INTERPOLATORY $\mathcal{H}_2$-OPTIMALITY CONDITIONS FOR STRUCTURED LINEAR TIME-INVARIANT SYSTEMS*

PREPRINT

Petar Mlinarić[†] Peter Benner[‡] Serkan Gugercin[§]

October 19, 2023

ABSTRACT

Interpolatory necessary optimality conditions for $\mathcal{H}_2$-optimal reduced-order modeling of unstructured linear time-invariant (LTI) systems are well-known. Based on previous work on $\mathcal{L}_2$-optimal reduced-order modeling of stationary parametric problems, in this paper we develop and investigate optimality conditions for $\mathcal{H}_2$-optimal reduced-order modeling of structured LTI systems, in particular, for second-order, port-Hamiltonian, and time-delay systems. We show that across all these different structured settings, bitangential Hermite interpolation is the common form for optimality, thus proving a unifying optimality framework for structured reduced-order modeling.

Keywords rational interpolation · approximation theory · model order reduction · linear time-invariant systems · optimization · $\mathcal{H}_2$ norm

1 Introduction

Linear time-invariant (LTI) dynamical systems are ubiquitous in many applications and can often be represented in a finite-dimensional, unstructured state-space form

$$E\dot{x}(t) = Ax(t) + Bu(t), \tag{1.1a}$$

$$y(t) = Cx(t), \tag{1.1b}$$

where $E, A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times n_{\mathrm{i}}}$, and $C \in \mathbb{R}^{n_{\mathrm{o}} \times n}$; $x(t) \in \mathbb{R}^n$ are the states (internal degrees of freedom) of the system, $u(t) \in \mathbb{R}^{n_{\mathrm{i}}}$ are the inputs (forcing), and $y(t) \in \mathbb{R}^{n_{\mathrm{o}}}$ are the outputs (quantities of interest). Given the LTI system (1.1) (called a full-order model (FOM)), the goal of (unstructured) $\mathcal{H}_2$-optimal reduced-order modeling is to find a reduced-order model (ROM) of order $r \ll n$ given in the state-space form as

$$\widehat{E}\dot{\widehat{x}}(t) = \widehat{A}\widehat{x}(t) + \widehat{B}u(t), \tag{1.2a}$$

$$\widehat{y}(t) = \widehat{C}\widehat{x}(t), \tag{1.2b}$$

with $\widehat{E}, \widehat{A} \in \mathbb{R}^{r \times r}$, $\widehat{B} \in \mathbb{R}^{r \times n_{\mathrm{i}}}$, and $\widehat{C} \in \mathbb{R}^{n_{\mathrm{o}} \times r}$ such that the ROM (1.2) minimizes the squared $\mathcal{H}_2$ error

$$\left\lVert H - \widehat{H} \right\rVert_{\mathcal{H}_2}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\lVert H(\imath\omega) - \widehat{H}(\imath\omega) \right\rVert_{\mathrm{F}}^2 \,\mathrm{d}\omega, \tag{1.3}$$

where H and $\widehat{H}$ are *the transfer functions* of the FOM and the ROM, respectively, and are given by

$$H(s) = C(sE - A)^{-1}B \quad \text{and} \quad \widehat{H}(s) = \widehat{C}\left(s\widehat{E} - \widehat{A}\right)^{-1}\widehat{B}.$$

*This work was partially funded by the U.S. National Science Foundation under grant DMS-1923221.

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

[‡]Computational Methods in Systems and Control Theory, Max Planck Institute for Dynamics of Complex Technical Systems, 39106 Magdeburg, Germany (benner@mpi-magdeburg.mpg.de).

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

In particular, transfer functions satisfy

$$Y(s) = H(s)U(s) \text{ and } \widehat{Y}(s) = \widehat{H}(s)U(s),$$

where $U, Y, \widehat{Y}$ are Laplace transforms of the input and output signals $u, y, \widehat{y}$. Additional assumptions are that H itself has a finite $\mathcal{H}_2$ norm and is analytic over $\mathbb{C}_+$ (the open left half-plane), and that $\widehat{E}$ is invertible and $\widehat{E}^{-1}\widehat{A}$ has eigenvalues in the open left half-plane, rendering the ROM asymptotically (exponentially, Lyapunov) stable and ensuring it has a finite $\mathcal{H}_2$ norm. Finding a ROM that minimizes the $\mathcal{H}_2$ error is motivated by the $\mathcal{L}_\infty$ output error bound, namely

$$\|y - \widehat{y}\|_{\mathcal{L}_\infty} \leqslant \left\|H - \widehat{H}\right\|_{\mathcal{H}_2} \|u\|_{\mathcal{L}_2}.$$

Thus, a small $\mathcal{H}_2$ error guarantees a small $\mathcal{L}_\infty$ output error (in the time-domain). The function H can be a transfer function of a finite-dimensional, LTI system of the same form as in (1.1), but this is not necessary (as for the case we consider in Section 6).

Assuming that $\widehat{E}^{-1}\widehat{A}$ is diagonalizable, we can write the transfer function $\widehat{H}$ in *the pole-residue form*

$$\widehat{H}(s) = \widehat{C}\left(s\widehat{E} - \widehat{A}\right)^{-1}\widehat{B} = \sum_{i=1}^{r} \frac{c_i b_i^*}{s - \lambda_i} \tag{1.4}$$

with pairwise distinct poles λ_i (here $(\cdot)^*$ denotes the complex conjugate). Then the well-known necessary $\mathcal{H}_2$-optimality conditions in interpolation form [ML67, GAB08] state that an $\mathcal{H}_2$-optimal ROM with the transfer function in the pole-residue form (1.4) needs to satisfy

$$H\left(-\overline{\lambda_i}\right)b_i = \widehat{H}\left(-\overline{\lambda_i}\right)b_i, \tag{1.5a}$$

$$c_i^* H\left(-\overline{\lambda_i}\right) = c_i^* \widehat{H}\left(-\overline{\lambda_i}\right), \tag{1.5b}$$

$$c_i^* H'\left(-\overline{\lambda_i}\right)b_i = c_i^* \widehat{H}'\left(-\overline{\lambda_i}\right)b_i, \tag{1.5c}$$

for $i = 1, 2, \ldots, r$. The necessary conditions (1.5a) and (1.5b) are, respectively, right and left tangential Lagrange interpolation conditions, and (1.5c) are bitangential Hermite conditions. These conditions simply state that the $\mathcal{H}_2$-optimal rational approximant $\widehat{H}$ needs to be a bitangential Hermite interpolant to H at the mirror of the poles of $\widehat{H}$ along the tangential directions resulting from the residues of $\widehat{H}$. In other words, the poles and residues of $\widehat{H}$ specify the optimal interpolation points and the tangent directions. The optimality conditions above assume simple poles, our main focus in this paper. For extension to poles with multiplicities, see [DGA10].

Interpolation conditions (1.5) can be enforced using either a Petrov-Galerkin projection (intrusively) or via data-driven Loewner matrices (non-intrusively). This observation has lead to iterative rational Krylov algorithm (IRKA) [GAB08] (projection-based formulation) and transfer-function IRKA (TF-IRKA) [BG12], an efficient framework for locally $\mathcal{H}_2$-optimal reduced-order modeling. For more details on interpolatory model reduction, we refer the reader to [ABG20].

The conditions (1.5) follow due to the (unstructured) pole-residue form (1.4) of the transfer function H. In this paper, we want to understand what happens to interpolatory conditions (1.5) when the ROM is a structured system, e.g., a second-order system or a time-delay system, and the pole-residue form needs to be modified accordingly. As the optimization set changes, it is expected that the necessary optimality conditions also change. The question then is, whether they are still in the form of bitangential Hermite conditions.

Recent work [MG23a, MG23b] introduced a unifying $\mathcal{L}_2$-optimal reduced-order modeling framework, covering both LTI systems and stationary parametric problems. In particular, [MG23b] showed that necessary optimality conditions in the form of bitangential Hermite interpolation appear often, demonstrated by recovering known conditions for (parametric) LTI systems and deriving new ones for discretized $\mathcal{H}_2$-optimal reduced-order modeling and for a class of stationary parametric problems. Here, we are interested in certain important classes of structured LTI systems: second-order systems, port-Hamiltonian (pH) systems, and time-delay systems.

Our main contributions are the following:

1. Extending the work of [MG23a, MG23b] to complex-valued models (Section 2).
2. Developing interpolatory necessary $\mathcal{H}_2$-optimality conditions for a class of diagonal structured systems (Section 3).

3. Deriving interpolatory optimality conditions for different structured LTI systems: (a) second-order systems (Section 4), (b) pH systems (Section 5), and (c) time-delay systems (Section 6).
4. In almost all cases, we show that bitangential Hermite interpolation is the key uniform framework behind optimality.

The rest of the paper is organized as follows. In Section 2, we recall some of the main results from [MG23a, MG23b] on $\mathcal{L}_2$-optimal reduced-order modeling and extend them to ROMs with complex and/or diagonal matrices. In Section 3, we specialize this discussion to $\mathcal{H}_2$-optimal reduced-order modeling of diagonal dynamical systems, and then use this result in Sections 4 to 6 to respectively cover second-order, pH, and time-delay systems. We conclude with Section 7.

2 $\mathcal{L}_2$-optimal Reduced-order Modeling

Here we recall the setting of [MG23a] and extend it to structured ROMs (StROMs) with complex matrices. Then we generalize the gradients of $\mathcal{J}$ (2.4) from [MG23a] and necessary $\mathcal{L}_2$-optimality conditions from [MG23b] to complex StROMs. The final section is on the extension to diagonal StROMs (D-StROMs).

2.1 Setting

Note that the transfer function of the ROM (1.2) can be reformulated as the output of a parametric stationary problem

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

where $s \in \imath\mathbb{R}$ is interpreted as a parameter. Generalizing the form of the above ROM and s to a parameter $\mathsf{p} \in \mathcal{P} \subseteq \mathbb{C}^{n_{\mathrm{p}}}$, for a positive integer n_{p}, brings us to the $\mathcal{L}_2$-optimal reduced-order modeling problem discussed in [MG23a]. There, given a parameter-to-output mapping

$$H \colon \mathcal{P} \to \mathbb{C}^{n_{\mathrm{o}} \times n_{\mathrm{i}}}, \tag{2.1}$$

the goal is to construct a StROM

$$\widehat{\mathcal{A}}(\mathsf{p})\widehat{X}(\mathsf{p}) = \widehat{\mathcal{B}}(\mathsf{p}), \tag{2.2a}$$
$$\widehat{H}(\mathsf{p}) = \widehat{\mathcal{C}}(\mathsf{p})\widehat{X}(\mathsf{p}), \tag{2.2b}$$

with a parameter-separable form

$$\widehat{\mathcal{A}}(\mathsf{p}) = \sum_{i=1}^{q_{\widehat{\mathcal{A}}}} \widehat{\alpha}_i(\mathsf{p})\widehat{A}_i, \quad \widehat{\mathcal{B}}(\mathsf{p}) = \sum_{j=1}^{q_{\widehat{\mathcal{B}}}} \widehat{\beta}_j(\mathsf{p})\widehat{B}_j, \quad \widehat{\mathcal{C}}(\mathsf{p}) = \sum_{k=1}^{q_{\widehat{\mathcal{C}}}} \widehat{\gamma}_k(\mathsf{p})\widehat{C}_k, \tag{2.3}$$

where $\widehat{X}(\mathsf{p}) \in \mathbb{C}^{r \times n_{\mathrm{i}}}$ is the reduced state, $\widehat{H}(\mathsf{p}) \in \mathbb{C}^{n_{\mathrm{o}} \times n_{\mathrm{i}}}$ 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 $\widehat{C}_k \in \mathbb{R}^{n_{\mathrm{o}} \times r}$. From here on, we also allow complex reduced matrices, i.e., $\widehat{A}_i \in \mathbb{C}^{r \times r}$, $\widehat{B}_j \in \mathbb{C}^{r \times n_{\mathrm{i}}}$, and $\widehat{C}_k \in \mathbb{C}^{n_{\mathrm{o}} \times r}$. We will use the notation $(\widehat{A}_i, \widehat{B}_j, \widehat{C}_k)$ to denote the StROM specified by (2.2) and (2.3).

