A UNIFYING FRAMEWORK FOR INTERPOLATORY $\mathcal{L}_2$-OPTIMAL REDUCED-ORDER MODELING*

PETAR MLINARIĆ† AND SERKAN GUGERCIN‡

Abstract. We develop a unifying framework for interpolatory $\mathcal{L}_2$-optimal reduced-order modeling for a wide classes of problems ranging from stationary models to parametric dynamical systems. We first show that the framework naturally covers the well-known interpolatory necessary conditions for $\mathcal{H}_2$-optimal model order reduction and leads to the interpolatory conditions for $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal model order reduction of multi-input/multi-output parametric dynamical systems. Moreover, we derive novel interpolatory optimality conditions for rational discrete least-squares minimization and for $\mathcal{L}_2$-optimal model order reduction of a class of parametric stationary models. We show that bitangential Hermite interpolation appears as the main tool for optimality across different domains. The theoretical results are illustrated on two numerical examples.

Key words. reduced-order modeling, parametric stationary problems, linear time-invariant systems, optimization, $\mathcal{L}_2$ norm, nonlinear least squares, interpolation

AMS subject classifications. 30E05, 30E10, 32E30, 41A20, 46N10, 93A15

1. Introduction.

Interpolatory methods have been one of the most commonly used model order reduction (MOR) techniques, see, e.g., [3, 6, 13]. For $\mathcal{H}_2$-optimal MOR of linear time-invariant (LTI) dynamical systems, the necessary optimality conditions are known and appear in the form of (bitangential) Hermite interpolation of the underlying transfer function [2, 29, 30, 36]. These interpolatory optimality conditions have formed the foundation of various algorithms and have been extended to different settings; see, e.g., [1, 8, 9, 15, 25, 26, 31, 32]. But in various other important settings, such as in the optimal approximation of stationary problems and discrete least-squares (LS) rational fitting, it is not yet established whether the optimality requires interpolation (as in the $\mathcal{H}_2$-case) and if so, what those interpolation conditions are. For instance, reduced basis methods use a greedy selection of sampling (interpolation) points to match the solution at these points [34, 40]. Is there an underlying framework for interpolatory optimality conditions?

The authors recently developed a data-driven framework for $\mathcal{L}_2$-optimal reduced-order modeling of parametric systems [38]. In this paper, we show how [38] provides a unifying framework for interpolatory optimal approximation both for dynamical systems and stationary problems. We prove that bitangential Hermite interpolation is the necessary condition for optimality not only for approximation of LTI systems in the $\mathcal{H}_2$ norm, but also in many other prominent cases, thus extending the optimal interpolation theory to a broader class of problems.

First we recall the $\mathcal{L}_2$-optimal reduced-order modeling problem discussed in [38]: Consider a parameter-to-output mapping

$$y\colon \mathcal{P} \to \mathbb{C}^{n_{\mathrm{o}} \times n_{\mathrm{i}}} \tag{1.1}$$

*Submitted to the editors September 1, 2022.

Funding: This work was partially funded by the U.S. National Science Foundation under grant DMS-1923221. Parts of this material are based upon work supported by the National Science Foundation under Grant No. DMS-1929284 while the authors were in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the Spring 2020 Reunion Event for Model and Dimension Reduction in Uncertain and Dynamic Systems program.

†Department of Mathematics, Virginia Tech, Blacksburg, VA 24061 (mlinaric@vt.edu).

‡Department of Mathematics and Division of Computational Modeling and Data Analytics, Academy of Data Science, Virginia Tech, Blacksburg, VA 24061 (gugercin@vt.edu).

where $\mathcal{P} \subseteq \mathbb{C}^{n_{\mathrm{p}}}$ and $n_{\mathrm{i}}, n_{\mathrm{o}}, n_{\mathrm{p}}$ are positive integers. Assume that evaluating $y(\mathsf{p})$ is expensive. Thus the goal is to construct a high-fidelity reduced-order mapping (approximation) $\widehat{y} \colon \mathcal{P} \to \mathbb{C}^{n_{\mathrm{o}} \times n_{\mathrm{i}}}$, which is much cheaper to evaluate than y. Inspired by the structures arising in projection-based (parametric) MOR, [38] constructs a structured reduced-order model (ROM)

$$\widehat{\mathcal{A}}(\mathsf{p})\widehat{x}(\mathsf{p}) = \widehat{\mathcal{B}}(\mathsf{p}), \tag{1.2a}$$

$$\widehat{y}(\mathsf{p}) = \widehat{\mathcal{C}}(\mathsf{p})\widehat{x}(\mathsf{p}), \tag{1.2b}$$

with a parameter-separable form

$$\widehat{\mathcal{A}}(\mathsf{p}) = \sum_{i=1}^{q_{\widehat{\mathcal{A}}}} \widehat{\alpha}_i(\mathsf{p})\widehat{A}_i, \quad \widehat{\mathcal{B}}(\mathsf{p}) = \sum_{j=1}^{q_{\widehat{\mathcal{B}}}} \widehat{\beta}_j(\mathsf{p})\widehat{B}_j, \quad \widehat{\mathcal{C}}(\mathsf{p}) = \sum_{k=1}^{q_{\widehat{\mathcal{C}}}} \widehat{\gamma}_k(\mathsf{p})\widehat{C}_k, \tag{1.3}$$

where $\widehat{x}(\mathsf{p}) \in \mathbb{C}^r$ is the reduced state, $\widehat{y}(\mathsf{p}) \in \mathbb{C}^{n_{\mathrm{o}} \times n_{\mathrm{i}}}$ is the approximate output, $\widehat{\mathcal{A}}(\mathsf{p}) \in \mathbb{C}^{r \times r}$, $\widehat{\mathcal{B}}(\mathsf{p}) \in \mathbb{C}^{r \times n_{\mathrm{i}}}$, $\widehat{\mathcal{C}}(\mathsf{p}) \in \mathbb{C}^{n_{\mathrm{o}} \times r}$, $\widehat{\alpha}_i, \widehat{\beta}_j, \widehat{\gamma}_k \colon \mathcal{P} \to \mathbb{C}$, $\widehat{A}_i \in \mathbb{R}^{r \times r}$, $\widehat{B}_j \in \mathbb{R}^{r \times n_{\mathrm{i}}}$, and $\widehat{C}_k \in \mathbb{R}^{n_{\mathrm{o}} \times r}$. Note that when the reduced-order dimension r is small, evaluating the structured ROM, i.e., evaluating $\widehat{y}(\mathsf{p})$, is cheap. In [38], we showed that the structure of the ROM in (1.2)–(1.3) covers a wide range of problems including (parametric) LTI systems and models arising from discretization of stationary parametric partial differential equations. We revisit some concrete choices of $\widehat{\alpha}_i, \widehat{\beta}_j, \widehat{\gamma}_k$ later in the paper. In [38], we developed numerical algorithms to construct the ROM (1.2) in a purely data-driven fashion and called it a data-driven ROM. In this paper, we simply call it a structured reduced-order model (StROM).

In order to judge the quality of a ROM, one needs an error measure. In [38], we constructed the StROM to minimize the squared $\mathcal{L}_2$ error

$$\mathcal{J}\left(\widehat{A}_i, \widehat{B}_j, \widehat{C}_k\right) = \|y - \widehat{y}\|_{\mathcal{L}_2(\mathcal{P},\mu)}^2 = \int_{\mathcal{P}} \|y(\mathsf{p}) - \widehat{y}(\mathsf{p})\|_{\mathrm{F}}^2 \,\mathrm{d}\mu(\mathsf{p}) \tag{1.4}$$

and derived the gradients of $\mathcal{J}$ with respect to the StROM matrices $\widehat{A}_i, \widehat{B}_j, \widehat{C}_k$. These gradient formulae, which we also recall in Section 2, were then used in developing an optimization-based reduced-order modeling algorithm.

Starting with the formulation of [38], our goals here are to develop a unifying framework for interpolatory $\mathcal{L}_2$-optimal reduced-order modeling that covers both stationary and dynamical problems and to prove that bitangential Hermite interpolation is the necessary conditions for optimality in a much broader classes of problems than previously studied. More specifically, our main contributions are as follows:

1. We show that the existing interpolatory optimality conditions for approximating LTI systems is a special case of our formulation and directly follows from it (Subsection 3.1).
2. We derive interpolatory optimality conditions for approximating parametric LTI systems (Subsection 3.2).
3. We derive interpolatory optimality conditions for rational discrete LS measure (Section 4).
4. We derive interpolatory optimality conditions for approximation of parametric stationary problems (Section 5).

The rest of paper is organized as follows. In Section 2, we recall some of the main results from [38] and give the necessary optimality conditions, which we use repeatedly throughout the paper. In Section 3, we show applications to LTI systems, both parametric and non-parametric, using the continuous, Lebesgue measure. We consider the discrete LS measure in Section 4, where we derive interpolatory conditions for the LS problem. In Section 5, we consider a class of stationary parametric systems and derive interpolatory conditions for the $\mathcal{L}_2$-optimal ROMs. Conclusions are given in Section 6.

2. Mathematical Preliminaries.

Here we recall one of the main results of [38], namely the gradients of $\mathcal{J}$ (1.4) with respect to the StROM matrices, and then present the necessary optimality conditions that immediately follow from this result.

2.1. Gradients of the Squared $\mathcal{L}_2$ Error.

We begin with the necessary assumptions.

Assumption 2.1. For the problem setup in (1.1)–(1.4), let the following hold:

1. The set $\mathcal{P} \subseteq \mathbb{C}^{n_{\mathsf{p}}}$ is closed under conjugation ($\overline{\mathsf{p}} \in \mathcal{P}$ for all $\mathsf{p} \in \mathcal{P}$).
2. The measure μ over $\mathcal{P}$ is closed under conjugation (for any measurable set S, $\overline{S}$ is measurable and $\mu(\overline{S}) = \mu(S)$).
3. The function $y\colon \mathcal{P} \to \mathbb{C}^{n_{\mathrm{o}} \times n_{\mathrm{i}}}$ is measurable, closed under conjugation ($\overline{y(\mathsf{p})} = y(\overline{\mathsf{p}})$ for all $\mathsf{p} \in \mathcal{P}$), and square-integrable ($\|y\|_{\mathcal{L}_2(\mathcal{P},\mu)} < \infty$).
4. The scalar functions $\widehat{\alpha}_i, \widehat{\beta}_j, \widehat{\gamma}_k\colon \mathcal{P} \to \mathbb{C}$, for $i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}}$, $j = 1, 2, \ldots, q_{\widehat{\mathcal{B}}}$, and $k = 1, 2, \ldots, q_{\widehat{\mathcal{C}}}$, are measurable, closed under conjugation, and

$$\int_{\mathcal{P}} \left(\frac{\sum_{j=1}^{q_{\widehat{\mathcal{B}}}} \left| \widehat{\beta}_j(\mathsf{p}) \right| \sum_{k=1}^{q_{\widehat{\mathcal{C}}}} |\widehat{\gamma}_k(\mathsf{p})|}{\sum_{i=1}^{q_{\widehat{\mathcal{A}}}} |\widehat{\alpha}_i(\mathsf{p})|} \right)^2 \mathrm{d}\mu(\mathsf{p}) < \infty. \tag{2.1}$$

5. The matrices $\widehat{A}_1, \widehat{A}_2, \ldots, \widehat{A}_{q_{\widehat{\mathcal{A}}}}$ are such that

$$\operatorname*{ess\,sup}_{\mathsf{p} \in \mathcal{P}} \left\| \widehat{\alpha}_i(\mathsf{p}) \widehat{\mathcal{A}}(\mathsf{p})^{-1} \right\|_{\mathrm{F}} < \infty, \quad i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}},$$

   where $\widehat{\mathcal{A}}$ is as in (1.3).

These assumptions trivially hold in many cases, see [38]. Note that item 5 requires $\widehat{\mathcal{A}}(\mathsf{p})$ be invertible for μ-almost all p in $\mathcal{P}$.

THEOREM 2.2 (Theorem 3.7 in [38]). *Let $\mathcal{P}$, μ, y, $\widehat{\alpha}_i$, $\widehat{\beta}_j$, $\widehat{\gamma}_k$, and $\widehat{A}_i$ satisfy Assumption 2.1. Then, the gradients of $\mathcal{J}$ with respect to the StROM matrices are*

$$\nabla_{\widehat{A}_i} \mathcal{J} = 2 \int_{\mathcal{P}} \widehat{\alpha}_i(\overline{\mathsf{p}}) \widehat{x}_d(\mathsf{p}) [y(\mathsf{p}) - \widehat{y}(\mathsf{p})] \widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}), \qquad i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}}, \tag{2.2a}$$

$$\nabla_{\widehat{B}_j} \mathcal{J} = 2 \int_{\mathcal{P}} \widehat{\beta}_j(\overline{\mathsf{p}}) \widehat{x}_d(\mathsf{p}) [\widehat{y}(\mathsf{p}) - y(\mathsf{p})] \,\mathrm{d}\mu(\mathsf{p}), \qquad j = 1, 2, \ldots, q_{\widehat{\mathcal{B}}}, \tag{2.2b}$$

$$\nabla_{\widehat{C}_k} \mathcal{J} = 2 \int_{\mathcal{P}} \widehat{\gamma}_k(\overline{\mathsf{p}}) [\widehat{y}(\mathsf{p}) - y(\mathsf{p})] \widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}), \qquad k = 1, 2, \ldots, q_{\widehat{\mathcal{C}}}, \tag{2.2c}$$

where $\widehat{\mathcal{A}}(\mathsf{p})^* \widehat{x}_d(\mathsf{p}) = \widehat{\mathcal{C}}(\mathsf{p})^*$ is the reduced dual state equation, $\widehat{x}_d(\mathsf{p}) \in \mathbb{C}^{r \times n_{\mathrm{o}}}$ is the reduced dual state, and $(\cdot)^*$ denotes the conjugate transpose.

Based on these gradients, in [38] we developed an $\mathcal{L}_2$-optimal reduced-order modeling algorithm and demonstrated it on various examples, both stationary parametric problems and LTI systems. In this paper, we are more interested in the theoretical implications of Theorem 2.2 than the algorithmic ones and show how it provides a unifying framework for interpolatory optimal approximation.

2.2. Necessary Conditions for $\mathcal{L}_2$-optimality. An important consequence of Theorem 2.2 is that, by setting the gradients to zero, it yields the necessary optimality conditions for $\mathcal{L}_2$-optimal reduced-order modeling using parameter-separable forms.

