter domain and instead parametrically combines intermediate subsystems that are non-parametric. Our approach guarantees uniform stability of the aggregated reduced model across the entire parameter domain in many cases. Beyond the computational examples we provide for illustration, we show in some detail how this approach can be deployed efficiently in parameter optimization problems as well

There are various future directions for this work. One is to employ this framework to parametric models resulting from stationary PDEs with low-rank parametric structure. In this case, our approach will correspond do converting a parametric stationary model to interconnection of four non-parametric ones. In this paper, our focus was to fully remove the parameter sampling, including the knowledge of the parameter domain. However, the knowledge of parameter domain can help produce even better reduced models. If one knows the parameter domain, this means that we have some information on the range of the term $D(\mathsf{p})$ in (13). In this case then, as opposed to reducing, for example, $\mathcal{H}_2(s)$, we can perform an (input) weighted model reduction on $\mathcal{H}_2(s)$ where this weight can reflect the range of $D(\mathsf{p})$. These issues will be investigated in future works.

Funding This work has been supported in parts by Croatian Science Foundation under the project "Vibration Reduction in Mechanical Systems" (IP-2019-04-6774) and project "Control of Dynamical Systems" (IP-2016-06-2468). The work of Beattie was supported in parts by NSF through Grant DMS-1819110. The work of Gugercin was supported in parts by NSF through Grants DMS-1720257 and DMS-1923221.

References

1. Alla, A., Hinze, M., Kolvenbach, P., Lass, O., Ulbrich, S.: A certified model reduction approach for robust parameter optimization with pde constraints. Adv. Comput. Math., 1–30 (2019)
2. Antil, H., Heinkenschloss, M., Hoppe, R.H.W.: Domain decomposition and balanced truncation model reduction for shape optimization of the Stokes system. Optim. Methods Softw. **26**(4–5), 643–669 (2011)
3. Antoulas, A., Beattie, C., Gugercin, S.: Interpolatory Methods for Model Reduction. SIAM Publications, Philadelphia (2020)
4. Antoulas, A.: Approximation of Large-Scale Dynamical Systems. SIAM Publications, Philadelphia (2005)
5. Antoulas, A., Anderson, B.: On the scalar rational interpolation problem. IMA J. Math. Control. Inf. **3**, 61–88 (1986)
6. Antoulas, A.C., Sorensen, D.C., Gugercin, S.: A survey of model reduction methods for large-scale systems. Contemp. Math. **280**, 193–219 (2001)
7. Arian, E., Fahl, M., Sachs, E.: Trust-region proper orthogonal decomposition models by optimization methods. In: Proceedings of the 41st IEEE Conference on Decision and Control, pp. 3300–3305. Las Vegas, NV (2002)
8. Baur, U., Beattie, C., Benner, P.: Mapping parameters across system boundaries: parameterized model reduction with low rank variability in dynamics. PAMM **14**(1), 19–22 (2014)
9. Baur, U., Beattie, C.A., Benner, P., Gugercin, S.: Interpolatory projection methods for parameterized model reduction. SIAM J. Sci. Comput. **33**(5), 2489–2518 (2011)
10. Baur, U., Benner, P.: Model reduction for parametric systems using balanced truncation and interpolation. at–Automatisierungstechnik **57**(8), 411–420 (2009)
11. Beattie, C., Gugercin, S.: Interpolatory projection methods for structure-preserving model reduction. Syst. Control Lett. **58**, 225–232 (2009)
12. Beattie, C., Gugercin, S.: A trust region method for optimal $\mathcal{H}_2$ model reduction. In: Proceedings of 48th IEEE Conference on Decision and Control, pp. 5370–5375 (2009)
13. Benner, P., Cohen, A., Ohlberger, M., Willcox, K.: Model Reduction and Approximation: Theory and Algorithms. Computational Science and Engineering. SIAM Publications, Philadelphia (2017)
14. Benner, P., Gugercin, S., Willcox, K.: A survey of projection-based model reduction methods for parametric dynamical systems. SIAM Rev. **57**(4), 483–531 (2015)
15. Benner, P., Kürschner, P., Tomljanović, Z., Truhar, N.: Semi-active damping optimization of vibrational systems using the parametric dominant pole algorithm. J. Appl. Math. Mech. 1–16. https://doi.org/10.1002/zamm201400158 (2015)
16. Benner, P., Mehrmann, V., Sorensen, D.: Dimension Reduction of Large-Scale Systems. Lecture Notes in Computational Science and Engineering. Springer, Berlin (2005)
17. Benner, P., Sachs, E., Volkwein, S.: Model order reduction for PDE constrained optimization. In: Leugering, G., Benner, P., Engell, S., Griewank, A., Harbrecht, H., Hinze, M., Rannacher, R., Ulbrich, S. (eds.) Trends in PDE Constrained Optimization, International Series of Numerical Mathematics, vol. 165, pp. 303–326. Springer International Publishing (2014)
18. Benner, P., Tomljanović, Z., Truhar, N.: Dimension reduction for damping optimization in linear vibrating systems. Z. Angew. Math. Mech. **91**(3), 179–191 (2011). https://doi.org/10.1002/zamm.201000077
19. Benner, P., Tomljanović, Z., Truhar, N.: Optimal damping of selected eigenfrequencies using dimension reduction. Numer. Linear Algebr. **20**(1), 1–17 (2013). https://doi.org/10.1002/nla.833
20. Berljafa, M., Güttel, S.: The RKFIT algorithm for nonlinear rational approximation. SIAM J. Sci. Comput. **39**(5), 2049–2071 (2017)
21. Blanchini, F., Casagrande, D., Gardonio, P., Miani, S.: Constant and switching gains in semi-active damping of vibrating structures. Int. J. Control **85**(12), 1886–1897 (2012)

