

INTERPOLATORY PROJECTION METHODS FOR PARAMETERIZED MODEL REDUCTION*

ULRIKE BAUR[†], CHRISTOPHER BEATTIE[‡], PETER BENNER[§], AND
SERKAN GUGERCIN[‡]

Abstract. We provide a unifying projection-based framework for structure-preserving interpolatory model reduction of parameterized linear dynamical systems, i.e., systems having a structured dependence on parameters that we wish to retain in the reduced-order model. The parameter dependence may be linear or nonlinear and is retained in the reduced-order model. Moreover, we are able to give conditions under which the gradient and Hessian of the system response with respect to the system parameters is matched in the reduced-order model. We provide a systematic approach built on established interpolatory \mathcal{H}_2 optimal model reduction methods that will produce parameterized reduced-order models having high fidelity throughout a parameter range of interest. For single input/single output systems with parameters in the input/output maps, we provide reduced-order models that are *optimal* with respect to an $\mathcal{H}_2 \otimes \mathcal{L}_2$ joint error measure. The capabilities of these approaches are illustrated by several numerical examples from technical applications.

Key words. parameterized model reduction, interpolation, rational Krylov

AMS subject classifications. 41A05, 93A15, 93C05, 35B30

DOI. 10.1137/090776925

1. Introduction. Numerical simulation has steadily increased in importance across virtually all scientific and engineering disciplines. In many application areas, experiments have been largely replaced by numerical simulation in order to save costs in design and development. High accuracy simulation requires high fidelity mathematical models which in turn induce dynamical systems of very large dimension. The ensuing demands on computational resources can be overwhelming and efficient model utilization becomes a necessity. It often is both possible and prudent to produce a lower dimension model that approximates the response of the original one to high accuracy. There are many model reduction strategies in use that are remarkably effective in the creation of compact, efficient, and high fidelity dynamical system models. Such a reduced model can then be used reliably as an efficient surrogate to the original system, replacing it as a component in larger simulations, for example, or in allied contexts that involve design optimization or the development of low-order, fast controllers suitable for real time applications.

Typically, a reduced-order model will represent a specific instance of the physical system under study and as a consequence will have high fidelity only for small variations around that base system instance. Significant modifications to the physical model such as geometric variations, changes in material properties, or alterations in

*Submitted to the journal's Computational Methods in Science and Engineering section November 12, 2009; accepted for publication (in revised form) June 28, 2011; published electronically October 13, 2011. This work was supported by DFG grant BE 2174/7-1, Automatic, Parameter-Preserving Model Reduction for Applications in Microsystems Technology, and NSF grants DMS-0505971 and DMS-0645347.

<http://www.siam.org/journals/sisc/33-5/77692.html>

[†]Fakultät für Mathematik, TU Chemnitz, D-09107 Chemnitz, Germany (ulrike.baur@mathematik.tu-chemnitz.de).

[‡]Department of Mathematics, Virginia Tech, Blacksburg, VA 24061-0123 (beattie@math.vt.edu, gugercin@math.vt.edu).

[§]Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany (benner@mpi-magdeburg.mpg.de).

boundary conditions generally necessitate generation of new reduced models. This can be particularly onerous in design optimization where parameters are changed in each optimization cycle. Since the generation of a high fidelity reduced model may be comparable in expense to a (brief) simulation of an instance of the original full-order model, the benefits of model reduction will be fully realized only if the parametric dependence found in the original dynamical system can be preserved in some fashion within the reduced model. This is the goal of *parameterized model reduction (PMOR)*: generate a reduced-order dynamical system retaining functional dependence on important design parameters and maintaining high fidelity with respect to the response of the original dynamical system, throughout the range of interest of design parameters.

Many design optimization approaches use surrogate models that are constructed using response surface modeling or Kriging [34, 33, 45]. These techniques are flexible and broadly applicable; they can be efficient for uncertain, unstructured, or empirically based models, but generally cannot exploit fully the character of time-dependent processes generated by an underlying dynamical system. PMOR is an approach that attempts to take direct account of structure in the underlying dynamical system creating the response data. It can be expected to produce more efficient and accurate models than general purpose approaches that provide ad hoc fits or regressions to observed input/output responses.

PMOR is at an early stage of development. Currently, there are developments based on multivariate Padé approximation [5, 12, 14, 15, 17, 18, 19, 27, 26, 37, 38, 41, 49]. These methods differ in the way moments are computed (implicitly vs. explicitly) and in the number of (mixed) moments that are matched. Approaches based on explicitly computed moments suffer from the same numerical instabilities as analogous methods for model reduction of nonparameterized systems. Implicit approaches appear to provide a robust resolution of these difficulties, at least for low dimensional parameter spaces. Moment-matching/interpolation properties can be proved (see, e.g., [18, 12, 49, 27]) analogously as for standard moment-matching methods such as Padé-via-Lanczos [16, 20]. Existing proofs appear either to be restrictive regarding the number of parameters or structure of dependence (e.g., only one additional parameter and linear parametric dependence in [49]; parameters only in some of the defining matrices in [49, 12, 27]), or they are formulated in terms of series expansions and term-by-term matching of moments. Explicit moment matching is conceptually simple (if painful), and indeed, [6] considers a framework of parametric dependence that is quite general; this approach could be extended also to the situations we consider. However, such approaches have led to strategies that are then based on explicit moment computation (e.g., [12, 6]) and so may again be susceptible to numerical instabilities as mentioned above, which could be further exacerbated in large scale settings. Note that to the extent that multi-input/multi-output (MIMO) cases have been considered at all in parameterized settings, full transfer matrix interpolation properties have been pursued, and these approaches also rapidly become infeasible already with modest input/output dimensionality.

In contrast to this, we aim at a broadly applicable implicit interpolation framework that is capable of treating parameter dependence in all matrices defining the parameterized system, possibly involving nonlinear parameter functions. Our approach allows for a numerically robust implementation and can handle an arbitrary number of interpolation points. We investigate here some first ideas in the direction of optimal selection of interpolation points. For this, we utilize algorithmic ideas for \mathcal{H}_2 -optimal model reduction developed for nonparameterized linear systems in [25]. For the MIMO case, this requires tangential Hermite interpolation. To the best of our

knowledge, this has not been considered in the literature thus far, and so we detail the use of tangential interpolation to match values of the transfer function and its gradient with respect to parameters. We also show that, as in standard moment-matching approaches, higher-order tangential interpolation is possible—this is demonstrated for the Hessian of a parametrized transfer function. This in turn may have interesting applications when using the reduced-order model as a surrogate in optimization methods, but this passes beyond the scope of this work.

Other PMOR approaches include interpolation of the full transfer function (see [3]) and reduced basis methods (see, e.g., [2, 22, 28, 32, 39]). Reduced-basis methods are successful in finding an information-rich set of global ansatz functions for spatial discretization of parameterized partial differential equations (PDEs). In the setting we consider here, we do not necessarily assume that a PDE is provided; we start instead from a parameterized state-space model. This is the case, e.g., when computer-aided engineering (CAE) tools for automatic model generation are used. In this situation, the spatial discretization of the PDE is performed inside the CAE tool and reduced basis methods are not directly applicable.

We lay out our basic problem setting, define notation, and describe precisely in what sense our model reduction methods are structure-preserving in section 2. In section 3, we review aspects of interpolatory model reduction in standard (nonparameterized) settings that are useful for us, focusing especially on the selection of interpolation points that lead to optimal reduced-order models. In section 4, we derive an interpolation-based approach to PMOR that is closely associated with rational Krylov methods developed by Grimme [24] and earlier work by Villemagne and Skelton [13]. As noted in these works, interpolation properties are governed by the range and cokernel of a (skew) projection associated with the model reduction process. Remarkably, similar conditions govern the matching of gradient and Hessian information of the system response with respect to the system parameters. Efficient numerical methods built on previously known \mathcal{H}_2 optimal model reduction methods are introduced in section 5, and we describe in section 5.1 how to find *optimal* parameterized reduced-order models for a special case of a parameterized single input/single output (SISO) system. The efficiency of the derived numerical algorithms for PMOR is illustrated using several real-world examples from microsystems technology in section 6.

2. Problem setting. Consider a multi-input/multi-output (MIMO) linear dynamical system parameterized with ν parameters $\mathbf{p} = [p_1, \dots, p_\nu]^T \in \mathbb{R}^\nu$, presented in state space form as

$$(2.1) \quad \begin{aligned} \mathbf{E}(\mathbf{p}) \dot{\mathbf{x}}(t) &= \mathbf{A}(\mathbf{p}) \mathbf{x}(t) + \mathbf{B}(\mathbf{p}) \mathbf{u}(t), & \text{with } \mathbf{x}(0) &= 0, \\ \mathbf{y}(t) &= \mathbf{C}(\mathbf{p}) \mathbf{x}(t), \end{aligned}$$

where $\mathbf{E}(\mathbf{p})$, $\mathbf{A}(\mathbf{p}) \in \mathbb{R}^{n \times n}$, $\mathbf{B}(\mathbf{p}) \in \mathbb{R}^{n \times m}$, and $\mathbf{C}(\mathbf{p}) \in \mathbb{R}^{\ell \times n}$. Our framework allows parameter dependency in all system matrices. Without loss of generality, assume the parametric dependence in the system matrices of (2.1) has the following form:

$$(2.2) \quad \begin{aligned} \mathbf{E}(\mathbf{p}) &= \mathbf{E}_0 + e_1(\mathbf{p})\mathbf{E}_1 + \dots + e_M(\mathbf{p})\mathbf{E}_M, \\ \mathbf{A}(\mathbf{p}) &= \mathbf{A}_0 + f_1(\mathbf{p})\mathbf{A}_1 + \dots + f_M(\mathbf{p})\mathbf{A}_M, \\ \mathbf{B}(\mathbf{p}) &= \mathbf{B}_0 + g_1(\mathbf{p})\mathbf{B}_1 + \dots + g_M(\mathbf{p})\mathbf{B}_M, \\ \mathbf{C}(\mathbf{p}) &= \mathbf{C}_0 + h_1(\mathbf{p})\mathbf{C}_1 + \dots + h_M(\mathbf{p})\mathbf{C}_M. \end{aligned}$$

We assume throughout that (2.1) is stable for all parameter choices \mathbf{p} considered.

The parameter dependence encoded in the functions e_j, f_j, g_j, h_j may be linear or nonlinear, but is assumed smooth enough to allow for approximation by interpolation.

The representation (2.2) is not unique; there may be many ways in which one may express system matrices, $\mathbf{E}(\mathbf{p})$, $\mathbf{A}(\mathbf{p})$, $\mathbf{B}(\mathbf{p})$, and $\mathbf{C}(\mathbf{p})$ in such a form, and the number of terms, M , as well as the particular parameter functions e_j, f_j, g_j, h_j may vary with the representation that one chooses. A desirable choice should produce as few terms as possible (M as small as possible) for reasons we describe below; the methods we propose will be most advantageous when $M \ll n$. Note also that the actual number of terms appearing may vary among the matrices $\mathbf{E}(\mathbf{p})$, $\mathbf{A}(\mathbf{p})$, $\mathbf{B}(\mathbf{p})$, and $\mathbf{C}(\mathbf{p})$.

A general projection framework for structure-preserving PMOR can be described as follows: suppose that (constant) matrices $\mathbf{V}_r, \mathbf{W}_r \in \mathbb{C}^{n \times r}$ are specified with $r \ll n$ and $\text{rank}(\mathbf{V}_r) = \text{rank}(\mathbf{W}_r) = r$ and define an associated reduced system (see, e.g., [18, 12, 17, 27, 49]):

$$\begin{aligned} \mathbf{E}_r(\mathbf{p}) \dot{\mathbf{x}}_r(t) &= \mathbf{A}_r(\mathbf{p}) \mathbf{x}_r(t) + \mathbf{B}_r(\mathbf{p}) \mathbf{u}(t), \\ \mathbf{y}_r(t) &= \mathbf{C}_r(\mathbf{p}) \mathbf{x}_r(t) \quad \text{with } \mathbf{x}_r(0) = 0, \end{aligned} \tag{2.3}$$

where $\mathbf{E}_r(\mathbf{p}) = \mathbf{W}_r^T \mathbf{E}(\mathbf{p}) \mathbf{V}_r$, $\mathbf{A}_r(\mathbf{p}) = \mathbf{W}_r^T \mathbf{A}(\mathbf{p}) \mathbf{V}_r$,
 $\mathbf{B}_r(\mathbf{p}) = \mathbf{W}_r^T \mathbf{B}(\mathbf{p})$, and $\mathbf{C}_r(\mathbf{p}) = \mathbf{C}(\mathbf{p}) \mathbf{V}_r$.

The parametric dependence of the original system (2.1) is retained in the reduced system (2.3) in the sense that

$$\begin{aligned} \mathbf{E}_r(\mathbf{p}) &= \mathbf{W}_r^T \mathbf{E}_0 \mathbf{V}_r + e_1(\mathbf{p}) \mathbf{W}_r^T \mathbf{E}_1 \mathbf{V}_r + \cdots + e_M(\mathbf{p}) \mathbf{W}_r^T \mathbf{E}_M \mathbf{V}_r, \\ \mathbf{A}_r(\mathbf{p}) &= \mathbf{W}_r^T \mathbf{A}_0 \mathbf{V}_r + f_1(\mathbf{p}) \mathbf{W}_r^T \mathbf{A}_1 \mathbf{V}_r + \cdots + f_M(\mathbf{p}) \mathbf{W}_r^T \mathbf{A}_M \mathbf{V}_r, \\ \mathbf{B}_r(\mathbf{p}) &= \mathbf{W}_r^T \mathbf{B}_0 + g_1(\mathbf{p}) \mathbf{W}_r^T \mathbf{B}_1 + \cdots + g_M(\mathbf{p}) \mathbf{W}_r^T \mathbf{B}_M, \\ \mathbf{C}_r(\mathbf{p}) &= \mathbf{C}_0 \mathbf{V}_r + h_1(\mathbf{p}) \mathbf{C}_1 \mathbf{V}_r + \cdots + h_M(\mathbf{p}) \mathbf{C}_M \mathbf{V}_r, \end{aligned} \tag{2.4}$$

which is evidently structurally similar to (2.2). Once the matrices \mathbf{V}_r and \mathbf{W}_r are specified, all the constituent matrices, $\mathbf{W}_r^T \mathbf{E}_k \mathbf{V}_r$, $\mathbf{W}_r^T \mathbf{A}_k \mathbf{V}_r$, $\mathbf{W}_r^T \mathbf{B}_k$, and $\mathbf{C}_k \mathbf{V}_r$ for $k = 0, \dots, M$ contributing to $\mathbf{E}_r(\mathbf{p})$, $\mathbf{A}_r(\mathbf{p})$, $\mathbf{B}_r(\mathbf{p})$, and $\mathbf{C}_r(\mathbf{p})$ can be *precomputed*, and this corresponds to the *offline* portion of the method. Although the order, r , of the dynamical system (2.3) is an obvious focus in judging the cost of using the reduced system, the size of M , as a measure of the complexity of the representation (2.2), may become a factor since for every new choice of parameter values, the cost of generating $\mathbf{E}_r(\mathbf{p})$, $\mathbf{A}_r(\mathbf{p})$, $\mathbf{B}_r(\mathbf{p})$, and $\mathbf{C}_r(\mathbf{p})$ obviously grows proportionally to M .

Whenever the input $\mathbf{u}(t)$ is exponentially bounded—that is, when there is a fixed $\gamma \in \mathbb{R}$ such that $\|\mathbf{u}(t)\| \sim \mathcal{O}(e^{\gamma t})$, then $\mathbf{x}(t)$ and $\mathbf{y}(t)$ from (2.1) and $\mathbf{x}_r(t)$ and $\mathbf{y}_r(t)$ from (2.3) will also be exponentially bounded, and the Laplace transform can be applied to (2.1) and (2.3) to obtain

$$\hat{\mathbf{y}}(s, \mathbf{p}) = \mathbf{C}(\mathbf{p}) (s \mathbf{E}(\mathbf{p}) - \mathbf{A}(\mathbf{p}))^{-1} \mathbf{B}(\mathbf{p}) \hat{\mathbf{u}}(s), \tag{2.5}$$

$$\hat{\mathbf{y}}_r(s, \mathbf{p}) = \mathbf{C}_r(\mathbf{p}) (s \mathbf{E}_r(\mathbf{p}) - \mathbf{A}_r(\mathbf{p}))^{-1} \mathbf{B}_r(\mathbf{p}) \hat{\mathbf{u}}(s), \tag{2.6}$$

where we have denoted Laplace transformed quantities with $\hat{\cdot}$. We define parameterized transfer functions accordingly:

$$\mathbf{H}(s, \mathbf{p}) = \mathbf{C}(\mathbf{p}) (s \mathbf{E}(\mathbf{p}) - \mathbf{A}(\mathbf{p}))^{-1} \mathbf{B}(\mathbf{p}) \tag{2.7}$$

and

$$\mathbf{H}_r(s, \mathbf{p}) = \mathbf{C}_r(\mathbf{p}) (s \mathbf{E}_r(\mathbf{p}) - \mathbf{A}_r(\mathbf{p}))^{-1} \mathbf{B}_r(\mathbf{p}). \tag{2.8}$$

The quality of the approximation $\hat{\mathbf{y}}_r(s, \mathbf{p}) \approx \hat{\mathbf{y}}(s, \mathbf{p})$ is tied directly to the quality of the approximation $\mathbf{H}_r(s, \mathbf{p}) \approx \mathbf{H}(s, \mathbf{p})$. The quality of this approximation in general, and interpolation properties in particular, depend entirely on how the matrices \mathbf{V}_r and \mathbf{W}_r are selected.