COROLLARY 2.3. *Let $\mathcal{P}$, μ, y, $\widehat{\alpha}_i$, $\widehat{\beta}_j$, $\widehat{\gamma}_k$, and $\widehat{A}_i$ satisfy Assumption 2.1. Furthermore, let $(\widehat{A}_i, \widehat{B}_j, \widehat{C}_k)$ be an $\mathcal{L}_2$-optimal StROM. Then*

$$\int_{\mathcal{P}} \widehat{\gamma}_k(\overline{\mathsf{p}}) y(\mathsf{p}) \widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}) = \int_{\mathcal{P}} \widehat{\gamma}_k(\overline{\mathsf{p}}) \widehat{y}(\mathsf{p}) \widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}), \qquad k = 1, 2, \ldots, q_{\widehat{\mathcal{C}}}, \tag{2.3a}$$

$$\int_{\mathcal{P}} \widehat{\beta}_j(\overline{\mathsf{p}}) \widehat{x}_d(\mathsf{p}) y(\mathsf{p}) \,\mathrm{d}\mu(\mathsf{p}) = \int_{\mathcal{P}} \widehat{\beta}_j(\overline{\mathsf{p}}) \widehat{x}_d(\mathsf{p}) \widehat{y}(\mathsf{p}) \,\mathrm{d}\mu(\mathsf{p}), \qquad j = 1, 2, \ldots, q_{\widehat{\mathcal{B}}}, \tag{2.3b}$$

$$\int_{\mathcal{P}} \widehat{\alpha}_i(\overline{\mathsf{p}}) \widehat{x}_d(\mathsf{p}) y(\mathsf{p}) \widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}) = \int_{\mathcal{P}} \widehat{\alpha}_i(\overline{\mathsf{p}}) \widehat{x}_d(\mathsf{p}) \widehat{y}(\mathsf{p}) \widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}), \quad i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}}. \tag{2.3c}$$

The optimality conditions in Corollary 2.3 are *interpolatory* in the sense that the quantities (integrals) in the left-hand sides of (2.3a)–(2.3c) involving the full-order model (FOM) output $y(\mathsf{p})$ need to be interpolated by the same integrals involving the StROM output $\widehat{y}(\mathsf{p})$. This result highlights that any $\mathcal{L}_2$-optimal StROM with the parameter-separable form (1.3) is interpolatory in the sense of Corollary 2.3. By carefully selecting, in (2.3), the scalar functions $\widehat{\alpha}_i, \widehat{\beta}_j, \widehat{\gamma}_k$, the parameter space $\mathcal{P}$, and the measure μ over $\mathcal{P}$, we derive concrete interpolatory optimality conditions (in the form of bitangential Hermite interpolation) for important classes of ROMs, including non-parametric and parametric LTI systems, stationary models, and for discrete LS problems, thus providing a unifying framework for interpolatory $\mathcal{L}_2$-optimal approximation across different domains.

3. Linear Time-invariant Systems: Continuous Measure. Our goal in this section is to illustrate (i) how Theorem 2.2 and Corollary 2.3 cover a wide range of settings arising in optimal MOR of dynamical systems and (ii) to develop new conditions for optimality. Furthermore, this analysis sets the stage for the interpolatory conditions we derive in Section 4 for discrete LS minimization and in Section 5 for stationary problems.

3.1. $\mathcal{H}_2$-optimal Model Order Reduction. Interpolatory necessary optimality conditions are known for $\mathcal{H}_2$-optimal MOR of LTI systems, both for the continuous-time case [2,29,36] and the discrete-time case [16,29]. In the following, we show that these conditions are a special case of the conditions in Corollary 2.3.

A continuous-time, finite-dimensional LTI system is given by

$$E\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = 0, \tag{3.1a}$$

$$y(t) = Cx(t), \tag{3.1b}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^{n_\mathrm{i}}$ is the input, $y(t) \in \mathbb{R}^{n_\mathrm{o}}$ is the output, $E, A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times n_\mathrm{i}}$, and $C \in \mathbb{R}^{n_\mathrm{o} \times n}$. We assume that E is invertible and all eigenvalues of $E^{-1}A$ have negative real parts. The rational function

$$H(s) = C(sE - A)^{-1}B \tag{3.2}$$

is the transfer function of (3.1) and satisfies $Y(s) = H(s)U(s)$ assuming Y and U, the Laplace transforms of y and u, exist; see, e.g., [42], for the conditions for the existence of the Laplace transform. Based on the assumptions above, H belongs to the $\mathcal{H}_2^{n_\mathrm{o} \times n_\mathrm{i}}(\mathbb{C}_+)$ Hardy space (where $\mathbb{C}_+$ is the open left half-plane), which is the set of holomorphic functions $F \colon \mathbb{C}_+ \to \mathbb{C}^{n_\mathrm{o} \times n_\mathrm{i}}$ such that $\sup_{\eta > 0} \int_{-\infty}^{\infty} \|F(\eta + \imath\omega)\|_\mathrm{F}^2 \,\mathrm{d}\omega < \infty$. It is known that F can be extended to $\overline{\mathbb{C}_+}$ and the $\mathcal{H}_2$ norm can be defined as

$$\|F\|_{\mathcal{H}_2} = \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \|F(\imath\omega)\|_\mathrm{F}^2 \,\mathrm{d}\omega \right)^{1/2}.$$

The analysis applies to any $H \in \mathcal{H}_2^{n_\mathrm{o} \times n_\mathrm{i}}(\mathbb{C}_+)$ as a FOM, including infinite-dimensional systems, and not just the finite-dimensional systems as in (3.1).

The goal of $\mathcal{H}_2$-optimal MOR is to find a ROM

$$\widehat{E}\dot{\widehat{x}}(t) = \widehat{A}\widehat{x}(t) + \widehat{B}u(t), \quad \widehat{x}(0) = 0, \tag{3.3a}$$

$$\widehat{y}(t) = \widehat{C}\widehat{x}(t), \tag{3.3b}$$

with the reduced state $\widehat{x}(t) \in \mathbb{R}^r$, the approximate output $\widehat{y}(t) \in \mathbb{R}^{n_\mathrm{o}}$, and the reduced quantities $\widehat{E}, \widehat{A} \in \mathbb{R}^{r \times r}$, $\widehat{B} \in \mathbb{R}^{r \times n_\mathrm{i}}$, and $\widehat{C} \in \mathbb{R}^{n_\mathrm{o} \times r}$ such that its transfer function $\widehat{H}(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B}$ minimizes the $\mathcal{H}_2$ error $\|H - \widehat{H}\|_{\mathcal{H}_2}$. The assumptions on the ROM are that $\widehat{E}$ be invertible and all eigenvalues of $\widehat{E}^{-1}\widehat{A}$ have negative real parts.

To state the $\mathcal{H}_2$-optimal interpolatory conditions, let $\widehat{H}(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B}$ have r distinct poles. Let $\widehat{T}, \widehat{S} \in \mathbb{C}^{r \times r}$ be invertible matrices such that $\widehat{S}^*\widehat{E}\widehat{T} = I$ and $\widehat{S}^*\widehat{A}\widehat{T} = \Lambda$ with $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$. Then, we can write $\widehat{H}(s)$ as

$$\widehat{H}(s) = \widehat{C}\widehat{T}(sI - \Lambda)^{-1}\widehat{S}^*\widehat{B} = \sum_{j=1}^{r} \frac{\widehat{C}\widehat{T}e_j e_j^\mathrm{T}\widehat{S}^*\widehat{B}}{s - \lambda_j} = \sum_{j=1}^{r} \frac{c_j b_j^*}{s - \lambda_j}, \tag{3.4}$$

where $c_j = \widehat{C}\widehat{T}e_j \in \mathbb{C}^{n_\mathrm{o}}$, $b_j = \widehat{B}^\mathrm{T}\widehat{S}e_j \in \mathbb{C}^{n_\mathrm{i}}$, and e_j denotes the jth unit vector (of appropriate size). The formulation of $\widehat{H}$ in (3.4) is called the *pole-residue form* where λ_j are the poles and $c_j b_j^*$ are the (rank-1) residues. If $\widehat{H}$ is an $\mathcal{H}_2$-optimal ROM of H, then it satisfies the interpolation conditions

$$H\big(-\overline{\lambda_k}\big)b_k = \widehat{H}\big(-\overline{\lambda_k}\big)b_k, \tag{3.5a}$$

$$c_k^* H\big(-\overline{\lambda_k}\big) = c_k^* \widehat{H}\big(-\overline{\lambda_k}\big), \tag{3.5b}$$

$$c_k^* H'\big(-\overline{\lambda_k}\big)b_k = c_k^* \widehat{H}'\big(-\overline{\lambda_k}\big)b_k, \tag{3.5c}$$

for $k = 1, 2, \ldots, r$, where H' and $\widehat{H}'$ denote the derivatives with respect to s; see [3,29]. More specifically, (3.5a) is called the right-tangential interpolation condition, (3.5b) the left-tangential interpolation condition, and (3.5c) the bitangential Hermite interpolation condition. We refer to all three conditions together as the bitangential Hermite interpolation conditions for $\mathcal{H}_2$-optimality. They state that the optimal reduced-order transfer function $\widehat{H}$ tangentially interpolates H (and $\widehat{H}'$ interpolates H') at the mirror images of its own poles, i.e., at $-\overline{\lambda_k}$, along the tangent directions c_k and b_k determined by its own rank-1 residues $c_k b_k^*$. These optimal interpolation conditions have led to effective numerical methods for optimal MOR and has been extended to various setting; for details we refer the reader to [3, 6, 29] and the references therein. We also refer the reader to [20–24] and the references therein for greedy-based selections of interpolation points in projection-based MOR of LTI systems.

Now we show how these optimal interpolatory conditions can be recovered from Corollary 2.3 as a special case. First we note that by setting $\mathcal{P} = \imath\mathbb{R}$, $\mathsf{p} = \imath\omega$,

$\mu = \frac{1}{2\pi}\lambda_{\imath\mathbb{R}}$ (where $\lambda_{\imath\mathbb{R}}$ is the Lebesgue measure over $\imath\mathbb{R}$), $y = H$, and $\widehat{y} = \widehat{H}$, we find $\|y\|_{\mathcal{L}_2(\mathcal{P},\mu)} = \|H\|_{\mathcal{H}_2}$ and thus $\|y-\widehat{y}\|_{\mathcal{L}_2(\mathcal{P},\mu)} = \|H-\widehat{H}\|_{\mathcal{H}_2}$. Furthermore, the reduced transfer function $\widehat{H}(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B}$ can be viewed as a StROM as in (1.2) by rewriting it as

$$\left(s\widehat{E} - \widehat{A}\right)\widehat{X}(s) = \widehat{B},$$
$$\widehat{H}(s) = \widehat{C}\widehat{X}(s),$$

and by selecting $\widehat{x} = \widehat{X}$, $\widehat{y} = \widehat{H}$, $q_{\widehat{\mathcal{A}}} = 2$, $\widehat{\alpha}_1(\mathsf{p}) = \mathsf{p}$, $\widehat{\alpha}_2(\mathsf{p}) = -1$, $q_{\widehat{\mathcal{B}}} = 1$, $\widehat{\beta}_1(\mathsf{p}) = 1$, $q_{\widehat{\mathcal{C}}} = 1$, and $\widehat{\gamma}_1(\mathsf{p}) = 1$ in (1.3). To obtain the optimality conditions (3.5) from Corollary 2.3, postmultiply the left-hand side of (2.3a) by $\widehat{T}^{-*}e_k$ to get

$$\begin{aligned}
&\int_{\mathcal{P}} \widehat{\gamma}_1(\overline{\mathsf{p}}) y(\mathsf{p}) \widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p})\, \widehat{T}^{-*}e_k = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\imath\omega)\widehat{B}^{\mathrm{T}}\left(\imath\omega\widehat{E} - \widehat{A}\right)^{-*} \mathrm{d}\omega\, \widehat{T}^{-*}e_k \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\imath\omega)\widehat{B}^{\mathrm{T}}\widehat{S}(\imath\omega I - \Lambda)^{-*}\widehat{T}^{*}\,\mathrm{d}\omega\, \widehat{T}^{-*}e_k = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{H(\imath\omega)b_k}{-\imath\omega - \overline{\lambda_k}}\,\mathrm{d}\omega.
\end{aligned}$$

Switching to a contour integral with $s = \imath\omega$ and $\mathrm{d}s = \imath\,\mathrm{d}\omega$, we find

$$\begin{aligned}
&\int_{\mathcal{P}} \widehat{\gamma}_1(\overline{\mathsf{p}}) y(\mathsf{p}) \widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p})\, \widehat{T}^{-*}e_k = \frac{1}{2\pi\imath}\oint_{\imath\mathbb{R}} \frac{H(s)b_k}{-s - \overline{\lambda_k}}\,\mathrm{d}s = -\frac{1}{2\pi\imath}\oint_{\imath\mathbb{R}} \frac{H(s)b_k}{s - (-\overline{\lambda_k})}\,\mathrm{d}s \\
&= \lim_{R\to\infty} -\frac{1}{2\pi\imath}\oint_{\Gamma_R} \frac{H(s)b_k}{s - (-\overline{\lambda_k})}\,\mathrm{d}s = H\left(-\overline{\lambda_k}\right)b_k,
\end{aligned}$$

where $\Gamma_R = [-\imath R, \imath R] \cup \{Re^{\imath\omega} : \omega \in [-\frac{\pi}{2}, \frac{\pi}{2}]\}$ is the clockwise contour of a semidisk as in [2, Lemma 1.1]. Applying the same manipulations to the right-hand side of (2.3a) yields $\widehat{H}(-\overline{\lambda_k})b_k$. Thus the first optimality conditions (2.3a) in Corollary 2.3 yield the right-tangential interpolation conditions (3.5a) for $\mathcal{H}_2$-optimality as a special case.

Similarly, premultiplying the left-hand side of (2.3b) by $e_k^{\mathrm{T}}\widehat{S}^{-1}$, we obtain

$$e_k^{\mathrm{T}}\widehat{S}^{-1}\int_{\mathcal{P}} \widehat{\beta}_1(\overline{\mathsf{p}})\widehat{x}_d(\mathsf{p}) y(\mathsf{p})\,\mathrm{d}\mu(\mathsf{p}) = c_k^* H\left(-\overline{\lambda_k}\right).$$

Thus, (2.3b) yields the left-tangential interpolation condition (3.5b) for $\mathcal{H}_2$-optimality (after applying the same manipulations to the right-hand side of (2.3b)). Finally, taking the left-hand side of (2.3c) in Corollary 2.3 related to $\widehat{A}$ with $\widehat{\alpha}_2(\mathsf{p}) = -1$, premultiplying it by $e_k^{\mathrm{T}}\widehat{S}^{-1}$, and postmultiplying it by $\widehat{T}^{-*}e_k$ gives

$$\begin{aligned}
&e_k^{\mathrm{T}}\widehat{S}^{-1}\int_{\mathcal{P}} \widehat{\alpha}_2(\overline{\mathsf{p}})\widehat{x}_d(\mathsf{p}) y(\mathsf{p}) \widehat{x}(\mathsf{p})^*\,\mathrm{d}\mu(\mathsf{p})\, \widehat{T}^{-*}e_k = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{c_k^* H(\imath\omega)b_k}{\left(-\imath\omega - \overline{\lambda_k}\right)^2}\,\mathrm{d}\omega \\
&= -\frac{1}{2\pi\imath}\oint_{\imath\mathbb{R}} \frac{c_k^* H(s)b_k}{\left(-s - \overline{\lambda_k}\right)^2}\,\mathrm{d}s = c_k^* H'\left(-\overline{\lambda_k}\right)b_k,
\end{aligned}$$

obtaining the final Hermite interpolation condition (3.5c). Thus, we recover the interpolatory $\mathcal{H}_2$-optimality conditions (3.5) as a special case of the more general $\mathcal{L}_2$-optimality conditions (2.3) in Corollary 2.3.