In [MG23a] the StROM (2.2) is constructed to minimize the squared $\mathcal{L}_2$ error

$$\mathcal{J}\left(\widehat{A}_i, \widehat{B}_j, \widehat{C}_k\right) = \left\lVert H - \widehat{H} \right\rVert_{\mathcal{L}_2(\mathcal{P},\mu)}^2 = \int_{\mathcal{P}} \left\lVert H(\mathsf{p}) - \widehat{H}(\mathsf{p}) \right\rVert_{\mathrm{F}}^2 \,\mathrm{d}\mu(\mathsf{p}), \tag{2.4}$$

where μ is a measure over $\mathcal{P}$.

The work [MG23a] derived the gradients of $\mathcal{J}$ with respect to the *real* StROM matrices $\widehat{A}_i, \widehat{B}_j, \widehat{C}_k$. To develop the interpolatory necessary optimality conditions of structured LTI systems, we will require diagonalizability of certain matrix pencils as we will explore in more detail in later sections. However, since diagonalization of real systems can lead to complex diagonal matrices, we would need to generalize the results in [MG23a, MG23b] to complex reduced-order matrices. The next section achieves this goal using Wirtinger calculus.

2.2 Wirtinger Calculus and Gradients of the Squared $\mathcal{L}_2$ Error

Since the (squared) $\mathcal{L}_2$ error is real-valued, it cannot be analytic with respect to complex variables (unless it is constant, which is not an interesting case). Wirtinger calculus generalizes the complex derivative to non-analytic functions; see [Vui14, Section 8.2.1] and references within. In particular, let the function f of a complex variable z be written as $f(z) = g(z, \overline{z}) = h(x, y)$ for a function g that is analytic with respect to both variables and for a function h of the real and imaginary part of $z = x + \imath y$. Define the following differential operators

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - \imath\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + \imath\frac{\partial}{\partial y}\right).$$

These differential operators allow to differentiate a non-analytic function f by treating z and $\overline{z}$ as independent variables.

This can be analogously generalized to gradients with respect to complex vectors and matrices. In particular, for a real-valued function f of a complex matrix X such that $f(X) = g(X, \overline{X})$, if we have that

$$f(X + \Delta X) = f(X) + \left\langle \nabla_X g(X), \overline{X} \right\rangle_{\mathrm{F}} + \left\langle \nabla_{\overline{X}} g(X), X \right\rangle_{\mathrm{F}} + o(\|\Delta X\|),$$

then we define $\nabla_X f(X) = \nabla_X g(X)$ and $\nabla_{\overline{X}} f(X) = \nabla_{\overline{X}} g(X)$. This allows us to directly generalize the results of [MG23a, MG23b] to the complex-valued case.

Theorem 2.1. *Suppose that $\mathcal{P} \subseteq \mathbb{C}^{n_{\mathrm{p}}}$, μ is a measure over $\mathcal{P}$, the function H is in $\mathcal{L}_2(\mathcal{P}, \mu; \mathbb{C}^{n_{\mathrm{o}} \times n_{\mathrm{i}}})$, functions $\widehat{\alpha}_i, \widehat{\beta}_j, \widehat{\gamma}_k \colon \mathcal{P} \to \mathbb{C}$ are measurable and satisfy*

$$\int_{\mathcal{P}} \left(\frac{\sum_{j=1}^{q_{\widehat{\mathcal{B}}}} \left|\widehat{\beta}_j(\mathsf{p})\right| \sum_{k=1}^{q_{\widehat{\mathcal{C}}}} |\widehat{\gamma}_k(\mathsf{p})|}{\sum_{i=1}^{q_{\widehat{\mathcal{A}}}} |\widehat{\alpha}_i(\mathsf{p})|} \right)^2 \mathrm{d}\mu(\mathsf{p}) < \infty, \tag{2.5}$$

$\widehat{A}_i \in \mathbb{C}^{r \times r}$, $\widehat{B}_j \in \mathbb{C}^{r \times n_{\mathrm{i}}}$, $\widehat{C}_k \in \mathbb{C}^{n_{\mathrm{o}} \times r}$, and

$$\operatorname*{ess\,sup}_{\mathsf{p} \in \mathcal{P}} \left\| \widehat{\alpha}_i(\mathsf{p}) \widehat{\mathcal{A}}(\mathsf{p})^{-1} \right\|_{\mathrm{F}} < \infty, \quad i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}}, \tag{2.6}$$

where $\widehat{\mathcal{A}}$ is as in (2.3). Then, for $\mathcal{J}$ in (2.4),

$$\nabla_{\overline{\widehat{A}_i}} \mathcal{J} = \int_{\mathcal{P}} \overline{\widehat{\alpha}_i(\mathsf{p})} \widehat{X}_d(\mathsf{p}) \Big(H(\mathsf{p}) - \widehat{H}(\mathsf{p})\Big) \widehat{X}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}), \qquad i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}}, \tag{2.7a}$$

$$\nabla_{\overline{\widehat{B}_j}} \mathcal{J} = \int_{\mathcal{P}} \overline{\widehat{\beta}_j(\mathsf{p})} \widehat{X}_d(\mathsf{p}) \Big(\widehat{H}(\mathsf{p}) - H(\mathsf{p})\Big) \,\mathrm{d}\mu(\mathsf{p}), \qquad j = 1, 2, \ldots, q_{\widehat{\mathcal{B}}}, \tag{2.7b}$$

$$\nabla_{\overline{\widehat{C}_k}} \mathcal{J} = \int_{\mathcal{P}} \overline{\widehat{\gamma}_k(\mathsf{p})} \Big(\widehat{H}(\mathsf{p}) - H(\mathsf{p})\Big) \widehat{X}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}), \qquad k = 1, 2, \ldots, q_{\widehat{\mathcal{C}}}, \tag{2.7c}$$

where $\widehat{X}_d(\mathsf{p}) = \widehat{\mathcal{A}}(\mathsf{p})^{-*} \widehat{\mathcal{C}}(\mathsf{p})^* \in \mathbb{C}^{r \times n_{\mathrm{o}}}$ is the dual reduced state.

Proof. The proof is similar to the proof of Theorem 3.7 in [MG23a]. The significant change is that $\langle H, \widehat{H} \rangle_{\mathcal{L}_2}$ is no longer necessarily real and not equal to $\langle \widehat{H}, H \rangle_{\mathcal{L}_2}$. Therefore, we have

$$\mathcal{J} = \left\| H - \widehat{H} \right\|_{\mathcal{L}_2}^2 = \|H\|_{\mathcal{L}_2}^2 - \left\langle H, \widehat{H} \right\rangle_{\mathcal{L}_2} - \left\langle \widehat{H}, H \right\rangle_{\mathcal{L}_2} + \left\| \widehat{H} \right\|_{\mathcal{L}_2}^2.$$

However, since $\langle H, \widehat{H} \rangle_{\mathcal{L}_2}$ does not depend directly on $\overline{\widehat{A}_i}, \overline{\widehat{B}_j}, \overline{\widehat{C}_k}$, only $\langle \widehat{H}, H \rangle_{\mathcal{L}_2}$ contributes to the gradients. Similarly, only one part in the product rule applied to $\langle \widehat{H}, \widehat{H} \rangle_{\mathcal{L}_2}$ contributes to the gradients. This leads to the missing factor of 2 compared to [MG23a]. Additionally, because the scalar functions $\widehat{\alpha}_i, \widehat{\beta}_j, \widehat{\gamma}_k$ are not assumed to be closed under conjugation, the conjugation remains outside of the function evaluation. ☐

These gradients can then be used to develop an $\mathcal{L}_2$-optimal reduced-order modeling algorithm, as done in [MG23a] for the case of real stationary parametric problems and LTI systems. In the following, we focus on necessary optimality conditions, especially in the interpolation form. In the next subsection, we start with the necessary conditions for StROMs. This will then be the basis for the structured problems we consider in later sections.

2.3 Necessary $\mathcal{L}_2$-optimality Conditions

As done in [MG23b] for the real-valued case, an important consequence of Theorem 2.1 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. The same holds for differential operators from Wirtinger calculus, i.e., local minima have zero gradient, leading to our next result.

Corollary 2.2. *Let the assumptions of Theorem 2.1 hold. Furthermore, let the reduced model $\widehat{H}$ defined by $(\widehat{A}_i, \widehat{B}_j, \widehat{C}_k)$ as in (2.2) and (2.3) be an $\mathcal{L}_2$-optimal StROM to H. Then,*

$$\int_{\mathcal{P}} \overline{\gamma_k(\mathsf{p})} H(\mathsf{p}) \widehat{X}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}) = \int_{\mathcal{P}} \overline{\gamma_k(\mathsf{p})} \widehat{H}(\mathsf{p}) \widehat{X}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}), \qquad k = 1, 2, \ldots, q_{\widehat{\mathcal{C}}}, \tag{2.8a}$$

$$\int_{\mathcal{P}} \overline{\beta_j(\mathsf{p})} \widehat{X}_d(\mathsf{p}) H(\mathsf{p}) \,\mathrm{d}\mu(\mathsf{p}) = \int_{\mathcal{P}} \overline{\beta_j(\mathsf{p})} \widehat{X}_d(\mathsf{p}) \widehat{H}(\mathsf{p}) \,\mathrm{d}\mu(\mathsf{p}), \qquad j = 1, 2, \ldots, q_{\widehat{\mathcal{B}}}, \tag{2.8b}$$

$$\int_{\mathcal{P}} \overline{\alpha_i(\mathsf{p})} \widehat{X}_d(\mathsf{p}) H(\mathsf{p}) \widehat{X}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}) = \int_{\mathcal{P}} \overline{\alpha_i(\mathsf{p})} \widehat{X}_d(\mathsf{p}) \widehat{H}(\mathsf{p}) \widehat{X}(\mathsf{p})^* \,\mathrm{d}\mu(\mathsf{p}), \qquad i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}}. \tag{2.8c}$$

In the next subsection, we specialize these conditions to D-StROMs.

2.4 $\mathcal{L}_2$-optimality for Diagonal Structured Dynamics

We are interested in deriving interpolatory optimality conditions and this will require a certain diagonalizability assumption (as also used in the unstructured $\mathcal{H}_2$-optimality conditions (1.5)). In particular, we require that there exist matrices $\widehat{S}, \widehat{T} \in \mathbb{C}^{r \times r}$ such that $\widehat{S}^* \widehat{A}_i \widehat{T}$ is diagonal for all $i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}}$.

This is generically true when $q_{\widehat{\mathcal{A}}} \leqslant 2$, as in the case of non-parametric LTI systems (1.2). Such systems were considered in [MG23b], as well as stationary parametric systems with $q_{\widehat{\mathcal{A}}} = 2$ and parametric LTI systems with a special tensor structure. Here, we consider systems with $q_{\widehat{\mathcal{A}}} \geqslant 3$, such as second-order and time-delay systems, where diagonalizability is no longer a generic property.

Therefore, we consider D-StROMs, i.e., StROMs $(\widehat{A}_i, \widehat{B}_j, \widehat{C}_k)$ defined by (2.2) and (2.3) but with diagonal $\widehat{A}_i$ (thus we are assuming that the transformation $\widehat{S}^* \widehat{A}_i \widehat{T}$ has been already applied). As mentioned, a generic example of such systems are those with $q_{\widehat{\mathcal{A}}} = 2$ after a diagonalizing transformation. An additional example are second-order systems with Rayleigh damping where $\widehat{\mathcal{A}}(s) = s^2 \widehat{M} + s(\alpha \widehat{M} + \beta \widehat{K}) + \widehat{K}$ and $\widehat{M}$ and $\widehat{K}$ are the mass and stiffness matrices, respectively.

Deriving the optimality conditions for D-StROMs will require computing gradients with respect to diagonal matrices. The following lemma will allow us to reuse the gradients with respect to full matrices to find gradients with respect to diagonal matrices.

Lemma 2.3. *Let V be $\mathbb{C}^m$ or $\mathbb{C}^{m \times n}$, and $f \colon V \to \mathbb{R}$ be a Wirtinger differentiable function. Furthermore, let $W \subseteq V$ be a subspace closed under conjugation and define $g = f|_W$ as the restriction of f onto W. Then for $w \in W$, the function g is Wirtinger differentiable at w and $\nabla_{\overline{w}} g(w) = \operatorname{Proj}_W \nabla_{\overline{w}} f(w)$ where $\operatorname{Proj}_W \colon V \to W$ is the orthogonal projector onto W.*

In particular, let $F \colon \mathbb{C}^{r \times r} \to \mathbb{R}$ be a Wirtinger differentiable function and $G = F|_D$ its restriction to diagonal matrices $D = \{X \in \mathbb{R}^{r \times r} : X_{ij} = 0,\ \forall i \neq j\}$. Then for $X \in D$, the function G is Wirtinger differentiable at X and $\nabla_{\overline{X}} G(X) = \operatorname{diag}(\nabla_{\overline{X}} F(X))$, where $\operatorname{diag}(Y) = \sum_{i=1}^r Y_{ii} e_i e_i^{\mathrm{T}}$ is the diagonal part of the matrix Y.