There is substantial flexibility in choosing \mathbf{V}_r and \mathbf{W}_r . We do require that both \mathbf{V}_r and \mathbf{W}_r have full rank but it is not necessary to require that either $\mathbf{W}_r^T \mathbf{V}_r$ or $\mathbf{W}_r^T \mathbf{E}(\mathbf{p}) \mathbf{V}_r$ be nonsingular. Note that if $\mathbf{E}(\mathbf{p})$ is nonsingular, then $\mathbf{H}(s, \mathbf{p})$ is a strictly proper transfer function and one may wish $\mathbf{H}_r(s, \mathbf{p})$ to be strictly proper as well—leading to the requirement that $\mathbf{E}_r(\mathbf{p}) = \mathbf{W}_r^T \mathbf{E}(\mathbf{p}) \mathbf{V}_r$ be nonsingular as well. This can be thought of as an interpolation condition since under these circumstances \mathbf{H}_r will interpolate \mathbf{H} at infinity: $\lim_{s \rightarrow \infty} \mathbf{H}(s) = \lim_{s \rightarrow \infty} \mathbf{H}_r(s) = 0$ (facilitating, in effect, a good match between true and reduced-order system response at high frequencies). Although we allow \mathbf{V}_r and \mathbf{W}_r to be complex in order to simplify the discussion, in most circumstances \mathbf{V}_r and \mathbf{W}_r can be chosen to be real so (2.3) represents a real dynamical system.

3. Interpolatory model reduction. To make the discussion largely self-contained, we briefly review the basic features of interpolatory model reduction for non-parameterized systems. Consider a full-order (nonparameterized) dynamical system described by

$$(3.1) \quad \mathbf{E} \dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{B} \mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{C} \mathbf{x}(t), \quad \text{with } \mathbf{x}(0) = 0,$$

where $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, and $\mathbf{C} \in \mathbb{R}^{\ell \times n}$ with the associated transfer function $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}$. We seek a reduced system with state-space form

$$(3.2) \quad \mathbf{E}_r \dot{\mathbf{x}}_r(t) = \mathbf{A}_r \mathbf{x}_r(t) + \mathbf{B}_r \mathbf{u}(t), \quad \mathbf{y}_r(t) = \mathbf{C}_r \mathbf{x}_r(t), \quad \text{with } \mathbf{x}_r(0) = 0,$$

and associated transfer function, $\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r$, where $\mathbf{A}_r, \mathbf{E}_r \in \mathbb{C}^{r \times r}$, $\mathbf{B}_r \in \mathbb{C}^{r \times m}$, $\mathbf{C}_r \in \mathbb{C}^{\ell \times r}$, and $r \ll n$, are such that $\mathbf{y}_r(t)$ approximates $\mathbf{y}(t)$ well. We adopt the projection framework described above, specifying matrices $\mathbf{V}_r \in \mathbb{C}^{n \times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n \times r}$, such that $\text{rank}(\mathbf{V}_r) = \text{rank}(\mathbf{W}_r) = r$, which then determine reduced system matrices $\mathbf{E}_r = \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r$, $\mathbf{A}_r = \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r$, $\mathbf{B}_r = \mathbf{W}_r^T \mathbf{B}$, and $\mathbf{C}_r = \mathbf{C} \mathbf{V}_r$.

Interpolatory model reduction is an approach introduced by Skelton et al. in [13, 52, 53], which was later placed into a numerically efficient framework by Grimme [24]. Gallivan, Vandendorpe, and Van Dooren [21] developed a more versatile version for MIMO systems, a variant of which we describe here and then adapt to parameterized systems: Starting with a full-order system as in (3.1) and selected interpolation points, σ_k , in the complex plane paired with corresponding left and right directions $\mathbf{c}_k \in \mathbb{C}^\ell$ and $\mathbf{b}_k \in \mathbb{C}^m$, we produce matrices $\mathbf{V}_r \in \mathbb{C}^{n \times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n \times r}$ that define a reduced-order system (3.2) in such a way that the reduced transfer function, $\mathbf{H}_r(s)$, is a *Hermite interpolant* of the full-order transfer function, $\mathbf{H}(s)$, at each σ_k along both left and right directions:

$$(3.3) \quad \begin{aligned} \mathbf{c}_i^T \mathbf{H}(\sigma_i) &= \mathbf{c}_i^T \mathbf{H}_r(\sigma_i), \quad \mathbf{H}(\sigma_i) \mathbf{b}_i = \mathbf{H}_r(\sigma_i) \mathbf{b}_i, \quad \text{and} \\ \mathbf{c}_i^T \mathbf{H}'(\sigma_i) \mathbf{b}_i &= \mathbf{c}_i^T \mathbf{H}'(\sigma_i) \mathbf{b}_i \quad \text{for } i = 1, \dots, r. \end{aligned}$$

$\mathbf{H}'(\sigma)$ here denotes the first derivative of $\mathbf{H}(s)$ with respect to s evaluated at σ . Since the matrix-valued function, $\mathbf{H}_r(s)$, consists of rational functions in s , (3.3) describes a *rational interpolation* problem. The following theorem gives elementary subspace criteria forcing interpolation.

THEOREM 3.1. *Let $\sigma \in \mathbb{C}$ be such that both $\sigma \mathbf{E} - \mathbf{A}$ and $\sigma \mathbf{E}_r - \mathbf{A}_r$ are invertible. If $\mathbf{b} \in \mathbb{C}^m$ and $\mathbf{c} \in \mathbb{C}^\ell$ are fixed nontrivial vectors, then*

- (a) if $(\sigma \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{b} \in \text{Ran}(\mathbf{V}_r)$, then $\mathbf{H}(\sigma) \mathbf{b} = \mathbf{H}_r(\sigma) \mathbf{b}$;
 (b) if $(\mathbf{c}^T \mathbf{C} (\sigma \mathbf{E} - \mathbf{A})^{-1})^T \in \text{Ran}(\mathbf{W}_r)$, then $\mathbf{c}^T \mathbf{H}(\sigma) = \mathbf{c}^T \mathbf{H}_r(\sigma)$; and
 (c) if both $(\sigma \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{b} \in \text{Ran}(\mathbf{V}_r)$ and $(\mathbf{c}^T \mathbf{C} (\sigma \mathbf{E} - \mathbf{A})^{-1})^T \in \text{Ran}(\mathbf{W}_r)$,
 then $\mathbf{c}^T \mathbf{H}'(\sigma) \mathbf{b} = \mathbf{c}^T \mathbf{H}'_r(\sigma) \mathbf{b}$.

Theorem 3.1 makes the solution of (3.3) straightforward. Given a set of distinct shifts $\{\sigma_k\}_{k=1}^r$, left-tangent directions $\{\mathbf{c}_k\}_{k=1}^r \subset \mathbb{C}^\ell$, and right-tangent directions $\{\mathbf{b}_k\}_{k=1}^r \subset \mathbb{C}^m$, construct full-rank matrices \mathbf{V}_r and \mathbf{W}_r such that

$$(3.4) \quad \text{Ran}(\mathbf{V}_r) \supseteq \text{span} \{[(\sigma_1 \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{b}_1, \dots, (\sigma_r \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{b}_r]\}$$

and

$$(3.5) \quad \text{Ran}(\mathbf{W}_r) \supseteq \text{span} \{[(\mathbf{c}_1^T \mathbf{C} (\sigma_1 \mathbf{E} - \mathbf{A})^{-1})^T, \dots, (\mathbf{c}_r^T \mathbf{C} (\sigma_r \mathbf{E} - \mathbf{A})^{-1})^T]\}.$$

If $\sigma_i \mathbf{E}_r - \mathbf{A}_r$ is nonsingular for each $i = 1, \dots, r$, then the reduced system $\mathbf{H}_r(s) = \mathbf{C}_r (s \mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r$ defined by $\mathbf{A}_r = \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r$, $\mathbf{E}_r = \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r$, $\mathbf{B}_r = \mathbf{W}_r^T \mathbf{B}$, and $\mathbf{C}_r = \mathbf{C} \mathbf{V}_r$ solves the tangential interpolation problem (3.3). In [4], Beattie and Gugercin showed how to solve the tangential interpolation problem posed in (3.3) for a substantially larger class of transfer functions—those having a coprime factorization of the form $\mathbf{H}(s) = \mathbf{C}(s) \mathbf{K}(s)^{-1} \mathbf{B}(s)$ with $\mathbf{B}(s)$, $\mathbf{C}(s)$, and $\mathbf{K}(s)$ given as meromorphic matrix-valued functions. This generalization lays the foundation of our present developments for parametrized model reduction described here.

The fidelity of the final reduced-order model must always be of central concern and clearly the selection of interpolation points and tangent directions becomes the main factor in determining success or failure. Until recently, selection of interpolation points was largely ad hoc. Recently however, Gugercin, Antoulas, and Beattie [25] showed an optimal shift selection strategy that produces reduced-order systems that are optimal \mathcal{H}_2 approximations to the original system. An optimal \mathcal{H}_2 approximant to the system $\mathbf{H}(s)$ is a system $\tilde{\mathbf{H}}_r(s)$ of reduced order, r , which solves

$$\min_{\tilde{\mathbf{H}}_r \text{ is stable}} \|\mathbf{H} - \tilde{\mathbf{H}}_r\|_{\mathcal{H}_2}, \quad \text{where} \quad \|\mathbf{H}\|_{\mathcal{H}_2} := \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} \|\mathbf{H}(j\omega)\|_{\text{F}}^2 d\omega \right)^{1/2},$$

and $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm of a matrix.

The set over which the optimization problem is posed, the set of all stable dynamical systems of order no greater than r , is nonconvex, so obtaining a global minimizer is at best a hard task, and indeed, it can be intractable. One moves instead toward a more modest goal and generally seeks “good” reduced models that satisfy first-order necessary optimality conditions, in principle allowing the possibility of having a local minimizer as an outcome. Many have worked on this problem; see [7, 29, 31, 36, 40, 44, 50, 51, 55]. Interpolation-based \mathcal{H}_2 optimality conditions were developed first by Meier and Luenberger [40] for SISO systems. Analogous \mathcal{H}_2 optimality conditions for MIMO systems have been placed within an interpolation framework recently in [10, 25, 46]. This is summarized in the next theorem.

THEOREM 3.2. *Suppose $\tilde{\mathbf{H}}_r(s) = \mathbf{C}_r (s \mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r$ minimizes $\|\mathbf{H} - \tilde{\mathbf{H}}_r\|_{\mathcal{H}_2}$ over all (stable) r th-order transfer functions and that the associated reduced-order pencil $s \mathbf{E}_r - \mathbf{A}_r$ has distinct eigenvalues $\{\tilde{\lambda}_i\}_{i=1}^r$. Let \mathbf{y}_i^* and \mathbf{x}_i denote left and right eigenvectors associated with $\tilde{\lambda}_i$ so that $\mathbf{A}_r \mathbf{x}_i = \tilde{\lambda}_i \mathbf{E}_r \mathbf{x}_i$, $\mathbf{y}_i^* \mathbf{A}_r = \tilde{\lambda}_i \mathbf{y}_i^* \mathbf{E}_r$, and $\mathbf{y}_i^* \mathbf{E}_r \mathbf{x}_j = \delta_{ij}$. Define $\tilde{\mathbf{c}}_i = \mathbf{C}_r \mathbf{x}_i$ and $\tilde{\mathbf{b}}_i^T = \mathbf{y}_i^* \mathbf{B}_r$.*

Then the residue of $\tilde{\mathbf{H}}_r(s)$ at $\tilde{\lambda}_i$ is matrix-valued and rank one: $\text{res}[\tilde{\mathbf{H}}_r(s), \tilde{\lambda}_i] = \tilde{\mathbf{c}}_i \tilde{\mathbf{b}}_i^T$. We can write $\tilde{\mathbf{H}}_r(s) = \sum_{i=1}^r \frac{1}{s-\tilde{\lambda}_i} \tilde{\mathbf{c}}_i \tilde{\mathbf{b}}_i^T$. Then, for $i = 1, \dots, r$,

$$(3.6) \quad \begin{aligned} \mathbf{H}(-\tilde{\lambda}_i) \tilde{\mathbf{b}}_i &= \tilde{\mathbf{H}}_r(-\tilde{\lambda}_i) \tilde{\mathbf{b}}_i, & \tilde{\mathbf{c}}_i^T \mathbf{H}(-\tilde{\lambda}_i) &= \tilde{\mathbf{c}}_i^T \tilde{\mathbf{H}}_r(-\tilde{\lambda}_i), \\ \text{and } \tilde{\mathbf{c}}_i^T \mathbf{H}'(-\tilde{\lambda}_i) \tilde{\mathbf{b}}_i &= \tilde{\mathbf{c}}_i^T \tilde{\mathbf{H}}_r'(-\tilde{\lambda}_i) \tilde{\mathbf{b}}_i. \end{aligned}$$

That is, first-order conditions for \mathcal{H}_2 optimality can be formulated as tangential interpolation conditions at reflected images of $\tilde{\lambda}_i$ through the origin.

Evidently, the \mathcal{H}_2 optimal interpolation points and associated tangent directions depend on knowledge of the reduced-order system and so will not be available a priori. An iterative algorithm was introduced in [25], called the *Iterative Rational Krylov Algorithm* (IRKA), built on successive substitution. Interpolation points used for the next step are chosen to be the reflected images of reduced-order poles for the current step: $\sigma \leftarrow -\tilde{\lambda}$ for eigenvalues, $\tilde{\lambda}_i$, of the pencil $\lambda \mathbf{E}_r - \mathbf{A}_r$ associated with reduced matrices of the current step. The tangent directions are corrected in a similar way, using residues of the previous reduced model successively until (3.6) is satisfied. A brief sketch of IRKA is described in Algorithm 3.1.

From steps 3(d) and 3(e), one sees that upon convergence, the reduced transfer function will satisfy, (3.6), first-order conditions for \mathcal{H}_2 optimality. The main computational cost involves solving $2r$ linear systems at every step to generate \mathbf{V}_r and \mathbf{W}_r . Computing the left and right eigenvectors \mathbf{y}_i and \mathbf{x}_i , and eigenvalues, $\lambda_i(\mathbf{A}_r, \mathbf{E}_r)$, of the reduced pencil $\lambda \mathbf{E}_r - \mathbf{A}_r$ is cheap since the dimension r is small.

ALGORITHM 3.1. MIMO \mathcal{H}_2 OPTIMAL TANGENTIAL INTERPOLATION METHOD.

1. Make an initial r -fold shift selection: $\{\sigma_1, \dots, \sigma_r\}$ that is closed under conjugation (i.e., $\{\sigma_1, \dots, \sigma_r\} \equiv \{\bar{\sigma}_1, \dots, \bar{\sigma}_r\}$ viewed as sets) and initial tangent directions $\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_r$ and $\tilde{\mathbf{c}}_1, \dots, \tilde{\mathbf{c}}_r$, also closed under conjugation.
2. $\mathbf{V}_r = \left[(\sigma_1 \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \tilde{\mathbf{b}}_1, \dots, (\sigma_r \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \tilde{\mathbf{b}}_r \right]$,
 $\mathbf{W}_r^T = \left[(\tilde{\mathbf{c}}_1^T \mathbf{C} (\sigma_1 \mathbf{E} - \mathbf{A})^{-1})^T, \dots, (\tilde{\mathbf{c}}_r^T \mathbf{C} (\sigma_r \mathbf{E} - \mathbf{A})^{-1})^T \right]^T$.
3. while (not converged)
 - (a) $\mathbf{A}_r = \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r$, $\mathbf{E}_r = \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r$, $\mathbf{B}_r = \mathbf{W}_r^T \mathbf{B}$, and $\mathbf{C}_r = \mathbf{C} \mathbf{V}_r$.
 - (b) Compute $\mathbf{A}_r \mathbf{x}_i = \lambda_i \mathbf{E}_r \mathbf{x}_i$ and $\mathbf{y}_i^* \mathbf{A}_r = \tilde{\lambda}_i \mathbf{y}_i^* \mathbf{E}_r$ with $\mathbf{y}_i^* \mathbf{E}_r \mathbf{x}_j = \delta_{ij}$ where \mathbf{y}_i^* and \mathbf{x}_i are left and right eigenvectors associated with λ_i .
 - (c) $\sigma_i \leftarrow -\lambda_i$, $\tilde{\mathbf{b}}_i^T \leftarrow \mathbf{y}_i^* \mathbf{B}_r$ and $\tilde{\mathbf{c}}_i \leftarrow \mathbf{C}_r \mathbf{x}_i$, for $i = 1, \dots, r$.
 - (d) $\mathbf{V}_r = \left[(\sigma_1 \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \tilde{\mathbf{b}}_1, \dots, (\sigma_r \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \tilde{\mathbf{b}}_r \right]$.
 - (e) $\mathbf{W}_r^T = \left[(\tilde{\mathbf{c}}_1^T \mathbf{C} (\sigma_1 \mathbf{E} - \mathbf{A})^{-1})^T, \dots, (\tilde{\mathbf{c}}_r^T \mathbf{C} (\sigma_r \mathbf{E} - \mathbf{A})^{-1})^T \right]^T$.
4. $\mathbf{A}_r = \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r$, $\mathbf{E}_r = \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r$, $\mathbf{B}_r = \mathbf{W}_r^T \mathbf{B}$, $\mathbf{C}_r = \mathbf{C} \mathbf{V}_r$.