Remark 3.1. The case of discrete-time LTI systems follows similarly. Discrete-time systems are obtained by replacing the derivative term $\dot{x}(t)$ in (3.1) with the time shift $x(t+1)$ and restricting the time t to integers. Transfer function H of a discrete-time LTI system has exactly the same form as (3.2), but we now assume that $E^{-1}A$ has eigenvalues in the open unit disk. The corresponding Hardy space is $h_2^{n_\mathrm{o} \times n_\mathrm{i}}(\overline{\mathbb{D}}^c)$ ($\mathbb{D}$ is the open unit disk and $\overline{\mathbb{D}}^c$ is the complement of the closed unit disk) containing functions F such that $\sup_{r>1} \int_0^{2\pi} \|F(re^{\imath\omega})\|_\mathrm{F}^2 \,\mathrm{d}\omega < \infty$ and the h_2 norm is given by

$$\|F\|_{h_2} = \left(\frac{1}{2\pi} \int_0^{2\pi} \|F(e^{\imath\omega})\|_\mathrm{F}^2 \,\mathrm{d}\omega \right)^{1/2}.$$

Then, we recover the interpolatory necessary optimality conditions from [16,29] analogously to the continuous-time case, in this case by setting $\mathcal{P} = \partial\mathbb{D}$, $\mathsf{p} = e^{\imath\omega}$, $\mu = \frac{1}{2\pi}\lambda_{\partial\mathbb{D}}$ (where $\lambda_{\partial\mathbb{D}}$ is the Lebesgue measure over $\partial\mathbb{D}$), and taking the contour integral over the unit circle. In particular, if $\widehat{H}(s) = \sum_{j=1}^{r} \frac{c_j b_j^*}{s-\lambda_j}$ is an h_2-optimal ROM for H, then

(3.6)
$$H\left(\frac{1}{\overline{\lambda_k}}\right) b_k = \widehat{H}\left(\frac{1}{\overline{\lambda_k}}\right) b_k, \quad c_k^* H\left(\frac{1}{\overline{\lambda_k}}\right) = c_k^* \widehat{H}\left(\frac{1}{\overline{\lambda_k}}\right), \quad c_k^* H'\left(\frac{1}{\overline{\lambda_k}}\right) b_k = c_k^* \widehat{H}'\left(\frac{1}{\overline{\lambda_k}}\right) b_k,$$

for $k = 1, 2, \ldots, r$. As in the continuous-time case, bitangential Hermite interpolation forms the necessary conditions for h_2-optimality where the interpolation points depend on the reduced-order poles and the tangent directions on the reduced-order residues. The difference is that the mirroring of the reduced-order poles to the interpolation points is now done with respect to the unit circle.

3.2. $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal Parametric Model Order Reduction. In Subsection 3.1, we considered non-parametric LTI systems as in (3.1) and showed how Corollary 2.3 recovers the well-known $\mathcal{H}_2$-optimality conditions. In this section, we focus on parametric LTI systems where the underlying dynamics depend on a set of parameters and thus consider *jointly optimal* approximation in the frequency and parameter space. We show how our $\mathcal{L}_2$-optimal modeling framework naturally covers this problem as well and extends the existing interpolatory necessary optimality conditions to a more general setting.

We consider FOMs of the form

$$\mathcal{E}(\xi)\dot{x}(t,\xi) = \mathcal{A}(\xi)x(t,\xi) + \mathcal{B}(\xi)u(t), \quad x(0,\xi) = 0, \tag{3.7a}$$
$$y(t,\xi) = \mathcal{C}(\xi)x(t,\xi), \tag{3.7b}$$

where $\xi \in \Xi \subset \mathbb{C}$ is the parameter, Ξ is the parameter space, $x(t,\xi) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}$ is the input, $y(t,\xi) \in \mathbb{R}$ is the output, $\mathcal{E}(\xi), \mathcal{A}(\xi) \in \mathbb{R}^{n \times n}$, $\mathcal{B}(\xi) \in \mathbb{R}^{n \times n_\mathrm{i}}$, and $\mathcal{C}(\xi) \in \mathbb{R}^{n_\mathrm{o} \times n}$. In practical applications, the variable ξ can correspond to, e.g., geometry, material properties (such as thickness), boundary conditions etc. We assume that $\mathcal{E}, \mathcal{A}, \mathcal{B}, \mathcal{C}$ are holomorphic, $\mathcal{E}(\xi)$ is invertible for all $\xi \in \Xi$, and $\mathcal{E}(\xi)^{-1}\mathcal{A}(\xi)$ has eigenvalues in the open left half-plane for all $\xi \in \Xi$. Then the corresponding parametric transfer function of (3.7) is $H(s,\xi) = \mathcal{C}(\xi)(s\mathcal{E}(\xi) - \mathcal{A}(\xi))^{-1}\mathcal{B}(\xi)$. Note that unlike in the non-parametric LTI case, transfer function now depends on both the frequency variable s and the parameter ξ. Thus a reduced-order transfer-function approximation $\widehat{H}(s,\xi) = \widehat{\mathcal{C}}(\xi)(s\widehat{\mathcal{E}}(\xi) - \widehat{\mathcal{A}}(\xi))^{-1}\widehat{\mathcal{B}}(\xi)$ to $H(s,\xi)$ should have high fidelity both in s and ξ. How should one choose the space in which to approximate $H(s,\xi)$?

Grimm [28] focused on a simplified problem and considered the single-input and single-output parametric LTI system, i.e., $n_\mathrm{i} = n_\mathrm{o} = 1$ in (3.7), and as the approximation space considered the Hardy space in two variables $\mathcal{H}_2(\mathbb{C}_+ \times \mathbb{D})$, which is the space of holomorphic functions $F\colon \mathbb{C}_+ \times \mathbb{D} \to \mathbb{C}$ such that

$$\sup_{\eta>0,\ 0<r<1} \int_0^{2\pi} \int_{-\infty}^{\infty} \lvert F(\eta + \imath\omega, r e^{\imath\omega_\xi}) \rvert^2 \,\mathrm{d}\omega \,\mathrm{d}\omega_\xi < \infty.$$

The corresponding $\mathcal{H}_2(\mathbb{C}_+ \times \mathbb{D})$ norm, refereed to as the $\mathcal{H}_2 \otimes \mathcal{L}_2$ norm, is given by

$$\lVert F \rVert_{\mathcal{H}_2 \otimes \mathcal{L}_2} = \left(\frac{1}{4\pi^2} \int_0^{2\pi} \int_{-\infty}^{\infty} \lvert F(\imath\omega, e^{\imath\omega_\xi}) \rvert^2 \,\mathrm{d}\omega \,\mathrm{d}\omega_\xi \right)^{1/2}.$$

The $\mathcal{H}_2 \otimes \mathcal{L}_2$ used in [28] is a special case of the more general $\mathcal{H}_2 \otimes \mathcal{L}_2$ norm definition introduced in [4]. Unlike the non-parametric LTI case considered in Subsection 3.1, the optimal interpolatory conditions for MOR in the $\mathcal{H}_2 \otimes \mathcal{L}_2$ norm are not known in general except for some special cases. For example, [4] shows that if the parametric dependency ξ only appears in $\mathcal{B}(\xi)$ and $\mathcal{C}(\xi)$ in (3.7), and $\mathcal{A}$ and $\mathcal{E}$ are non-parametric, then one can convert the resulting problem into an equivalent $\mathcal{H}_2$-optimal approximation problem and obtain interpolatory optimality conditions. However, this is restrictive since in most parametric problems, $\mathcal{A}$ and $\mathcal{E}$ vary with a parameter.

Grimm [28], instead, considers a simplification in the form of the ROM. Recall that the pole-residue form (3.4) has proved vital in deriving the $\mathcal{H}_2$-optimality conditions. Inspired by (3.4), [28] considered ROMs with the form

$$\widehat{H}(s, \xi) = \sum_{i=1}^{r} \sum_{j=1}^{r_\xi} \frac{\phi_{ij}}{(s - \lambda_i)(\xi - \pi_j)}, \tag{3.8}$$

where $\phi_{ij} \in \mathbb{C}$, $\lambda_i \in \mathbb{C}_-$, and $\pi_j \in \overline{\mathbb{D}}^c$ ($\mathbb{C}_-$ is the open left half-plane). In (3.8), one may view λ_i's as the frequency poles (in the s variable) and π_j's as the parameter poles (in the ξ variable). Then [28, Thm 3.3.4] shows that the interpolatory necessary conditions for $\widehat{H}$ in (3.8) to be an $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal ROM are

$$H\left(-\lambda_k, \frac{1}{\pi_\ell}\right) = \widehat{H}\left(-\lambda_k, \frac{1}{\pi_\ell}\right), \tag{3.9a}$$

$$\sum_{j=1}^{r_\xi} \frac{\phi_{kj}}{\pi_j} \frac{\partial H}{\partial s}\left(-\lambda_k, \frac{1}{\pi_j}\right) = \sum_{j=1}^{r_\xi} \frac{\phi_{kj}}{\pi_j} \frac{\partial \widehat{H}}{\partial s}\left(-\lambda_k, \frac{1}{\pi_j}\right), \tag{3.9b}$$

$$\sum_{i=1}^{r} \phi_{i\ell} \frac{\partial H}{\partial \xi}\left(-\lambda_i, \frac{1}{\pi_\ell}\right) = \sum_{i=1}^{r} \phi_{i\ell} \frac{\partial \widehat{H}}{\partial \xi}\left(-\lambda_i, \frac{1}{\pi_\ell}\right), \tag{3.9c}$$

for $k = 1, 2, \ldots, r$ and $\ell = 1, 2, \ldots, r_\xi$. In the rest of this section, we show that the interpolation conditions in Corollary 2.3 cover the $\mathcal{H}_2 \otimes \mathcal{L}_2$ framework of [28] and at the same time extend the analysis to $\mathcal{H}_2 \otimes \mathcal{L}_2$ approximation of *multiple-input/multiple-output* parametric LTI systems. Therefore, in (3.7) we do not need to assume $n_{\mathrm{i}} = n_{\mathrm{o}} = 1$ as done in [28].

For matrix-valued transfer functions H and $\widehat{H}$, the squared $\mathcal{H}_2 \otimes \mathcal{L}_2$ error is defined as

$$\left\lVert H - \widehat{H} \right\rVert_{\mathcal{H}_2 \otimes \mathcal{L}_2}^2 = \frac{1}{4\pi^2} \int_0^{2\pi} \int_{-\infty}^{\infty} \left\lVert H(\imath\omega, e^{\imath\omega_\xi}) - \widehat{H}(\imath\omega, e^{\imath\omega_\xi}) \right\rVert_{\mathrm{F}}^2 \,\mathrm{d}\omega \,\mathrm{d}\omega_\xi. \tag{3.10}$$

Observe that the ROM (3.8) (for the single-input/single-output case) can be written as a StROM in (1.2) using the representation

$$\begin{aligned}\left[(sI_r - \Lambda) \otimes (\xi I_{r_\xi} - \Pi)\right] \widehat{X}(s,\xi) &= \mathbb{1}, \\ \widehat{H}(s,\xi) &= \phi^{\mathrm{T}} \widehat{X}(s,\xi),\end{aligned}$$

where $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$, $\Pi = \operatorname{diag}(\pi_1, \pi_2, \ldots, \pi_{r_\xi})$, $\mathbb{1} = [1, \ldots, 1]^{\mathrm{T}}$, and $\phi = [\phi_{1,1}, \ldots, \phi_{r,r_\xi}]^{\mathrm{T}}$. To extend this formulation to multiple-input/multiple-output problems, we then consider

$$\left[\left(s\widehat{E} - \widehat{A}\right) \otimes \left(\xi \widehat{E}_\xi - \widehat{A}_\xi\right)\right] \widehat{X}(s,\xi) = \widehat{B}, \tag{3.11a}$$

$$\widehat{H}(s,\xi) = \widehat{C}\widehat{X}(s,\xi), \tag{3.11b}$$

where $\widehat{E}, \widehat{A} \in \mathbb{R}^{r \times r}$, $\widehat{E}_\xi, \widehat{A}_\xi \in \mathbb{R}^{r_\xi \times r_\xi}$, $\widehat{B} \in \mathbb{R}^{r r_\xi \times n_{\mathrm{i}}}$, and $\widehat{C} \in \mathbb{R}^{n_{\mathrm{o}} \times r r_\xi}$. Note that (3.11) is a StROM as in (1.2) with $\mathsf{p} = (s,\xi)$ and $\widehat{\mathcal{A}}(\mathsf{p}) = (s\widehat{E} - \widehat{A}) \otimes (\xi \widehat{E}_\xi - \widehat{A}_\xi)$. By expanding $\widehat{\mathcal{A}}(\mathsf{p})$, we find that the form (3.11) corresponds to $q_{\widehat{\mathcal{A}}} = 4$ and

$$\widehat{\alpha}_1(s,\xi) = s\xi, \quad \widehat{\alpha}_2(s,\xi) = -s, \quad \widehat{\alpha}_3(s,\xi) = -\xi, \quad \widehat{\alpha}_4(s,\xi) = 1, \tag{3.12a}$$

$$\widehat{A}_1 = \widehat{E} \otimes \widehat{E}_\xi, \quad \widehat{A}_2 = \widehat{E} \otimes \widehat{A}_\xi, \quad \widehat{A}_3 = \widehat{A} \otimes \widehat{E}_\xi, \quad \widehat{A}_4 = \widehat{A} \otimes \widehat{A}_\xi. \tag{3.12b}$$

Therefore, the ROM (3.11) is a StROM (1.2) fitting into our $\mathcal{L}_2$-optimal modeling framework where the StROM matrices have an additional Kronecker structure. The following lemma will help us in computing the gradients of the cost function with respect to $\widehat{A}$ and $\widehat{A}_\xi$ having this specific Kronecker structure.

LEMMA 3.2. *Let $F\colon \mathbb{R}^{nm \times nm} \to \mathbb{R}$ be a differentiable function at $X = A \otimes B$ where $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$ are nonzero matrices. Let $G\colon \mathbb{R}^{n \times n} \to \mathbb{R}$ be defined as $G(A_1) = F(A_1 \otimes B)$. Then G is differentiable at A and*

$$\nabla G(A) = \sum_{j=1}^{m} \left(I_n \otimes e_j^{\mathrm{T}} B_L^*\right) \nabla F(X) (I_n \otimes B_R^* e_j),$$

for any $B_L, B_R \in \mathbb{C}^{m \times m}$ such that $B = B_L B_R$. Similarly, let $H\colon \mathbb{R}^{m \times m} \to \mathbb{R}$ be defined as $H(B_1) = F(A \otimes B_1)$. Then H is differentiable at B and

$$\nabla H(B) = \sum_{i=1}^{n} \left(e_i^{\mathrm{T}} A_L^* \otimes I_m\right) \nabla F(X) (A_R^* e_i \otimes I_m),$$

for any $A_L, A_R \in \mathbb{C}^{n \times n}$ such that $A = A_L A_R$.