Proof. Since f is Wirtinger differentiable at w, we have that, for $\Delta w \in W$,

$$f(w + \Delta w) = f(w) + \left\langle \nabla_w f(w), \overline{\Delta w} \right\rangle + \left\langle \nabla_{\overline{w}} f(w), \Delta w \right\rangle + o(\|\Delta w\|).$$

Therefore, since f and g are equal over W, we have

$$g(w + \Delta w) = g(w) + \left\langle \nabla_w f(w), \overline{\Delta w} \right\rangle + \left\langle \nabla_{\overline{w}} f(w), \Delta w \right\rangle + o(\|\Delta w\|).$$

We observe that

$$\begin{aligned} \left\langle \nabla_w f(w), \overline{\Delta w} \right\rangle &= \left\langle \operatorname{Proj}_W \nabla_w f(w), \overline{\Delta w} \right\rangle + \left\langle \nabla_w f(w) - \operatorname{Proj}_W \nabla_w f(w), \overline{\Delta w} \right\rangle \\ &= \left\langle \operatorname{Proj}_W \nabla_w f(w), \overline{\Delta w} \right\rangle \end{aligned}$$

and

$$\begin{aligned}\langle \nabla_{\overline{w}} f(w), \Delta w\rangle &= \langle \operatorname{Proj}_W \nabla_{\overline{w}} f(w), \Delta w\rangle + \langle \nabla_{\overline{w}} f(w) - \operatorname{Proj}_W \nabla_{\overline{w}} f(w), \Delta w\rangle \\ &= \langle \operatorname{Proj}_W \nabla_{\overline{w}} f(w), \Delta w\rangle.\end{aligned}$$

Since $\operatorname{Proj}_W \nabla_w f(w)$ and $\operatorname{Proj}_W \nabla_{\overline{w}} f(w)$ are elements of W, g is Wirtinger differentiable at w, $\nabla_w g(w) = \operatorname{Proj}_W \nabla_w f(w)$, and $\nabla_{\overline{w}} g(w) = \operatorname{Proj}_W \nabla_{\overline{w}} f(w)$, which proves the first part.

Next, we notice that the set of diagonal matrices D is a subspace of the space of matrices $\mathbb{C}^{r\times r}$, it is closed under conjugation, and $\operatorname{Proj}_D(Y) = \operatorname{diag}(Y)$. Thus, the second part of the result directly follows from the first. □

Lemma 2.3 will now help us derive the necessary $\mathcal{L}_2$-optimality conditions for D-StROMs. Since in this case all $\widehat{A}_i$'s are diagonal (and thus in return so is $\widehat{\mathcal{A}}(\mathsf{p})$ in (2.2)), $\widehat{H}$ for a D-StROM has a "pole-residue" form, i.e.,

$$\widehat{H}(\mathsf{p}) = \widehat{\mathcal{C}}(\mathsf{p})\widehat{\mathcal{A}}(\mathsf{p})^{-1}\widehat{\mathcal{B}}(\mathsf{p}) = \sum_{\ell=1}^{r} \frac{c_\ell(\mathsf{p}) b_\ell(\mathsf{p})^*}{a_\ell(\mathsf{p})}, \tag{2.9}$$

where $a_\ell(\mathsf{p})$ is the ℓth diagonal entry of $\widehat{\mathcal{A}}(\mathsf{p})$, $b_\ell(\mathsf{p}) \in \mathbb{C}^{n_{\mathrm{i}}}$ is the ℓth column of $\widehat{\mathcal{B}}(\mathsf{p})^*$, and $c_\ell(\mathsf{p}) \in \mathbb{C}^{n_{\mathrm{o}}}$ is the ℓth column of $\widehat{\mathcal{C}}(\mathsf{p})$. With this pole-residue form in hand, we now can investigate the optimality conditions for D-StROMs.

Corollary 2.4. *Let the assumptions in Theorem 2.1 hold. Furthermore, let $(\widehat{A}_i, \widehat{B}_j, \widehat{C}_k)$ be an $\mathcal{L}_2$-optimal D-StROM of H with $\widehat{H}$ as in (2.9). Then*

$$\int_{\mathcal{P}} \frac{\overline{\widehat{\gamma}_k(\mathsf{p})} H(\mathsf{p}) b_\ell(\mathsf{p})}{\overline{a_\ell(\mathsf{p})}} \,\mathrm{d}\mu(\mathsf{p}) = \int_{\mathcal{P}} \frac{\overline{\widehat{\gamma}_k(\mathsf{p})} \widehat{H}(\mathsf{p}) b_\ell(\mathsf{p})}{\overline{a_\ell(\mathsf{p})}} \,\mathrm{d}\mu(\mathsf{p}), \qquad k = 1, 2, \ldots, q_{\widehat{\mathcal{C}}}, \tag{2.10a}$$

$$\int_{\mathcal{P}} \frac{\overline{\widehat{\beta}_j(\mathsf{p})} c_\ell(\mathsf{p})^* H(\mathsf{p})}{\overline{a_\ell(\mathsf{p})}} \,\mathrm{d}\mu(\mathsf{p}) = \int_{\mathcal{P}} \frac{\overline{\widehat{\beta}_j(\mathsf{p})} c_\ell(\mathsf{p})^* \widehat{H}(\mathsf{p})}{\overline{a_\ell(\mathsf{p})}} \,\mathrm{d}\mu(\mathsf{p}), \qquad j = 1, 2, \ldots, q_{\widehat{\mathcal{B}}}, \tag{2.10b}$$

$$\int_{\mathcal{P}} \frac{\overline{\widehat{\alpha}_i(\mathsf{p})} c_\ell(\mathsf{p})^* H(\mathsf{p}) b_\ell(\mathsf{p})}{\overline{a_\ell(\mathsf{p})}^2} \,\mathrm{d}\mu(\mathsf{p}) = \int_{\mathcal{P}} \frac{\overline{\widehat{\alpha}_i(\mathsf{p})} c_\ell(\mathsf{p})^* \widehat{H}(\mathsf{p}) b_\ell(\mathsf{p})}{\overline{a_\ell(\mathsf{p})}^2} \,\mathrm{d}\mu(\mathsf{p}), \qquad i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}}, \tag{2.10c}$$

for $\ell = 1, 2, \ldots, r$ where the coefficients $\widehat{\alpha}_i$, $\widehat{\beta}_j$, and $\widehat{\gamma}_k$ are as defined in (2.3).

Proof. The right and left tangential conditions (2.10a) and (2.10b) follow, respectively, from conditions (2.8a) and (2.8b) of Corollary 2.2, using the facts that

$$\widehat{X}(\mathsf{p})^* = \widehat{\mathcal{B}}(\mathsf{p})^* \widehat{\mathcal{A}}(\mathsf{p})^{-*} = \begin{bmatrix} \frac{b_1(\mathsf{p})}{a_1(\mathsf{p})^*} & \cdots & \frac{b_r(\mathsf{p})}{a_r(\mathsf{p})^*} \end{bmatrix}, \quad \widehat{X}_d(\mathsf{p}) = \widehat{\mathcal{A}}(\mathsf{p})^{-*} \widehat{\mathcal{C}}(\mathsf{p})^* = \begin{bmatrix} \frac{c_1(\mathsf{p})^*}{a_1(\mathsf{p})^*} \\ \vdots \\ \frac{c_r(\mathsf{p})^*}{a_r(\mathsf{p})^*} \end{bmatrix},$$

since $\widehat{\mathcal{A}}(\mathsf{p})$ is diagonal. For the bitangential conditions (2.10c), we cannot directly use the condition (2.8c), but we can use the gradient (2.7a) together with Lemma 2.3, to conclude that the gradient with respect to the diagonal is the diagonal part of the gradient. Setting the diagonal entries to zeros gives us (2.10c). □

Theorem 2.1 and Corollary 2.4 will form the foundation of our analysis and allow us to derive the interpolatory optimality conditions for structured LTI systems by choosing the variables (such as $a_\ell(\mathsf{p})$, $\widehat{\alpha}_i(\mathsf{p})$ etc.) suitably based on the various LTI system structures we consider. In particular, these results will be used for the non-parametric structured $\mathcal{H}_2$ cases in Sections 3 to 6.

Remark 2.5. *The necessary optimality conditions in Corollary 2.4 can be interpreted as orthogonality conditions:*

$$\left\langle \frac{\widehat{\gamma}_k(\mathsf{p}) e_{\ell_2} b_\ell(\mathsf{p})^*}{a_\ell(\mathsf{p})}, H - \widehat{H} \right\rangle_{\mathcal{L}_2} = 0, \qquad k = 1, 2, \ldots, q_{\widehat{\mathcal{C}}},$$

$$\left\langle \frac{\widehat{\beta}_j(\mathsf{p}) c_\ell(\mathsf{p}) e_{\ell_2}^{\mathrm{T}}}{a_\ell(\mathsf{p})}, H - \widehat{H} \right\rangle_{\mathcal{L}_2} = 0, \qquad j = 1, 2, \ldots, q_{\widehat{\mathcal{B}}},$$

$$\left\langle \frac{\widehat{\alpha}_i(\mathsf{p})c_\ell(\mathsf{p})b_\ell(\mathsf{p})^*}{a_\ell(\mathsf{p})^2}, H - \widehat{H} \right\rangle_{\mathcal{L}_2} = 0, \qquad i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}},$$

for $\ell, \ell_2 = 1, 2, \ldots, r$. This directly generalizes the same property for unstructured LTI systems.

3 Interpolatory $\mathcal{H}_2$-optimality Conditions for Diagonal Structured Dynamics

We now turn our attention back to non-parametric (yet structured) LTI systems where, in the language of Section 2, the parameter space is $\mathcal{P} = \imath\mathbb{R}$ and the parameter is $\mathsf{p} = s = \imath\omega$. For simplicity, we only consider LTI systems with s-independent $\widehat{\mathcal{B}}$ and $\widehat{\mathcal{C}}$. Additionally, as in Section 2.4, we are interested in LTI systems with diagonal structure.

Recall that the well-established interpolatory optimality conditions (1.5) for unstructured LTI systems (1.2) assume diagonalizability of $\widehat{E}^{-1}\widehat{A}$, which leads to the pole-residue form (1.4) with the denominator of the simple form $s - \lambda_i$. Here, we derive the interpolatory $\mathcal{H}_2$-optimality conditions for structured transfer functions with a much more general (structured) denominator, which will be needed for particular types of structured LTI systems in later sections.

The pole-residue decomposition of a rational function plays the fundamental role in deriving the $\mathcal{H}_2$-optimality conditions for the unstructured problem. For structured problems, the residue computations are more involved. The next result will be useful in residue computation and then in simplifying the integrals in Corollary 2.4 for the structured problems under consideration.

Lemma 3.1. *Let g and h be analytic in a neighborhood of $c \in \mathbb{C}$ such that $h(c) = 0$ and $h'(c) \neq 0$. Define f_1 and f_2 as $f_1(z) = \frac{g(z)}{h(z)}$ and $f_2(z) = \frac{g(z)}{h(z)^2}$ in a neighborhood of c. Then*

$$\operatorname{Res}(f_1, c) = \frac{g(c)}{h'(c)} \quad \text{and} \quad \operatorname{Res}(f_2, c) = \frac{g'(c)}{h'(c)^2} - \frac{g(c)h''(c)}{h'(c)^3}. \tag{3.1}$$

Proof. See Appendix A for the proof. □

Applying Lemma 3.1 to Corollary 2.4 yields the following result.