4. Interpolatory model reduction of parameterized systems. We are able to extend the results of the previous section in a natural way to an interpolation framework for applying PMOR to the parameterized system (2.1)–(2.2) in order to produce a parameterized reduced system (2.3)–(2.4). In addition to the basic interpolation conditions for the transfer function as in (3.6), we develop conditions that also will guarantee matching of both the gradient and Hessian of the transfer function *with respect to the parameters*. Our framework allows parameter dependency (linear or nonlinear) in all state-space quantities.

THEOREM 4.1. *Suppose $\sigma \in \mathbb{C}$ and $\hat{\mathbf{p}} \in \mathbb{C}^\nu$ is such that both $\sigma \mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}})$ and $\sigma \mathbf{E}_r(\hat{\mathbf{p}}) - \mathbf{A}_r(\hat{\mathbf{p}})$ are invertible. Suppose $\mathbf{b} \in \mathbb{C}^m$ and $\mathbf{c} \in \mathbb{C}^\ell$ are fixed nontrivial vectors.*

(4.1) (a) *If $(\sigma \mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}}))^{-1} \mathbf{B}(\hat{\mathbf{p}})\mathbf{b} \in \text{Ran}(\mathbf{V}_r)$, then $\mathbf{H}(\sigma, \hat{\mathbf{p}})\mathbf{b} = \mathbf{H}_r(\sigma, \hat{\mathbf{p}})\mathbf{b}$.*

(4.2) (b) *If $(\mathbf{c}^T \mathbf{C}(\hat{\mathbf{p}}) (\sigma \mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}}))^{-1})^T \in \text{Ran}(\mathbf{W}_r)$, then $\mathbf{c}^T \mathbf{H}(\sigma, \hat{\mathbf{p}}) = \mathbf{c}^T \mathbf{H}_r(\sigma, \hat{\mathbf{p}})$.*

Proof. Define $\mathcal{A}(s, \mathbf{p}) = s\mathbf{E}(\mathbf{p}) - \mathbf{A}(\mathbf{p})$ and $\mathcal{A}_r(s, \mathbf{p}) = s\mathbf{E}_r(\mathbf{p}) - \mathbf{A}_r(\mathbf{p}) = \mathbf{W}_r^T \mathcal{A}(s, \mathbf{p}) \mathbf{V}_r$, and consider the (skew) projections

$$\mathcal{P}_r(s, \mathbf{p}) = \mathbf{V}_r \mathcal{A}_r(s, \mathbf{p})^{-1} \mathbf{W}_r^T \mathcal{A}(s, \mathbf{p}) \quad \text{and} \quad \mathcal{Q}_r(s, \mathbf{p}) = \mathcal{A}(s, \mathbf{p}) \mathbf{V}_r \mathcal{A}_r(s, \mathbf{p})^{-1} \mathbf{W}_r^T.$$

Define $\mathbf{f}(s, \mathbf{p}) = \mathcal{A}(s, \mathbf{p})^{-1} \mathbf{B}(\mathbf{p})\mathbf{b}$ and $\mathbf{g}^T(s, \mathbf{p}) = \mathbf{c}^T \mathbf{C}(\mathbf{p}) \mathcal{A}(s, \mathbf{p})^{-1}$. Then observe that the hypotheses of (4.1) means $\mathbf{f}(\sigma, \hat{\mathbf{p}}) \in \text{Ran}(\mathcal{P}_r(\sigma, \hat{\mathbf{p}}))$ and thus

$$\mathbf{H}(\sigma, \hat{\mathbf{p}})\mathbf{b} - \mathbf{H}_r(\sigma, \hat{\mathbf{p}})\mathbf{b} = \mathbf{C}(\hat{\mathbf{p}}) (\mathbf{I} - \mathcal{P}_r(\sigma, \hat{\mathbf{p}})) \mathbf{f}(\sigma, \hat{\mathbf{p}}) = 0,$$

proving (a). Analogously, the hypotheses of (4.2) means $\mathbf{g}(\sigma, \hat{\mathbf{p}}) \perp \text{Ker}(\mathcal{Q}_r(\sigma, \hat{\mathbf{p}}))$ and

$$\mathbf{c}^T \mathbf{H}(\sigma, \hat{\mathbf{p}}) - \mathbf{c}^T \mathbf{H}_r(\sigma, \hat{\mathbf{p}}) = \mathbf{g}^T(\sigma, \hat{\mathbf{p}}) (\mathbf{I} - \mathcal{Q}_r(\sigma, \hat{\mathbf{p}})) \mathbf{B}(\hat{\mathbf{p}}) = 0,$$

yielding (b). \square

Next, we show how to construct an interpolatory reduced-order model whose transfer function not only interpolates the original one, but which also forces matching of parameter-gradient values.

THEOREM 4.2. *Assume the hypotheses of Theorem 4.1. Suppose, in addition, that $\mathbf{E}(\mathbf{p})$, $\mathbf{A}(\mathbf{p})$, $\mathbf{B}(\mathbf{p})$, and $\mathbf{C}(\mathbf{p})$ are continuously differentiable in a neighborhood of $\hat{\mathbf{p}}$. Then both $\mathbf{c}^T \mathbf{H}(\sigma, \mathbf{p})\mathbf{b}$ and $\mathbf{c}^T \mathbf{H}_r(\sigma, \mathbf{p})\mathbf{b}$ are differentiable with respect to \mathbf{p} in a neighborhood of $\hat{\mathbf{p}}$ as well.*

(4.3) *If both $(\sigma \mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}}))^{-1} \mathbf{B}(\hat{\mathbf{p}})\mathbf{b} \in \text{Ran}(\mathbf{V}_r)$ and $(\mathbf{c}^T \mathbf{C}(\hat{\mathbf{p}}) (\sigma \mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}}))^{-1})^T \in \text{Ran}(\mathbf{W}_r)$,*

(4.4) *then $\nabla_{\mathbf{p}} \mathbf{c}^T \mathbf{H}(\sigma, \hat{\mathbf{p}})\mathbf{b} = \nabla_{\mathbf{p}} \mathbf{c}^T \mathbf{H}_r(\sigma, \hat{\mathbf{p}})\mathbf{b}$.*

From Theorem 3.1, these conditions also guarantee that $\mathbf{c}^T \mathbf{H}'(\sigma, \hat{\mathbf{p}})\mathbf{b} = \mathbf{c}^T \mathbf{H}'_r(\sigma, \hat{\mathbf{p}})\mathbf{b}$ (where again, ' denotes differentiation with respect to the frequency parameter, s).

Proof. Fix an arbitrary nontrivial direction $\mathbf{n} = [n_1, \dots, n_\nu]^T \in \mathbb{C}^\nu$ and denote the associated directional derivative as

$$\mathbf{n} \cdot \nabla_{\mathbf{p}} = \sum_{i=1}^{\nu} n_i \frac{\partial}{\partial p_i}.$$

Note that for all s and \mathbf{p} at which \mathcal{P}_r and \mathcal{Q}_r are continuous, we have: $\text{Ran}((\mathbf{n} \cdot \nabla_{\mathbf{p}}) \mathcal{P}_r(s, \mathbf{p})) \subset \text{Ran}(\mathcal{P}_r(s, \mathbf{p}))$ and $\text{Ker}((\mathbf{n} \cdot \nabla_{\mathbf{p}}) \mathcal{Q}_r(s, \mathbf{p})) \supset \text{Ker}(\mathcal{Q}_r(s, \mathbf{p}))$. Thus

$$(4.5) \quad (\mathbf{I} - \mathcal{P}_r(s, \mathbf{p})) [(\mathbf{n} \cdot \nabla_{\mathbf{p}}) \mathcal{P}_r(s, \mathbf{p})] = 0 \quad \text{and} \quad [(\mathbf{n} \cdot \nabla_{\mathbf{p}}) \mathcal{Q}_r(s, \mathbf{p})] (\mathbf{I} - \mathcal{Q}_r(s, \mathbf{p})) = 0.$$

As a consequence,

$$(\mathbf{n} \cdot \nabla_{\mathbf{p}}) [(\mathbf{I} - \mathcal{Q}_r(s, \mathbf{p})) \mathcal{A}(s, \mathbf{p}) (\mathbf{I} - \mathcal{P}_r(s, \mathbf{p}))] = (\mathbf{I} - \mathcal{Q}_r(s, \mathbf{p})) [(\mathbf{n} \cdot \nabla_{\mathbf{p}}) \mathcal{A}(s, \mathbf{p})] (\mathbf{I} - \mathcal{P}_r(s, \mathbf{p})).$$

Observe that

$$\mathbf{c}^T \mathbf{H}(s, \mathbf{p}) \mathbf{b} - \mathbf{c}^T \mathbf{H}_r(s, \mathbf{p}) \mathbf{b} = \mathbf{g}^T(s, \mathbf{p}) (\mathbf{I} - \mathbf{Q}_r(s, \mathbf{p})) \mathcal{A}(s, \mathbf{p}) (\mathbf{I} - \mathcal{P}_r(s, \mathbf{p})) \mathbf{f}(s, \mathbf{p}).$$

Thus, we may calculate a directional derivative and evaluate at $s = \sigma$ and $\mathbf{p} = \hat{\mathbf{p}}$:

$$\begin{aligned} & (\mathbf{n} \cdot \nabla \mathbf{p}) [\mathbf{c}^T \mathbf{H}(\sigma, \mathbf{p}) \mathbf{b} - \mathbf{c}^T \mathbf{H}_r(\sigma, \mathbf{p}) \mathbf{b}] \Big|_{\mathbf{p}=\hat{\mathbf{p}}} \\ &= [(\mathbf{n} \cdot \nabla \mathbf{p}) \mathbf{g}^T(\sigma, \hat{\mathbf{p}})] (\mathbf{I} - \mathbf{Q}_r(\sigma, \hat{\mathbf{p}})) \mathcal{A}(\sigma, \hat{\mathbf{p}}) (\mathbf{I} - \mathcal{P}_r(\sigma, \hat{\mathbf{p}})) \mathbf{f}(\sigma, \hat{\mathbf{p}}) \\ &\quad + \mathbf{g}^T(\sigma, \hat{\mathbf{p}}) (\mathbf{I} - \mathbf{Q}_r(\sigma, \hat{\mathbf{p}})) [(\mathbf{n} \cdot \nabla \mathbf{p}) \mathcal{A}(\sigma, \hat{\mathbf{p}})] (\mathbf{I} - \mathcal{P}_r(\sigma, \hat{\mathbf{p}})) \mathbf{f}(\sigma, \hat{\mathbf{p}}) \\ &\quad + \mathbf{g}^T(\sigma, \hat{\mathbf{p}}) (\mathbf{I} - \mathbf{Q}_r(\sigma, \hat{\mathbf{p}})) \mathcal{A}(\sigma, \hat{\mathbf{p}}) (\mathbf{I} - \mathcal{P}_r(\sigma, \hat{\mathbf{p}})) [(\mathbf{n} \cdot \nabla \mathbf{p}) \mathbf{f}(\sigma, \hat{\mathbf{p}})]. \end{aligned}$$

The hypothesis of (4.3) implies $\mathbf{f}(\sigma, \hat{\mathbf{p}}) \in \text{Ran}(\mathcal{P}_r(\sigma, \hat{\mathbf{p}}))$ and $\mathbf{g}(\sigma, \hat{\mathbf{p}}) \perp \text{Ker}(\mathbf{Q}_r(\sigma, \hat{\mathbf{p}}))$ so $(\mathbf{n} \cdot \nabla \mathbf{p}) [\mathbf{c}^T \mathbf{H}(\sigma, \mathbf{p}) \mathbf{b} - \mathbf{c}^T \mathbf{H}_r(\sigma, \mathbf{p}) \mathbf{b}] \Big|_{\mathbf{p}=\hat{\mathbf{p}}} = 0$. Since \mathbf{n} was arbitrarily chosen the conclusion follows. \square

Notice that for SISO systems (where tangent directions play no role), we create a parameterized reduced system, $H_r(s, \mathbf{p})$; not only is it a Hermite interpolant (with respect to frequency) to $H(s, \mathbf{p})$ at $(\sigma, \hat{\mathbf{p}})$ but the \mathbf{p} -gradients of H_r and H also match at $(\sigma, \hat{\mathbf{p}})$. Furthermore, we can guarantee this additional matching for essentially no greater cost and *without* computing the \mathbf{p} -gradient of either $H_r(s, \mathbf{p})$ or $H(s, \mathbf{p})$. This is a significant feature with regard to sensitivity analysis [11]: notice that the parameterized reduced-order model may be used to compute parameter sensitivities more cheaply than the original model and will exactly match the original model sensitivities at every parameter interpolation point, $\hat{\mathbf{p}}$. See also [30, 48] for recent methods that use sensitivity data and PMOR type methods.

There are also interesting consequences for optimization with respect to \mathbf{p} of objective functions depending on $\mathbf{H}(s, \mathbf{p})$ (or on the output $\hat{\mathbf{y}}(s, \mathbf{p})$ for a fixed input $\hat{\mathbf{u}}$). Under natural auxiliary conditions, reduced-order models satisfying the conditions (4.3) of Theorem 4.2 will lead to, in the terminology of [1], *first-order accurate approximate models* for the objective function and this feature is sufficient in establishing robust convergence behavior of related trust region methods utilizing reduced-order models as surrogate models.

In the context of optimization, the next obvious question is under what conditions will a reduced-order model retain the same *curvature* or *Hessian* information with respect to parameters as the original model?

THEOREM 4.3. *Assume the hypotheses of Theorem 4.2 including (4.3) and suppose that $\mathbf{E}(\mathbf{p})$, $\mathbf{A}(\mathbf{p})$, $\mathbf{B}(\mathbf{p})$, and $\mathbf{C}(\mathbf{p})$ are twice continuously differentiable in a neighborhood of $\hat{\mathbf{p}}$. Then $\mathbf{c}^T \mathbf{H}(\sigma, \mathbf{p}) \mathbf{b}$ and $\mathbf{c}^T \mathbf{H}_r(\sigma, \mathbf{p}) \mathbf{b}$ are each twice continuously differentiable at $\hat{\mathbf{p}}$.*

(a) *Let $\{\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_\nu\}$ be a basis for \mathbb{C}^ν with related quantities*

$$\begin{aligned} \mathbf{f}_i(\sigma, \hat{\mathbf{p}}) &= (\mathbf{n}_i \cdot \nabla \mathbf{p}) (\sigma \mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}}))^{-1} \mathbf{B}(\hat{\mathbf{p}}) \mathbf{b} \quad \text{and} \\ \mathbf{g}_i^T(\sigma, \hat{\mathbf{p}}) &= (\mathbf{n}_i \cdot \nabla \mathbf{p}) \mathbf{c}^T \mathbf{C}(\hat{\mathbf{p}}) (\sigma \mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}}))^{-1}. \end{aligned}$$

(4.6) *If either $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_\nu\} \subset \text{Ran}(\mathbf{V}_r)$*

(4.7) *or $\{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_\nu\} \subset \text{Ran}(\mathbf{W}_r)$,*

then $\nabla_{\mathbf{p}}^2 [\mathbf{c}^T \mathbf{H}(\sigma, \hat{\mathbf{p}}) \mathbf{b}] = \nabla_{\mathbf{p}}^2 [\mathbf{c}^T \mathbf{H}_r(\sigma, \hat{\mathbf{p}}) \mathbf{b}]$.