Proof. See Appendix A. ∎

Before stating the next theorem, we need to derive a pole-residue form of (3.11). Let $\widehat{T}, \widehat{S} \in \mathbb{C}^{r \times r}$ and $\widehat{T}_\xi, \widehat{S}_\xi \in \mathbb{C}^{r_\xi \times r_\xi}$ be invertible matrices such that $\widehat{S}^* \widehat{E} \widehat{T} = I_r$, $\widehat{S}^* \widehat{A} \widehat{T} = \Lambda$, $\widehat{S}_\xi^* \widehat{E}_\xi \widehat{T}_\xi = I_{r_\xi}$, and $\widehat{S}_\xi^* \widehat{A}_\xi \widehat{T}_\xi = \Pi$. Then

$$\begin{aligned}\widehat{H}(s,\xi) &= \widehat{C}\left(\widehat{T} \otimes \widehat{T}_\xi\right)\left[(sI_r - \Lambda) \otimes \left(\xi I_{r_\xi} - \Pi\right)\right]^{-1}\left(\widehat{S}^* \otimes \widehat{S}_\xi^*\right)\widehat{B} \\ &= \sum_{i=1}^{r} \sum_{j=1}^{r_\xi} \frac{c_{ij} b_{ij}^*}{(s-\lambda_i)(\xi - \pi_j)},\end{aligned} \tag{3.13}$$

where $c_{ij} = \widehat{C}(\widehat{T} \otimes \widehat{T}_\xi)(e_i \otimes e_j)$ and $b_{ij} = \widehat{B}^{\mathrm{T}}(\widehat{S} \otimes \widehat{S}_\xi)(e_i \otimes e_j)$.

THEOREM 3.3. *Let $\widehat{H}$ be a StROM as in (3.11) with the pole-residue form (3.13) where $\lambda_i \in \mathbb{C}_-$ and $\pi_j \in \overline{\mathbb{D}}^c$ are pairwise distinct. If $\widehat{H}$ is an $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal approximation to $H \in \mathcal{H}_2^{n_{\mathrm{o}} \times n_{\mathrm{i}}}(\mathbb{C}_+ \times \mathbb{D})$, then*

$$H\left(-\overline{\lambda_k}, \frac{1}{\overline{\pi_\ell}}\right) b_{k\ell} = \widehat{H}\left(-\overline{\lambda_k}, \frac{1}{\overline{\pi_\ell}}\right) b_{k\ell}, \tag{3.14a}$$

$$c_{k\ell}^* H\left(-\overline{\lambda_k}, \frac{1}{\overline{\pi_\ell}}\right) = c_{k\ell}^* \widehat{H}\left(-\overline{\lambda_k}, \frac{1}{\overline{\pi_\ell}}\right), \tag{3.14b}$$

$$\sum_{j=1}^{r_\xi} \frac{1}{\overline{\pi_j}} c_{kj}^* \frac{\partial H}{\partial s}\left(-\overline{\lambda_k}, \frac{1}{\overline{\pi_j}}\right) b_{kj} = \sum_{j=1}^{r_\xi} \frac{1}{\overline{\pi_j}} c_{kj}^* \frac{\partial \widehat{H}}{\partial s}\left(-\overline{\lambda_k}, \frac{1}{\overline{\pi_j}}\right) b_{kj}, \tag{3.14c}$$

$$\sum_{i=1}^{r} c_{i\ell}^* \frac{\partial H}{\partial \xi}\left(-\overline{\lambda_i}, \frac{1}{\overline{\pi_\ell}}\right) b_{i\ell} = \sum_{i=1}^{r} c_{i\ell}^* \frac{\partial \widehat{H}}{\partial \xi}\left(-\overline{\lambda_i}, \frac{1}{\overline{\pi_\ell}}\right) b_{i\ell}, \tag{3.14d}$$

for $k = 1, 2, \ldots, r$ and $\ell = 1, 2, \ldots, r_\xi$.

Proof. First observe that we recover the $\mathcal{H}_2 \otimes \mathcal{L}_2$ norm by setting $\mathcal{P} = \imath\mathbb{R} \times \partial\mathbb{D}$, $\mathsf{p} = (\imath\omega, e^{\imath\omega_\xi})$, $\mu = \frac{1}{4\pi^2}\lambda_{\imath\mathbb{R}} \times \lambda_{\partial\mathbb{D}}$, and $y(\mathsf{p}) = H(s, \xi)$ in the definition of the $\mathcal{L}_2(\mathcal{P}, \mu)$ norm. Also we have already shown that $\widehat{H}$ in (3.11) is a StROM as in (1.2) with $\widehat{x}(\mathsf{p}) = \widehat{X}(s, \xi)$, $\widehat{y}(\mathsf{p}) = \widehat{H}(s, \xi)$, and the Kronecker structure (3.12). We start by postmultiplying the left-hand side of the first condition (2.3a) in Corollary 2.3 by $(\widehat{T} \otimes \widehat{T}_\xi)^{-*}(e_k \otimes e_\ell)$ to obtain

$$\begin{aligned} &\int_{\mathcal{P}} y(\mathsf{p})\widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}) \left(\widehat{T} \otimes \widehat{T}_\xi\right)^{-*} (e_k \otimes e_\ell) \\ &= \frac{1}{4\pi^2} \int_0^{2\pi} \int_{-\infty}^{\infty} \frac{H(\imath\omega, e^{\imath\omega_\xi}) b_{k\ell}}{(-\imath\omega - \overline{\lambda_k})(e^{-\imath\omega_\xi} - \overline{\pi_\ell})} \,\mathrm{d}\omega \,\mathrm{d}\omega_\xi. \end{aligned}$$

Substituting $s = \imath\omega$, $\mathrm{d}s = \imath \,\mathrm{d}\omega$, $\xi = e^{\imath\omega_\xi}$, and $\mathrm{d}\xi = \imath\xi \,\mathrm{d}\omega_\xi$, we find

$$\begin{aligned} \int_{\mathcal{P}} y(\mathsf{p})\widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}) \left(\widehat{T} \otimes \widehat{T}_\xi\right)^{-*} (e_k \otimes e_\ell) &= \frac{1}{4\pi^2\imath^2} \oint_{\partial\mathbb{D}} \oint_{\imath\mathbb{R}} \frac{\frac{1}{\xi} H(s, \xi) b_{k\ell}}{\left(-s - \overline{\lambda_k}\right)\left(\frac{1}{\xi} - \overline{\pi_\ell}\right)} \,\mathrm{d}s \,\mathrm{d}\xi \\ = \frac{1}{4\pi^2\imath^2\overline{\pi_\ell}} \oint_{\partial\mathbb{D}} \oint_{\imath\mathbb{R}} \frac{H(s, \xi) b_{k\ell}}{\left(s - (-\overline{\lambda_k})\right)\left(\xi - \frac{1}{\overline{\pi_\ell}}\right)} \,\mathrm{d}s \,\mathrm{d}\xi &= -\frac{1}{\overline{\pi_\ell}} H\left(-\overline{\lambda_k}, \frac{1}{\overline{\pi_\ell}}\right) b_{k\ell}, \end{aligned}$$

by using the Cauchy integral formula twice in the last equality. Performing similar operations to the right-hand side of (2.3a), we obtain the condition (3.14a). Similarly, premultiplying the left-hand side of (2.3b) by $(e_k \otimes e_\ell)^{\mathrm{T}}(\widehat{S} \otimes \widehat{S}_\xi)^{-1}$, we obtain

$$(e_k \otimes e_\ell)^{\mathrm{T}} \left(\widehat{S} \otimes \widehat{S}_\xi\right)^{-1} \int_{\mathcal{P}} \widehat{x}_d(\mathsf{p}) y(\mathsf{p}) \,\mathrm{d}\mu(\mathsf{p}) = -\frac{1}{\overline{\pi_\ell}} c_{k\ell}^* H\left(-\overline{\lambda_k}, \frac{1}{\overline{\pi_\ell}}\right).$$

Doing the same for the right-hand side of (2.3b), we obtain the condition (3.14b).

Similar to recovering the bitangential Hermite condition for $\mathcal{H}_2$-optimality (3.5c) where we used the gradient of the squared $\mathcal{H}_2$ error with respect to $\widehat{A}$, in this setting, we need to differentiate the squared $\mathcal{H}_2 \otimes \mathcal{L}_2$ error (3.10) with respect to $\widehat{A}$ and $\widehat{A}_\xi$. We start by computing the gradient with respect to $\widehat{A}$. Using Lemma 3.2 with

$\widehat{E}_\xi = \widehat{S}_\xi^{-*}\widehat{T}_\xi^{-1}$ and $\widehat{A}_\xi = (\widehat{S}_\xi^{-*}\Pi)\widehat{T}_\xi^{-1}$, we see that

$$\begin{aligned}
&\frac{1}{2}\nabla_{\widehat{A}}\left\|H-\widehat{H}\right\|_{\mathcal{H}_2\otimes\mathcal{L}_2}^2\\
&=\sum_{j=1}^{r_\xi}\left(I_r\otimes e_j^{\mathrm{T}}\widehat{S}_\xi^{-1}\right)\int_{\mathcal{P}}\widehat{\alpha}_3(\overline{\mathsf{p}})\widehat{x}_d(\mathsf{p})[y(\mathsf{p})-\widehat{y}(\mathsf{p})]\widehat{x}(\mathsf{p})^*\,\mathrm{d}\mu(\mathsf{p})\left(I_r\otimes\widehat{T}_\xi^{-*}e_j\right)\\
&\quad+\sum_{j=1}^{r_\xi}\left(I_r\otimes e_j^{\mathrm{T}}\Pi^*\widehat{S}_\xi^{-1}\right)\int_{\mathcal{P}}\widehat{\alpha}_4(\overline{\mathsf{p}})\widehat{x}_d(\mathsf{p})[y(\mathsf{p})-\widehat{y}(\mathsf{p})]\widehat{x}(\mathsf{p})^*\,\mathrm{d}\mu(\mathsf{p})\left(I_r\otimes\widehat{T}_\xi^{-*}e_j\right)\\
&=-\sum_{j=1}^{r_\xi}\int_{\mathcal{P}}\left(\overline{\xi}-\overline{\pi_j}\right)\left(I_r\otimes e_j^{\mathrm{T}}\widehat{S}_\xi^{-1}\right)\widehat{x}_d(\mathsf{p})[y(\mathsf{p})-\widehat{y}(\mathsf{p})]\widehat{x}(\mathsf{p})^*\left(I_r\otimes\widehat{T}_\xi^{-*}e_j\right)\mathrm{d}\mu(\mathsf{p}).
\end{aligned}$$

Then,

$$\begin{aligned}
&\frac{1}{2}\left(e_k^{\mathrm{T}}\widehat{S}^{-1}\otimes I_{r_\xi}\right)\nabla_{\widehat{A}}\left\|H-\widehat{H}\right\|_{\mathcal{H}_2\otimes\mathcal{L}_2}^2\left(\widehat{T}^{-*}e_k\otimes I_{r_\xi}\right)\\
&=-\frac{1}{4\pi^2}\sum_{j=1}^{r_\xi}\int_0^{2\pi}\int_{-\infty}^{\infty}\left(e^{-\imath\omega_\xi}-\overline{\pi_j}\right)\frac{c_{kj}^*\left(H(\imath\omega,e^{\imath\omega_\xi})-\widehat{H}(\imath\omega,e^{\imath\omega_\xi})\right)b_{kj}}{\left(-\imath\omega-\overline{\lambda_k}\right)^2\left(e^{-\imath\omega_\xi}-\overline{\pi_j}\right)^2}\,\mathrm{d}\omega\,\mathrm{d}\omega_\xi\\
&=-\frac{1}{4\pi^2}\sum_{j=1}^{r_\xi}\int_0^{2\pi}\int_{-\infty}^{\infty}\frac{c_{kj}^*\left(H(\imath\omega,e^{\imath\omega_\xi})-\widehat{H}(\imath\omega,e^{\imath\omega_\xi})\right)b_{kj}}{\left(-\imath\omega-\overline{\lambda_k}\right)^2\left(e^{-\imath\omega_\xi}-\overline{\pi_j}\right)}\,\mathrm{d}\omega\,\mathrm{d}\omega_\xi\\
&=\frac{1}{4\pi^2\imath^2}\sum_{j=1}^{r_\xi}\frac{1}{\overline{\pi_j}}\oint_{\partial\mathbb{D}}\oint_{\imath\mathbb{R}}\frac{c_{kj}^*\left(H(s,\xi)-\widehat{H}(s,\xi)\right)b_{kj}}{\left(s-(-\overline{\lambda_k})\right)^2\left(\xi-\frac{1}{\overline{\pi_j}}\right)}\,\mathrm{d}s\,\mathrm{d}\xi\\
&=-\sum_{j=1}^{r_\xi}\frac{1}{\overline{\pi_j}}c_{kj}^*\left(\frac{\partial H}{\partial s}\left(-\overline{\lambda_k},\frac{1}{\overline{\pi_j}}\right)-\frac{\partial\widehat{H}}{\partial s}\left(-\overline{\lambda_k},\frac{1}{\overline{\pi_j}}\right)\right)b_{kj},
\end{aligned}$$

where we used the Cauchy integral formula twice in the last equality. Setting this equation equal to zero gives the condition (3.14c).

Lastly, computing the gradient with respect to $\widehat{A}_\xi$ using Lemma 3.2 with $\widehat{E} = \widehat{S}^{-*}\widehat{T}^{-1}$ and $\widehat{A} = (\widehat{S}^{-*}\Lambda)\widehat{T}^{-1}$, we obtain (similar to the gradient with respect to $\widehat{A}$)

$$\begin{aligned}
&\frac{1}{2}\nabla_{\widehat{A}_\xi}\left\|H-\widehat{H}\right\|_{\mathcal{H}_2\otimes\mathcal{L}_2}^2\\
&=-\sum_{i=1}^{r}\int_{\mathcal{P}}\left(\overline{s}-\overline{\lambda_i}\right)\left(e_i^{\mathrm{T}}\widehat{S}^{-1}\otimes I_{r_\xi}\right)\widehat{x}_d(\mathsf{p})[y(\mathsf{p})-\widehat{y}(\mathsf{p})]\widehat{x}(\mathsf{p})^*\left(\widehat{T}^{-*}e_i\otimes I_{r_\xi}\right)\mathrm{d}\mu(\mathsf{p})
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{2}\left(I_r\otimes e_\ell^{\mathrm{T}}\widehat{S}_\xi^{-1}\right)\nabla_{\widehat{A}_\xi}\left\|H-\widehat{H}\right\|_{\mathcal{H}_2\otimes\mathcal{L}_2}^2\left(I_r\otimes\widehat{T}_\xi^{-*}e_\ell\right)\\
&=-\frac{1}{4\pi^2}\sum_{i=1}^{r}\int_0^{2\pi}\int_{-\infty}^{\infty}\frac{c_{i\ell}^*\left(H(\imath\omega,e^{\imath\omega_\xi})-\widehat{H}(\imath\omega,e^{\imath\omega_\xi})\right)b_{i\ell}}{\left(-\imath\omega-\overline{\lambda_i}\right)\left(e^{-\imath\omega_\xi}-\overline{\pi_\ell}\right)^2}\,\mathrm{d}\omega\,\mathrm{d}\omega_\xi\\
&=\frac{1}{4\pi^2\imath^2}\sum_{i=1}^{r}\frac{1}{\overline{\pi_\ell}^2}\oint_{\partial\mathbb{D}}\oint_{\imath\mathbb{R}}\frac{\xi c_{i\ell}^*\left(H(s,\xi)-\widehat{H}(s,\xi)\right)b_{i\ell}}{\left(s-(-\overline{\lambda_i})\right)\left(\xi-\frac{1}{\overline{\pi_\ell}}\right)^2}\,\mathrm{d}s\,\mathrm{d}\xi
\end{aligned}$$

$$= -\sum_{i=1}^{r} \frac{1}{\overline{\pi_\ell}^3} c_{i\ell}^* \left(\frac{\partial H}{\partial \xi} \left(-\overline{\lambda_i}, \frac{1}{\overline{\pi_\ell}} \right) - \frac{\partial \widehat{H}}{\partial \xi} \left(-\overline{\lambda_i}, \frac{1}{\overline{\pi_\ell}} \right) \right) b_{i\ell},$$

which gives the final condition (3.14d). ∎

Therefore, using Corollary 2.3 we are not only able recover the optimality conditions (3.9) for single-input single-output systems, but also generalize them to systems with multiple inputs and outputs. One can see parallels to the LTI case considered in Subsection 3.1. The first two conditions (3.14a) and (3.14b) are analogous to the left- and right-tangential (Lagrange) interpolation conditions of (3.5a) and (3.5b), respectively. Furthermore, (3.14c) and (3.14d) resemble the bitangential Hermite conditions in (3.5c). However, since we have a multivariate function H in the $\mathcal{H}_2 \otimes \mathcal{L}_2$ case, ordinary derivatives with respect to s are now replaced by partial derivatives respect to s and ξ. And a bigger distinction is that in the Hermite conditions (3.14c) and (3.14d), interpolated function is not simply a derivative of H. Rather a weighted sum of partial derivatives are interpolated. Optimal interpolation points still result from *mirroring* of the poles. While the mirroring of s-poles is done with respect to the imaginary axis, the mirroring of ξ-poles is with respect to the unit circle. This is not surprising since we used $\mathcal{P} = \imath\mathbb{R} \times \partial\mathbb{D}$. Therefore, the mirroring of s-poles resembles the continuous-time $\mathcal{H}_2$ conditions (3.5) and the mirroring of ξ-poles resembles the discrete-time h_2 conditions (3.6).