Theorem 3.2. *Let the assumptions in Corollary 2.4 hold with $\mathcal{P} = \imath\mathbb{R}$ and $\widehat{\mathcal{B}}, \widehat{\mathcal{C}}$ being non-parametric. Let $\Lambda_\ell \subset \mathbb{C}_-$ be the set of zeros of a_ℓ where a_ℓ is the ℓth diagonal entry of $\widehat{\mathcal{A}}(s)$. Assume that all the zeros are simple and Λ_ℓ are pairwise disjoint. Further assume that the coefficients $\widehat{\alpha}_i$ in the parameter separable form (2.3) are analytic over $\mathbb{C}_+$ (and therefore so are a_ℓ). Then*

$$\widehat{H}(s) = \widehat{\mathcal{C}}\widehat{\mathcal{A}}(s)^{-1}\widehat{\mathcal{B}} = \sum_{\ell=1}^{r} \frac{c_\ell b_\ell^*}{a_\ell(s)} \tag{3.2}$$

and satisfies the interpolatory optimality conditions given by

$$\sum_{\lambda \in \Lambda_\ell} \frac{H\left(-\overline{\lambda}\right)b_\ell}{\overline{a_\ell'(\lambda)}} = \sum_{\lambda \in \Lambda_\ell} \frac{\widehat{H}\left(-\overline{\lambda}\right)b_\ell}{\overline{a_\ell'(\lambda)}}, \tag{3.3a}$$

$$\sum_{\lambda \in \Lambda_\ell} \frac{c_\ell^* H\left(-\overline{\lambda}\right)}{\overline{a_\ell'(\lambda)}} = \sum_{\lambda \in \Lambda_\ell} \frac{c_\ell^* \widehat{H}\left(-\overline{\lambda}\right)}{\overline{a_\ell'(\lambda)}}, \tag{3.3b}$$

$$\begin{aligned} &\sum_{\lambda \in \Lambda_\ell} c_\ell^* \left(\overline{\left(\frac{\widehat{\alpha}_i(\lambda)}{a_\ell'(\lambda)^2} \right)} H'\left(-\overline{\lambda}\right) - \overline{\left(\frac{\widehat{\alpha}_i'(\lambda)}{a_\ell'(\lambda)^2} - \frac{\widehat{\alpha}_i(\lambda)a_\ell''(\lambda)}{a_\ell'(\lambda)^3} \right)} H\left(-\overline{\lambda}\right) \right) b_\ell \\ &= \sum_{\lambda \in \Lambda_\ell} c_\ell^* \left(\overline{\left(\frac{\widehat{\alpha}_i(\lambda)}{a_\ell'(\lambda)^2} \right)} \widehat{H}'\left(-\overline{\lambda}\right) - \overline{\left(\frac{\widehat{\alpha}_i'(\lambda)}{a_\ell'(\lambda)^2} - \frac{\widehat{\alpha}_i(\lambda)a_\ell''(\lambda)}{a_\ell'(\lambda)^3} \right)} \widehat{H}\left(-\overline{\lambda}\right) \right) b_\ell. \end{aligned} \tag{3.3c}$$

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

Proof. The "pole-residue" form (3.2) directly follows from (2.9).

Then, the condition (2.10a) becomes

$$\int_{-\infty}^{\infty} \frac{H(\imath\omega)b_\ell}{\overline{a_\ell}(-\imath\omega)} \,\mathrm{d}\omega = \int_{-\infty}^{\infty} \frac{\widehat{H}(\imath\omega)b_\ell}{\overline{a_\ell}(-\imath\omega)} \,\mathrm{d}\omega,$$

where $\overline{a_\ell}(s) = \overline{a_\ell(\overline{s})}$ is analytic. Converting to a contour integral with $s = \imath\omega$, we obtain

$$\oint_{\imath\mathbb{R}} \frac{H(s)b_\ell}{\overline{a_\ell}(-s)} \,\mathrm{d}s = \oint_{\imath\mathbb{R}} \frac{\widehat{H}(s)b_\ell}{\overline{a_\ell}(-s)} \,\mathrm{d}s.$$

Since $H, \widehat{H}$ are analytic over $\mathbb{C}_+$ and $\overline{a_\ell}(-s)$ has poles at $-\overline{\Lambda_\ell}$, using the Residue Theorem and Lemma 3.1 gives

$$\sum_{\lambda\in\Lambda_\ell} \frac{H(-\overline{\lambda})b_\ell}{\overline{a_\ell}'(\overline{\lambda})} = \sum_{\lambda\in\Lambda_\ell} \frac{\widehat{H}(-\overline{\lambda})b_\ell}{\overline{a_\ell}'(\overline{\lambda})},$$

which implies condition (3.3a). Condition (3.3b) is obtained in a similar way.

Condition (2.10c), after substitution, becomes

$$\oint_{\imath\mathbb{R}} \frac{\overline{\alpha_i}(-s)c_\ell^* H(s)b_\ell}{\overline{a_\ell}(-s)^2} \,\mathrm{d}s = \oint_{\imath\mathbb{R}} \frac{\overline{\alpha_i}(-s)c_\ell^* \widehat{H}(s)b_\ell}{\overline{a_\ell}(-s)^2} \,\mathrm{d}s, \qquad i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}},$$

where $\overline{\alpha_i}(s) = \overline{\alpha_i(\overline{s})}$ is analytic. Using the Residue Theorem and Lemma 3.1 gives

$$\begin{aligned}
&\sum_{\lambda\in\Lambda_\ell} c_\ell^* \left(\frac{-\overline{\alpha_i}'(\overline{\lambda})H(-\overline{\lambda}) + \overline{\alpha_i}(\overline{\lambda})H'(-\overline{\lambda})}{\overline{a_\ell}'(\overline{\lambda})^2} + \frac{\overline{\alpha_i}(\overline{\lambda})\overline{a_\ell}''(\overline{\lambda})H(-\overline{\lambda})}{\overline{a_\ell}'(\overline{\lambda})^3} \right) b_\ell \\
&= \sum_{\lambda\in\Lambda_\ell} c_\ell^* \left(\frac{-\overline{\alpha_i}'(\overline{\lambda})\widehat{H}(-\overline{\lambda}) + \overline{\alpha_i}(\overline{\lambda})\widehat{H}'(-\overline{\lambda})}{\overline{a_\ell}'(\overline{\lambda})^2} + \frac{\overline{\alpha_i}(\overline{\lambda})\overline{a_\ell}''(\overline{\lambda})\widehat{H}(-\overline{\lambda})}{\overline{a_\ell}'(\overline{\lambda})^3} \right) b_\ell,
\end{aligned}$$

for $i = 1, 2, \ldots, q_{\widehat{\mathcal{A}}}$. After simplifying,

$$\begin{aligned}
&\sum_{\lambda\in\Lambda_\ell} c_\ell^* \left(\frac{\overline{\alpha_i}(\overline{\lambda})H'(-\overline{\lambda})}{\overline{a_\ell}'(\overline{\lambda})^2} - \left(\frac{\overline{\alpha_i}'(\overline{\lambda})}{\overline{a_\ell}'(\overline{\lambda})^2} - \frac{\overline{\alpha_i}(\overline{\lambda})\overline{a_\ell}''(\overline{\lambda})}{\overline{a_\ell}'(\overline{\lambda})^3} \right) H(-\overline{\lambda}) \right) b_\ell \\
&= \sum_{\lambda\in\Lambda_\ell} c_\ell^* \left(\frac{\overline{\alpha_i}(\overline{\lambda})\widehat{H}'(-\overline{\lambda})}{\overline{a_\ell}'(\overline{\lambda})^2} - \left(\frac{\overline{\alpha_i}'(\overline{\lambda})}{\overline{a_\ell}'(\overline{\lambda})^2} - \frac{\overline{\alpha_i}(\overline{\lambda})\overline{a_\ell}''(\overline{\lambda})}{\overline{a_\ell}'(\overline{\lambda})^3} \right) \widehat{H}(-\overline{\lambda}) \right) b_\ell.
\end{aligned}$$

This gives condition (3.3c). ☐

We note that in the classical unstructured case (1.2), we have $a_\ell(s) = s - \lambda_\ell$. Since $a_\ell'(s) = 1$ and $a_\ell''(s) = 0$, and we see that Theorem 3.2 recovers the known result (1.5) as a special case. Thus, we have extended the well-known interpolatory $\mathcal{H}_2$-optimality conditions for classical LTI systems with the pole-residue form (1.4) to structured LTI systems with a generalized "pole-residue"-like form (3.2), allowing structured forms to be encoded in the denominator $a_\ell(s)$. In the following sections, by adjusting $a_\ell(s)$ to the structures of interest, we will develop new interpolatory $\mathcal{H}_2$ optimality conditions for important classes of structured LTI systems.

4 Second-order Systems

In this section, we consider a prominent class of StROMs, namely the LTI second-order systems of the form

$$\widehat{M}\ddot{\widehat{x}}(t) + \widehat{E}\dot{\widehat{x}}(t) + \widehat{K}\widehat{x}(t) = \widehat{B}u(t), \tag{4.1a}$$

$$\widehat{y}(t) = \widehat{C}\widehat{x}(t), \tag{4.1b}$$

where $\widehat{M}, \widehat{E}, \widehat{K} \in \mathbb{R}^{n\times n}$ are, respectively, the mass, damping, and stiffness matrices; $\widehat{B} \in \mathbb{R}^{n\times n_{\mathrm{i}}}$ is the input-to-state map; $\widehat{C} \in \mathbb{R}^{n_{\mathrm{o}}\times n}$ is the state-to-output map; $\widehat{x}(t) \in \mathbb{R}^n$ is the internal state; $u(t) \in \mathbb{R}^{n_{\mathrm{i}}}$ are the inputs, and $\widehat{y}(t) \in \mathbb{R}^{n_o}$ are outputs. These systems appear frequently, e.g., in analyzing mechanical or electrical systems (see, e.g., [RS08] and references within). The transfer function of (4.1) is given by

$$\widehat{H}(s) = \widehat{C}\left(s^2\widehat{M} + s\widehat{E} + \widehat{K}\right)^{-1}\widehat{B}.$$

We assume that the mass matrix $\widehat{M}$ is invertible and that the matrix pencil $\lambda^2\widehat{M} + \lambda\widehat{E} + \widehat{K}$ is asymptotically stable. These assumptions are sufficient for (2.5) and (2.6) to hold. Again, the transfer function H of the FOM can be associated to a finite-dimensional, second-order systems, but that is not necessary in the following (it is enough that it has a finite $\mathcal{H}_2$ norm).

There are various approaches to model reduction of second-order systems using different error measures. We refer the reader to [Wer21, SSW19] for an overview. Here our focus is on interpolatory methods, more precisely in establishing the interpolatory conditions for minimizing the $\mathcal{H}_2$ error (1.3).

Structure-preserving interpolatory reduced-order modeling of second-order systems has been studied in detail, see, e.g., [BFSZ16, BG09, CGVV05, BS05, Bai02, BG05] and the references therein. Inspired by IRKA [GAB08] for $\mathcal{H}_2$-optimal reduced-order modeling of unstructured systems, Wyatt [Wya12] proposed several iterative methods for reducing second-order dynamics with a focus on the $\mathcal{H}_2$ norm. However, the second-order $\mathcal{H}_2$-optimality conditions were not established and thus the resulting reduced models did not satisfy the true optimality conditions. The interpolatory $\mathcal{H}_2$-optimality conditions for reducing second-order systems were derived by Beattie and Benner [BB14] where the ROM was assumed to be modally-damped, i.e., $\widehat{M}, \widehat{E}, \widehat{K}$ are symmetric positive definite and $\widehat{M}^{-1}\widehat{E}$ commutes with $\widehat{M}^{-1}\widehat{K}$.

Starting with the general optimality conditions in Corollary 2.4 and by appropriately choosing the parameters to reflect the second-order dynamics, we first derive the interpolatory optimality conditions for $\mathcal{H}_2$-optimal reduced-order modeling of second-order systems for the special case of modally-damped systems. In other words, we show that the optimality conditions of [BB14, Section 5] follow as a special case of our general framework from Theorem 3.2. These interpolatory conditions involve mixed terms and do not follow the optimal bitangential Hermite interpolation formulation of the unstructured case (1.5). But we then show that these conditions can be interpreted as a bitangential Hermite interpolation of not the original transfer function H but of a modified multivariate transfer function; thus showing that the classical bitangential Hermite interpolation forms the optimality in the second-order dynamics case as well.

As in [BB14], we assume that the reduced second-order model is modally damped; in other words $\widehat{M}, \widehat{E}, \widehat{K}$ are symmetric positive definite and $\widehat{M}^{-1}\widehat{E}$ and $\widehat{M}^{-1}\widehat{K}$ are simultaneously diagonalizable. Therefore, using a state space transformation, the reduced model can be brought to a form with $\widehat{M} = I$ and $\widehat{E}, \widehat{K}$ are real diagonal, thus making $\widehat{H}$ a D-StROM. In particular, the matrix $\lambda^2\widehat{M} + \lambda\widehat{E} + \widehat{K}$ is diagonal with quadratic polynomials on its diagonal, and thus can be decomposed into $(\lambda I - \Lambda^+)(\lambda I - \Lambda^-)$ for some complex diagonal matrices Λ^+ and Λ^-. This is precisely what we use in the next result.