(b) Let \mathbf{n} be a fixed nontrivial vector in \mathbb{C}^ν and suppose that

$$\begin{aligned} (\mathbf{n} \cdot \nabla_{\mathbf{p}})(\sigma \mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}}))^{-1} \mathbf{B}(\hat{\mathbf{p}})\mathbf{b} &\in \text{Ran}(\mathbf{V}_r) \text{ and} \\ (\mathbf{n} \cdot \nabla_{\mathbf{p}})\left(\mathbf{c}^T \mathbf{C}(\hat{\mathbf{p}})(\sigma \mathbf{E}(\hat{\mathbf{p}}) - \mathbf{A}(\hat{\mathbf{p}}))^{-1}\right)^T &\in \text{Ran}(\mathbf{W}_r). \end{aligned}$$

$$(4.8) \quad \text{Then} \quad \nabla_{\hat{\mathbf{p}}}^2[\mathbf{c}^T \mathbf{H}(\sigma, \hat{\mathbf{p}})\mathbf{b}] \mathbf{n} = \nabla_{\hat{\mathbf{p}}}^2[\mathbf{c}^T \mathbf{H}_r(\sigma, \hat{\mathbf{p}})\mathbf{b}] \mathbf{n}.$$

Proof. Let $\mathbf{n} = [n_1, \dots, n_\nu]^T$ and $\mathbf{m} = [m_1, \dots, m_\nu]^T$ be arbitrary vectors in \mathbb{C}^ν and consider the composition of the associated directional derivatives:

$$(\mathbf{m} \cdot \nabla_{\mathbf{p}})(\mathbf{n} \cdot \nabla_{\mathbf{p}})\left[\mathbf{c}^T \mathbf{H}(\sigma, \mathbf{p})\mathbf{b} - \mathbf{c}^T \mathbf{H}_r(\sigma, \mathbf{p})\mathbf{b}\right]\Big|_{\mathbf{p}=\hat{\mathbf{p}}} = \mathbf{m}^T \nabla_{\hat{\mathbf{p}}}^2[\mathbf{c}^T \mathbf{H}(\sigma, \hat{\mathbf{p}})\mathbf{b} - \mathbf{c}^T \mathbf{H}_r(\sigma, \hat{\mathbf{p}})\mathbf{b}] \mathbf{n}.$$

Using (4.5), one may calculate

$$\begin{aligned} &(\mathbf{m} \cdot \nabla_{\mathbf{p}})(\mathbf{n} \cdot \nabla_{\mathbf{p}})[(\mathbf{I} - \mathbf{Q}_r(s, \mathbf{p}))\mathcal{A}(s, \mathbf{p})(\mathbf{I} - \mathcal{P}_r(s, \mathbf{p}))] \\ &= -[(\mathbf{m} \cdot \nabla_{\mathbf{p}})\mathbf{Q}_r(s, \mathbf{p})][(\mathbf{n} \cdot \nabla_{\mathbf{p}})\mathcal{A}(s, \mathbf{p})](\mathbf{I} - \mathcal{P}_r(s, \mathbf{p})) \\ &\quad + (\mathbf{I} - \mathbf{Q}_r(s, \mathbf{p}))[(\mathbf{m} \cdot \nabla_{\mathbf{p}})(\mathbf{n} \cdot \nabla_{\mathbf{p}})\mathcal{A}(s, \mathbf{p})](\mathbf{I} - \mathcal{P}_r(s, \mathbf{p})) \\ &\quad - (\mathbf{I} - \mathbf{Q}_r(s, \mathbf{p}))[(\mathbf{n} \cdot \nabla_{\mathbf{p}})\mathcal{A}(s, \mathbf{p})][(\mathbf{m} \cdot \nabla_{\mathbf{p}})\mathcal{P}_r(s, \mathbf{p})]. \end{aligned}$$

Then with (4.3), one finds

$$\begin{aligned} &\mathbf{m}^T \nabla_{\hat{\mathbf{p}}}^2[\mathbf{c}^T \mathbf{H}(\sigma, \hat{\mathbf{p}})\mathbf{b} - \mathbf{c}^T \mathbf{H}_r(\sigma, \hat{\mathbf{p}})\mathbf{b}] \mathbf{n} \\ &= \left[[(\mathbf{m} \cdot \nabla_{\mathbf{p}})\mathbf{c}^T \mathbf{C}(\mathbf{p})\mathcal{A}(s, \mathbf{p})^{-1}] \cdot (\mathbf{I} - \mathbf{Q}_r(s, \mathbf{p}))\mathcal{A}(s, \mathbf{p})(\mathbf{I} - \mathcal{P}_r(s, \mathbf{p})) \cdot \right. \\ &\quad \left. [(\mathbf{n} \cdot \nabla_{\mathbf{p}})\mathcal{A}(s, \mathbf{p})^{-1}\mathbf{B}(\mathbf{p})\mathbf{b}] \right. \\ &\quad \left. + [(\mathbf{n} \cdot \nabla_{\mathbf{p}})\mathbf{c}^T \mathbf{C}(\mathbf{p})\mathcal{A}(s, \mathbf{p})^{-1}] \cdot (\mathbf{I} - \mathbf{Q}_r(s, \mathbf{p}))\mathcal{A}(s, \mathbf{p})(\mathbf{I} - \mathcal{P}_r(s, \mathbf{p})) \cdot \right. \\ &\quad \left. [(\mathbf{m} \cdot \nabla_{\mathbf{p}})\mathcal{A}(s, \mathbf{p})^{-1}\mathbf{B}(\mathbf{p})\mathbf{b}] \right]\Big|_{\mathbf{p}=\hat{\mathbf{p}}}. \end{aligned}$$

If (4.6) holds, then both vectors $(\mathbf{m} \cdot \nabla_{\mathbf{p}})\mathcal{A}(s, \hat{\mathbf{p}})^{-1}\mathbf{B}(\hat{\mathbf{p}})\mathbf{b}$ and $(\mathbf{n} \cdot \nabla_{\mathbf{p}})\mathcal{A}(s, \hat{\mathbf{p}})^{-1}\mathbf{B}(\hat{\mathbf{p}})\mathbf{b}$ are in $\text{Ran}(\mathcal{P}_r(\sigma, \hat{\mathbf{p}}))$, leading to the conclusion of (a), since \mathbf{m} and \mathbf{n} could be arbitrarily chosen. A similar argument holds if (4.7) is true.

If the hypotheses of (b) hold, then observe that

$$\mathbf{m}^T \nabla_{\hat{\mathbf{p}}}^2[\mathbf{c}^T \mathbf{H}(\sigma, \hat{\mathbf{p}})\mathbf{b} - \mathbf{c}^T \mathbf{H}_r(\sigma, \hat{\mathbf{p}})\mathbf{b}] \mathbf{n} = 0,$$

independent of how \mathbf{m} is chosen, which then yields the conclusion (4.8). \square

A generic implementation of PMOR using interpolatory projections as described in Theorem 4.1 is provided in Algorithm 4.1, where we continue to use the notation $\mathcal{A}(s, \mathbf{p}) := s\mathbf{E}(\mathbf{p}) - \mathbf{A}(\mathbf{p})$ as we have above. Note that the number of interpolation frequencies, K , and the number of interpolation points for parameter vectors, L , needs to be chosen a priori; the total model order is (nominally) $r = LK$.

If we were to attempt interpolation of the full transfer function using the same interpolation points, we would need

$$\text{Ran}(\mathbf{V}_r) \supset \text{span}_{\substack{i=1, \dots, K \\ j=1, \dots, L}} \left\{ \mathcal{A}(\sigma_i, \mathbf{p}^{(j)})^{-1} \mathbf{B}(\mathbf{p}^{(j)}) \right\}.$$

$\text{Ran}(\mathbf{V}_r)$ could thus have dimension as large as mLK , and there exist many applications where the system input dimension m indeed is rather large, perhaps on the order

ALGORITHM 4.1. PMOR WITH INTERPOLATORY PROJECTIONS.

1. Select “frequencies” $\sigma_1, \dots, \sigma_K \in \mathbb{C}$, parameter vectors $\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(L)} \in \mathbb{R}^\nu$, left tangent directions $\{\mathbf{c}_{11}, \dots, \mathbf{c}_{1,L}, \mathbf{c}_{21}, \dots, \mathbf{c}_{K,L}\} \subset \mathbb{C}^\ell$, and right tangent directions $\{\mathbf{b}_{11}, \dots, \mathbf{b}_{1,L}, \mathbf{b}_{21}, \dots, \mathbf{b}_{K,L}\} \subset \mathbb{C}^m$.
The order of the reduced model will be $r = K \cdot L$.
2. Compute a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ for

$$\mathcal{V}_r = \text{span}_{\substack{i=1, \dots, K \\ j=1, \dots, L}} \left\{ \mathcal{A} \left(\sigma_i, \mathbf{p}^{(j)} \right)^{-1} \mathbf{B}(\mathbf{p}^{(j)}) \mathbf{b}_{ij} \right\}.$$

3. Compute a basis $\{\mathbf{w}_1, \dots, \mathbf{w}_r\}$ for

$$\mathcal{W}_r = \text{span}_{\substack{i=1, \dots, K \\ j=1, \dots, L}} \left\{ \left(\mathbf{c}_{ij}^T \mathbf{C}(\mathbf{p}^{(j)}) \mathcal{A} \left(\sigma_i, \mathbf{p}^{(j)} \right)^{-1} \right)^T \right\}.$$

4. Set $\mathbf{V}_r := [\mathbf{v}_1, \dots, \mathbf{v}_r]$ and $\mathbf{W}_r := [\mathbf{w}_1, \dots, \mathbf{w}_r]$.
5. (Pre)compute from (2.4): $\mathbf{A}_r(\mathbf{p}) = \mathbf{W}_r^T \mathbf{A}(\mathbf{p}) \mathbf{V}_r$, $\mathbf{E}_r(\mathbf{p}) = \mathbf{W}_r^T \mathbf{E}(\mathbf{p}) \mathbf{V}_r$,
 $\mathbf{B}_r(\mathbf{p}) = \mathbf{W}_r^T \mathbf{B}(\mathbf{p})$, $\mathbf{C}_r(\mathbf{p}) = \mathbf{C}(\mathbf{p}) \mathbf{V}_r$.

of hundreds, leading then to a forbiddingly large reduced-order dimension. Tangential interpolation by contrast is more frugal in its use of interpolation information. For full matrix interpolation, every interpolation point adds m columns to \mathbf{V}_r , while for tangential interpolation each interpolation point will add only a single column.

Certainly, the performance of the procedure depends strongly on the choice of interpolation data. A first refinement of this basic approach is to compute frequency points for a fixed selection of parameter vectors that are locally optimal with respect to \mathcal{H}_2 error measures using the IRKA as in [25]. Choosing both the frequency and the parameter interpolation data as well as the tangent directions in an optimal way will be discussed in the next section.

5. An \mathcal{H}_2 -based approach to parameterized model reduction. Algorithm 4.1 will produce a parameterized reduced-order model that interpolates the original system in the tangent directions \mathbf{b}_i and \mathbf{c}_i^T at the (complex) frequency σ_i and parameter values, $\hat{\mathbf{p}}_j$. In many problem scenarios, there will be a natural choice of parameter vectors that will be representative of the parameter ranges within which the original system must operate. Sometimes designers will specify important parameter sets in the neighborhood of which reduced-order models should be particularly accurate. In other cases, the physics of the problem will provide some insight to where parameters should be chosen. In all these circumstances, the choice of interpolation data for parameter vectors has been made, leaving open the question of how best to choose the frequency interpolation data. We will give a heuristic approach to resolve this problem using methods for nonparameterized systems that can yield optimal \mathcal{H}_2 frequency interpolation points.

Given a full-order parameterized system $\mathbf{H}(s, \mathbf{p})$, suppose L different parameter vectors $\{\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(L)}\}$ are selected as parameter interpolation points. For each $\mathbf{p}^{(i)}$, define $\mathbf{H}^{(i)}(s) = \mathbf{H}(s, \mathbf{p}^{(i)})$. For each $i = 1, \dots, L$, $\mathbf{H}^{(i)}$ can be viewed as a (nonparameterized) full-order model and we may apply Algorithm 3.1 to each $\mathbf{H}^{(i)}(s)$ to obtain an \mathcal{H}_2 optimal reduced-order model (say, of order r_i) and corresponding projection

subspaces $\mathbf{V}^{(i)} \in \mathbb{R}^{n \times r_i}$ and $\mathbf{W}^{(i)} \in \mathbb{R}^{n \times r_i}$. Let $r = r_1 + r_2 + \dots + r_L$. We concatenate these matrices to get

$$\mathbf{V}_r = [\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \dots, \mathbf{V}^{(L)}] \in \mathbb{R}^{n \times r} \text{ and } \mathbf{W}_r = [\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \dots, \mathbf{W}^{(L)}] \in \mathbb{R}^{n \times r}.$$

This leads to the final parameterized reduced-order model, $\mathbf{H}_r(s, \mathbf{p})$, as in (2.3). Note that the $\mathbf{H}_r(s, \mathbf{p})$ will not be an \mathcal{H}_2 optimal system approximation to $\mathbf{H}(s, \mathbf{p})$ for any parameter choice although it contains L smaller \mathcal{H}_2 optimal submodels that can be recovered by truncation of \mathbf{H}_r evaluated at each of the L given parameter vectors. In any case, \mathbf{H}_r still interpolates \mathbf{H} at all parameter choices. A brief sketch of the method is given in Algorithm 5.1. Notice that the exact interpolation properties would be lost if we were to use a truncated SVD in step 4; even so, linear dependencies are removed only up to thresholds associated with machine precision. The construction of truncation matrices is similar to the trajectory piecewise approximation methods suggested in [43, 47]. Effectiveness of this algorithm is illustrated with several numerical examples in section 6.

ALGORITHM 5.1. PIECEWISE \mathcal{H}_2 OPTIMAL INTERPOLATORY PMOR.

1. Select L parameter vectors $\{\mathbf{p}^{(1)}, \mathbf{p}^{(2)}, \dots, \mathbf{p}^{(L)}\}$ and reduction orders $\{r_1, r_2, \dots, r_L\}$.
2. For each $i = 1, 2, \dots, L$
Define the i th system instance: $\mathbf{H}^{(i)}(s) = \mathbf{H}(s, \mathbf{p}^{(i)})$ and apply the optimal \mathcal{H}_2 reduction of Algorithm 3.1 to $\mathbf{H}^{(i)}(s)$, constructing interpolating spaces of dimension r_i spanned by $\mathbf{V}^{(i)}$ and $\mathbf{W}^{(i)}$.
3. Concatenate $\mathbf{V}^{(i)}$ and $\mathbf{W}^{(i)}$ for $i = 1, \dots, L$ to obtain the final projection matrices \mathbf{V}_r and \mathbf{W}_r of dimension $r = r_1 + \dots + r_L$:

$$\mathbf{V}_r = [\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \dots, \mathbf{V}^{(L)}] \text{ and } \mathbf{W}_r = [\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \dots, \mathbf{W}^{(L)}].$$

4. Use an SVD or rank-revealing QR factorization to remove rank-deficient components from \mathbf{V}_r and \mathbf{W}_r .

The final parameterized reduced model is determined by \mathbf{V}_r and \mathbf{W}_r from (2.3).

The situation becomes harder if we have no a priori knowledge of particular parameter values that are important but instead have perhaps only information about allowable parameter ranges within the parameter space. There are methods to address this difficulty. One possible approach is the so-called greedy selection algorithm of Bui-Thanh, Willcox, and Ghattas [8]. Even though the final reduced-order model of [8] proves to be a high quality approximation, the optimization algorithm that needs to be solved at each step could be computationally expensive, possibly prohibitively so. Another strategy for an effective and representative choice of parameter points in higher dimensional parameter spaces (for example, say, with $\nu = 10$) comes through the use of sparse grids [9, 23, 54]. This approach is based on a hierarchical basis and a sparse tensor product construction. The dimension of the sparse grid space is of order $\mathcal{O}(2^n n^{\nu-1})$ compared to the dimension of the corresponding full grid space given by $\mathcal{O}(2^{\nu n})$. See [3, 42] for other approaches to parameterized model reduction using sparse grids.

Heuristics such as these can provide effective choices for interpolation points. However, in the absence of compelling heuristic choices there is value in considering approaches that can lead to *optimal parameter selection points* that are chosen so as

to minimize error measures appropriate to parameterized systems. We consider this problem below and provide a solution for SISO systems having a specific parameter dependence.

5.1. Optimal interpolation for special SISO parameterizations. In the particular case that $\mathbf{H}(s, \mathbf{p})$ is a SISO system with the parametric dependence occurring solely in $\mathbf{C}(\mathbf{p})$ and $\mathbf{B}(\mathbf{p})$, we are able to produce reduced-order systems that are optimal with respect to a composite error measure that is an \mathcal{L}_2 error relative to parameters and \mathcal{H}_2 error relative to the system response. To illustrate, we consider a simple two-parameter case for a system having the form

$$(5.1) \quad H(s, \mathbf{p}) = \mathbf{c}^T(p) (s \mathbf{E} - \mathbf{A})^{-1} \mathbf{b}(q),$$

with $\mathbf{c}(p) = \mathbf{c}_0 + p \mathbf{c}_1$ and $\mathbf{b}(q) = \mathbf{b}_0 + q \mathbf{b}_1$,

where $\mathbf{p} = [p, q]^T$ and $0 \leq p, q \leq 1$. This setting can be generalized in many directions but serves to illustrate the main points.

Denoting $\mathcal{D} = [0, 1] \times [0, 1]$, define a norm for systems having the form (5.1):

$$(5.2) \quad \|H\|_{\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})}^2 \stackrel{\text{def}}{=} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \iint_{\mathcal{D}} |H(i\omega, \mathbf{p})|^2 dA(\mathbf{p}) d\omega.$$

Obviously other choices for \mathcal{D} and other measures aside from Lebesgue measure $dA(\mathbf{p})$ are possible (e.g., p and q can be random variables jointly distributed according to $dA(\mathbf{p})$). We seek an *optimal* reduced-order parameterized model, $\tilde{H}_r(s, \mathbf{p})$, having the same form as $H(s, \mathbf{p})$,

$$(5.3) \quad \tilde{H}_r(s, \mathbf{p}) = (\mathbf{c}_{0,r} + p \mathbf{c}_{1,r})^T (s \mathbf{E}_r - \mathbf{A}_r)^{-1} (\mathbf{b}_{0,r} + q \mathbf{b}_{1,r}),$$

such that

$$(5.4) \quad \|H - \tilde{H}_r\|_{\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})} = \min_{\substack{H_r \text{ stable} \\ \text{for all } \mathbf{p} \in \mathcal{D}}} \|H - H_r\|_{\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})}.$$

THEOREM 5.1. *Let $H(s, \mathbf{p})$ be given as in (5.1) and let $\mathcal{D} = [0, 1] \times [0, 1]$. Define the auxiliary MIMO transfer function:*

$$(5.5) \quad \mathcal{H}(s) = [\mathbf{c}_0, \mathbf{c}_1]^T (s \mathbf{E} - \mathbf{A})^{-1} [\mathbf{b}_0, \mathbf{b}_1].$$

Then $\|H\|_{\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})} = \|\mathbf{L}^T \mathcal{H} \mathbf{L}\|_{\mathcal{H}_2}$, where

$$\mathbf{L} = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2\sqrt{3}} \end{bmatrix}.$$

In particular, the norm we have defined on $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ for the parameterized system $H(s, \mathbf{p})$ is equivalent to a (weighted) MIMO \mathcal{H}_2 norm for $\mathcal{H}(s)$.