4. Linear Time-invariant Systems: Discrete Measure. In Subsection 3.1, we focused on $\mathcal{L}_2$-optimal modeling of LTI systems using a continuous $\mathcal{L}_2$ measure in the frequency domain leading to various systems-theoretic norms such as the $\mathcal{H}_2$ norm. In this section, we change our focus to a discrete measure and investigate the resulting discrete LS problem.

4.1. Necessary Conditions for Discrete LS Problem. Let H be the transfer function of a continuous-time LTI system (e.g., as in (3.2)). Assume we only have access to the samples of H at the sampling frequencies $\{\imath\omega_i\}_{i=1}^N$ where $\omega_i \in \mathbb{R}$. Let $H_i = H(\imath\omega_i) \in \mathbb{C}^{n_o \times n_i}$, for $i = 1, 2, \ldots, N$, denote the corresponding frequency response data. We assume that the sampling frequencies are closed under conjugation; i.e., if $\imath\omega_k$ is a sampling point, then so is $-\imath\omega_k$. In most cases, as in (3.2), H has a real state-space realization, thus leading to $H(-\imath\omega_k) = \overline{H(\imath\omega_k)}$. Therefore, the frequency response data $\{(\imath\omega_i, H_i)\}_{i=1}^N$ is closed under conjugation.

Given the sampling data $\{(\imath\omega_i, H_i)\}_{i=1}^N$, the goal is to find a ROM (3.3) with transfer function $\widehat{H}(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B}$ that minimizes the LS error

$$\mathcal{J}(\widehat{H}) = \sum_{i=1}^{N} \rho_i \left\| H_i - \widehat{H}(\imath\omega_i) \right\|_{\mathrm{F}}^2, \tag{4.1}$$

where $\rho_i > 0$ are the weights (equal for complex conjugate pairs of sampling frequencies). We note that the LS error in (4.1) is a special case of the $\mathcal{L}_2$ error (1.4) with the choices of $\mathcal{P} = \{\imath\omega_i\}_{i=1}^N$, $y(\imath\omega_i) = H_i$, $\widehat{y} = \widehat{H}$, and $\mu = \sum_{i=1}^N \rho_i \delta_{\imath\omega_i}$, where $\delta_{\imath\omega_i}$ is the Dirac measure at $\imath\omega_i$. Thus, the rational LS minimization problem (4.1) directly fits under our $\mathcal{L}_2$-optimal reduced-order modeling framework.

In [38], we have already considered this problem, i.e., the problem of approximating LTI systems from their frequency-domain data using the discrete LS measure. We have devised a gradient-based optimization algorithm to minimize the LS cost (4.1). Our goal here is not algorithmic. Here, using Corollary 2.3, we derive new interpolatory necessary conditions for LS reduced-order modeling, the first such conditions to the best of our knowledge, for the rational LS minimization problem.

Rational LS fitting problem, i.e., minimizing the LS cost using a rational function, is an important and widely studied problem and there are various approaches to tackling it, see, e.g., [14, 18, 19, 33, 35, 39], and the references therein. What we show here is that regardless of the underlying numerical algorithm, a solution of the nonlinear rational LS minimization problem is interpolatory and satisfies specific bitangential Hermite interpolation conditions.

THEOREM 4.1. *Given the sampling data* $\{(\imath\omega_i, H_i)\}_{i=1}^N$, *let the StROM* $\widehat{H}(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B}$ *having the pole-residue form* $\widehat{H}(s) = \sum_{j=1}^r \frac{c_j b_j^*}{s-\lambda_j}$ *with pairwise distinct poles be a local minimum of* $\mathcal{J}$ (4.1). *Then*

$$\sum_{i=1}^N \rho_i \frac{H_i b_k}{-\imath\omega_i - \overline{\lambda_k}} = \sum_{i=1}^N \rho_i \frac{\widehat{H}(\imath\omega_i) b_k}{-\imath\omega_i - \overline{\lambda_k}}, \tag{4.2a}$$

$$\sum_{i=1}^N \rho_i \frac{c_k^* H_i}{-\imath\omega_i - \overline{\lambda_k}} = \sum_{i=1}^N \rho_i \frac{c_k^* \widehat{H}(\imath\omega_i)}{-\imath\omega_i - \overline{\lambda_k}}, \tag{4.2b}$$

$$\sum_{i=1}^N \rho_i \frac{c_k^* H_i b_k}{\left(-\imath\omega_i - \overline{\lambda_k}\right)^2} = \sum_{i=1}^N \rho_i \frac{c_k^* \widehat{H}(\imath\omega_i) b_k}{\left(-\imath\omega_i - \overline{\lambda_k}\right)^2}, \tag{4.2c}$$

for $k = 1, 2, \ldots, r$.

Proof. Let $\widehat{H}(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B}$ and $\widehat{T}, \widehat{S} \in \mathbb{C}^{r\times r}$ be invertible matrices, with $c_j = \widehat{C}\widehat{S}^{-1}e_j$ and $b_j = \widehat{B}^{\mathrm{T}}\widehat{T}^{-*}e_j$, as in Subsection 3.1, yielding the pole-residue form of $\widehat{H}$. Then, using $\mathcal{P} = \{\imath\omega_i\}_{i=1}^N$, $y(\imath\omega_i) = H_i$, $\widehat{y} = \widehat{H}$, $\widehat{x} = \widehat{X}$, and $\mu = \sum_{i=1}^N \rho_i \delta_{\imath\omega_i}$ in the left-hand side term in the $\mathcal{L}_2$-optimality condition (2.3a) gives

$$\int_{\mathcal{P}} y(\mathsf{p})\widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p})\, \widehat{T}^{-*}e_k = \sum_{i=1}^N \rho_i H_i \widehat{B}^{\mathrm{T}} \left(\imath\omega_i\widehat{E} - \widehat{A}\right)^{-*} \widehat{T}^{-*}e_k = \sum_{i=1}^N \rho_i \frac{H_i b_k}{-\imath\omega_i - \overline{\lambda_k}}.$$

Similarly, using the left-hand side term in (2.3b), we obtain

$$e_k^{\mathrm{T}}\widehat{S}^{-1} \int_{\mathcal{P}} \widehat{x}_d(\mathsf{p}) y(\mathsf{p}) \,\mathrm{d}\mu(\mathsf{p}) = \sum_{i=1}^N \rho_i e_k^{\mathrm{T}} \widehat{S}^{-1} \left(\imath\omega_i\widehat{E} - \widehat{A}\right)^{-*} \widehat{C}^{\mathrm{T}} H_i = \sum_{i=1}^N \rho_i \frac{c_k^* H_i}{-\imath\omega_i - \overline{\lambda_k}}.$$

Lastly, the left-hand side term in (2.3c) corresponding to $\widehat{A}$ shows

$$\begin{aligned} &-e_k^{\mathrm{T}}\widehat{S}^{-1} \int_{\mathcal{P}} \widehat{\alpha}_2(\mathsf{p}) \widehat{x}_d(\mathsf{p}) y(\mathsf{p}) \widehat{x}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p})\, \widehat{T}^{-*}e_k \\ &= \sum_{i=1}^N \rho_i e_k^{\mathrm{T}} \widehat{S}^{-1} \left(\imath\omega_i\widehat{E} - \widehat{A}\right)^{-*} \widehat{C}^{\mathrm{T}} H_i \widehat{B}^{\mathrm{T}} \left(\imath\omega_i\widehat{E} - \widehat{A}\right)^{-*} \widehat{T}^{-*}e_k = \sum_{i=1}^N \rho_i \frac{c_k^* H_i b_k}{\left(-\imath\omega_i - \overline{\lambda_k}\right)^2}. \end{aligned}$$

Analogous calculations for the right-hand sides directly gives the conditions (4.2). ∎

We can rewrite the conditions (4.2) to give a more immediate interpolatory interpretation.

Corollary 4.2. *Given the sampling data $\{(\imath\omega_i, H_i)\}_{i=1}^N$, let $\widehat{H}(s) = \sum_{j=1}^r \frac{c_j b_j^*}{s-\lambda_j}$ have pairwise distinct poles and be a local minimizer of the LS error (4.1). Furthermore, define the transfer functions*

$$G(s) = \sum_{i=1}^N \rho_i \frac{H_i}{s - \imath\omega_i} \quad \text{and} \quad \widehat{G}(s) = \sum_{i=1}^N \rho_i \frac{\widehat{H}(\imath\omega_i)}{s - \imath\omega_i}. \tag{4.3}$$

Then

$$G\bigl(-\overline{\lambda_k}\bigr) b_k = \widehat{G}\bigl(-\overline{\lambda_k}\bigr) b_k, \tag{4.4a}$$

$$c_k^* G\bigl(-\overline{\lambda_k}\bigr) = c_k^* \widehat{G}\bigl(-\overline{\lambda_k}\bigr), \tag{4.4b}$$

$$c_k^* G'\bigl(-\overline{\lambda_k}\bigr) b_k = c_k^* \widehat{G}'\bigl(-\overline{\lambda_k}\bigr) b_k, \tag{4.4c}$$

for $k = 1, 2, \ldots, r$.

Proof. The conditions (4.4a) and (4.4b) follow from (4.2a) and (4.2b), respectively, based on the definitions of G and $\widehat{G}$ in (4.3). The condition (4.2c) yields (4.4c) after observing $G'(s) = -\sum_{i=1}^N \rho_i \frac{H_i}{(s-\imath\omega_i)^2}$ and $\widehat{G}'(s) = -\sum_{i=1}^N \rho_i \frac{\widehat{H}(\imath\omega_i)}{(s-\imath\omega_i)^2}$. ∎

The conditions (4.4) illustrate that bitangential Hermite interpolation is the necessary condition for the discrete $\mathcal{L}_2$ cost function as well; the interpolatory $\mathcal{L}_2$-optimal modeling framework equally applies. The optimal approximant is still a bitangential Hermite interpolant, but what is interpolated is different. Here, two order-N rational functions G and $\widehat{G}$ interpolate each other where G depends on the evaluation of H and $\widehat{G}$ on the evaluations of $\widehat{H}$. Yet, the interpolation points and directions are still determined by the poles and residues of the optimal rational approximant $\widehat{H}$. Mirror images of the reduced-order poles still appear as the interpolation points.

Remark 4.3. Alternatively, we can view the optimality conditions (4.2) as discretized $\mathcal{H}_2$-optimality conditions (3.5). In particular, note that the interpolatory conditions (3.5) are equivalent to (using the Cauchy integral formula)

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{H(\imath\omega) b_k}{-\imath\omega - \overline{\lambda_k}} \,\mathrm{d}\omega &= \int_{-\infty}^{\infty} \frac{\widehat{H}(\imath\omega) b_k}{-\imath\omega - \overline{\lambda_k}} \,\mathrm{d}\omega, \\
\int_{-\infty}^{\infty} \frac{c_k^* H(\imath\omega)}{-\imath\omega - \overline{\lambda_k}} \,\mathrm{d}\omega &= \int_{-\infty}^{\infty} \frac{c_k^* \widehat{H}(\imath\omega)}{-\imath\omega - \overline{\lambda_k}} \,\mathrm{d}\omega, \\
\int_{-\infty}^{\infty} \frac{c_i^* H(\imath\omega) b_k}{\bigl(-\imath\omega - \overline{\lambda_k}\bigr)^2} \,\mathrm{d}\omega &= \int_{-\infty}^{\infty} \frac{c_i^* \widehat{H}(\imath\omega) b_k}{\bigl(-\imath\omega - \overline{\lambda_k}\bigr)^2} \,\mathrm{d}\omega,
\end{aligned}$$

for $k = 1, 2, \ldots, r$. Approximating these integrals (representing the $\mathcal{H}_2$-optimality conditions) using a numerical quadrature with nodes ω_i and weights ρ_i leads to the new optimality conditions (4.2) for the cost function (4.1).

4.2. Numerical Example. To demonstrate the new interpolatory conditions from Corollary 4.2 for the rational discrete LS minimization problem, we use the Penzl's FOM model from the Niconet benchmark collection [17]. The model is an LTI system of order $n = 1006$ with $n_{\mathrm{i}} = 1$ input and $n_{\mathrm{o}} = 1$ output. We chose $r = 2$ as the reduced order to make the illustration clear. Clearly a high-fidelity approximation requires a higher order; but our goal is here to illustrate the theory of Corollary 4.2. For the data, we take 50 logarithmically-spaced frequencies between 10^0 and 10^4 on

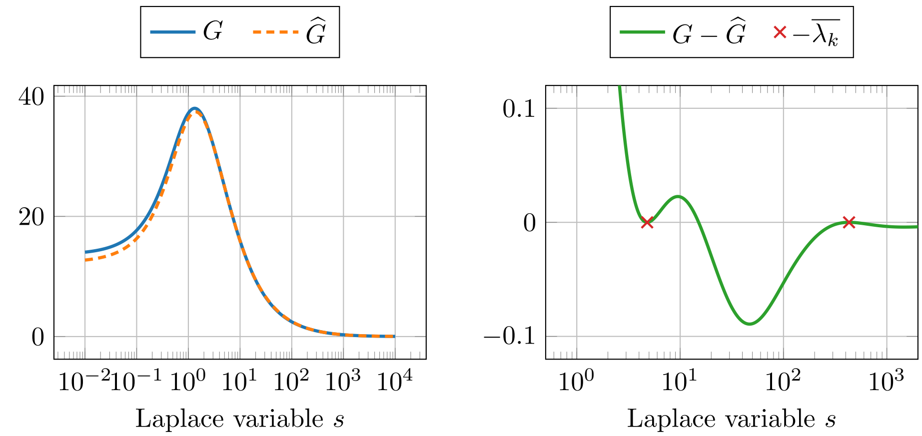

FIG. 1. *Left: Transfer functions corresponding to the FOM and ROM data for the Penzl example ($n = 1006$, $r = 2$). Right: The difference between the modified outputs and the mirrored poles of the ROM.*

the imaginary axis, including the endpoints. With the inclusion of complex conjugate points, we obtain $N = 100$ data points in the LS problem (4.1). We use the gradient-based optimization algorithm $\mathcal{L}_2$-Opt-PSF in [38] (initialized with the ROM from iterative rational Krylov algorithm [29]) and obtain a ROM of order $r = 2$ with poles at $\lambda_1 \approx -4.7984$ and $\lambda_2 \approx -431.00$.