Theorem 4.1. *Let (4.1) be a locally $\mathcal{H}_2$-optimal ROM with $\widehat{M} = I$ and $\widehat{E}, \widehat{K}$ are real diagonal matrices. Let $\Lambda^+ = \operatorname{diag}(\lambda_i^+)$ and $\Lambda^- = \operatorname{diag}(\lambda_i^-)$ be complex diagonal matrices such that $\widehat{E} = -(\Lambda^+ + \Lambda^-)$ and $\widehat{K} = \Lambda^+\Lambda^-$. Additionally, let all λ_i^+ and λ_j^- be pairwise distinct. Define $c_i = \widehat{C}e_i$ and $b_i = \widehat{B}^{\mathrm{T}}e_i$. Then the transfer function $\widehat{H}$ of the ROM can be written as*

$$\widehat{H}(s) = \widehat{C}\left(s^2\widehat{M} + s\widehat{E} + \widehat{K}\right)^{-1}\widehat{B} = \sum_{i=1}^{r} \frac{c_i b_i^{\mathrm{T}}}{\left(s - \lambda_i^+\right)\left(s - \lambda_i^-\right)} \tag{4.2}$$

and satisfies the interpolatory optimality conditions

$$\left(H\left(-\overline{\lambda_i^+}\right) - H\left(-\overline{\lambda_i^-}\right)\right)b_i = \left(\widehat{H}\left(-\overline{\lambda_i^+}\right) - \widehat{H}\left(-\overline{\lambda_i^-}\right)\right)b_i, \tag{4.3a}$$

$$c_i^{\mathrm{T}}\left(H\left(-\overline{\lambda_i^+}\right) - H\left(-\overline{\lambda_i^-}\right)\right) = c_i^{\mathrm{T}}\left(\widehat{H}\left(-\overline{\lambda_i^+}\right) - \widehat{H}\left(-\overline{\lambda_i^-}\right)\right), \tag{4.3b}$$

$$c_i^{\mathrm{T}}H'\left(-\overline{\lambda_i^+}\right)b_i = c_i^{\mathrm{T}}\widehat{H}'\left(-\overline{\lambda_i^+}\right)b_i, \tag{4.3c}$$

$$c_i^{\mathrm{T}}H'\left(-\overline{\lambda_i^-}\right)b_i = c_i^{\mathrm{T}}\widehat{H}'\left(-\overline{\lambda_i^-}\right)b_i, \tag{4.3d}$$

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

Proof. We start by using $s^2\widehat{M} + s\widehat{E} + \widehat{K} = (sI - \Lambda^+)(sI - \Lambda^-)$ to obtain

$$\begin{aligned}\widehat{H}(s) &= \widehat{C}\left(s^2\widehat{M} + s\widehat{E} + \widehat{K}\right)^{-1}\widehat{B} = \widehat{C}\left(sI - \Lambda^+\right)^{-1}\left(sI - \Lambda^-\right)^{-1}\widehat{B} \\ &= \sum_{i=1}^{r} \frac{c_i b_i^{\mathrm{T}}}{\left(s - \lambda_i^+\right)\left(s - \lambda_i^-\right)},\end{aligned}$$

which proves (4.2). The more general form in (2.3) recovers the transfer function $\widehat{H}(s) = \widehat{C}(s^2\widehat{M} + s\widehat{E} + \widehat{K})^{-1}\widehat{B}$ by defining $\mathsf{p} = s$ and

$$\begin{gathered}\widehat{\mathcal{A}}(s) = s^2\widehat{M} + s\widehat{E} + \widehat{K}, \quad \widehat{\mathcal{B}}(s) = \widehat{B}, \quad \widehat{\mathcal{C}}(s) = \widehat{C}, \\ \widehat{\alpha}_1(s) = s^2, \quad \widehat{\alpha}_2(s) = s, \quad \widehat{\alpha}_3(s) = 1, \quad \widehat{\beta}_1(s) = 1, \quad \widehat{\gamma}_1(s) = 1.\end{gathered}$$

The $\mathcal{L}_2(\mathcal{P}, \mu)$ norm recovers the $\mathcal{H}_2$-norm with the choices of $\mathcal{P} = \imath\mathbb{R}$ and Lebesgue measure over the imaginary axis $\mu = \lambda_{\imath\mathbb{R}}$. Therefore, we can apply Theorem 3.2 with these choices to recover explicit optimality conditions where $a_i(s) = (s - \lambda_i^+)(s - \lambda_i^-)$. Note that $a_i'(s) = (s - \lambda_i^+) + (s - \lambda_i^-)$ and $a_i''(s) = 2$. To simplify the notation, let us denote $\kappa_i = \frac{1}{\lambda_i^+ - \lambda_i^-}$.

From condition (3.3a) in Theorem 3.2, we obtain

$$\left(\overline{\kappa_i}H\left(-\overline{\lambda_i^+}\right) - \overline{\kappa_i}H\left(-\overline{\lambda_i^-}\right)\right)b_i = \left(\overline{\kappa_i}\widehat{H}\left(-\overline{\lambda_i^+}\right) - \overline{\kappa_i}\widehat{H}\left(-\overline{\lambda_i^-}\right)\right)b_i.$$

Dividing by $\overline{\kappa_i}$ gives (4.3a). Similarly, condition (3.3b) in Theorem 3.2 yields (4.3b).

Using condition (3.3c) with $\widehat{\alpha}_3$ (corresponding to $\widehat{K}$) in Theorem 3.2, we obtain

$$\begin{aligned}&c_i^{\mathrm{T}}\left(\overline{\kappa_i}^2H'\left(-\overline{\lambda_i^+}\right) - 2\overline{\kappa_i}^3H\left(-\overline{\lambda_i^+}\right) + \overline{\kappa_i}^2H'\left(-\overline{\lambda_i^-}\right) + 2\overline{\kappa_i}^3H\left(-\overline{\lambda_i^-}\right)\right)b_i \\ &= c_i^{\mathrm{T}}\left(\overline{\kappa_i}^2\widehat{H}'\left(-\overline{\lambda_i^+}\right) - 2\overline{\kappa_i}^3\widehat{H}\left(-\overline{\lambda_i^+}\right) + \overline{\kappa_i}^2\widehat{H}'\left(-\overline{\lambda_i^-}\right) + 2\overline{\kappa_i}^3\widehat{H}\left(-\overline{\lambda_i^-}\right)\right)b_i.\end{aligned}$$

Dividing by $\overline{\kappa_i}^2$ gives

$$\begin{aligned}&c_i^{\mathrm{T}}\left(\left(H'\left(-\overline{\lambda_i^+}\right) + H'\left(-\overline{\lambda_i^-}\right)\right) - 2\overline{\kappa_i}\left(H\left(-\overline{\lambda_i^+}\right) - H\left(-\overline{\lambda_i^-}\right)\right)\right)b_i \\ &= c_i^{\mathrm{T}}\left(\left(\widehat{H}'\left(-\overline{\lambda_i^+}\right) + \widehat{H}'\left(-\overline{\lambda_i^-}\right)\right) - 2\overline{\kappa_i}\left(\widehat{H}\left(-\overline{\lambda_i^+}\right) - \widehat{H}\left(-\overline{\lambda_i^-}\right)\right)\right)b_i.\end{aligned}$$

Since the third and fourth terms cancel out using (4.3a), this simplifies to

$$c_i^{\mathrm{T}}H'\left(-\overline{\lambda_i^+}\right)b_i + c_i^{\mathrm{T}}H'\left(-\overline{\lambda_i^-}\right)b_i = c_i^{\mathrm{T}}\widehat{H}'\left(-\overline{\lambda_i^+}\right)b_i + c_i^{\mathrm{T}}\widehat{H}'\left(-\overline{\lambda_i^-}\right)b_i. \tag{4.4}$$

From condition (3.3c) with $\widehat{\alpha}_2$ (corresponding to $\widehat{E}$) in Theorem 3.2, we obtain (using $\widehat{\alpha}_2(s) = s$) using the Residue Theorem and Lemma 3.1

$$\begin{aligned}&\overline{\lambda_i^+}\,\overline{\kappa_i}^2c_i^{\mathrm{T}}H'\left(-\overline{\lambda_i^+}\right)b_i - \left(\overline{\kappa_i}^2 - 2\overline{\lambda_i^+}\,\overline{\kappa_i}^3\right)c_i^{\mathrm{T}}H\left(-\overline{\lambda_i^+}\right)b_i \\ &+ \overline{\lambda_i^-}\,\overline{\kappa_i}^2c_i^{\mathrm{T}}H'\left(-\overline{\lambda_i^-}\right)b_i - \left(\overline{\kappa_i}^2 + 2\overline{\lambda_i^-}\,\overline{\kappa_i}^3\right)c_i^{\mathrm{T}}H\left(-\overline{\lambda_i^-}\right)b_i \\ &= \overline{\lambda_i^+}\,\overline{\kappa_i}^2c_i^{\mathrm{T}}\widehat{H}'\left(-\overline{\lambda_i^+}\right)b_i - \left(\overline{\kappa_i}^2 - 2\overline{\lambda_i^+}\,\overline{\kappa_i}^3\right)c_i^{\mathrm{T}}\widehat{H}\left(-\overline{\lambda_i^+}\right)b_i \\ &\quad + \overline{\lambda_i^-}\,\overline{\kappa_i}^2c_i^{\mathrm{T}}\widehat{H}'\left(-\overline{\lambda_i^-}\right)b_i - \left(\overline{\kappa_i}^2 + 2\overline{\lambda_i^-}\,\overline{\kappa_i}^3\right)c_i^{\mathrm{T}}\widehat{H}\left(-\overline{\lambda_i^-}\right)b_i.\end{aligned}$$

Dividing by $\overline{\kappa_i}^3$ gives

$$\begin{aligned}&\left(\overline{\lambda_i^+} - \overline{\lambda_i^-}\right)\overline{\lambda_i^+}c_i^{\mathrm{T}}H'\left(-\overline{\lambda_i^+}\right)b_i + \left(\overline{\lambda_i^+} + \overline{\lambda_i^-}\right)c_i^{\mathrm{T}}H\left(-\overline{\lambda_i^+}\right)b_i \\ &+ \left(\overline{\lambda_i^+} - \overline{\lambda_i^-}\right)\overline{\lambda_i^-}c_i^{\mathrm{T}}H'\left(-\overline{\lambda_i^-}\right)b_i - \left(\overline{\lambda_i^+} + \overline{\lambda_i^-}\right)c_i^{\mathrm{T}}H\left(-\overline{\lambda_i^-}\right)b_i \\ &= \left(\overline{\lambda_i^+} - \overline{\lambda_i^-}\right)\overline{\lambda_i^+}c_i^{\mathrm{T}}\widehat{H}'\left(-\overline{\lambda_i^+}\right)b_i + \left(\overline{\lambda_i^+} + \overline{\lambda_i^-}\right)c_i^{\mathrm{T}}\widehat{H}\left(-\overline{\lambda_i^+}\right)b_i \\ &\quad + \left(\overline{\lambda_i^+} - \overline{\lambda_i^-}\right)\overline{\lambda_i^-}c_i^{\mathrm{T}}\widehat{H}'\left(-\overline{\lambda_i^-}\right)b_i - \left(\overline{\lambda_i^+} + \overline{\lambda_i^-}\right)c_i^{\mathrm{T}}\widehat{H}\left(-\overline{\lambda_i^-}\right)b_i.\end{aligned}$$

Since the second and fourth terms above cancel out using (4.3a), this simplifies to

$$\overline{\lambda_i^+} c_i^{\mathrm{T}} H'\left(-\overline{\lambda_i^+}\right) b_i + \overline{\lambda_i^-} c_i^{\mathrm{T}} H'\left(-\overline{\lambda_i^-}\right) b_i = \overline{\lambda_i^+} c_i^{\mathrm{T}} \widehat{H}'\left(-\overline{\lambda_i^+}\right) b_i + \overline{\lambda_i^-} c_i^{\mathrm{T}} \widehat{H}'\left(-\overline{\lambda_i^-}\right) b_i. \tag{4.5}$$

Note that (4.4) and (4.5) can be merged together to give

$$\begin{bmatrix} 1 & 1 \\ \overline{\lambda_i^+} & \overline{\lambda_i^-} \end{bmatrix} \begin{bmatrix} c_i^{\mathrm{T}} H'\left(-\overline{\lambda_i^+}\right) b_i \\ c_i^{\mathrm{T}} H'\left(-\overline{\lambda_i^-}\right) b_i \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \overline{\lambda_i^+} & \overline{\lambda_i^-} \end{bmatrix} \begin{bmatrix} c_i^{\mathrm{T}} \widehat{H}'\left(-\overline{\lambda_i^+}\right) b_i \\ c_i^{\mathrm{T}} \widehat{H}'\left(-\overline{\lambda_i^-}\right) b_i \end{bmatrix}.$$

Since we assumed that $\lambda_i^+ \neq \lambda_i^-$, it follows that

$$c_i^{\mathrm{T}} H'\left(-\overline{\lambda_i^+}\right) b_i = c_i^{\mathrm{T}} \widehat{H}'\left(-\overline{\lambda_i^+}\right) b_i \quad \text{and} \quad c_i^{\mathrm{T}} H'\left(-\overline{\lambda_i^-}\right) b_i = c_i^{\mathrm{T}} \widehat{H}'\left(-\overline{\lambda_i^-}\right) b_i,$$

proving (4.3c) and (4.3d), and completing the proof. □

Therefore, our general framework for optimality as given in Theorem 3.2 and Corollary 2.2 recovers the interpolatory optimality conditions from [BB14] as a special case. Indeed, our framework is able to derive the interpolation conditions for the case of additional velocity measurements as well, i.e., $\widehat{y}(t) = C_p\widehat{x}(t) + C_v\dot{\widehat{x}}(t)$, which was not covered in [BB14]. However, the resulting equations are rather tedious and for the sake of conciseness, are not included here.