Proof. Observe that

$$(5.6) \quad H(s, \mathbf{p}) = [1, p] \mathcal{H}(s) \begin{bmatrix} 1 \\ q \end{bmatrix}.$$

Substitute this expression into (5.2), rearrange the integrand, and note that \mathbf{L} is the Cholesky factor of

$$\int_0^1 \begin{bmatrix} 1 \\ q \end{bmatrix} [1, q] dq = \int_0^1 \begin{bmatrix} 1 \\ p \end{bmatrix} [1, p] dp = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix} = \mathbf{L} \mathbf{L}^T. \quad \square$$

Although the model system we consider in (5.1) has a parameter range restricted to $\mathbf{p} = [p, q]^T \in \mathcal{D}$, interpolation is well-defined for parameter values outside of \mathcal{D} . Indeed, parameter interpolation will be well-defined even for $p = \infty$ or $q = \infty$: consider for nonzero (but finite) p, q the interpolation condition,

$$\begin{aligned} H(\sigma, \mathbf{p}) &= pq \left(\frac{1}{p} \mathbf{c}_0 + \mathbf{c}_1 \right)^T (\sigma \mathbf{E} - \mathbf{A})^{-1} \left(\frac{1}{q} \mathbf{b}_0 + \mathbf{b}_1 \right) \\ &= pq \left(\frac{1}{p} \mathbf{c}_{0,r} + \mathbf{c}_{1,r} \right)^T (\sigma \mathbf{E}_r - \mathbf{A}_r)^{-1} \left(\frac{1}{q} \mathbf{b}_{0,r} + \mathbf{b}_{1,r} \right) = H_r(\sigma, \mathbf{p}), \end{aligned}$$

and then let p or q (or both) approach ∞ . We interpret the interpolation condition $H(\sigma, [p, q]) = H_r(\sigma, [p, q])$ for such extended complex values for $\mathbf{p} = [p, q]$ as follows:

- $H(\sigma, [\infty, q]) = H_r(\sigma, [\infty, q])$ with q fixed and finite is interpreted as $\mathbf{c}_1^T (\sigma \mathbf{E} - \mathbf{A})^{-1} (\mathbf{b}_0 + q \mathbf{b}_1) = \mathbf{c}_{1,r}^T (\sigma \mathbf{E}_r - \mathbf{A}_r)^{-1} (\mathbf{b}_{0,r} + q \mathbf{b}_{1,r})$;
- $H(\sigma, [p, \infty]) = H_r(\sigma, [p, \infty])$ with p fixed and finite is interpreted as $(\mathbf{c}_0 + p \mathbf{c}_1)^T (\sigma \mathbf{E} - \mathbf{A})^{-1} \mathbf{b}_1 = (\mathbf{c}_{0,r} + p \mathbf{c}_{1,r})^T (\sigma \mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{b}_{1,r}$;
- $H(\sigma, [\infty, \infty]) = H_r(\sigma, [\infty, \infty])$ is interpreted as $\mathbf{c}_1^T (\sigma \mathbf{E} - \mathbf{A})^{-1} \mathbf{b}_1 = \mathbf{c}_{1,r}^T (\sigma \mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{b}_{1,r}$.

Similar extensions can be made for derivative interpolation conditions.

Theorem 5.1 shows that the least-squares error measure in the $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ norm for the SISO parametric system is indeed a MIMO \mathcal{H}_2 norm for a nonparameterized linear system. This means we can solve the parametric $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ optimization problem (5.4) by solving an equivalent nonparameterized MIMO \mathcal{H}_2 optimization problem which we know how to solve using Theorem 3.2 and Algorithm 3.1. This leads to the following result.

THEOREM 5.2. *Let $H(s, \mathbf{p})$ be given as in (5.1). Suppose a parameterized reduced-order model $\tilde{H}_r(s, \mathbf{p})$ of the form (5.3) minimizes $\|H - \tilde{H}_r\|_{\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})}$ over all (stable) r th-order transfer functions and that the associated reduced-order pencil $s\mathbf{E}_r - \mathbf{A}_r$ has only simple eigenvalues $\{\tilde{\lambda}_i\}_{i=1}^r$. Then there are optimal frequency shifts, $\{-\tilde{\lambda}_i\}_{i=1}^r$, and optimal parameter interpolation vectors, $\{\tilde{\mathbf{p}}_i\}_{i=1}^r$ such that*

$$(5.7) \quad \begin{aligned} H(-\tilde{\lambda}_i, \tilde{\mathbf{p}}_i) &= \tilde{H}_r(-\tilde{\lambda}_i, \tilde{\mathbf{p}}_i), & H'(-\tilde{\lambda}_i, \tilde{\mathbf{p}}_i) &= \tilde{H}'_r(-\tilde{\lambda}_i, \tilde{\mathbf{p}}_i), \\ \text{and } \nabla_{\mathbf{p}} H(-\tilde{\lambda}_i, \tilde{\mathbf{p}}_i) &= \nabla_{\mathbf{p}} \tilde{H}_r(-\tilde{\lambda}_i, \tilde{\mathbf{p}}_i) \end{aligned}$$

for $i = 1, \dots, r$ ($'$ denotes differentiation with respect to the frequency parameter, s).

Proof. Define a reduced-order MIMO system associated with \tilde{H}_r :

$$\tilde{\mathcal{H}}_r(s) = [\mathbf{c}_{0,r}, \mathbf{c}_{1,r}]^T (s \mathbf{E}_r - \mathbf{A}_r)^{-1} [\mathbf{b}_{0,r}, \mathbf{b}_{1,r}].$$

Analogously to (5.6), we have $\tilde{H}_r(s, \mathbf{p}) = [1, p] \tilde{\mathcal{H}}_r(s) \begin{bmatrix} 1 \\ p \end{bmatrix}$.

Since $\tilde{H}_r(s, \mathbf{p})$ minimizes the $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ error from the original system $H(s, \mathbf{p})$, we find an equivalent weighted \mathcal{H}_2 approximation problem:

$$\begin{aligned} \|\mathbf{L}^T \mathcal{H} \mathbf{L} - \mathbf{L}^T \tilde{\mathcal{H}}_r \mathbf{L}\|_{\mathcal{H}_2} &= \|H - \tilde{H}_r\|_{\mathcal{H}_2 \otimes \mathcal{L}_2} = \min_{\tilde{H}_r \text{ is stable}} \|H - \tilde{H}_r\|_{\mathcal{H}_2 \otimes \mathcal{L}_2} \\ &= \min_{\tilde{\mathcal{H}}_r \text{ is stable}} \|\mathbf{L}^T \mathcal{H} \mathbf{L} - \mathbf{L}^T \tilde{\mathcal{H}}_r \mathbf{L}\|_{\mathcal{H}_2}. \end{aligned}$$

Thus, $\mathbf{L}^T \tilde{\mathcal{H}}_r(s) \mathbf{L}$ is an \mathcal{H}_2 optimal reduced-order approximation to the associated MIMO system $\mathbf{L}^T \mathcal{H}(s) \mathbf{L}$. Since the reduced-order pencil $s\mathbf{E}_r - \mathbf{A}_r$ has only simple

eigenvalues, $\mathbf{L}^T \tilde{\mathcal{H}}_r(s) \mathbf{L}$ has a partial fraction expansion,

$$\mathbf{L}^T \tilde{\mathcal{H}}_r(s) \mathbf{L} = \sum_{i=1}^r \frac{1}{s - \tilde{\lambda}_i} \tilde{\mathbf{c}}_i \tilde{\mathbf{b}}_i^T,$$

with $\tilde{\mathbf{c}}_i, \tilde{\mathbf{b}}_i \in \mathbb{C}^2$ for $i = 1, \dots, r$. This reduced-order MIMO system must satisfy tangential interpolation conditions that are necessary consequences for \mathcal{H}_2 optimality:

$$\begin{aligned} \mathbf{L}^T \mathcal{H}(-\tilde{\lambda}_i) \mathbf{L} \tilde{\mathbf{b}}_i &= \mathbf{L}^T \tilde{\mathcal{H}}_r(-\tilde{\lambda}_i) \mathbf{L} \tilde{\mathbf{b}}_i, \\ \tilde{\mathbf{c}}_i^T \mathbf{L}^T \mathcal{H}(-\tilde{\lambda}_i) \mathbf{L} &= \tilde{\mathbf{c}}_i^T \mathbf{L}^T \tilde{\mathcal{H}}_r(-\tilde{\lambda}_i) \mathbf{L}, \\ (5.8) \quad \text{and} \quad \tilde{\mathbf{c}}_i^T \mathbf{L}^T \mathcal{H}'(-\tilde{\lambda}_i) \mathbf{L} \tilde{\mathbf{b}}_i &= \tilde{\mathbf{c}}_i^T \mathbf{L}^T \tilde{\mathcal{H}}_r'(-\tilde{\lambda}_i) \mathbf{L} \tilde{\mathbf{b}}_i \end{aligned}$$

($'$ denotes differentiation with respect to the frequency parameter, s).

Define for $i = 1, \dots, r$,

$$(5.9) \quad \tilde{\mathbf{c}}_i^T \mathbf{L}^T = [\mu_i, \alpha_i] \quad \text{and} \quad \mathbf{L} \tilde{\mathbf{b}}_i = \begin{bmatrix} \nu_i \\ \beta_i \end{bmatrix},$$

and associated optimal parameter values:

$$(5.10) \quad p_i = \alpha_i / \mu_i \quad \text{and} \quad q_i = \beta_i / \nu_i.$$

For $\mu_i \neq 0$ and $\nu_i \neq 0$, we may simplify (5.8) as

$$\begin{aligned} \nu_i \mathcal{H}(-\tilde{\lambda}_i) \begin{bmatrix} 1 \\ q_i \end{bmatrix} &= \nu_i \tilde{\mathcal{H}}_r(-\tilde{\lambda}_i) \begin{bmatrix} 1 \\ q_i \end{bmatrix}, \\ \mu_i [1, p_i] \mathcal{H}(-\tilde{\lambda}_i) &= \mu_i [1, p_i] \tilde{\mathcal{H}}_r(-\tilde{\lambda}_i), \quad \text{and} \\ \mu_i \nu_i [1, p_i] [\mathcal{H}'(-\tilde{\lambda}_i)] \begin{bmatrix} 1 \\ q_i \end{bmatrix} &= \mu_i \nu_i [1, p_i] [\tilde{\mathcal{H}}_r'(-\tilde{\lambda}_i)] \begin{bmatrix} 1 \\ q_i \end{bmatrix}, \end{aligned}$$

which leads immediately to the conditions (5.7). If either $\mu_i = 0$ or $\nu_i = 0$ (or both), then either p_i or q_i (or both) could take the value ∞ and the interpolation conditions (5.8) are equivalent to interpolation conditions given above for extended complex values of parameter values. \square

ALGORITHM 5.2. OPTIMAL INTERPOLATION FOR SISO PARAMETERIZATIONS WITH $H(s, \mathbf{p}) = (\mathbf{c}_0 + p \mathbf{c}_1)^T (s \mathbf{E} - \mathbf{A})^{-1} (\mathbf{b}_0 + q \mathbf{b}_1)$.

1. Construct $\tilde{\mathbf{H}}(s)$ as in (5.5) and \mathbf{L} as in Theorem 5.1.
 2. Apply Algorithm 3.1 to find an \mathcal{H}_2 optimal r th-order approximant to $\mathbf{L}^T \tilde{\mathbf{H}}(s) \mathbf{L}$. Let $\tilde{\mathbf{c}}_i$ and $\tilde{\mathbf{b}}_i$, for $i = 1, \dots, r$ denote the resulting optimal left and right tangent directions, respectively. Also, let $\tilde{\lambda}_i$ denote the resulting reduced-order poles.
 3. Compute p_i and q_i for $i = 1, \dots, r$ using (5.9), (5.10).
 4. Construct \mathbf{V}_r and \mathbf{W}_r as in lines 2–4 in Algorithm 4.1 using $\tilde{\mathbf{p}}_i = [p_i, q_i]^T$ as optimal parameter interpolation points, $\sigma_i = -\tilde{\lambda}_i$ as frequency interpolation points, $\tilde{\mathbf{c}}_i$ and $\tilde{\mathbf{b}}_i$ as left and right tangent directions for $i = 1, \dots, r$.
- The final optimal parameterized reduced-order model is determined from (2.3).

Note that the optimal parameter interpolation points $\tilde{\mathbf{p}}_i = [p_i, q_i]^T$ in Theorem 5.2 are *not* necessarily contained in \mathcal{D} , although if $\tilde{H}_r(s, [0, 0])$ is a minimal realization, then at least all of them can be guaranteed to be finite.

The definitions in (5.9) and (5.10) will be used in Algorithm 5.2 for the computation of an optimal parameterized reduced-order SISO system having the special form (5.3). Using the results of Theorem 5.2, Algorithm 5.2 first converts the SISO parameterized model reduction problem in $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ to an equivalent (nonparameterized) MIMO \mathcal{H}_2 model reduction problem. Algorithm 3.1 provides frequency interpolation points and tangent directions. Optimal parameter interpolation points are then recovered using (5.9) and (5.10), yielding in the end an optimal parameterized reduced model for the original problem with respect to the $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ norm. To the best of our knowledge, this is the first interpolatory parametric model reduction approach that *jointly* chooses frequency and the parameter interpolation points to minimize an associated system theoretic error norm.

6. Numerical examples.

6.1. Convection-diffusion flow. We consider a convection-diffusion equation on the unit square $\Omega = (0, 1)^2$:

$$\frac{\partial \mathbf{x}}{\partial t}(t, \xi) = \Delta \mathbf{x}(t, \xi) + \mathbf{p} \cdot \nabla \mathbf{x}(t, \xi) + b(\xi)u(t), \quad \xi \in \Omega, \quad t \in (0, \infty),$$

with homogeneous Dirichlet boundary conditions $\mathbf{x}(t, \xi) = 0$, $\xi \in \partial\Omega$.

The parameter vector $\mathbf{p} = [p_1, p_2]^T$ determines convective transport in both coordinate directions, whereas the function $b(\cdot)$ is the characteristic function of the domain where the input function $u(\cdot)$ acts.

We discretize the convection-diffusion equation with finite differences to obtain a parameterized linear system in state-space form:

$$(6.1) \quad \dot{\mathbf{x}}(t) = (\mathbf{A}_0 + p_1 \mathbf{A}_1 + p_2 \mathbf{A}_2) \mathbf{x}(t) + \mathbf{B} u(t), \quad y(t) = \mathbf{C} \mathbf{x}(t),$$

with $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{400 \times 400}$, $\mathbf{B} \in \mathbb{R}^{400 \times 1}$, and $\mathbf{C} \in \mathbb{R}^{1 \times 400}$. We assume $\mathbf{B} = \mathbf{e}_1$ (first unit vector) and $\mathbf{C} = \mathbf{e}^T$ (all ones). The parameter range considered is $p_1, p_2 \in [0, 1]$.

In this example, the physics of the problem does not provide particular insight to what parameter values might be important. The range of parameter values we consider keep the behavior of the system diffusion-dominated, so we don't take into account the possible desirability of changing the discretization for different parameter values so as to maintain an upwind bias in the discretization. Motivated by sparse-grid point selection in two dimensions, we use the following level-1 sparse-grid points, $\mathbf{p} = [p_1, p_2]^T$, to discretize the parameter space:

$$\mathbf{p}^{(1)} = [0.5, 0.5]^T, \quad \mathbf{p}^{(2)} = [0, 0.5]^T, \quad \mathbf{p}^{(3)} = [1, 0.5]^T, \quad \mathbf{p}^{(4)} = [0.5, 0]^T, \quad \mathbf{p}^{(5)} = [0.5, 1]^T.$$

We further simplify this selection by removing the $\mathbf{p}^{(4)}$ and $\mathbf{p}^{(5)}$ due to symmetry of the problem. Hence, our parameter set becomes $\{\mathbf{p}^{(1)}, \mathbf{p}^{(2)}, \mathbf{p}^{(3)}\}$. We apply Algorithm 5.1 with $r_1 = r_2 = r_3 = 4$ for $\mathbf{p}^{(i)}$, $i = 1, 2, 3$; the final parameterized reduced-order system as defined in (2.3) has dimension $r = 12$.

A good parameterized reduced-order model needs to represent the full parameterized model with high fidelity for a wide range of parameter values; certainly not just for those values chosen as the interpolation parameters. To illustrate the quality of our parameterized reduced-order models, we evaluate the full-order model, $\mathbf{H}(\cdot, \mathbf{p})$, varying parameter values, $\mathbf{p} = [p_1, p_2]^T$, across the full parameter range $[0, 1] \times [0, 1]$,

and compute

$$(6.2) \quad \text{the relative } \mathcal{H}_2 \text{ error at } \mathbf{p} \equiv \frac{\|\mathbf{H}(\cdot, \mathbf{p}) - \mathbf{H}_r(\cdot, \mathbf{p})\|_{\mathcal{H}_2}}{\|\mathbf{H}(\cdot, \mathbf{p})\|_{\mathcal{H}_2}} \text{ and}$$

$$(6.3) \quad \text{the relative } \mathcal{H}_\infty \text{ error at } \mathbf{p} \equiv \frac{\|\mathbf{H}(\cdot, \mathbf{p}) - \mathbf{H}_r(\cdot, \mathbf{p})\|_{\mathcal{H}_\infty}}{\|\mathbf{H}(\cdot, \mathbf{p})\|_{\mathcal{H}_\infty}}.$$

The corresponding mesh plots of relative error are shown in Figures 6.1 and 6.2. With a model of order $r = 12$, the maximum relative \mathcal{H}_∞ errors and \mathcal{H}_2 errors are, respectively, 5.21×10^{-3} and 1.86×10^{-3} . In terms of either error measure, the reduced-order model is accurate to an order of at least 10^{-3} , and we are able to capture the full-order dynamics accurately throughout the whole parameter range.