Figure 1 shows the transfer functions G and $\widehat{G}$, defined in Corollary 4.2, when evaluated over the positive real axis. Even though in the left plot, we see overlap around $-\overline{\lambda_1}$ and $-\overline{\lambda_2}$, it is not clear whether Hermite interpolation is achieved. (Since H is a single-input/single-output LTI system, tangential interpolation boils downs to scalar interpolation, i.e., (4.4a) and (4.4b) coincide.) To make the illustration clearer, the right plot shows the difference between G and $\widehat{G}$, and here we see (from the shape and curvature of the error plot) that $\widehat{G}$ is indeed a Hermite interpolant of G at $-\overline{\lambda_1}$ and $-\overline{\lambda_2}$ as the theory predicts. We note that there is another interpolation point around $s \approx 15$. However the derivative is not matched at this point; thus no Hermite interpolation occurs at this point.

The Python code used to compute the presented results can be obtained from [37].

5. Stationary Parametric Problems. In Sections 3 and 4, we have focused on approximating LTI systems. Now we turn our attention to *stationary* models and prove that bitangential Hermite interpolation forms the necessary conditions for $\mathcal{L}_2$-optimal MOR in this case as well. Even though what is interpolated and the optimal interpolation points differ from the LTI system case, the optimality still requires bitangential Hermite interpolation.

5.1. Necessary Conditions for Stationary Models. Let us consider the stationary FOM (resulting from, e.g., discretization of a stationary parametric partial differential equation)

$$(A_1 + \mathsf{p}A_2)x(\mathsf{p}) = B, \tag{5.1a}$$
$$y(\mathsf{p}) = Cx(\mathsf{p}), \tag{5.1b}$$

where $x(\mathsf{p}) \in \mathbb{R}^n$ is the state, $y(\mathsf{p}) \in \mathbb{R}^{n_{\mathrm{o}} \times n_{\mathrm{i}}}$ is the output, $A_1, A_2 \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times n_{\mathrm{i}}}$, $C \in \mathbb{R}^{n_{\mathrm{o}} \times n}$, and $\mathcal{P} = [a, b] \subset \mathbb{R}$ for some $a < b$. Next, let

$$\left(\widehat{A}_1 + \mathsf{p}\widehat{A}_2\right)\widehat{x}(\mathsf{p}) = \widehat{B}, \tag{5.2a}$$

$$\widehat{y}(\mathsf{p}) = \widehat{C}\widehat{x}(\mathsf{p}), \tag{5.2b}$$

be a StROM of order r; in particular, $\widehat{x}(\mathsf{p}) \in \mathbb{R}^r$ is the reduced state, $\widehat{y}(\mathsf{p}) \in \mathbb{R}^{n_{\mathrm{o}} \times n_{\mathrm{i}}}$ is the approximate output, $\widehat{A}_1, \widehat{A}_2 \in \mathbb{R}^{r \times r}$, $\widehat{B} \in \mathbb{R}^{r \times n_{\mathrm{i}}}$, and $\widehat{C} \in \mathbb{R}^{n_{\mathrm{o}} \times r}$. Our goal in this section is to show that the $\mathcal{L}_2$-optimal StROM (5.2) satisfies special interpolation conditions. More specially, we show that a modified reduced-order output $\widehat{y}$ interpolates a modified full-order output y at special parameter points λ_k.

Note that both y and $\widehat{y}$ are rational functions of the parameter p. Motivated by the interpolation-based optimality conditions from Subsection 3.1 and those for frequency-limited $\mathcal{H}_2$-optimal MOR [43], we develop the interpolatory conditions for the $\mathcal{L}_2$-optimal approximation of FOMs of the form (5.1) using the pole-residue forms of y and $\widehat{y}$.

The pole-residue form of $\widehat{y}$ (5.2) can be obtained similarly as done for LTI system. Let $\widehat{A}_2$ be invertible and $\widehat{A}_2^{-1}\widehat{A}_1$ have r distinct real eigenvalues. Then, proceeding as in Subsection 3.1, let $\widehat{T}, \widehat{S} \in \mathbb{R}^{r \times r}$ be invertible matrices such that $\widehat{S}^{\mathrm{T}}\widehat{A}_1\widehat{T} = -\Lambda = \operatorname{diag}(-\lambda_1, -\lambda_2, \ldots, -\lambda_r)$ and $\widehat{S}^{\mathrm{T}}\widehat{A}_2\widehat{T} = I$. Then we obtain that $\widehat{y}(\mathsf{p}) = \sum_{j=1}^{r} \frac{c_j b_j^{\mathrm{T}}}{\mathsf{p} - \lambda_j}$, where $c_j = \widehat{C}\widehat{T}e_j \in \mathbb{R}^{n_{\mathrm{o}}}$, and $b_j = \widehat{B}^{\mathrm{T}}\widehat{S}^{\mathrm{T}}e_j \in \mathbb{R}^{n_{\mathrm{i}}}$ for $j = 1, 2, \ldots, r$.

Inspired by the structure of a FOM in the numerical example in Subsection 5.2, we allow a slightly more general pole-residue form for y, namely $y(\mathsf{p}) = \Phi_0 + \sum_{i=1}^{n} \frac{\Phi_i}{\mathsf{p} - \nu_i}$, where Φ_0 is a constant term, ν_i are the poles and Φ_i the corresponding residues for $i = 1, 2, \ldots, n$. The constant term Φ_0 results from allowing A_2 to be a singular matrix (as in the case of the numerical example in Subsection 5.2). The details of the derivation of this pole-residue form are given in Appendix B.

THEOREM 5.1. *Let $\widehat{y}(\mathsf{p}) = \sum_{j=1}^{r} \frac{c_j b_j^{\mathrm{T}}}{\mathsf{p} - \lambda_j}$ be the output of the StROM (5.2) with pairwise distinct $\lambda_j \in \mathbb{R} \setminus [a, b]$. Furthermore, let $y(\mathsf{p}) = \Phi_0 + \sum_{i=1}^{n} \frac{\Phi_i}{\mathsf{p} - \nu_i}$ be the output of the FOM (5.1), also with pairwise distinct $\nu_i \in \mathbb{R} \setminus [a, b]$. For any $\sigma \in \mathbb{R} \setminus \{a, b\}$, define the function $f_\sigma \colon \mathbb{R} \setminus \{a, b\} \to \mathbb{R}$ as*

$$f_\sigma(\mathsf{p}) = \begin{cases} \left(\ln\left|\frac{\mathsf{p} - b}{\mathsf{p} - a}\right| - \ln\left|\frac{\sigma - b}{\sigma - a}\right|\right)\frac{1}{\mathsf{p} - \sigma}, & \text{if } \mathsf{p} \neq \sigma, \\ \frac{b - a}{(\sigma - a)(\sigma - b)}, & \text{if } \mathsf{p} = \sigma. \end{cases}$$

Furthermore, define the modified output functions $Y, \widehat{Y} \colon \mathbb{R} \setminus \{a, b\} \to \mathbb{R}$ as

$$Y(\mathsf{p}) = \ln\left|\frac{\mathsf{p} - b}{\mathsf{p} - a}\right|\Phi_0 + \sum_{i=1}^{n} f_{\nu_i}(\mathsf{p})\Phi_i \quad \text{and} \quad \widehat{Y}(\mathsf{p}) = \sum_{j=1}^{r} f_{\lambda_j}(\mathsf{p})c_j b_j^{\mathrm{T}}. \tag{5.3}$$

Let $\widehat{y}$ be an $\mathcal{L}_2$-optimal structured approximation of y. Then,

$$Y(\lambda_k)b_k = \widehat{Y}(\lambda_k)b_k, \tag{5.4a}$$

$$c_k^{\mathrm{T}}Y(\lambda_k) = c_k^{\mathrm{T}}\widehat{Y}(\lambda_k), \tag{5.4b}$$

$$c_k^{\mathrm{T}}Y'(\lambda_k)b_k = c_k^{\mathrm{T}}\widehat{Y}'(\lambda_k)b_k, \tag{5.4c}$$

for $k = 1, 2, \ldots, r$.

Remark 5.2. Some remarks are in order before we prove Theorem 5.1. The optimality conditions (5.4) show that bitangential Hermite interpolation forms the necessary conditions for $\mathcal{L}_2$-optimal approximation over an interval $[a, b]$; thus extending the theory from LTI systems to stationary problems. However, what is to be interpolated is no longer the original function y itself, instead the modified output Y in (5.3) needs to be interpolated. Another major difference here is that the interpolation occurs at the reduced system poles, as opposed to at *the mirror images of the poles* in the LTI system case. (Both modified outputs Y and $\widehat{Y}$ are well defined at the reduced poles as shown in the proof below.)

Proof. First note that f_σ is continuously differentiable and its derivative is

$$f'_\sigma(\mathsf{p}) = \begin{cases} \left(\frac{(b-a)(\mathsf{p}-\sigma)}{(\mathsf{p}-a)(\mathsf{p}-b)} - \ln\left|\frac{\mathsf{p}-b}{\mathsf{p}-a}\right| + \ln\left|\frac{\sigma-b}{\sigma-a}\right|\right)\frac{1}{(\mathsf{p}-\sigma)^2}, & \text{if } \mathsf{p} \neq \sigma, \\ \frac{(b-a)(a+b-2\sigma)}{2(\sigma-a)^2(\sigma-b)^2}, & \text{if } \mathsf{p} = \sigma. \end{cases}$$

Since f_σ is continuously differentiable, so are Y and $\widehat{Y}$.

Next, compare the simple StROM (5.2) to the general case (1.3) to observe that we have $q_{\widehat{\mathcal{A}}} = 2$ with $\widehat{\alpha}_1(\mathsf{p}) = 1$ and $\widehat{\alpha}_2(\mathsf{p}) = \mathsf{p}$, $q_{\widehat{\mathcal{B}}} = 1$ with $\widehat{\beta}_1(\mathsf{p}) = 1$, and $q_{\widehat{\mathcal{C}}} = 1$ with $\widehat{\gamma}_1(\mathsf{p}) = 1$. It follows from the left-hand side of (2.3a) in Corollary 2.3 that

$$\int_a^b y(\mathsf{p})\widehat{x}(\mathsf{p})^{\mathrm{T}} \,\mathrm{d}\mathsf{p} = \int_a^b \left(\Phi_0 + \sum_{i=1}^n \frac{\Phi_i}{\mathsf{p} - \nu_i}\right)\widehat{B}^{\mathrm{T}}\left(\widehat{A}_1 + \mathsf{p}\widehat{A}_2\right)^{-\mathrm{T}} \mathrm{d}\mathsf{p}.$$

Then, for $k = 1, 2, \ldots, r$, we have

$$\begin{aligned}\int_a^b y(\mathsf{p})\widehat{x}(\mathsf{p})^{\mathrm{T}} \,\mathrm{d}\mathsf{p}\, \widehat{T}^{-\mathrm{T}} e_k &= \int_a^b \left(\Phi_0 + \sum_{i=1}^n \frac{\Phi_i}{\mathsf{p} - \nu_i}\right)\widehat{B}^{\mathrm{T}}\widehat{S}(-\Lambda + \mathsf{p}I)^{-\mathrm{T}} e_k \,\mathrm{d}\mathsf{p} \\ &= \int_a^b \frac{\Phi_0 b_k}{\mathsf{p} - \lambda_k}\,\mathrm{d}\mathsf{p} + \sum_{i=1}^n \int_a^b \frac{\Phi_i b_k}{(\mathsf{p} - \nu_i)(\mathsf{p} - \lambda_k)}\,\mathrm{d}\mathsf{p}.\end{aligned}$$

If $\nu_i = \lambda_k$, then

$$\int_a^b \frac{\Phi_i b_k}{(\mathsf{p} - \nu_i)(\mathsf{p} - \lambda_k)}\,\mathrm{d}\mathsf{p} = \int_a^b \frac{\Phi_i b_k}{(\mathsf{p} - \nu_i)^2}\,\mathrm{d}\mathsf{p} = \frac{(b-a)\Phi_i b_k}{(\nu_i - a)(\nu_i - b)} = f_{\nu_i}(\lambda_k)\Phi_i b_k.$$

Otherwise,

$$\begin{aligned}&\int_a^b \frac{\Phi_i b_k}{(\mathsf{p} - \nu_i)(\mathsf{p} - \lambda_k)}\,\mathrm{d}\mathsf{p} = \frac{1}{\lambda_k - \nu_i}\int_a^b \left(\frac{1}{\mathsf{p} - \lambda_k} - \frac{1}{\mathsf{p} - \nu_i}\right)\mathrm{d}\mathsf{p}\, \Phi_i b_k \\ &= \left(\ln\left|\frac{\lambda_k - b}{\lambda_k - a}\right| - \ln\left|\frac{\nu_i - b}{\nu_i - a}\right|\right)\frac{\Phi_i}{\lambda_k - \nu_i} b_k = f_{\nu_i}(\lambda_k)\Phi_i b_k.\end{aligned}$$

Therefore, we obtain

$$\int_a^b y(\mathsf{p})\widehat{x}(\mathsf{p})^{\mathrm{T}} \,\mathrm{d}\mathsf{p}\, \widehat{T}^{-\mathrm{T}} e_k = \ln\left|\frac{\lambda_k - b}{\lambda_k - a}\right|\Phi_0 b_k + \sum_{i=1}^n f_{\nu_i}(\lambda_k)\Phi_i b_k = Y(\lambda_k) b_k.$$

From the right-hand side of (2.3a), we similarly find

$$\begin{aligned}&\int_a^b \widehat{y}(\mathsf{p})\widehat{x}(\mathsf{p})^{\mathrm{T}} \,\mathrm{d}\mathsf{p}\, \widehat{T}^{-\mathrm{T}} e_k \\ &= \sum_{\substack{j=1\\ j\neq k}}^{r} \left(\ln\left|\frac{\lambda_k - b}{\lambda_k - a}\right| - \ln\left|\frac{\lambda_j - b}{\lambda_j - a}\right|\right) \frac{c_j b_j^{\mathrm{T}}}{\lambda_k - \lambda_j} b_k + \frac{(b-a)c_k b_k^{\mathrm{T}}}{(\lambda_k - a)(\lambda_k - b)} b_k = \widehat{Y}(\lambda_k) b_k.\end{aligned}$$

Therefore, we get the right-tangential interpolation conditions (5.4a).