At a first glance, the interpolatory optimality conditions (4.3) for second-order structures are different from those of the classical bitangential Hermite interpolation conditions of the unstructured $\mathcal{H}_2$ approximation problem (1.5). More specifically, even though the bitangential Hermite conditions (4.3c)–(4.3d) resemble the classical case, the left- and right-tangential Lagrange conditions (4.3a)–(4.3b) are rather different since they impose interpolating a difference as opposed to the original transfer function. However, these new interpolatory conditions can still be interpreted as bitangential Hermite conditions for a modified transfer function as we show next.

Corollary 4.2. *Let the assumptions in Theorem 4.1 hold. Define the 2D full-order and reduced-order transfer functions*

$$G(s_1, s_2) = H(s_1) - H(s_2) \quad \textit{and} \quad \widehat{G}(s_1, s_2) = \widehat{H}(s_1) - \widehat{H}(s_2). \tag{4.6}$$

Then, the optimality conditions (4.3a)–(4.3d) are, respectively, equivalent to

$$G\left(-\overline{\lambda_i^+}, -\overline{\lambda_i^-}\right) b_i = \widehat{G}\left(-\overline{\lambda_i^+}, -\overline{\lambda_i^-}\right) b_i, \tag{4.7a}$$

$$c_i^{\mathrm{T}} G\left(-\overline{\lambda_i^+}, -\overline{\lambda_i^-}\right) = c_i^{\mathrm{T}} \widehat{G}\left(-\overline{\lambda_i^+}, -\overline{\lambda_i^-}\right), \tag{4.7b}$$

$$c_i^{\mathrm{T}} \frac{\partial G}{\partial s_1}\left(-\overline{\lambda_i^+}, -\overline{\lambda_i^-}\right) b_i = c_i^{\mathrm{T}} \frac{\partial \widehat{G}}{\partial s_1}\left(-\overline{\lambda_i^+}, -\overline{\lambda_i^-}\right) b_i, \tag{4.7c}$$

$$c_i^{\mathrm{T}} \frac{\partial G}{\partial s_2}\left(-\overline{\lambda_i^+}, -\overline{\lambda_i^-}\right) b_i = c_i^{\mathrm{T}} \frac{\partial \widehat{G}}{\partial s_2}\left(-\overline{\lambda_i^+}, -\overline{\lambda_i^-}\right) b_i, \tag{4.7d}$$

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

Proof. The condition (4.7a) and (4.7b) directly follow from, respectively, (4.3a) and (4.3b) based on the definitions of G and $\widehat{G}$ in (4.6). Also note that $\frac{\partial G}{\partial s_1}(s_1, s_2) = H'(s_1)$ and $\frac{\partial G}{\partial s_2}(s_1, s_2) = -H'(s_2)$; and similarly for $\widehat{G}$. These observations immediately reveal that (4.7c) and (4.7d) are equivalent to, respectively, (4.3c) and (4.3d). □

Therefore, as in the unstructured case, $\mathcal{H}_2$-optimal reduced-order modeling of second-order systems requires bitangential Hermite interpolation. The optimal interpolation points are still the mirror images of the reduced-order poles and the tangential directions result from the residues of the ROM. However, what needs to be interpolated is a modified 2D transfer function G. In our earlier work [MG23b], we showed that in addition to the classical $\mathcal{H}_2$-optimal approximation problem, the bitangential Hermite interpolation formed the necessary conditions for optimality for the rational nonlinear least-squares fitting of LTI systems and for the $\mathcal{L}_2$-optimal approximation of stationary problems. In both cases, the interpolation had to be enforced on a modified mapping rather than the original one. Corollary 4.2 proves this to be the case for structure-preserving $\mathcal{H}_2$-optimal approximation of second-order LTI systems as well. Thus, bitangential Hermite interpolation remains the unifying framework even for a larger class of problems.

5 Port-Hamiltonian Systems

LTI pH systems naturally arise in modeling a wide range of physical, engineering, and biological systems. They generalize the classical Hamiltonian structure by including inputs and outputs, thus allowing interaction with the environment via their *ports*. The pH systems are important in energy-based modeling as their form ensures passivity and allows energy-preserving interconnections. We refer the reader to, e.g., [vdSJ14, MU22, JZ12], for more details on pH systems.

Thus, given an LTI system H, in this section we consider StROMs that have the LTI pH form

$$\dot{\widehat{x}}(t) = \left(\widehat{J} - \widehat{R}\right)\widehat{x}(t) + \widehat{B}u(t), \tag{5.1a}$$

$$\widehat{y}(t) = \widehat{B}^{\mathrm{T}}\widehat{x}(t), \tag{5.1b}$$

where $\widehat{J}, \widehat{R} \in \mathbb{R}^{r\times r}$, $\widehat{B} \in \mathbb{R}^{r\times n_i}$, $\widehat{J}^{\mathrm{T}} = -\widehat{J}$, and $\widehat{R} \succcurlyeq 0$[5]. Additionally, we assume $\widehat{J} - \widehat{R}$ is asymptotically stable, which is guaranteed when $\widehat{R} \succ 0$.

Given an LTI system H, our goal is to find a pH ROM as in (5.1) that minimizes the $\mathcal{H}_2$ error (1.3). Several approaches have been proposed for this problem. In [GPBvdS12], the authors propose an iterative algorithm similar to IRKA, called IRKA-PH, which finds a ROM that, upon convergence, satisfies one, namely (1.5a), out of the three necessary conditions for *unstructured* first-order systems (1.5). However, it is important to point out that the condition (1.5a) that IRKA-PH satisfies does not correspond to true optimality conditions for a pH system; it simply works with the conditions for the unstructured case. Deriving the true structured optimality conditions for pH systems is the main goal of this section. Exploiting the minimal solution of an algebraic Riccati equation related to pH systems, Breiten and Unger [BU22] significantly improved the performance of IRKA-PH. Optimal-$\mathcal{H}_2$ model reduction with pH structure was also considered, e.g., in [ML20, SMMV22, MSMV22] where gradient-based optimization with structure preservation is used to construct the ROMs.

These aforementioned methods provide high-quality ROMs with pH structure, yet they do not derive or work with interpolatory optimality conditions, which is our main focus here. Interpolatory $\mathcal{H}_2$ conditions with pH structure were first studied in [BB14] where the authors derived (a subset) of the necessary conditions for $\mathcal{H}_2$-optimality, shown in (5.2) below. We first show in Theorem 5.1 that these conditions follow from our general framework, Corollary 2.2, directly. But more importantly what this derivation will reveal is that the interpolatory conditions in [BB14] are not complete in the sense that they correspond to only (2.8c) in Corollary 2.2 (corresponding to the gradient with respect to $\widehat{J} - \widehat{R}$) without the other necessary conditions. Then based on this observation, in Theorem 5.2, we provide the remaining (realization-dependent) conditions for $\mathcal{H}_2$-optimal model reduction with pH structure. With an additional structural assumption, in Theorem 5.3, we provide a realization-independent set of interpolatory optimality conditions.

Theorem 5.1. *Suppose that $\widehat{H}$ is an $\mathcal{H}_2$-optimal ROM with $\widehat{R} \succ 0$, has r distinct poles, and is represented as $\widehat{H}(s) = \sum_{i=1}^{r} \frac{c_i b_i^*}{s-\lambda_i}$. Then*

$$c_i^*\left(H(-\overline{\lambda_i}) - H(-\overline{\lambda_j})\right)b_j = c_i^*\left(\widehat{H}(-\overline{\lambda_i}) - \widehat{H}(-\overline{\lambda_j})\right)b_j, \tag{5.2a}$$

$$c_i^* H'(-\overline{\lambda_i})b_i = c_i^* \widehat{H}'(-\overline{\lambda_i})b_i, \tag{5.2b}$$

for $i, j = 1, 2, \ldots, r$.

Proof. The matrix $\widehat{J} - \widehat{R}$ satisfies $(\widehat{J} - \widehat{R}) + (\widehat{J} - \widehat{R})^{\mathrm{T}} = -2\widehat{R} \prec 0$. Conversely, any matrix $\widehat{A}$ such that $\widehat{A} + \widehat{A}^{\mathrm{T}} \prec 0$ can be decomposed as $\widehat{A} = \widehat{J} - \widehat{R}$ where $\widehat{J}$ is skew-symmetric and $\widehat{R}$ is symmetric positive definite. Since the set of matrices $S = \{\widehat{A} \in \mathbb{R}^{r\times r} : \widehat{A} + \widehat{A}^{\mathrm{T}} \prec 0\}$ is open in $\mathbb{R}^{r\times r}$ and $\widehat{J} - \widehat{R} \in S$, the condition (2.8c) corresponding to the gradient with respect to $\widehat{J} - \widehat{R}$ holds, i.e.,

$$\int_{-\infty}^{\infty} \left(\imath\omega I - \widehat{J} + \widehat{R}\right)^{-*} \widehat{B}\left(H(\imath\omega) - \widehat{H}(\imath\omega)\right)\widehat{B}^{\mathrm{T}}\left(\imath\omega I - \widehat{J} + \widehat{R}\right)^{-*} \mathrm{d}\omega = 0. \tag{5.3}$$

[5]There are slightly more general forms of the pH systems. For the conciseness of the presentation, we focus on the form (5.1).

Let $\widehat{T} \in \mathbb{C}^{r\times r}$ be an invertible matrix such that $\widehat{T}^{-1}(\widehat{J}-\widehat{R})\widehat{T} = \Lambda$. Then, after multiplying from the left by $\widehat{T}^*$ and from the right by $\widehat{T}^{-*}$, (5.3) becomes

$$\int_{-\infty}^{\infty} (\imath\omega I - \Lambda)^{-*}\widehat{T}^*\widehat{B}\Big(H(\imath\omega) - \widehat{H}(\imath\omega)\Big)\widehat{B}^{\mathrm{T}}\widehat{T}^{-*}(\imath\omega I - \Lambda)^{-*}\,\mathrm{d}\omega = 0. \tag{5.4}$$

Note that $b_j = \widehat{B}^{\mathrm{T}}\widehat{T}^{-*}e_j$ and $c_i^* = e_i^{\mathrm{T}}\widehat{T}^*\widehat{B}$. Multiplying (5.4) from the left by e_i^{T} and from the right by e_j gives

$$\int_{-\infty}^{\infty} \frac{c_i^*\Big(H(\imath\omega) - \widehat{H}(\imath\omega)\Big)b_j}{\big(-\imath\omega - \overline{\lambda_i}\big)\big(-\imath\omega - \overline{\lambda_j}\big)}\,\mathrm{d}\omega = 0.$$

Substituting $s = \imath\omega$ in this last equality yields

$$\oint_{\imath\mathbb{R}} \frac{c_i^*\Big(H(s) - \widehat{H}(s)\Big)b_j}{\big(s + \overline{\lambda_i}\big)\big(s + \overline{\lambda_j}\big)}\,\mathrm{d}s = 0. \tag{5.5}$$

When $i \neq j$, (5.5) immediately leads to the conditions in (5.2a). And finally, for $i = j$, we obtain (5.2b). □

Thus, we are able to recover the interpolatory conditions from [BB14] as a special case of our general framework. Additionally, and more importantly, the proof reveals that in the derivation of these conditions only the information in $\widehat{J}-\widehat{R}$ is used, but not in $\widehat{B}$. Thus, these conditions do not provide the full set of interpolatory conditions for optimality. To illustrate this, for simplicity, assume we have a SISO LTI system. Even though (5.2) seems to correspond to $r^2 + r$ conditions, that is not the case. If the interpolatory conditions

$$H\big(-\overline{\lambda_i}\big) - H\big(-\overline{\lambda_j}\big) = \widehat{H}\big(-\overline{\lambda_i}\big) - \widehat{H}\big(-\overline{\lambda_j}\big)$$

and

$$H\big(-\overline{\lambda_j}\big) - H\big(-\overline{\lambda_k}\big) = \widehat{H}\big(-\overline{\lambda_j}\big) - \widehat{H}\big(-\overline{\lambda_k}\big)$$

hold, then by adding them, we automatically obtain

$$H\big(-\overline{\lambda_i}\big) - H\big(-\overline{\lambda_k}\big) = \widehat{H}\big(-\overline{\lambda_i}\big) - \widehat{H}\big(-\overline{\lambda_k}\big).$$

Then, for the SISO case, one has a total $2r-1$ conditions, which is not enough to uniquely specify a rational function of order r. The following theorem provides the remaining conditions.