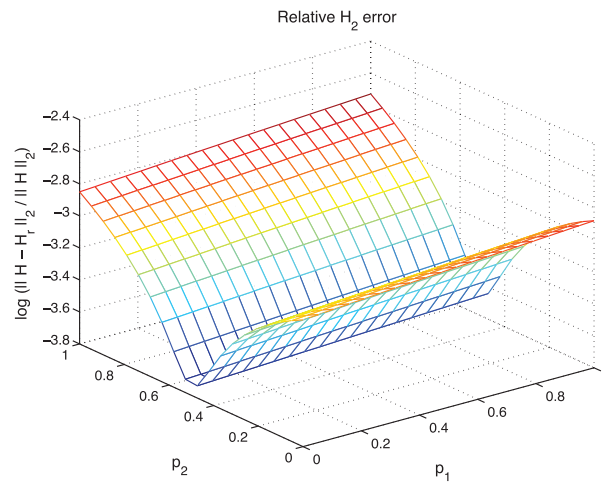


FIG. 6.1. Example 6.1 with $\nu = 2$: relative \mathcal{H}_2 error as p_1 and p_2 vary.

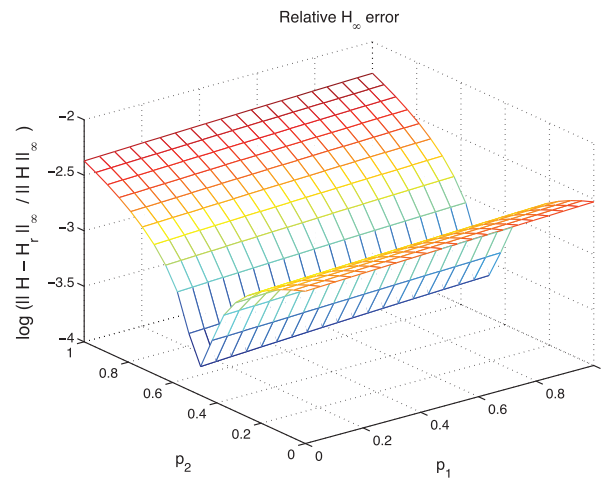


FIG. 6.2. Example 6.1 with $\nu = 2$: relative \mathcal{H}_∞ error as p_1 and p_2 vary.

Next, we add a third parameter p_0 to the model (6.1) in order to vary the diffusion:

$$(6.4) \quad \dot{\mathbf{x}}(t) = (p_0 \mathbf{A}_0 + p_1 \mathbf{A}_1 + p_2 \mathbf{A}_2) \mathbf{x}(t) + \mathbf{B} u(t), \quad y(t) = \mathbf{C} \mathbf{x}(t).$$

The diffusion coefficient p_0 varies in $[0.1, 1]$ and becomes the crucial parameter for smaller values in that range. Hence, we weight our parameter selection as follows. The problem approaches the previous case as p_0 increases to 1. Thus, we keep the same choice for p_1 and p_2 as above for $p_0 = 0.8$ and add three more choices for p_1 and p_2 for the case $p_0 = 0.1$. Overall, our parameter selection for $\mathbf{p} = [p_0, p_1, p_2]^T$ becomes

$$\begin{aligned} \mathbf{p}^{(1)} &= [0.8, 0.5, 0.5]^T, & \mathbf{p}^{(2)} &= [0.8, 0, 0.5]^T, & \mathbf{p}^{(3)} &= [0.8, 1, 0.5]^T, \\ \mathbf{p}^{(4)} &= [0.1, 0.5, 0.5]^T, & \mathbf{p}^{(5)} &= [0.1, 0, 1]^T, & \mathbf{p}^{(6)} &= [0.1, 1, 1]^T. \end{aligned}$$

As in the two parameter case, we apply Algorithm 5.1 by reducing the order at parameter values $\mathbf{p}^{(i)}$, $i = 1, \dots, 6$, using \mathcal{H}_2 optimal frequency interpolants with orders $r_1 = r_2 = r_3 = 3$ and $r_4 = r_5 = r_6 = 4$. To illustrate the performance of the reduced-order model, we fix p_0 at a specific value, vary the parameters p_1 and p_2 over the full parameter space $[0, 1] \times [0, 1]$, and compute relative \mathcal{H}_∞ error (6.3) and relative \mathcal{H}_2 error (6.2) at each grid point. We choose the values $p_0 = 0.1$ and $p_0 = 0.5$. Note that $p_0 = 0.5$ is not in the parameter selection set. The error plots for $p_0 = 0.1$ are shown in Figures 6.3 and 6.4. As in the two-parameter case, the reduced models approximate the full-order dynamics accurately. The resulting maximum relative \mathcal{H}_∞ error and relative \mathcal{H}_2 error for $p_0 = 0.1$ are 2.66×10^{-3} and 2.13×10^{-3} , respectively.

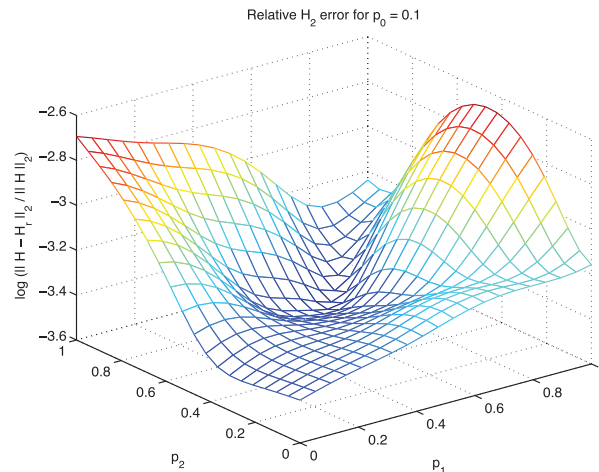


FIG. 6.3. Example 6.1 with $\nu = 3$: relative \mathcal{H}_2 error as p_1 and p_2 vary.

To better illustrate the quality of the approximation attained, we provide in Figure 6.5 amplitude Bode plots for $\mathbf{H}(s, \mathbf{p})$, $\mathbf{H}_r(s, \mathbf{p})$, and for the error system $\mathbf{H}(s, \mathbf{p}) - \mathbf{H}_r(s, \mathbf{p})$ using three different choices of $[p_1, p_2]$ values keeping $p_0 = 0.1$ fixed. The behavior of the reduced model is indistinguishable from that of the full-order model across the full parameter range.

The errors over the full range of p_1 and p_2 are even smaller with $p_0 = 0.5$, as can be seen in Figures 6.6 and 6.7. The maximum relative \mathcal{H}_∞ error and relative \mathcal{H}_2 error are, respectively, 3.62×10^{-4} and 1.44×10^{-4} , i.e., one order of magnitude smaller than for $p_0 = 0.1$.

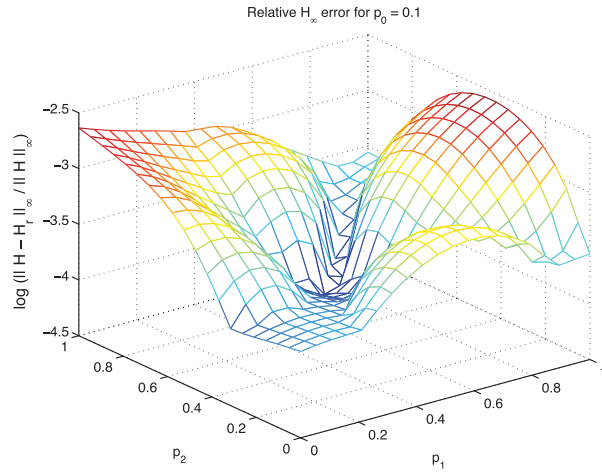


FIG. 6.4. Example 6.1 with $\nu = 3$: relative \mathcal{H}_∞ error as p_1 and p_2 vary.

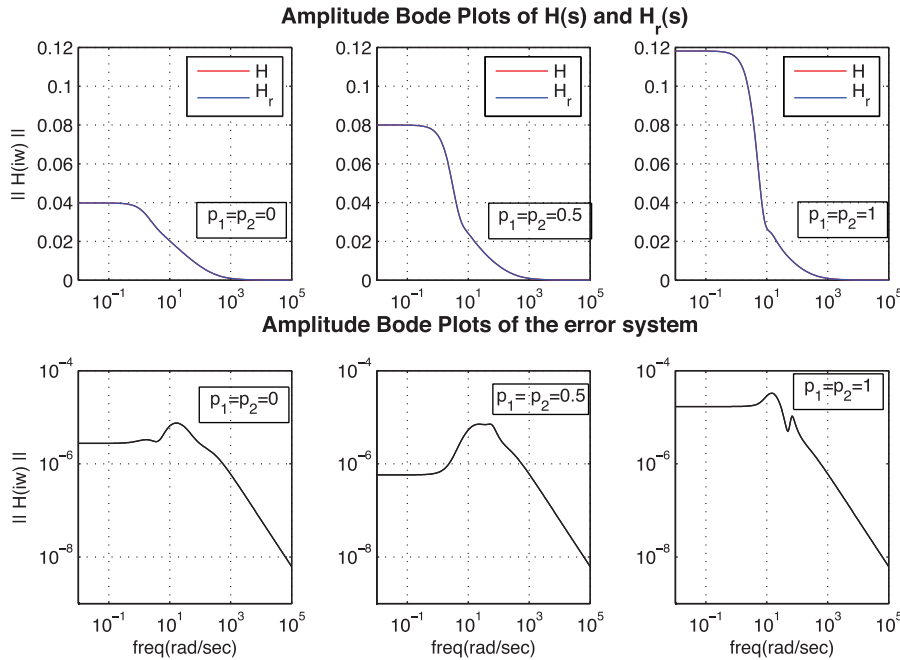


FIG. 6.5. Example 6.1 with $\nu = 3$: the amplitude Bode plots as p_1 and p_2 vary (with $p_0 = 0.1$).

6.1.1. Comparison with other model reduction approaches. To illustrate the superiority of our piecewise \mathcal{H}_2 optimal approach as described in Algorithm 5.1, we compare it with assorted generic interpolatory model reduction methods where the interpolation points do not have the (local) \mathcal{H}_2 optimality that Algorithm 5.1 produces. We proceed as follows: For the same parameter sets as above, $\{\mathbf{p}^{(i)}\}_{i=1}^6$, we obtain the projection matrices $\mathbf{V}^{(i)}$ and $\mathbf{W}^{(i)}$ using the frequency interpolation points that are used to initiate the optimal \mathcal{H}_2 reduction process at each $\mathbf{p}^{(i)}$. In effect, we apply only one-step interpolatory model reduction as opposed to the iterative

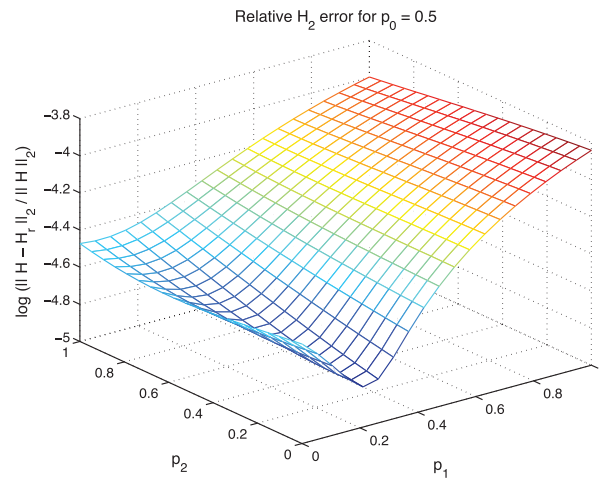


FIG. 6.6. Example 6.1 with $\nu = 3$: relative \mathcal{H}_2 error as p_1 and p_2 vary.

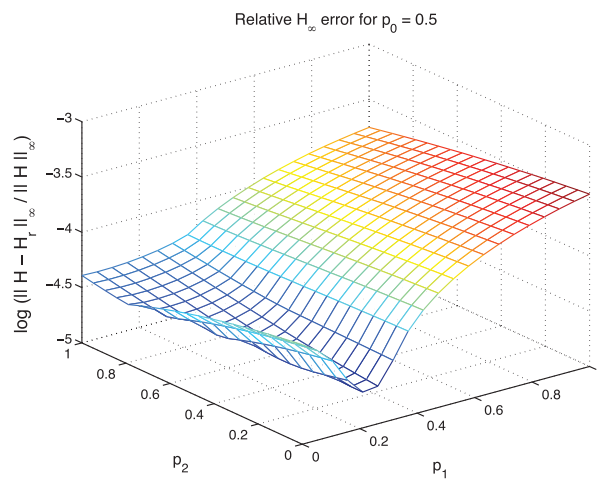


FIG. 6.7. Example 6.1 with $\nu = 3$: relative \mathcal{H}_∞ error as p_1 and p_2 vary.

\mathcal{H}_2 -optimal (IRKA) process. This is what one would do in a generic interpolation setting by choosing some interpolation points and obtaining the reduced model. We have concatenated $\mathbf{V}^{(i)}$ and $\mathbf{W}^{(i)}$ for $i = 1, \dots, L$ as Algorithm 5.1 does and then obtained the corresponding parameterized reduced model. For comparison, we calculate the error at the same grid points used before by fixing p_0 at 0.1 and display the resulting relative \mathcal{H}_2 errors and relative \mathcal{H}_∞ error in Figures 6.8 and 6.9. The maximum relative \mathcal{H}_∞ errors and relative \mathcal{H}_2 errors are, respectively, 4.98×10^{-1} and 2.19×10^{-1} . Note that these relative errors are two orders of magnitude higher than those obtained by the piecewise \mathcal{H}_2 optimal approach that we propose. This illustrates clearly the importance of optimal \mathcal{H}_2 shift selection in our algorithm. It is useful to note that we have initialized Algorithm 5.1 with the same interpolation points, and IRKA (Algorithm 3.1) adjusted these points iteratively without any user intervention, yielding in the end very accurate parameterized reduced models. Since

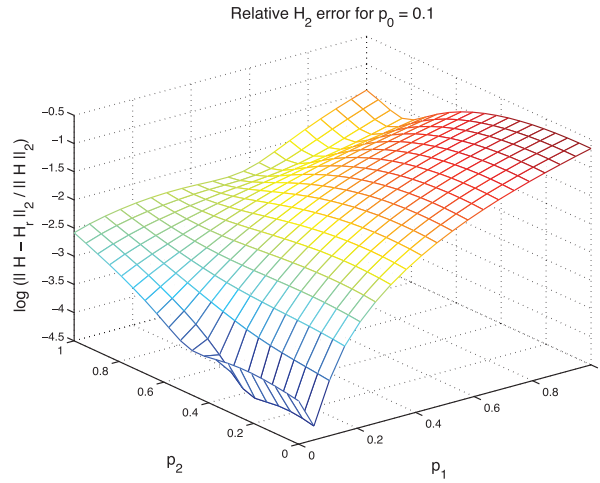


FIG. 6.8. Example 6.1 ($\nu = 3$) without optimal \mathcal{H}_2 shift selection: relative \mathcal{H}_2 error.

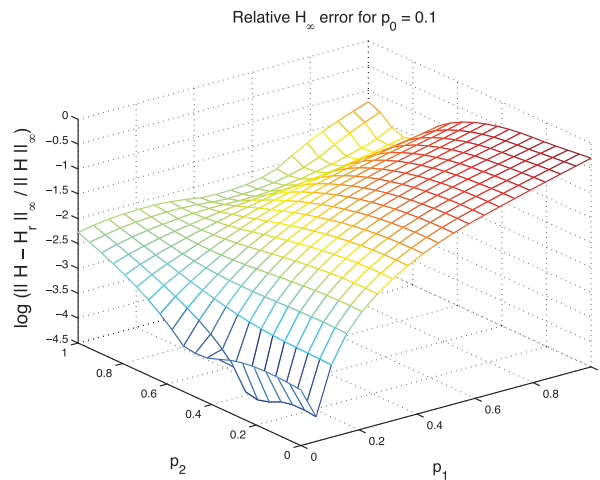
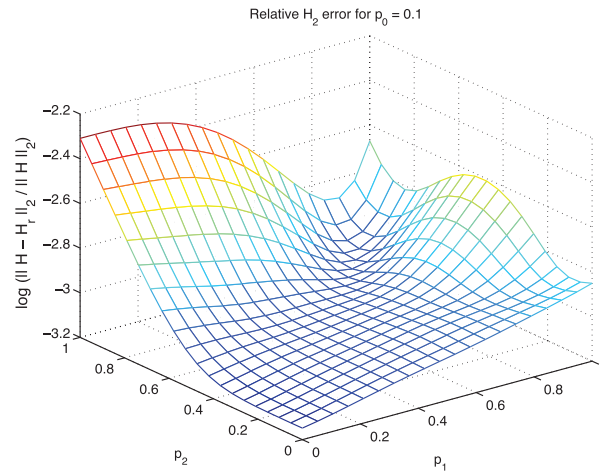
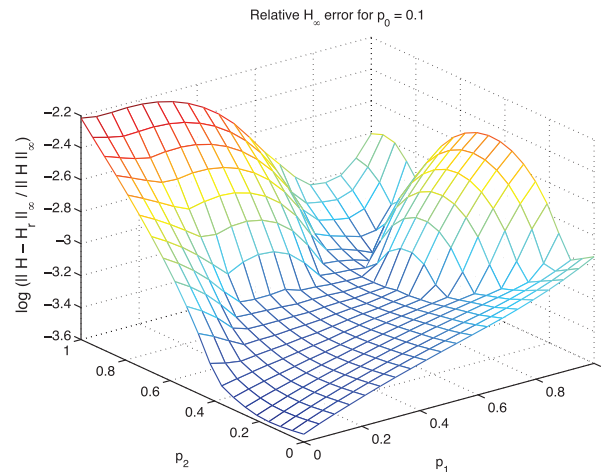


FIG. 6.9. Example 6.1 ($\nu = 3$) without optimal \mathcal{H}_2 shift selection: relative \mathcal{H}_∞ error.

the IRKA iteration generally converges very quickly (see [25]), the additional sparse linear systems that must be solved do not significantly increase cost, yet additional iterations increase the accuracy of the reduced model by two orders of magnitude.