22. Bui-Thanh, T., Willcox, K., Ghattas, O.: Model reduction for large-scale systems with high-dimensional parametric input space. SIAM J. Sci. Comput. **30**(6), 3270–3288 (2008)
23. Carracedo Rodriguez, A., Gugercin, S.: The p-AAA algorithm for data driven modeling of parametricdynamical systems. Technical report, arXiv preprint available at arXiv:2003.06536 (2020)
24. Chinea, A., Grivet-Talocia, S.: On the parallelization of vector fitting algorithms. IEEE Trans. Compon. Packaging Manuf. Technol. **1**(11), 1761–1773 (2011)
25. Desai, U., Pal, D.: A transformation approach to stochastic model reduction. IEEE Trans. Autom. Control **29**(12), 1097–1100 (1984)
26. Drmač, Z., Gugercin, S., Beattie, C.: Quadrature-based vector fitting for discretized $\mathcal{H}_2$ approximation. SIAM J. Sci. Comp. **37**(2), A625–A652 (2015)
27. Drmač, Z., Gugercin, S., Beattie, C.: Vector fitting for matrix-valued rational approximation. SIAM J. Sci. Comput. **37**(5), A2346–A2379 (2015)
28. Egger, H., Kugler, T., Liljegren-Sailer, B., Marheineke, N., Mehrmann, V.: On structure-preserving model reduction for damped wave propagation in transport networks. SIAM J. Sci. Comp. **40**(1), A331–A365 (2018)
29. Feng, L.: Parameter independent model order reduction. Math. Comput. Simulation **68**, 221–234 (2005)
30. Feng, L., Rudnyi, E.B., Korvink, J.G.: Parametric model reduction to generate boundary condition independent compact thermal model. Technical report, IMTEK-Institute for Microsystem Technology. http://modelreduction.com/doc/papers/feng04THERMINIC.pdf (2004)
31. Glover, K.: All optimal Hankel-norm approximations of linear multivariable systems and their l_∞-error bounds. Int. J. Control **39**(6), 1115–1193 (1984)
32. Golub, G.H., Loan, C.V.F.: Matrix Computations. The Johns Hopkins University Press, Baltimore (1998)
33. Gonnet, P., Pachón, R., Trefethen, L.N.: Robust rational interpolation and least-squares. Electron. Trans. Numer. Anal. **38**, 146–167 (2011)
34. Grimm, A.R.: Parametric dynamical systems: transient analysis and data driven modeling. Ph.D. thesis, Virginia Tech (2018)
35. Grivet-Talocia, S., Gustavsen, B.: Passive Macromodeling: Theory and Applications, vol. 239. Wiley, Hoboken (2015)
36. Gugercin, S.: An iterative SVD-Krylov based method for model reduction of large-scale dynamical systems. Linear Algebra Appl. **428**(8-9), 1964–1986 (2008)
37. Gugercin, S., Antoulas, A., Beattie, C.: $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. SIAM J. Matrix Anal. Appl. **30**(2), 609–638 (2008)
38. Gugercin, S., Polyuga, R., Beattie, C., Schaft, A.V.: Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems. Automatica **48**, 1963–1974 (2012)
39. Gustavsen, B.: Improving the pole relocating properties of vector fitting. IEEE Trans. Power Deliv. **21**(3), 1587–1592 (2006)
40. Gustavsen, B., Semlyen, A.: Rational approximation of frequency domain responses by vector fitting. IEEE Trans. Power Deliv. **14**(3), 1052–1061 (1999)
41. Heinkenschloss, M., Jando, D.: Reduced order modeling for time-dependent optimization problems with initial value controls. SIAM J. Sci. Comput. **40**(1), A22–A51 (2018)
42. Hokanson, J.M.: Projected nonlinear least squares for exponential fitting. SIAM J. Sci. Comput. **39**(6), A3107–A3128 (2017)
43. Hund, M., Mlinarić, P., Saak, J.: An $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal model order reduction approach for parametric linear time-invariant systems. Proc. Appl. Math. Mech. **18**(1), e201800084 (2018). https://doi.org/10.1002/pamm.201800084
44. Ionita, A.C., Antoulas, A.C.: Data-driven parametrized model reduction in the Loewner framework. SIAM J. Sci. Comput. **36**(3), 984–1007 (2014). https://doi.org/10.1137/130914619
45. Kunisch, K., Volkwein, S.: Proper orthogonal decomposition for optimality systems. ESAIM Math. Model. Numer. Anal. **42**(1), 1–23 (2008)
46. Kuzmanović, I., Tomljanović, Z., Truhar, N.: Optimization of material with modal damping. Appl. Math. Comput. **218**, 7326–7338 (2012)
47. Mayo, A., Antoulas, A.: A framework for the solution of the generalized realization problem. Linear Algebra Appl. **425**(2-3), 634–662 (2007)
48. Moore, B.: Principal component analysis in linear systems: controllability, observability, and model reduction. IEEE Trans. Autom. Control **26**(1), 17–32 (1981)