Theorem 5.2. *Let the assumptions in Theorem 5.1 hold. Furthermore, let $\widehat{T}$ be an eigenvector matrix of $\widehat{J}-\widehat{R}$ and denote $\widehat{T}e_i = t_i$, $\widehat{T}^{-*}e_i = s_i$, $\widehat{B}^{\mathrm{T}}t_i = c_i$, and, $\widehat{B}^{\mathrm{T}}s_i = b_i$. Then, in addition to (5.2), $\widehat{H}$ also satisfies*

$$\sum_{i=1}^{r}\Big(H\big(-\overline{\lambda_i}\big)b_i t_i^* + H\big(-\overline{\lambda_i}\big)^* c_i s_i^*\Big) = \sum_{i=1}^{r}\Big(\widehat{H}\big(-\overline{\lambda_i}\big)b_i t_i^* + \widehat{H}\big(-\overline{\lambda_i}\big)^* c_i s_i^*\Big). \tag{5.6}$$

Proof. The gradient with respect to $\widehat{B}$ follows from Theorem 2.1. In particular, since $\widehat{B}$ in the pH system (5.1) appears in both $\widehat{\mathcal{B}}$ and $\widehat{\mathcal{C}}$, the gradient of the squared $\mathcal{H}_2$ error with respect to $\widehat{B}$ is

$$\begin{aligned}
\nabla_{\overline{\widehat{B}}}\mathcal{J} &= \int_{-\infty}^{\infty}\Big(\imath\omega I - \widehat{J} + \widehat{R}\Big)^{-*}\widehat{B}\Big(\widehat{H}(\imath\omega) - H(\imath\omega)\Big)\,\mathrm{d}\omega \\
&\quad + \left(\int_{-\infty}^{\infty}\Big(\widehat{H}(\imath\omega) - H(\imath\omega)\Big)\widehat{B}^{\mathrm{T}}\Big(\imath\omega I - \widehat{J} + \widehat{R}\Big)^{-*}\,\mathrm{d}\omega\right)^{\mathrm{T}} \\
&= \int_{-\infty}^{\infty}\Big(\imath\omega I - \widehat{J} + \widehat{R}\Big)^{-*}\widehat{B}\Big(\widehat{H}(\imath\omega) - H(\imath\omega)\Big)\,\mathrm{d}\omega \\
&\quad + \int_{-\infty}^{\infty}\Big(-\imath\omega I - \widehat{J} + \widehat{R}\Big)^{-1}\widehat{B}\Big(\widehat{H}(\imath\omega)^{\mathrm{T}} - H(\imath\omega)^{\mathrm{T}}\Big)\,\mathrm{d}\omega.
\end{aligned}$$

Using $\widehat{T}^{-1}(\widehat{J}-\widehat{R})\widehat{T} = \Lambda$, the gradient becomes

$$\begin{aligned}
\nabla_{\overline{\widehat{B}}}\mathcal{J} &= \widehat{T}^{-*}\int_{-\infty}^{\infty}(\imath\omega I - \Lambda)^{-*}\widehat{T}^*\widehat{B}\Big(\widehat{H}(\imath\omega) - H(\imath\omega)\Big)\,\mathrm{d}\omega \\
&\quad + \widehat{T}\int_{-\infty}^{\infty}(-\imath\omega I - \Lambda)^{-1}\widehat{T}^{-1}\widehat{B}\Big(\widehat{H}(\imath\omega)^{\mathrm{T}} - H(\imath\omega)^{\mathrm{T}}\Big)\,\mathrm{d}\omega.
\end{aligned}$$

Using $I = \sum_{i=1}^{r} e_i e_i^{\mathrm{T}}$, we find

$$\begin{aligned}\nabla_{\overline{\widehat{B}}}\mathcal{J} &= \widehat{T}^{-*}\sum_{i=1}^{r} e_i e_i^{\mathrm{T}} \int_{-\infty}^{\infty} (\imath\omega I - \Lambda)^{-*}\widehat{T}^{*}\widehat{B}\Big(\widehat{H}(\imath\omega) - H(\imath\omega)\Big)\,\mathrm{d}\omega \\ &\quad + \widehat{T}\sum_{i=1}^{r} e_i e_i^{\mathrm{T}} \int_{-\infty}^{\infty} (-\imath\omega I - \Lambda)^{-1}\widehat{T}^{-1}\widehat{B}\Big(\widehat{H}(\imath\omega)^{\mathrm{T}} - H(\imath\omega)^{\mathrm{T}}\Big)\,\mathrm{d}\omega,\end{aligned}$$

which simplifies to

$$\begin{aligned}\nabla_{\overline{\widehat{B}}}\mathcal{J} &= \sum_{i=1}^{r} s_i c_i^{*} \int_{-\infty}^{\infty} \frac{\widehat{H}(\imath\omega) - H(\imath\omega)}{-\imath\omega - \overline{\lambda_i}}\,\mathrm{d}\omega \\ &\quad + \sum_{i=1}^{r} t_i b_i^{*} \int_{-\infty}^{\infty} \frac{\widehat{H}(\imath\omega)^{\mathrm{T}} - H(\imath\omega)^{\mathrm{T}}}{-\imath\omega - \lambda_i}\,\mathrm{d}\omega.\end{aligned}$$

Using the substitution $s = \imath\omega$ and the Cauchy integral formula, and equating the gradient to zero, we obtain

$$0 = \sum_{i=1}^{r} s_i c_i^{*}\Big(\widehat{H}(-\overline{\lambda_i}) - H(-\overline{\lambda_i})\Big) + \sum_{i=1}^{r} t_i b_i^{*}\Big(\widehat{H}(-\lambda_i)^{\mathrm{T}} - H(-\lambda_i)^{\mathrm{T}}\Big).$$

Applying conjugate transpose gives the tangential interpolatory condition (5.6). □

Now, the optimality conditions (5.2) in Theorem 5.1 together with the new optimality condition (5.6) in Theorem 5.2 describe *the full set of interpolatory conditions* for $\mathcal{H}_2$-optimal approximation of pH systems.

However, note the major difference in this new additional condition (5.6). In the earlier condition (5.2) (and the other conditions derived in the earlier sections), interpolation is realization-independent, i.e., it does not depend on a specific state-space form and only uses realization-independent quantities such as poles and residues. However, the new condition (5.6) is realization-dependent due to its dependence on the eigenvectors of $\widehat{J} - \widehat{R}$. To remove this dependence on eigenvectors, we require an additional assumption, namely the normality of $\widehat{J} - \widehat{R}$, and obtain a full set of realization-independent interpolatory optimality conditions.

Theorem 5.3. *Let the assumptions in Theorem 5.1 hold. Furthermore, let $\widehat{J} - \widehat{R}$ be normal. Then*

$$\widehat{H}(s) = \sum_{i=1}^{r} \frac{b_i b_i^{*}}{s - \lambda_i} \tag{5.7}$$

and satisfies the interpolatory optimality conditions

$$\Big(H(-\overline{\lambda_i}) + H(-\overline{\lambda_i})^{*}\Big)b_i = \Big(\widehat{H}(-\overline{\lambda_i}) + \widehat{H}(-\overline{\lambda_i})^{*}\Big)b_i, \tag{5.8a}$$

$$b_i^{*}H'(-\overline{\lambda_i})b_i = b_i^{*}\widehat{H}'(-\overline{\lambda_i})b_i, \tag{5.8b}$$

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

Proof. Since $\widehat{J} - \widehat{R}$ is assumed to be normal, the transformation matrix $\widehat{T}$ in the proof of Theorem 5.1 can be taken to be unitary, which leads to $c_i = b_i$ and thus to (5.7). The bitangential Hermite condition (5.8b) follows directly from (5.2b) in Theorem 5.1. The tangential Lagrange condition (5.8a) follows from (5.6), noting that $t_i = s_i$ since $\widehat{J} - \widehat{R}$ is normal, and multiplying (5.6) from the right by $\widehat{T}$. □

Therefore, using our general framework for $\mathcal{L}_2$-optimality and with the additional assumption of normality, we obtain more familiar bitangential Hermite interpolation conditions for pH systems, better mimicking the unstructured case. Moreover, we obtain a full set of conditions. Similar to second-order systems, we can reformulate these conditions as true bitangential Hermite conditions for a modified transfer function as we show next.

Corollary 5.4. *Let the assumptions in Theorem 5.3 hold. Define the transfer functions $G(s) = H(s) + H(s)^{*}$ and $\widehat{G}(s) = \widehat{H}(s) + \widehat{H}(s)^{*}$. Then*

$$G(-\overline{\lambda_i})b_i = \widehat{G}(-\overline{\lambda_i})b_i, \tag{5.9a}$$

$$b_i^* G(-\overline{\lambda_i}) = b_i^* \widehat{G}(-\overline{\lambda_i}), \tag{5.9b}$$

$$b_i^* \frac{\partial G}{\partial s}(-\overline{\lambda_i}) b_i = b_i^* \frac{\partial \widehat{G}}{\partial s}(-\overline{\lambda_i}) b_i, \tag{5.9c}$$

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

Proof. The right tangential Lagrange conditions (5.9a) follows directly from Theorem 5.3 and from the definitions of G and $\widehat{G}$. Due to symmetry, we also get the left tangential Lagrange conditions (5.9b). The bitangential Hermite condition (5.9c) also follows directly by using Wirtinger calculus. □

Thus once again, bitangential Hermite interpolation (of a modified transfer function) appears as the necessary condition for $\mathcal{H}_2$-optimality.

6 Time-delay Systems

Beyond standard LTI systems (1.1), dynamical systems with internal and/or input/output delays appear in many control systems (see, e.g., [Fri14]). An LTI time-delay system with a single internal delay has the state-space form

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + A_\tau x(t-\tau) + Bu(t), \\ y(t) &= Cx(t), \end{aligned}$$

where $\tau > 0$ is the internal delay. The system can contain more than one delay, but we focus on the single-delay case for simplicity. In these situations, where the original model to be approximated has a delay, it will be natural to consider a time-delay StROM of the form

$$\widehat{E}\dot{\widehat{x}}(t) = \widehat{A}\widehat{x}(t) + \widehat{A}_\tau \widehat{x}(t-\tau) + \widehat{B}u(t), \tag{6.1a}$$

$$\widehat{y}(t) = \widehat{C}\widehat{x}(t), \tag{6.1b}$$

for $\widehat{A}_\tau \in \mathbb{R}^{r\times r}$ and $\tau > 0$. We denote this system by $(\widehat{E}, \widehat{A}, \widehat{A}_\tau, \widehat{B}, \widehat{C})$. The transfer function of the time-delay StROM (6.1) is given by

$$\widehat{H}(s) = \widehat{C}\left(s\widehat{E} - \widehat{A} - e^{-\tau s}\widehat{A}_\tau\right)^{-1}\widehat{B}. \tag{6.2}$$

We assume that $\widehat{E}$ is invertible and $\lambda\widehat{E} - \widehat{A} - e^{-\tau\lambda}\widehat{A}_\tau$ is asymptotically stable. These assumptions are sufficient for (2.5) and (2.6) to hold. The goal is the same as before: find an optimal reduced time-delay system as in (6.2) that minimizes the $\mathcal{H}_2$ distance between the original transfer function H and the reduced $\widehat{H}$. Even though we assumed above that the full-order transfer function H also corresponds to a time-delay system, this is not needed analytically (although it is usually the case in practice).