Next, we compare our piecewise \mathcal{H}_2 optimal method with an approach where *balanced truncation* is used to reduce the order at each parameter set, $\mathbf{p}^{(i)}$. Towards this goal, we chose a reduced order of four at each parameter value and obtained corresponding $\mathbf{V}^{(i)}$ and $\mathbf{W}^{(i)}$ for $i = 1, \dots, 6$. Then as before, we concatenate the subspaces obtained by balanced truncation to form a final parameterized reduced-order model; since it is similar in structure to our piecewise \mathcal{H}_2 optimal method, we call this “piecewise balanced truncation.” (Note that this approach differs from the hybrid interpolation balanced truncation method described in [3].) For a fixed $p_0 = 0.1$, the maximum relative \mathcal{H}_∞ error calculated on the same grid for p_1 and p_2 is 6.10×10^{-3} ; the maximum relative \mathcal{H}_2 error is 4.91×10^{-3} . Plots for the relative

FIG. 6.10. Example 6.1 ($\nu = 3$) with piecewise balanced truncation: relative \mathcal{H}_2 error.FIG. 6.11. Example 6.1 ($\nu = 3$) with piecewise balanced truncation: relative \mathcal{H}_∞ error.

\mathcal{H}_2 error and relative \mathcal{H}_∞ error are shown in Figures 6.10 and 6.11, respectively. We note that both errors are somewhat higher than the results obtained by the proposed approach in section 5. This result is not surprising. Even though $\mathbf{V}^{(i)}$ and $\mathbf{W}^{(i)}$ are the balancing subspaces at the parameter values $\mathbf{p}^{(i)}$, once they are concatenated, the resulting reduced-order model is *no longer* balanced even when evaluated at the parameter set $\mathbf{p}^{(i)}$. On the other hand, if $\mathbf{V}^{(i)}$ and $\mathbf{W}^{(i)}$ are interpolating spaces that are obtained by forcing interpolation at some interpolation points $\{\sigma_{ik}\}_{k=1}^{r_i}$ for the parameter set described by $\mathbf{p}^{(i)}$, even after the subspaces are concatenated, the final reduced-order parameterized model would still interpolate the original model at the same interpolation points $\{\sigma_{ik}\}_{k=1}^{r_i}$ for the parameter set $\mathbf{p}^{(i)}$. In short, our piecewise \mathcal{H}_2 optimal algorithm has two important properties. First, due to the interpolatory structure, the final parameterized reduced-order model interpolates the original one even after the subspaces augmented. Second, the interpolation points at each parameter set are chosen in an \mathcal{H}_2 optimal way yielding accurate reduced-order models.

To more thoroughly compare the two approaches, we used many different values for r_1, r_2, \dots, r_6 at the corresponding parameter values $\mathbf{p}^{(i)}$ and computed the corresponding reduced-order models both by balanced truncation and by IRKA, producing projection subspaces $\mathbf{V}^{(i)}$ and $\mathbf{W}^{(i)}$. The final reduced-order parameterized systems are obtained as in (2.3) and their quality is compared by computing the maximum relative \mathcal{H}_2 error and \mathcal{H}_∞ error again varying p_1 and p_2 over the full parameter range of $[0, 1]$. The results are tabulated in Table 6.1. In this table, the ∞ entries indicate that some unstable reduced-order models were encountered for some choices of p_1 and p_2 . The table shows that except for cases where the approach using balanced truncation results in unstable reduced-order models, both approaches are comparable yielding similar quality reduced-order models. Note that in the PMOR approach combining balanced truncation and interpolation [3], the computation of unstable systems is avoided. Unfortunately, the method of [3] does not provide a reduced-order model in parameterized state-space form for more than one parameter. As we are focusing here on structure-preserving methods, we provide comparisons only with structure-preserving balancing-based methods such as described above.

6.2. Thermal conduction in a semiconductor chip. We consider now a model representing thermal conduction in a semiconductor chip described in [35]. An important requirement for a compact and efficient model of thermal conduction in this context is that it should allow flexibility in specifying boundary conditions in order to allow independent designers to evaluate how changes in the environment can influence the temperature distribution in the chip. The thermal problem is modeled as homogeneous heat diffusion with heat exchange occurring at three device interfaces modeled with convection boundary conditions. These conditions introduce film coefficients, p_1, p_2 , and p_3 , describing the heat exchange on the three device interfaces. Discretization leads to a system of ordinary differential equations

$$\mathbf{E}\dot{\mathbf{x}}(t) = \left(\mathbf{A} + \sum_{i=1}^3 p_i \mathbf{A}_i \right) \mathbf{x}(t) + \mathbf{B}u(t), \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t),$$

where $\mathbf{E} \in \mathbb{R}^{4257 \times 4257}$ and $\mathbf{A} \in \mathbb{R}^{4257 \times 4257}$ are system matrices, $\mathbf{A}_i \in \mathbb{R}^{4257 \times 4257}$, $i = 1, \dots, 3$, are diagonal matrices arising from the discretization of the convection boundary condition on the i th interface, and $\mathbf{B} \in \mathbb{R}^{4257}$ and $\mathbf{C} \in \mathbb{R}^{7 \times 4257}$; i.e., the system has a single input and seven outputs. The range for each parameter is the interval $[1, 10^4]$. Four important parameter vectors in $[1, 10^4]^3$ are given in Table 6.2 below: We use two of them $\mathbf{p}^{(1)} = [10^4, 10^4, 1]^T$ and $\mathbf{p}^{(2)} = [1, 1, 1]^T$ and apply Algorithm 5.1 as follows: In step 2, we reduce the order of the systems to $r_1 = 8$ and $r_2 = 7$ using Algorithm 3.1; i.e., projection subspaces $\mathbf{V}^{(i)} \in \mathbb{R}^{4257 \times r_i}$ and $\mathbf{W}^{(i)} \in \mathbb{R}^{4257 \times r_i}$ were computed for $i = 1, 2$. We concatenate these matrices to build the final projection matrices

$$\mathbf{V}_r = [\mathbf{V}^{(1)}, \mathbf{V}^{(2)}] \in \mathbb{R}^{4257 \times 15} \quad \text{and} \quad \mathbf{W}_r = [\mathbf{W}^{(1)}, \mathbf{W}^{(2)}] \in \mathbb{R}^{4257 \times 15}.$$

Having removed the rank-deficient components from \mathbf{V}_r and \mathbf{W}_r , our final parameterized reduced-order model has order $r = 14$ and is given by

$$\begin{aligned} \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r \dot{\mathbf{x}}_r(t) &= \left(\mathbf{W}_r^T \mathbf{A} \mathbf{V}_r + \sum_{i=1}^3 p_i \mathbf{W}_r^T \mathbf{A}_i \mathbf{V}_r \right) \mathbf{x}_r(t) + \mathbf{W}_r^T \mathbf{B}u(t), \\ \mathbf{y}_r(t) &= \mathbf{C} \mathbf{V}_r \mathbf{x}_r(t). \end{aligned}$$

TABLE 6.1
 \mathbf{H}_r : Piecewise \mathcal{H}_2 optimal reduced model; \mathbf{H}_{bal} : Piecewise balanced truncation reduced model.

Case:	A	B	C	D	E	F	G	H	I	J
dim. of proj. subsp. at $\mathbf{p}^{(z)}$	r_1	0	1	0	0	0	1	1	2	2
	r_2	1	1	0	0	1	0	1	2	2
	r_3	0	1	0	0	0	1	1	1	2
	r_4	1	1	2	2	2	2	2	2	2
	r_5	1	1	2	2	2	2	2	2	2
	r_6	1	1	2	2	2	2	2	2	2
Total dim r : $\mathbf{H}_r/\mathbf{H}_{bal}$	4/4	6/6	5/5	7/7	7/7	7/7	7/7	8/7	9/9	12/12
Max. rel. \mathbf{H}_r	2.85E-1	9.54E-2	1.21E-1	4.30E-2	4.65E-2	4.04E-2	3.91E-2	3.72E-2	5.23E-2	1.81E-2
\mathcal{H}_∞ -err. \mathbf{H}_{bal}	4.32E-2	∞	2.11E-2	2.26E-2	2.29E-2	2.57E-2	2.29E-2	2.51E-2	1.28E-2	∞
Max. rel. \mathbf{H}_r	1.25E-1	7.79E-2	9.39E-2	4.23E-2	4.54E-2	3.94E-2	4.28E-2	2.43E-2	3.76E-2	1.49E-2
\mathcal{H}_2 -err. \mathbf{H}_{bal}	9.89E-2	∞	3.91E-2	1.39E-1	1.28E-1	1.73E-1	1.41E-1	1.03E-1	1.27E-2	∞

Case:	K	L	M	N	O	P	Q	R	S	T
dim. of proj. subsp. at $\mathbf{p}^{(z)}$	r_1	1	1	1	1	2	0	1	1	2
	r_2	0	1	1	2	2	0	1	2	3
	r_3	0	0	1	1	2	0	1	2	3
	r_4	3	3	3	3	3	4	4	4	4
	r_5	3	3	3	3	3	4	4	4	4
	r_6	3	3	3	3	3	4	4	4	4
Total dim. r : $\mathbf{H}_r/\mathbf{H}_{bal}$	10/10	10/11	11/12	12/13	14/13	12/12	14/15	16/16	16/16	17/17
Max. rel. \mathbf{H}_r	7.66E-2	7.84E-2	7.64E-2	1.34E-2	1.16E-2	3.65E-2	1.46E-2	8.53E-3	2.12E-2	2.66E-3
\mathcal{H}_∞ -err. \mathbf{H}_{bal}	∞	∞	∞	∞	4.01E-2	1.41E-2	9.95E-3	1.74E-2	1.51E-3	6.10E-3
Max. rel. \mathbf{H}_r	3.22E-2	3.23E-2	3.23E-2	6.54E-3	1.09E-2	1.74E-2	8.92E-3	4.46E-3	1.09E-2	2.13E-3
\mathcal{H}_2 -err. \mathbf{H}_{bal}	∞	∞	∞	∞	2.19E-2	1.00E-2	4.91E-3	9.11E-3	8.05E-3	4.91E-3

TABLE 6.2

Example 6.2: parameter vectors (with $p_3 = 1$).

	$\mathbf{p}^{(1)}$	$\mathbf{p}^{(2)}$	$\mathbf{p}^{(3)}$	$\mathbf{p}^{(4)}$
p_1	10^4	1	10	10^4
p_2	10^4	1	10^4	10
p_3	1	1	1	1

To illustrate the quality of this reduced-order model, we fix $p_3 = 1$ and vary both p_1 and p_2 between 1 and 10^4 . For each mesh point (i.e., for each triple of parameter values in this range), we compute both the corresponding full-order model and the reduced-order model; and evaluate the relative \mathcal{H}_∞ errors. The resulting mesh plot of \mathcal{H}_∞ errors is given in Figure 6.12. The maximum relative \mathcal{H}_∞ error is 2.16×10^{-2} . The parameterized reduced model $\mathbf{H}_r(s, \mathbf{p})$ has system order smaller than 4% of the original system order, yet is able to maintain high fidelity and a small relative error of around 2% or less over the full range of variation of p_1 and p_2 .

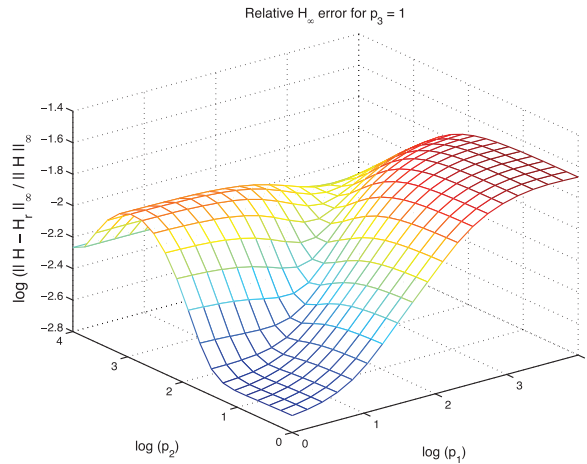


FIG. 6.12. Example 6.2: relative \mathcal{H}_∞ error as p_1 and p_2 vary.

In Figure 6.13, we give amplitude Bode plots for $\mathbf{H}(s, \mathbf{p})$, $\mathbf{H}_r(s, \mathbf{p})$, and for the error system $\mathbf{H}(s, \mathbf{p}) - \mathbf{H}_r(s, \mathbf{p})$ using three different choices of $[p_1, p_2]$ values keeping $p_3 = 1$ fixed. Once again, the reduced model almost exactly replicates the full-order model across the full parameter range.

6.2.1. Comparison with piecewise balanced truncation. As in the previous example, we present a comparison between our piecewise \mathcal{H}_2 optimal approach and piecewise balanced truncation that concatenates the projection matrices that are obtained by using balanced truncation for the fixed parameter vectors $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$. To give an overall picture, we use many different combinations of r_1 and r_2 values and then compute maximum relative \mathcal{H}_∞ errors encountered while varying p_1 and p_2 over the full parameter range of $[1, 10^4]$. The results are tabulated in Table 6.3, where ∞ corresponds to encountering some unstable reduced-order models while p_1 and p_2 vary. One obvious conclusion is that the proposed \mathcal{H}_2 -based method consistently yields results that are as accurate as those obtained by the balancing-based approach. Note that the error values are computed using the \mathcal{H}_∞ norm. Hence, the proposed \mathcal{H}_2 -based approach yields accurate reduced-model not only in the \mathcal{H}_2 norm

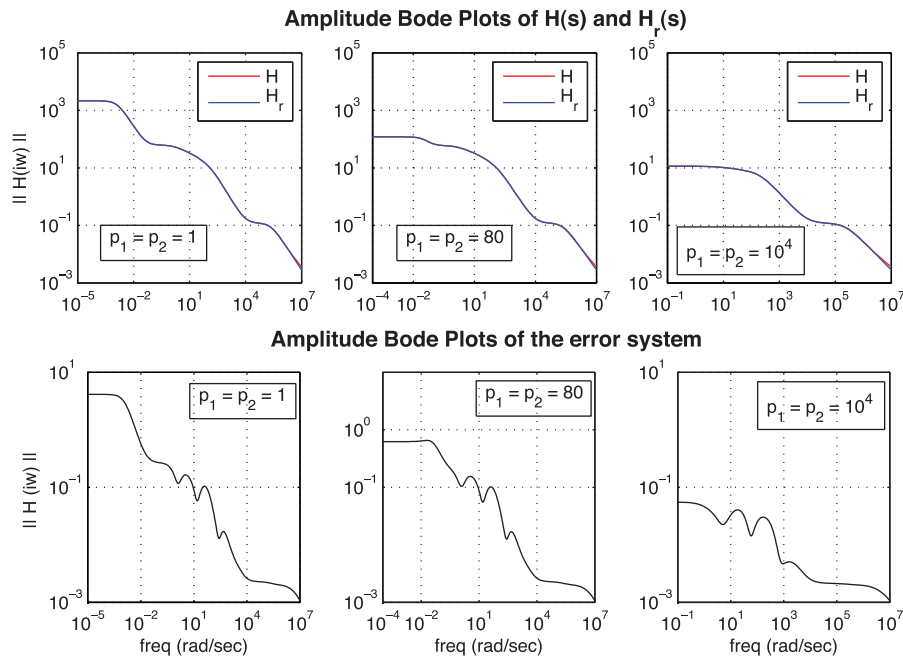
FIG. 6.13. Example 6.2: The amplitude Bode plots as p_1 and p_2 vary (with $p_3 = 1$).

TABLE 6.3

 \mathbf{H}_r : Piecewise \mathcal{H}_2 optimal reduced model; \mathbf{H}_{bal} : Piecewise balanced truncation reduced model.

Cases:		A	B	C	D	E	F	G
dim. at $\mathbf{p}^{(i)}$	r_1	4	5	6	7	7	6	8
	r_2	4	5	4	4	5	6	7
Total dim: r		6/6-8	9/9	9/10	10/11	10/12	11/12	14/14
Max. rel. \mathcal{H}_∞ -err.	\mathbf{H}_r	1.87 E-1	6.69 E-2	9.75 E-2	8.29 E-2	6.88 E-2	3.50 E-2	2.16 E-2
	\mathbf{H}_{bal}	∞	2.65 E-2	5.23 E-2	5.09 E-2	4.73 E-2	2.47 E-2	4.56 E-2

but also in the \mathcal{H}_∞ norm. This is not surprising since the optimal \mathcal{H}_2 method described in Algorithm 3.1 for nonparameterized systems is known to yield both good \mathcal{H}_∞ performance and \mathcal{H}_2 performance; see [25].