Using (2.3b), we similarly find

$$e_k^{\mathrm{T}} \widehat{S}^{-1} \int_a^b \widehat{x}_d(\mathsf{p}) y(\mathsf{p}) \,\mathrm{d}\mathsf{p} = c_k^{\mathrm{T}} Y(\lambda_k) \quad \text{and} \quad e_k^{\mathrm{T}} \widehat{S}^{-1} \int_a^b \widehat{x}_d(\mathsf{p}) \widehat{y}(\mathsf{p}) \,\mathrm{d}\mathsf{p} = c_k^{\mathrm{T}} \widehat{Y}(\lambda_k),$$

which yield the left-tangential interpolation conditions (5.4b).

Using (2.3c) corresponding to $\widehat{A}_1$, we find

$$\begin{aligned}e_k^{\mathrm{T}} \widehat{S}^{-1} \int_a^b \widehat{x}_d(\mathsf{p}) y(\mathsf{p}) \widehat{x}(\mathsf{p})^{\mathrm{T}} \,\mathrm{d}\mathsf{p}\, \widehat{T}^{-\mathrm{T}} e_k &= \int_a^b \frac{c_k^{\mathrm{T}} \Phi_0 b_k}{(\mathsf{p} - \lambda_k)^2} \,\mathrm{d}\mathsf{p} \\ &\quad + \sum_{i=1}^{n} \int_a^b \frac{c_k^{\mathrm{T}} \Phi_i b_k}{(\mathsf{p} - \nu_i)(\mathsf{p} - \lambda_k)^2} \,\mathrm{d}\mathsf{p}.\end{aligned}$$

If $\nu_i = \lambda_k$, then

$$\begin{aligned}\int_a^b \frac{c_k^{\mathrm{T}} \Phi_i b_k}{(\mathsf{p} - \nu_i)(\mathsf{p} - \lambda_k)^2} \,\mathrm{d}\mathsf{p} = \int_a^b \frac{c_k^{\mathrm{T}} \Phi_i b_k}{(\mathsf{p} - \nu_i)^3} \,\mathrm{d}\mathsf{p} &= \frac{(b-a)(a+b-2\nu_i) c_k^{\mathrm{T}} \Phi_i b_k}{2(\nu_i - a)^2 (\nu_i - b)^2} \\ &= f_{\nu_i}'(\lambda_k) c_k^{\mathrm{T}} \Phi_i b_k.\end{aligned}$$

Otherwise,

$$\begin{aligned}\int_a^b \frac{c_k^{\mathrm{T}} \Phi_i b_k}{(\mathsf{p} - \nu_i)(\mathsf{p} - \lambda_k)^2} \,\mathrm{d}\mathsf{p} &= \frac{1}{(\lambda_k - \nu_i)^2} \int_a^b \left(\frac{1}{\mathsf{p} - \nu_i} - \frac{1}{\mathsf{p} - \lambda_k} + \frac{\lambda_k - \nu_i}{(\mathsf{p} - \lambda_k)^2}\right) \mathrm{d}\mathsf{p}\, c_k^{\mathrm{T}} \Phi_i b_k \\ &= \left(\ln\left|\frac{\nu_i - b}{\nu_i - a}\right| - \ln\left|\frac{\lambda_k - b}{\lambda_k - a}\right| + \frac{(b-a)(\lambda_k - \nu_i)}{(\lambda_k - a)(\lambda_k - b)}\right) \frac{c_k^{\mathrm{T}} \Phi_i b_k}{(\lambda_k - \nu_i)^2} \\ &= f_{\nu_i}'(\lambda_k) c_k^{\mathrm{T}} \Phi_i b_k.\end{aligned}$$

Therefore,

$$\begin{aligned}&e_k^{\mathrm{T}} \widehat{S}^{-1} \int_a^b \widehat{x}_d(\mathsf{p}) y(\mathsf{p}) \widehat{x}(\mathsf{p})^{\mathrm{T}} \,\mathrm{d}\mathsf{p}\, \widehat{T}^{-\mathrm{T}} e_k \\ &= \left(\frac{1}{\lambda_k - b} - \frac{1}{\lambda_k - a}\right) c_k^{\mathrm{T}} \Phi_0 b_k + \sum_{i=1}^{n} f_{\nu_i}'(\lambda_k) c_k^{\mathrm{T}} \Phi_i b_k = c_k^{\mathrm{T}} Y'(\lambda_k) b_k.\end{aligned}$$

Similarly, we find that

$$e_k^{\mathrm{T}} \widehat{S}^{-1} \int_a^b \widehat{x}_d(\mathsf{p}) \widehat{y}(\mathsf{p}) \widehat{x}(\mathsf{p})^{\mathrm{T}} \,\mathrm{d}\mathsf{p}\, \widehat{T}^{-\mathrm{T}} e_k = c_k^{\mathrm{T}} \widehat{Y}'(\lambda_k) b_k,$$

which proves the final bitangential Hermite interpolation conditions (5.4c). □

Theorem 5.1 has shown that bitangential Hermite interpolation, which is at the core of $\mathcal{H}_2$-optimal approximation of LTI systems, also naturally appears in the approximation of parametric stationary problems. In the LTI system setting, these interpolatory conditions have been at the core of many algorithmic developments and extended to various different settings; see, e.g., [1, 2, 4, 5, 7, 11, 12, 15, 25, 26, 29–32, 41], and the references therein. Similar potential extensions and algorithmic developments for the stationary parametric case will be a topic of future research.

5.2. Numerical Example. As we did for Corollary 4.2 in Subsection 4.2, we demonstrate the new interpolatory results of Theorem 5.1 using a numerical example. Following [38], we consider the Poisson equation over the unit square $\Omega = (0,1)^2$ with homogeneous Dirichlet boundary conditions:

$$\begin{aligned} -\nabla \cdot (d(z, \mathsf{p}) \nabla x(z, \mathsf{p})) &= 1, && z \in \Omega, \\ x(z, \mathsf{p}) &= 0, && z \in \partial\Omega, \end{aligned}$$

where $d(z, \mathsf{p}) = z_1 + \mathsf{p}(1 - z_1)$ and $\mathcal{P} = [0.1, 10]$. After a finite element discretization, we obtain the FOM of the form (5.1) with $n = 1089$ and $n_{\mathrm{i}} = 1$, and the choice of $C = B^{\mathrm{T}}$ (and $n_{\mathrm{o}} = 1$). We use the gradient-based optimization algorithm $\mathcal{L}_2$-Opt-PSF in [38] (initialized with the StROM resulting from applying a reduced-basis approach [10]) and obtain a StROM of order $r = 2$ with poles at $\lambda_1 \approx -3.2777$ and $\lambda_2 \approx -0.30509$.

To numerically verify the interpolatory conditions, we need the pole-residue forms of y and $\widehat{y}$ as stated in Theorem 5.1. Note that these computations are not needed by the $\mathcal{L}_2$-Opt-PSF algorithm and are done here just to illustrate the interpolation theory. Since $\widehat{A}_2$ is invertible, the pole-residue form of $\widehat{y}$ directly follows as explained above, right before Theorem 5.1. The situation is more involved for y since A_2 in (5.1) is rank-deficient; it has numerical rank $n_2 = 961$. We use the procedure explained in Appendix B to compute the pole-residue form of y. With the pole-residue forms of y and $\widehat{y}$, we can now numerically evaluate the modified outputs Y and $\widehat{Y}$ defined in (5.3) and illustrate the interpolation result.

The left plot in Figure 2 shows that Y and $\widehat{Y}$ almost overlap, making it unclear whether Hermite interpolation is achieved. To illustrate the results better, the right plot in Figure 2 depicts the difference $Y - \widehat{Y}$ around the location of the poles of the StROM, clearly demonstrating (based on the shape and curvature of the error plot) that the StROM satisfies the Hermite interpolation conditions (5.4) of Theorem 5.1.

It is interesting to note that the interpolation points λ_1 and λ_2 are outside the parameter space $\mathcal{P} = [a, b]$. This is in agreement with the $\mathcal{H}_2$-optimality conditions (3.5), where interpolation is enforced away from the imaginary axis, in particular, in the open right half-plane. In contrast to the reduced-basis methods that choose the greedy sampling points in the parameter interval of interest, $\mathcal{L}_2$-optimal reduced-order modeling necessitates interpolation of a modified output Y outside the domain of interest.

The Python code used to compute the presented results can be obtained from [37].

6. Conclusion. We developed a unifying framework for $\mathcal{L}_2$-optimal interpolatory reduced-order modeling. In particular, we showed that the $\mathcal{L}_2$-optimality conditions resulting from this framework naturally cover known interpolatory conditions for $\mathcal{H}_2$-optimal MOR of LTI systems, both for continuous-time and discrete-time cases. Furthermore, they lead to interpolatory conditions for $\mathcal{H}_2 \otimes \mathcal{L}_2$-optimal MOR of multi-input/multi-output parametric LTI systems. We also derived novel bitangential Hermite interpolation conditions for rational LS problems and for a class of

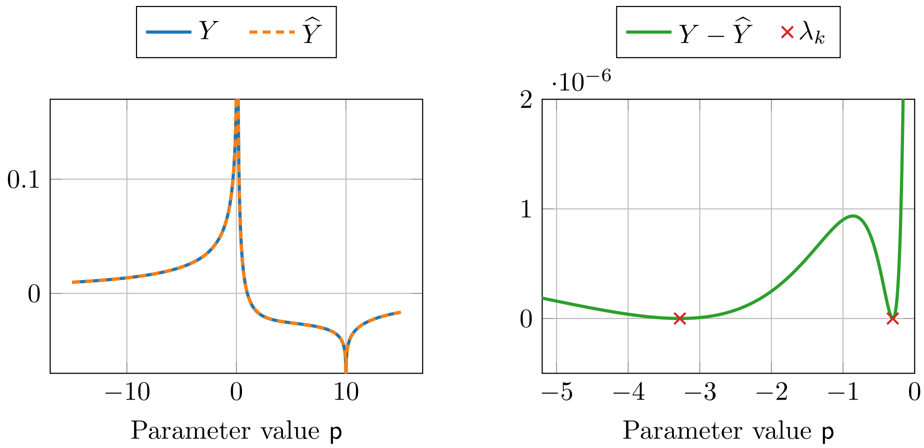

Fig. 2. *Left: Modified outputs of the FOM and ROM for the Poisson example ($n = 1089$, $r = 2$). Right: Difference between the modified outputs and the poles of the ROM.*

stationary parametric problems. These results illustrate that bitangential Hermite interpolation appears as the main tool for $\mathcal{L}_2$-optimality across different domains.

Interpolatory conditions for $\mathcal{H}_2$-optimal MOR lead to various numerical algorithms for MOR, such as the iterative rational Krylov algorithm. Algorithmic implications of the new interpolatory conditions for the discrete LS measure and the stationary problems are interesting avenues to investigate.

Appendix A. Proof of Lemma 3.2. We have that for any $\Delta A \in \mathbb{R}^{n \times n}$,

$$\begin{aligned} G(A + \Delta A) &= F((A + \Delta A) \otimes B) = F(A \otimes B + \Delta A \otimes B) \\ &= G(A) + \langle \nabla F(A \otimes B), \Delta A \otimes B \rangle_{\mathrm{F}} + o(\|\Delta A\|_{\mathrm{F}}). \end{aligned}$$

Let $P_A \colon \mathbb{R}^{nm \times nm} \to \mathbb{R}^{n \times n}$ be defined as

$$P_A(Y) = \operatorname*{arg\,min}_{A_1 \in \mathbb{R}^{n \times n}} \|Y - A_1 \otimes B\|_{\mathrm{F}}^2.$$

Using the orthogonality property of the least squares approximation, it follows that $\langle Y - P_A(Y) \otimes B, A_1 \otimes B \rangle_{\mathrm{F}} = 0$ for all $A_1 \in \mathbb{R}^{n \times n}$. Then,

$$\begin{aligned} &\langle \nabla F(A \otimes B), \Delta A \otimes B \rangle_{\mathrm{F}} = \langle P_A(\nabla F(A \otimes B)) \otimes B, \Delta A \otimes B \rangle_{\mathrm{F}} \\ &= \operatorname{tr}\big((P_A(\nabla F(A \otimes B)) \otimes B)^{\mathrm{T}} (\Delta A \otimes B)\big) = \operatorname{tr}\big(P_A(\nabla F(A \otimes B))^{\mathrm{T}} \Delta A \otimes B^{\mathrm{T}} B\big) \\ &= \operatorname{tr}\big(P_A(\nabla F(A \otimes B))^{\mathrm{T}} \Delta A\big) \operatorname{tr}\big(B^{\mathrm{T}} B\big) = \langle P_A(\nabla F(A \otimes B)), \Delta A \rangle_{\mathrm{F}} \|B\|_{\mathrm{F}}^2 \\ &= \big\langle \|B\|_{\mathrm{F}}^2 P_A(\nabla F(A \otimes B)), \Delta A \big\rangle_{\mathrm{F}}. \end{aligned}$$

Therefore, it follows that G is differentiable at A and

$$\nabla G(A) = \|B\|_{\mathrm{F}}^2 P_A(\nabla F(A \otimes B)). \tag{A.1}$$

To find $P_A(Y)$, define $V_A = \{A_1 \otimes B : A_1 \in \mathbb{R}^{n \times n}\}$. Note that V_A is a subspace of $\mathbb{R}^{nm \times nm}$. We see that $\{\frac{1}{\|B\|_{\mathrm{F}}} e_k e_\ell^{\mathrm{T}} \otimes B : k, \ell \in \{1, 2, \ldots, n\}\}$ is an orthonormal basis for V_A. Therefore

$$P_A(Y) \otimes B = \sum_{k,\ell=1}^{n} \left\langle Y, \frac{1}{\|B\|_{\mathrm{F}}} e_k e_\ell^{\mathrm{T}} \otimes B \right\rangle_{\mathrm{F}} \left(\frac{1}{\|B\|_{\mathrm{F}}} e_k e_\ell^{\mathrm{T}} \otimes B \right)$$

$$= \frac{1}{\|B\|_{\mathrm{F}}^2} \sum_{k,\ell=1}^{n} \operatorname{tr}\Big(\big(e_k e_\ell^{\mathrm{T}} \otimes B\big)^{\mathrm{T}} Y\Big)\big(e_k e_\ell^{\mathrm{T}} \otimes B\big).$$

Note that

$$\begin{aligned}
&\operatorname{tr}\Big(\big(e_k e_\ell^{\mathrm{T}} \otimes B\big)^{\mathrm{T}} Y\Big) = \operatorname{tr}\big(\big(e_\ell e_k^{\mathrm{T}} \otimes B^{\mathrm{T}}\big)Y\big) = \operatorname{tr}\big((e_\ell \otimes B_R^*)\big(e_k^{\mathrm{T}} \otimes B_L^*\big)Y\big) \\
&= \operatorname{tr}\big(\big(e_k^{\mathrm{T}} \otimes B_L^*\big)Y(e_\ell \otimes B_R^*)\big) = \sum_{j=1}^{m} e_j^{\mathrm{T}}\big(e_k^{\mathrm{T}} \otimes B_L^*\big)Y(e_\ell \otimes B_R^*)e_j \\
&= \sum_{j=1}^{m} \big(e_k^{\mathrm{T}} \otimes e_j^{\mathrm{T}} B_L^*\big)Y(e_\ell \otimes B_R^* e_j) = \sum_{j=1}^{m} e_k^{\mathrm{T}}\big(I_n \otimes e_j^{\mathrm{T}} B_L^*\big)Y(I_n \otimes B_R^* e_j)e_\ell \\
&= \left[\sum_{j=1}^{m} \big(I_n \otimes e_j^{\mathrm{T}} B_L^*\big)Y(I_n \otimes B_R^* e_j)\right]_{k\ell}.
\end{aligned}$$

Now,

$$\begin{aligned}
P_A(Y) \otimes B &= \frac{1}{\|B\|_{\mathrm{F}}^2}\left(\sum_{k,\ell=1}^{n}\left[\sum_{j=1}^{m}\big(I_n \otimes e_j^{\mathrm{T}} B_L^*\big)Y(I_n \otimes B_R^* e_j)\right]_{k\ell} e_k e_\ell^{\mathrm{T}}\right) \otimes B \\
&= \left(\frac{1}{\|B\|_{\mathrm{F}}^2}\sum_{j=1}^{m}\big(I_n \otimes e_j^{\mathrm{T}} B_L^*\big)Y(I_n \otimes B_R^* e_j)\right) \otimes B.
\end{aligned}$$

Using that B is nonzero, we obtain

$$P_A(Y) = \frac{1}{\|B\|_{\mathrm{F}}^2}\sum_{j=1}^{m}\big(I_n \otimes e_j^{\mathrm{T}} B_L^*\big)Y(I_n \otimes B_R^* e_j),$$

and the expression for $\nabla G(A)$ follows using (A.1). The expression for $\nabla H(B)$ can be found analogously.