49. Müller, P., Schiehlen, W.: Linear Vibrations. Martinus Nijhoff Publishers, Leiden (1985)
50. Mullis, C., Roberts, R.: Synthesis of minimum roundoff noise fixed point digital filters. EEE Trans. Circuits Syst. **23**(9), 551–562 (1976)
51. Nakatsukasa, Y., Sète, O., Trefethen, L.N.: The AAA algorithm for rational approximation. SIAM J. Sci. Comput. **40**(3), A1494–A1522 (2018)
52. Nakić, I., Tomljanović, Z., Truhar, N.: Mixed control of vibrational systems. J. Appl. Math. Mech. **99**(9), 1–15 (2019)
53. Ober, R.: Balanced parametrization of classes of linear systems. SIAM J. Control. Optim. **29**(6), 1251–1287 (1991)
54. Oberwolfach Benchmark Collection: Thermal model. hosted at MORwiki – Model Order Reduction Wiki (20XX). http://modelreduction.org/index.php/Thermal_Model
55. Opdenacker, P., Jonckheere, E.: A contraction mapping preserving balanced reduction scheme and its infinity norm error bounds. IEEE Trans. Circuits Syst. **35**(2), 184–189 (1988)
56. van Ophem, S., Deckers, E., Desmet, W.: Parametric model order reduction without a priori sampling for low rank changes in vibro-acoustic systems. Mech. Syst. Signal Process. **130**, 597–609 (2019)
57. Penzl, T.: Numerische simulation auf massiv parallelen rechnern. Ph.D. thesis, TU Chemnitz. Algorithms for Model Reduction of Large Dynamical Systems, Tech. Rep. SFB393/99-40 (1999)
58. Polyuga, R., van der Schaft, A.: Structure preserving moment matching for port-hamiltonian systems: Arnoldi and Lanczos. IEEE Trans. Autom. Control **56**(6), 1458–1462 (2011)
59. Quarteroni, A., Manzoni, A., Negri, F.: Reduced basis methods for partial differential equations: an introduction. R. UNITEXT. Springer, Cham (2016)
60. Rommes, J., Martins, N.: Computing transfer function dominant poles of large-scale second-order dynamical systems. SIAM J. Sci. Comput. **30**(4), 2137–2157 (2008)
61. Sanathanan, C., Koerner, J.: Transfer function synthesis as a ratio of two complex polynomials. IEEE Trans. Autom. Control **8**(1), 56–58 (1963)
62. Tomljanović, Z., Beattie, C., Gugercin, S.: Damping optimization of parameter dependent mechanical systems by rational interpolation. Adv. Comput. Math. **44**, 1797–1820 (2018)
63. Tomljanović, Z., Voigt, M.: Semi-active $\mathcal{H}_\infty$ damping optimization by adaptive interpolation. Numer. Linear Algebra Appl **27**(4), 1–17 (2020)
64. Veselić, K.: Damped Oscillations of Linear Systems. Springer Lecture Notes in Mathematics. Springer, Berlin (2011)
65. Yue, Y., Meerbergen, K.: Accelerating optimization of parametric linear systems by model order reduction. SIAM J. Optim. **23**(2), 1344–1370 (2013)
66. Yue, Y., Meerbergen, K.: Parametric model order reduction of damped mechanical systems via the block Arnoldi process. Appl. Math. Lett. **26**(6), 643–648 (2013)
67. Zhou, K., Doyle, J.C., Glover K.: Robust and Optimal Control, vol. 40. Prentice Hall, New Jersey (1996)

Affiliations

Christopher Beattie[1] · Serkan Gugercin[1] · Zoran Tomljanović[2]

Christopher Beattie
beattie@vt.edu

Serkan Gugercin
gugercin@vt.edu

[1] Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

[2] Department of Mathematics, J.J. Strossmayer University of Osijek, Osijek, Croatia