6.3. Optimal SISO parameterized model reduction example. We illustrate here the approach introduced in section 5.1. A full-order model of the form (5.1) represents the evolution of the temperature distribution on a plate as described by the heat equation. A model of order 197 is obtained by a finite difference discretization. The vectors \mathbf{b}_0 and $\mathbf{b}_0 + \mathbf{b}_1$ could be interpreted as the spatial distribution of two heat sources. As the parameter q varies from 0 to 1, the input shifts from one heat source distribution \mathbf{b}_0 to the other $\mathbf{b}_0 + \mathbf{b}_1$. Similarly, the vectors \mathbf{c}_0 and $\mathbf{c}_0 + \mathbf{c}_1$ can represent temperature sensing profiles so that as the parameter p varies from 0 to 1, the sensing profile shifts from \mathbf{c}_0 to $\mathbf{c}_0 + \mathbf{c}_1$.

We minimize the $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ error between the full-order and the reduced-order transfer functions as shown in Theorem 5.2 by applying Algorithm 5.2. The corresponding MIMO nonparameterized systems in line 2 of the algorithm are reduced to order $r = 10$ by \mathcal{H}_2 optimal model reduction in Algorithm 3.1. The resulting optimal

frequency shifts, $\{-\tilde{\lambda}_i\}_{i=1}^r$, and *optimal parameter interpolation vectors*, $\{\tilde{\mathbf{p}}^{(i)}\}_{i=1}^r$ are given below:

$-\tilde{\lambda}_i$	$\tilde{\mathbf{p}}^{(i)}$	
	p_i	q_i
0.0152	0.559	0.344
0.142	0.246	0.351
0.416	-0.516	0.359
0.862	0.454	0.337
0.102	0.620	0.310
0.184	0.549	0.385
0.419	0.512	0.366
28.9	0.349	0.319
$7.24 - \iota 1.10$	$0.435 + \iota 0.0404$	$0.406 - \iota 0.0778$
$7.24 + \iota 1.10$	$0.435 - \iota 0.0404$	$0.406 + \iota 0.0778$

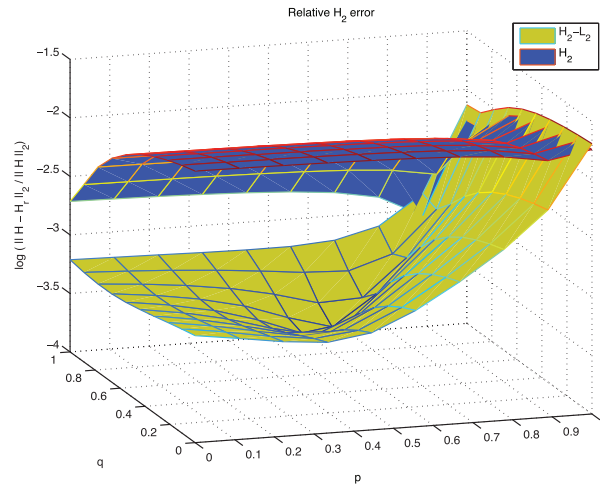


FIG. 6.14. Example C: relative \mathcal{H}_2 error as p_1 and p_2 vary.

An interesting observation is that even though both parameters p and q are contained in the interval $[0, 1]$, some of the optimal parameter values lie outside this region; indeed some of the optimal points are even complex. This example is a perfect illustration of the fact that the best parameter selection does not necessarily lies in the parameter range; i.e., one can obtain a better performance by including complex parameter points or at least parameter values outside the region of interest. The 10th-order optimal parameterized reduced-order model yields an extremely satisfactory relative $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ error of 7.54×10^{-4} .

To show the superiority of this optimal selection for the introduced $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ measure, we compare the results with those obtained by the \mathcal{H}_2 -based method in Algorithm 5.1; i.e., we choose $[0, 0]^T$, $[0, 0.5]^T$, $[0.5, 0]^T$ and $[1, 1]^T$ as parameter vectors, use \mathcal{H}_2 optimal reduced-order models at each parameter set, and then combine the resulting subspaces together. The resulting reduced-order model of order $r = 10$ leads to a relative $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ error of 2.09×10^{-2} . Even though this is a satisfactory relative error, the result using the optimal points is two order of magnitudes better, illustrating the superiority of the $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ optimal point selection.

Even though the $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ optimal approach does not minimize the \mathcal{H}_2 error

at every point in the parameter range, we compare the quality of the derived results by computing the relative \mathcal{H}_2 error (6.2) over the full parameter range. The results are shown in Figure 6.14. The $\mathcal{H}_2 \otimes \mathcal{L}_2(\mathcal{D})$ optimal approach yields much smaller \mathcal{H}_2 -errors for most of the grid points with a maximum error of 2.04×10^{-2} . On the other hand, the maximum \mathcal{H}_2 -error due to Algorithm 5.1 is 2.09×10^{-2} .

7. Conclusions. We have introduced a unifying projection-based framework for structure-preserving interpolatory model reduction of parameterized linear dynamical systems. Analogous to the nonparameterized case, we provide conditions under which the transfer functions of original and reduced-order model coincide in given directions at interpolation points in the parameter domain. Furthermore, we give conditions under which the gradient and Hessian of the system response model with respect to the system parameters is matched by the reduced-order response model. A systematic approach built on established interpolatory \mathcal{H}_2 -optimal model reduction methods is provided that produces parameterized reduced-order models having high fidelity throughout a parameter range of interest. For single input/single output systems with parameters in the input/output maps, we offer an approach that yields reduced-order models that are optimal with respect to an $\mathcal{H}_2 \otimes \mathcal{L}_2$ joint error measure. The capabilities of these approaches are illustrated by several numerical examples from technical applications.

REFERENCES

- [1] N. ALEXANDROV, J. E. DENNIS, JR., R. M. LEWIS, AND V. TORCZON, *A trust region framework for managing the use of approximation models in optimization*, Structural Optimization, 15 (1998), pp. 16–23.
- [2] M. BARRAULT, Y. MADAY, N. C. NGUYEN, AND A. T. PATERA, *An ‘empirical interpolation’ method: Application to efficient reduced-basis discretization of partial differential equations*, C. R. Math. Acad. Sci. Paris, 339 (2004), pp. 667–672.
- [3] U. BAUR AND P. BENNER, *Model reduction for parametric systems using balanced truncation and interpolation*, at-Automatisierungstechnik, 57 (2009), pp. 411–420 (in German).
- [4] C. A. BEATTIE AND S. GUGERCIN, *Interpolatory projection methods for structure-preserving model reduction*, Systems Control Lett., 58 (2008), pp. 225–232.
- [5] B. N. BOND AND L. DANIEL, *Parameterized model order reduction of nonlinear dynamical systems*, in IEEE/ACM International Conference on Computer-Aided Design (ICCAD-2005), 2005, pp. 487–494.
- [6] B. N. BOND AND L. DANIEL, *A piecewise-linear moment-matching approach to parameterized model-order reduction for highly nonlinear systems*, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., 26 (2007), pp. 2116–2129.
- [7] A. BRYSON, *Second-order algorithm for optimal model order reduction*, J. Guidance Control Dynamics, 13 (1990), pp. 887–892.
- [8] T. BUI-THANH, K. WILLCOX, AND O. GHATTAS, *Model reduction for large-scale systems with high-dimensional parametric input space*, SIAM J. Sci. Comput., 30 (2008), pp. 3270–3288.
- [9] H.-J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, Acta Numerica, 13 (2004), pp. 147–269.
- [10] A. BUNSE-GERSTNER, D. KUBALINSKA, G. VOSSEN, AND D. WILCZEK, *H_2 -norm optimal model reduction for large scale discrete dynamical mimo systems*, J. Comput. Appl. Math., 233 (2010), pp. 1202–1216.
- [11] D. G. CACUCI, *Sensitivity and Uncertainty Analysis, Vol. I: Theory*, Chapman & Hall/CRC, Boca Raton, FL, 2003.
- [12] L. DANIEL, O. C. SIONG, L. S. CHAY, K. H. LEE, AND J. WHITE, *A multiparameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models*, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., 23 (2004), pp. 678–693.
- [13] C. DE VILLEMAGNE AND R. E. SKELTON, *Model reductions using a projection formulation*, Internat. J. Control, 46 (1987), pp. 2141–2169.
- [14] R. EID, B. SALIMBAHRAMI, B. LOHMANN, E. B. RUDNYI, AND J. G. KORVINK, *Parametric order reduction of proportionally damped second-order systems*, Sensors and Materials, 19

- (2007), pp. 149–164.
- [15] O. FARLE, V. HILL, P. INGELSTRÖM, AND R. DYCZIJ-EDLINGER, *Multi-parameter polynomial order reduction of linear finite element models*, Math. Comput. Model. Dyn. Syst., 14 (2008), pp. 421–434.
- [16] P. FELDMANN AND R. W. FREUND, *Efficient linear circuit analysis by Padé approximation via the Lanczos process*, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., 14 (1995), pp. 639–649.
- [17] L. FENG, *Parameter independent model order reduction*, Math. Comput. Simulation, 68 (2005), pp. 221–234.
- [18] L. FENG AND P. BENNER, *A robust algorithm for parametric model order reduction*, Proc. Appl. Math. Mech., 7 (2008), pp. 1021501–1021502; available online from <http://dx.doi.org/10.1002/pamm.200700749>.
- [19] L. FENG, E. B. RUDNYI, AND J. G. KORVINK, *Preserving the film coefficient as a parameter in the compact thermal model for fast electrothermal simulation*, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., 24 (2005), pp. 1838–1847.
- [20] K. GALLIVAN, E. J. GRIMME, AND P. VAN DOOREN, *Asymptotic waveform evaluation via a Lanczos method*, Appl. Math. Lett., 7 (1994), pp. 75–80.
- [21] K. GALLIVAN, A. VANDENDORPE, AND P. VAN DOOREN, *Model reduction of MIMO systems via tangential interpolation*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 328–349.
- [22] M. A. GREPL, Y. MADAY, N. C. NGUYEN, AND A. T. PATERA, *Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations*, ESAIM Math. Model. Numer. Anal., 41 (2007), pp. 575–605.
- [23] M. GRIEBEL, *Sparse grids and related approximation schemes for higher dimensional problems*, in Foundations of Computational Mathematics, Santander 2005, L. M. Pardo et al., eds., London Math. Soc. Lecture Note Ser. 331, Cambridge University Press, Cambridge, 2006, pp. 106–161.
- [24] E. J. GRIMME, *Krylov projection methods for model reduction*, Ph.D. thesis, University of Illinois at Urbana-Champaign, 1997.
- [25] S. GUGERCIN, A. C. ANTOUNAS, AND C. BEATTIE, \mathcal{H}_2 model reduction for large-scale linear dynamical systems, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 609–638.
- [26] P. K. GUNUPUDI, R. KHAZAKA, AND M. S. NAKHLA, *Analysis of transmission line circuits using multidimensional model reduction techniques*, IEEE Trans. Adv. Packaging, 25 (2002), pp. 174–180.
- [27] P. K. GUNUPUDI, R. KHAZAKA, M. S. NAKHLA, T. SMY, AND D. CELO, *Passive parameterized time-domain macromodels for high-speed transmission-line networks*, IEEE Trans. Microwave Theory and Techniques, 51 (2003), pp. 2347–2354.
- [28] B. HAASDONK, M. OHLBERGER, AND G. ROZZA, *A reduced basis method for evolution schemes with parameter-dependent explicit operators*, Electron. Trans. Numer. Anal., 32 (2008), pp. 145–161.
- [29] Y. HALEVI, *Frequency weighted model reduction via optimal projection*, IEEE Trans. Automat. Control, 37 (1992), pp. 1537–1542.
- [30] K. R. HELOUE, S. ONAISSI, AND F. N. NAJM, *Efficient block-based parameterized timing analysis covering all potentially critical paths*, in IEEE/ACM International Conference on Computer-Aided Design (ICCAD 2008), 2008, pp. 173–180.
- [31] D. HYLAND AND D. BERNSTEIN, *The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore*, IEEE Trans. Automat. Control, 30 (1985), pp. 1201–1211.
- [32] K. ITO AND S. S. RAVINDRAN, *A reduced basis method for control problems governed by PDEs*, in Control and Estimation of Distributed Parameter Systems, W. Desch et al., eds., Int. Ser. Numer. Math. 126, Birkhäuser, Basel, 1998, pp. 153–168.
- [33] D. R. JONES, *A taxonomy of global optimization methods based on response surfaces*, J. Global Optim., 21 (2001), pp. 345–383.
- [34] D. G. KRIGE, *A statistical approach to some basic mine valuation problems on the Witwatersrand*, J. Chemical Metallurgical and Mining Society of South Africa, 52 (1951), pp. 119–139.
- [35] C. J. M. LASANCE, *Two benchmarks to facilitate the study of compact thermal modeling phenomena*, IEEE Trans. Components and Packaging Technologies, 24 (2001), pp. 559–565.
- [36] A. LEPSCHY, G. A. MIAN, G. PINATO, AND U. VIARO, *Rational L^2 approximation: A non-gradient algorithm*, in Proceedings of the 30th IEEE Conference on Decision and Control, 1991, pp. 2321–2323.
- [37] A. T.-M. LEUNG AND R. KHAZAKA, *Parametric model order reduction technique for design optimization*, in IEEE Internat. Symposium on Circuits and Systems (ISCAS), 2005, pp. 1290–1293.

- [38] M. MA, A. T.-M. LEUNG, AND R. KHAZAKA, *Sparse macromodels for parametric networks*, in IEEE Internat. Symposium on Circuits and Systems (ISCAS), 2006, pp. 2545–2548.
- [39] Y. MADAY AND E. M. RÖNQUIST, *A reduced-basis element method*, C. R. Math. Acad. Sci. Paris, 335 (2002), pp. 195–200.
- [40] L. MEIER III AND D. LUENBERGER, *Approximation of linear constant systems*, IEEE Trans. Automat. Control, 12 (1967), pp. 585–588.
- [41] K. MOOSMANN AND J. G. KORVINK, *Automatic parametric MOR for MEMS design*, in Tagungsband GMA-FA 1.30 “Modellbildung, Identifikation und Simulation in der Automatisierungstechnik,” Workshop am Bostalsee, 27.–29.9.2006, B. Lohmann and A. Kugi, eds., 2006, pp. 89–99.
- [42] J. R. PHILLIPS, *Variational interconnect analysis via PMTBR*, in Proceedings of ICCAD, 2004, pp. 872–879.
- [43] M. REWIEŃSKI AND J. WHITE, *Model order reduction for nonlinear dynamical systems based on trajectory piecewise-linear approximations*, Linear Algebra Appl., 415 (2006), pp. 426–454.
- [44] J. T. SPANOS, M. H. MILMAN, AND D. L. MINGORI, *A new algorithm for L^2 optimal model reduction*, Automatica J. IFAC, 28 (1992), pp. 897–909.
- [45] M. L. STEIN, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer-Verlag, New York, 1999.
- [46] P. VAN DOOREN, K. A. GALLIVAN, AND P.-A. ABSIL, *\mathcal{H}_2 -optimal model reduction of MIMO systems*, Appl. Math. Lett., 21 (2008), pp. 1267–1273.
- [47] D. VASILYEV, M. REWIEŃSKI, AND J. WHITE, *A TBR-based trajectory piecewise-linear algorithm for generating accurate low-order models for nonlinear analog circuits and MEMS*, in Proceedings of Design Automation Conference (DAC), 2003, pp. 490–495.
- [48] C. VISWESWARIAH, K. RAVINDRAN, K. KALAFALA, S. G. WALKER, S. NARAYAN, D. K. BEECE, J. PIAGET, N. VENKATESWARAN, AND J. G. HEMMETT, *First-order incremental block-based statistical timing analysis*, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., 25 (2006), pp. 2170–2180.
- [49] D. S. WEILE, E. MICHELSEN, E. J. GRIMME, AND K. A. GALLIVAN, *A method for generating rational interpolant reduced order models of two-parameter linear systems*, Appl. Math. Lett., 12 (1999), pp. 93–102.
- [50] D. A. WILSON, *Optimum solution of model-reduction problem*, Proc. Inst. Elec. Eng., 117 (1970), pp. 1161–1165.
- [51] W. Y. YAN AND J. LAM, *An approximate approach to h^2 optimal model reduction*, IEEE Trans. Automat. Control, 44 (1999), pp. 1341–1358.
- [52] A. YOUSUFF AND R. E. SKELTON, *Covariance equivalent realizations with applications to model reduction of large-scale systems*, Control and Dynamic Systems, 22 (1985), pp. 273–348.
- [53] A. YOUSUFF, D. WAGIE, AND R. E. SKELTON, *Linear system approximation via covariance equivalent realizations*, J. Math. Anal. Appl., 106 (1985), pp. 91–115.
- [54] C. ZENGER, *Sparse grids*, in Parallel Algorithms for Partial Differential Equations (Kiel, 1990), Notes Numer. Fluid Mech. 31, Vieweg, Braunschweig, 1991, pp. 241–251.
- [55] D. ZIGIC, L. T. WATSON, AND C. A. BEATTIE, *Contragredient transformations applied to the optimal projection equations*, Linear Algebra Appl., 188 (1993), pp. 665–676.