Appendix B. Pole-residue Form with a Constant Term. To convert the form in (5.1) to the pole-residue form in Theorem 5.1, write $A_2 = UV^{\mathrm{T}}$ for some $U, V \in \mathbb{R}^{n \times n_2}$ of full column rank. Then, using the Sherman-Morrison-Woodbury formula [27], we obtain

$$\begin{aligned}
y(\mathsf{p}) &= C\big(A_1 + \mathsf{p}UV^{\mathrm{T}}\big)^{-1}B \\
&= C\Big(A_1^{-1} - \mathsf{p}A_1^{-1}U\big(I_{n_2} + \mathsf{p}V^{\mathrm{T}}A_1^{-1}U\big)^{-1}V^{\mathrm{T}}A_1^{-1}\Big)B \\
&= CA_1^{-1}B - \mathsf{p}CA_1^{-1}U\big(I_{n_2} + \mathsf{p}V^{\mathrm{T}}A_1^{-1}U\big)^{-1}V^{\mathrm{T}}A_1^{-1}B.
\end{aligned}$$

Next, let $T \in \mathbb{C}^{n_2 \times n_2}$ be an invertible matrix such that $V^{\mathrm{T}}A_1^{-1}U = TDT^{-1}$ for $D = \operatorname{diag}(d_1, d_2, \ldots, d_{n_2})$. Furthermore, define $C_U = CA_1^{-1}U$ and $B_V = V^{\mathrm{T}}A_1^{-1}B$. Continuing the above derivation, we obtain

$$\begin{aligned}
y(\mathsf{p}) &= CA_1^{-1}B - \mathsf{p}C_U\big(I_{n_2} + \mathsf{p}TDT^{-1}\big)^{-1}B_V \\
&= CA_1^{-1}B - \mathsf{p}C_U T(I_{n_2} + \mathsf{p}D)^{-1}T^{-1}B_V
\end{aligned}$$

$$= CA_1^{-1}B - \sum_{i=1}^{n_2} \frac{\mathsf{p} C_U T e_i e_i^{\mathrm{T}} T^{-1} B_V}{1 + \mathsf{p} d_i}.$$

Using $\frac{\mathsf{p}}{1+\mathsf{p}d_i} = \frac{1}{d_i} - \frac{\frac{1}{d_i^2}}{\mathsf{p}+\frac{1}{d_i}}$, $\sum_{i=1}^{n_2} \frac{1}{d_i} e_i e_i^{\mathrm{T}} = D^{-1}$, and $TD^{-1}T^{-1} = (V^{\mathrm{T}} A_1^{-1} U)^{-1}$ yields

$$y(\mathsf{p}) = CA_1^{-1}B - C_U \left(V^{\mathrm{T}} A_1^{-1} U\right)^{-1} B_V + \sum_{i=1}^{n_2} \frac{\frac{1}{d_i^2} C_U T e_i e_i^{\mathrm{T}} T^{-1} B_V}{\mathsf{p} + \frac{1}{d_i}}.$$

Therefore, in the pole-residue form Theorem 5.1, we have

$$\begin{aligned}
\Phi_0 &= CA_1^{-1}B - CA_1^{-1}U\left(V^{\mathrm{T}} A_1^{-1} U\right)^{-1} V^{\mathrm{T}} A_1^{-1} B, \\
\Phi_i &= CA_1^{-1} U T e_i e_i^{\mathrm{T}} T^{-1} V^{\mathrm{T}} A_1^{-1} B / d_i^2, \\
\nu_i &= -1/d_i,
\end{aligned}$$

for $i = 1, 2, \ldots, n_2$.

REFERENCES

[1] B. Anić, C. Beattie, S. Gugercin, and A. C. Antoulas, *Interpolatory weighted-$\mathcal{H}_2$ model reduction*, Automatica, 49 (2013), pp. 1275–1280, https://doi.org/10.1016/j.automatica.2013.01.040.

[2] A. C. Antoulas, C. A. Beattie, and S. Gugercin, *Interpolatory model reduction of large-scale dynamical systems*, in Efficient Modeling and Control of Large-Scale Systems, J. Mohammadpour and K. M. Grigoriadis, eds., Springer US, Boston, MA, 2010, pp. 3–58, https://doi.org/10.1007/978-1-4419-5757-3_1.

[3] A. C. Antoulas, C. A. Beattie, and S. Güğercin, *Interpolatory methods for model reduction*, Computational Science and Engineering 21, SIAM, Philadelphia, PA, 2020, https://doi.org/10.1137/1.9781611976083.

[4] U. Baur, C. A. Beattie, P. Benner, and S. Gugercin, *Interpolatory projection methods for parameterized model reduction*, SIAM J. Sci. Comput., 33 (2011), pp. 2489–2518, https://doi.org/10.1137/090776925.

[5] C. Beattie and S. Gugercin, *Interpolatory projection methods for structure-preserving model reduction*, Systems Control Lett., 58 (2009), pp. 225–232, https://doi.org/10.1016/j.sysconle.2008.10.016.

[6] C. Beattie and S. Gugercin, *Chapter 7: Model Reduction by Rational Interpolation*, SIAM, 2017, pp. 297–334, https://doi.org/10.1137/1.9781611974829.ch7.

[7] C. A. Beattie and P. Benner, *$\mathcal{H}_2$-optimality conditions for structured dynamical systems*. Preprint MPIMD/14-18, Max Planck Institute Magdeburg, 2014, https://csc.mpi-magdeburg.mpg.de/preprints/2014/18/.

[8] P. Benner and T. Breiten, *Interpolation-based $\mathcal{H}_2$-model reduction of bilinear control systems*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 859–885, https://doi.org/10.1137/110836742.

[9] P. Benner, P. Goyal, and S. Gugercin, *$\mathcal{H}_2$-quasi-optimal model order reduction for quadratic-bilinear control systems*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 983–1032, https://doi.org/10.1137/16M1098280.

[10] P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. Schilders, and L. M. Silveira, eds., *Model Order Reduction: Volume 2: Snapshot-Based Methods and Algorithms*, De Gruyter, Berlin, Boston, 2020, https://doi.org/10.1515/9783110671490.

[11] P. Benner, S. Gugercin, and S. W. R. Werner, *Structure-preserving interpolation for model reduction of parametric bilinear systems*, Automatica, 132 (2021), pp. 1–9, https://doi.org/10.1016/j.automatica.2021.109799.

[12] P. Benner, S. Gugercin, and S. W. R. Werner, *Structure-preserving interpolation of bilinear control systems*, Adv. Comput. Math., 47 (2021), pp. 1–38, https://doi.org/10.1007/s10444-021-09863-w.

[13] P. Benner, M. Ohlberger, A. Cohen, and K. Willcox, *Model Reduction and Approximation*, SIAM, Philadelphia, PA, 2017, https://doi.org/10.1137/1.9781611974829.

[14] M. Berljafa and S. Güttel, *The RKFIT algorithm for nonlinear rational approximation*, SIAM J. Sci. Comput., 39 (2017), pp. A2049–A2071, https://doi.org/10.1137/15M1025426.

[15] T. Breiten, C. Beattie, and S. Gugercin, *Near-optimal frequency-weighted interpolatory model reduction*, Systems Control Lett., 78 (2015), pp. 8–18, https://doi.org/10.1016/j.sysconle.2015.01.005.

[16] 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, https://doi.org/10.1016/j.cam.2008.12.029.

[17] Y. Chahlaoui and P. Van Dooren, *A collection of benchmark examples for model reduction of linear time invariant dynamical systems*, Tech. Report 2002-2, SLICOT Working Note, 2002. Available from www.slicot.org.

[18] Z. Drmač, S. Gugercin, and C. Beattie, *Quadrature-based vector fitting for discretized $\mathcal{H}_2$ approximation*, SIAM J. Sci. Comput., 37 (2015), pp. A625–A652, https://doi.org/10.1137/140961511.

[19] Z. Drmač, S. Gugercin, and C. Beattie, *Vector fitting for matrix-valued rational approximation*, SIAM J. Sci. Comput., 37 (2015), pp. A2346–A2379, https://doi.org/10.1137/15M1010774.

[20] V. Druskin and V. Simoncini, *Adaptive rational Krylov subspaces for large-scale dynamical systems*, Systems Control Lett., 60 (2011), pp. 546–560, https://doi.org/10.1016/j.sysconle.2011.04.013.

[21] V. Druskin, V. Simoncini, and M. Zaslavsky, *Adaptive tangential interpolation in rational Krylov subspaces for MIMO dynamical systems*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 476–498, https://doi.org/10.1137/120898784.

[22] L. Feng, A. C. Antoulas, and P. Benner, *Some a posteriori error bounds for reduced order modelling of (non-)parametrized linear systems*, ESAIM: Math. Model. Numer. Anal., 51 (2017), pp. 2127–2158, https://doi.org/10.1051/m2an/2017014.

[23] L. Feng and P. Benner, *A new error estimator for reduced-order modeling of linear parametric systems*, IEEE Trans. Microw. Theory Techn., 67 (2019), pp. 4848–4859, https://doi.org/10.1109/TMTT.2019.2948858.

[24] L. Feng and P. Benner, *On error estimation for reduced-order modeling of linear non-parametric and parametric systems*, ESAIM: Math. Model. Numer. Anal., 55 (2021), pp. 561–594, https://doi.org/10.1051/m2an/2021001.

[25] G. Flagg, C. A. Beattie, and S. Gugercin, *Interpolatory H_∞ model reduction*, Systems Control Lett., 62 (2013), pp. 567–574, https://doi.org/10.1016/j.sysconle.2013.03.006.

[26] G. Flagg and S. Gugercin, *Multipoint Volterra series interpolation and $\mathcal{H}_2$ optimal model reduction of bilinear systems*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 549–579, https://doi.org/10.1137/130947830.

[27] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, fourth ed., 2013.

[28] A. R. Grimm, *Parametric Dynamical Systems: Transient Analysis and Data Driven Modeling*, PhD thesis, Virginia Polytechnic Institute and State University, 2018, http://hdl.handle.net/10919/83840.

[29] S. Gugercin, A. C. Antoulas, and C. Beattie, *$\mathcal{H}_2$ model reduction for large-scale linear dynamical systems*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 609–638, https://doi.org/10.1137/060666123.

[30] S. Gugercin, A. C. Antoulas, and C. A. Beattie, *A rational Krylov iteration for optimal $\mathcal{H}_2$ model reduction*, in Proc. of the 17th International Symposium on Mathematical Theory of Networks and Systems, 2006, pp. 1665–1667.

[31] S. Gugercin, R. V. Polyuga, C. Beattie, and A. van der Schaft, *Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems*, Automatica, 48 (2012), pp. 1963–1974, https://doi.org/10.1016/j.automatica.2012.05.052.

[32] S. Gugercin, T. Stykel, and S. Wyatt, *Model reduction of descriptor systems by interpolatory projection methods*, SIAM J. Sci. Comput., 35 (2013), pp. B1010–B1033, https://doi.org/10.1137/130906635.

[33] B. Gustavsen and A. Semlyen, *Rational approximation of frequency domain responses by vector fitting*, IEEE Trans. Power Del., 14 (1999), pp. 1052–1061, https://doi.org/10.1109/61.772353.

[34] J. S. Hesthaven, G. Rozza, and B. Stamm, *Certified reduced basis methods for parametrized partial differential equations*, Springer Briefs in Mathematics, Springer, Switzerland, 2016, https://doi.org/10.1007/978-3-319-22470-1.

[35] J. M. Hokanson and C. C. Magruder, *$\mathcal{H}_2$-optimal model reduction using projected nonlinear least squares*, SIAM J. Sci. Comput., 42 (2020), pp. A4017–A4045, https://doi.org/10.

1137/19M1247863.

[36] L. Meier and D. Luenberger, *Approximation of linear constant systems*, IEEE Trans. Autom. Control, 12 (1967), pp. 585–588, https://doi.org/10.1109/TAC.1967.1098680.

[37] P. Mlinarić, *$\mathcal{L}_2$-optimal interpolation experiments*, Aug. 2022, https://github.com/pmli/l2-opt-interp-ex/tree/v1.

[38] P. Mlinarić and S. Gugercin, *$\mathcal{L}_2$-optimal reduced-order modeling using parameter-separable forms.* arXiv preprint 2206.02929, 2022, https://doi.org/10.48550/arXiv.2206.02929.

[39] Y. Nakatsukasa, O. Sète, and L. N. Trefethen, *The AAA algorithm for rational approximation*, SIAM J. Sci. Comput., 40 (2018), pp. A1494–A1522, https://doi.org/10.1137/16M1106122.

[40] A. Quarteroni, A. Manzoni, and F. Negri, *Reduced basis methods for partial differential equations*, UNITEXT, Springer Cham, Switzerland, 2016, https://doi.org/10.1007/978-3-319-15431-2.

[41] A. C. Rodriguez, S. Gugercin, and J. Borggaard, *Interpolatory model reduction of parameterized bilinear dynamical systems*, Adv. Comput. Math., 44 (2018), pp. 1887–1916, https://doi.org/10.1007/s10444-018-9611-y.

[42] E. D. Sontag, *Mathematical Control Theory*, Springer New York, 1998, https://doi.org/10.1007/978-1-4612-0577-7.

[43] P. Vuillemin, C. Poussot-Vassal, and D. Alazard, *Poles residues descent algorithm for optimal frequency-limited $\mathcal{H}_2$ model approximation*, in European Control Conference (ECC), 2014, pp. 1080–1085, https://doi.org/10.1109/ECC.2014.6862152.