The works [SGB16, AW23] proposed IRKA-type algorithms to construct StROMs, including time-delay systems. Even though both algorithms work well in practice, they do not guarantee $\mathcal{H}_2$-optimality and are not based on true $\mathcal{H}_2$-optimality conditions. An $\mathcal{H}_2$-optimal method for time-delay system is proposed in [GEMM19], but this approach uses gradient-based optimization based on the system Gramians and does not reveal or consider optimal interpolatory conditions. That is precisely what we establish in this section.

For SISO systems and for the simple case of $r = 1$ and $\widehat{A} = 0$ in (6.1), the interpolatory necessary $\mathcal{H}_2$-optimality conditions were derived in [PGB+16] and more generally by [Pon23]. Below using our generalized optimality framework from Theorem 3.2, we derive the interpolatory $\mathcal{H}_2$-optimality conditions for general MIMO time-delay systems for an arbitrary reduced order $r \geqslant 1$.

For the time-delay system (6.1), the diagonal structure that we assume for ROMs corresponds to assuming $\widehat{E}^{-1}\widehat{A}$ and $\widehat{E}^{-1}\widehat{A}_\tau$ are simultaneously diagonalizable. Therefore, there exist invertible matrices $\widehat{T}, \widehat{S} \in \mathbb{C}^{r\times r}$ such that $\widehat{S}^*\widehat{E}\widehat{T} = I$, $\widehat{S}^*\widehat{A}\widehat{T} = \widehat{M}$, and $\widehat{S}^*\widehat{A}_\tau\widehat{T} = \widehat{\Sigma}$, with $\widehat{M} = \operatorname{diag}(\mu_1, \mu_2, \ldots, \mu_r)$ and $\widehat{\Sigma} = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$. Then, the transfer function $\widehat{H}$ in (6.2) can be equivalently rewritten as

$$\begin{aligned} \widehat{H}(s) &= \widehat{C}\left(s\widehat{S}^{-*}\widehat{T}^{-1} - \widehat{S}^{-*}\widehat{M}\widehat{T}^{-1} - e^{-\tau s}\widehat{S}^{-*}\widehat{\Sigma}\widehat{T}^{-1}\right)^{-1}\widehat{B} \\ &= \widehat{C}\widehat{T}\left(sI - \widehat{M} - e^{-\tau s}\widehat{\Sigma}\right)^{-1}\widehat{S}^*\widehat{B} \end{aligned}$$

$$= \sum_{i=1}^{r} \frac{c_i b_i^*}{s - \mu_i - e^{-\tau s}\sigma_i}, \tag{6.3}$$

where $c_i = \widehat{C}\widehat{T}e_i$ and $b_i = \widehat{B}^{\mathrm{T}}\widehat{S}e_i$. To find the zeros of $s - \mu_i - e^{-\tau s}\sigma_i$, we can use the Lambert W function

$$\lambda_{ij} = \mu_i + \frac{1}{\tau} W_j\left(\tau\sigma_i e^{-\tau\mu_i}\right),$$

where λ_{ij} is the pole corresponding to the jth branch of the Lambert W function (see [CGM15]).

Note that the reformulation of $\widehat{H}$ as in (6.3) perfectly aligns with the form of transfer functions with general denominators in Theorem 3.2. The next result uses this observation and applies Theorem 3.2 to the transfer function (6.3) to develop the interpolatory necessary $\mathcal{H}_2$-optimality conditions for MIMO time-delay systems.

Theorem 6.1. *Let $H \in \mathcal{H}_2$ be real. Suppose that $(I, \widehat{M}, \widehat{\Sigma}, \widehat{B}, \widehat{C})$ is an $\mathcal{H}_2$-optimal diagonal time-delay system with the transfer function $\widehat{H}$ in (6.3). If $\widehat{H}$ has pairwise distinct simple poles, then*

$$\left(\sum_{j=-\infty}^{\infty} \overline{\phi_{ij}} H\left(-\overline{\lambda_{ij}}\right)\right) b_i = \left(\sum_{j=-\infty}^{\infty} \overline{\phi_{ij}} \widehat{H}\left(-\overline{\lambda_{ij}}\right)\right) b_i, \tag{6.4a}$$

$$c_i^*\left(\sum_{j=-\infty}^{\infty} \overline{\phi_{ij}} H\left(-\overline{\lambda_{ij}}\right)\right) = c_i^*\left(\sum_{j=-\infty}^{\infty} \overline{\phi_{ij}} \widehat{H}\left(-\overline{\lambda_{ij}}\right)\right), \tag{6.4b}$$

$$\begin{aligned} & c_i^*\left(\sum_{j=-\infty}^{\infty} \left(\overline{\psi_{ij}} H'\left(-\overline{\lambda_{ij}}\right) - \overline{\rho_{ij}} H\left(-\overline{\lambda_{ij}}\right)\right)\right) b_i \\ & = c_i^*\left(\sum_{j=-\infty}^{\infty} \left(\overline{\psi_{ij}} \widehat{H}'\left(-\overline{\lambda_{ij}}\right) - \overline{\rho_{ij}} \widehat{H}\left(-\overline{\lambda_{ij}}\right)\right)\right) b_i, \text{ and} \end{aligned} \tag{6.4c}$$

$$\begin{aligned} & c_i^*\left(\sum_{j=-\infty}^{\infty} \left(\left(\overline{\phi_{ij}} - \overline{\psi_{ij}}\right) H'\left(-\overline{\lambda_{ij}}\right) + \overline{\rho_{ij}} H\left(-\overline{\lambda_{ij}}\right)\right)\right) b_i \\ & = c_i^*\left(\sum_{j=-\infty}^{\infty} \left(\left(\overline{\phi_{ij}} - \overline{\psi_{ij}}\right) \widehat{H}'\left(-\overline{\lambda_{ij}}\right) + \overline{\rho_{ij}} \widehat{H}\left(-\overline{\lambda_{ij}}\right)\right)\right) b_i, \end{aligned} \tag{6.4d}$$

for $i = 1, 2, \ldots, r$, where

$$\phi_{ij} = \frac{1}{1 + \tau(\lambda_{ij} - \mu_i)}, \ \psi_{ij} = \frac{1}{\left(1 + \tau(\lambda_{ij} - \mu_i)\right)^2}, \text{ and } \rho_{ij} = \frac{\tau^2(\lambda_{ij} - \mu_i)}{\left(1 + \tau(\lambda_{ij} - \mu_i)\right)^3}.$$

Proof. In terms of the parameter-separable form (2.3), we have

$$\begin{gathered} \mathsf{p} = s, \quad \widehat{H}(s) = \widehat{\mathcal{C}}(s)\widehat{\mathcal{A}}(s)^{-1}\widehat{\mathcal{B}}(s), \\ \widehat{\mathcal{A}}(s) = \widehat{\alpha}_1(s)I + \widehat{\alpha}_2(s)\widehat{M} + \widehat{\alpha}_3(s)\widehat{\Sigma}, \quad \widehat{\mathcal{B}}(s) = \widehat{B}, \quad \widehat{\mathcal{C}}(s) = \widehat{C}, \\ \widehat{\alpha}_1(s) = s, \quad \widehat{\alpha}_2(s) = -1, \quad \text{and} \quad \widehat{\alpha}_3(s) = -e^{-\tau s}. \end{gathered}$$

Furthermore, using the notation of Theorem 3.2, time-delay transfer function $\widehat{H}$ in (6.3) corresponds to $a_i(s) = s - \mu_i - e^{-\tau s}\sigma_i$. Thus, we have identified all the necessary ingredients to apply Theorem 3.2.

In the following, we will use the expressions

$$a_i(\lambda_{ij}) = \lambda_{ij} - \mu_i - e^{-\tau\lambda_{ij}}\sigma_i = 0, \tag{6.5a}$$

$$a_i'(\lambda_{ij}) = 1 + \tau e^{-\tau\lambda_{ij}}\sigma_i = 1 + \tau(\lambda_{ij} - \mu_i), \tag{6.5b}$$

$$a_i''(\lambda_{ij}) = -\tau^2 e^{-\tau\lambda_{ij}}\sigma_i = -\tau^2(\lambda_{ij} - \mu_i), \tag{6.5c}$$

$$\widehat{\alpha}_3(\lambda_{ij}) = -e^{-\tau\lambda_{ij}} = -\frac{\lambda_{ij} - \mu_i}{\sigma_i}, \tag{6.5d}$$

$$\widehat{\alpha}_3'(\lambda_{ij}) = \tau e^{-\tau\lambda_{ij}} = \frac{\tau(\lambda_{ij} - \mu_i)}{\sigma_i}, \tag{6.5e}$$

where we used (6.5a) in the later expressions.

From (6.5b) we get $\frac{1}{a_i'(\lambda_{ij})} = \phi_{ij}$. Inserting this into the conditions (3.3a) and (3.3b) yield (6.4a) and (6.4b), respectively.

From (6.5b) we also get $\frac{1}{a_i'(\lambda_{ij})^2} = \psi_{ij}$. Next, from (6.5b) and (6.5c) we find $-\frac{a_i''(\lambda_{ij})}{a_i'(\lambda_{ij})^3} = \rho_{ij}$. Together with $\widehat{\alpha}_2(\lambda_{ij}) = -1$ and $\widehat{\alpha}_2'(\lambda_{ij}) = 0$, the condition (3.3c) with $\widehat{\alpha}_2$ implies (6.4c).

For the final condition, together with $\phi_{ij} - \psi_{ij} = \frac{\tau(\lambda_{ij}-\mu_i)}{a_i'(\lambda_{ij})^2}$, expression (6.5d) gives

$$\frac{\widehat{\alpha}_3(\lambda_{ij})}{a_i'(\lambda_{ij})^2} = -\frac{\lambda_{ij} - \mu_i}{\sigma_i a_i'(\lambda_{ij})^2} = -\frac{1}{\tau\sigma_i}(\phi_{ij} - \psi_{ij}). \tag{6.6}$$

Using (6.5b), (6.5c), (6.5d), and (6.5e) gives

$$\begin{aligned}
\frac{\widehat{\alpha}_3'(\lambda_{ij})}{a_i'(\lambda_{ij})^2} - \frac{\widehat{\alpha}_3(\lambda_{ij})a_i''(\lambda_{ij})}{a_i'(\lambda_{ij})^3} &= \frac{\tau(\lambda_{ij} - \mu_i)}{\sigma_i(1 + \tau(\lambda_{ij} - \mu_i))^2} - \frac{\tau^2(\lambda_{ij} - \mu_i)^2}{\sigma_i(1 + \tau(\lambda_{ij} - \mu_i))^3} \\
&= \frac{\tau}{\sigma_i} \cdot \frac{(\lambda_{ij} - \mu_i)(1 + \tau(\lambda_{ij} - \mu_i)) - \tau(\lambda_{ij} - \mu_i)^2}{(1 + \tau(\lambda_{ij} - \mu_i))^3} \\
&= \frac{\tau}{\sigma_i} \cdot \frac{\lambda_{ij} - \mu_i}{(1 + \tau(\lambda_{ij} - \mu_i))^3} \\
&= \frac{1}{\tau\sigma_i}\rho_{ij}.
\end{aligned} \tag{6.7}$$

Inserting (6.6) and (6.7) into the condition (3.3c) with $\widehat{\alpha}_3$ implies (6.4d). □

Not surprisingly, the interpolatory optimality conditions for time-delay systems appear much more involved than the earlier ones for second-order systems and pH systems. The single interpolation conditions are replaced by an (infinite) weighted sum, reflecting the fact that time-delay systems have infinitely many poles (identified via the Lambert W function). Moreover, the Hermite conditions in this case appear as mixed conditions where a linear combination of $\widehat{H}$ and $\widehat{H}'$ needs to be interpolated. However, one aspects stays the same: interpolation needs to happen at the mirror images of the poles.

Note that we can replace the condition (6.4d) by the addition of (6.4d) and (6.4c), which simply becomes

$$c_i^*\left(\sum_{j=-\infty}^{\infty} \overline{\phi_{ij}} H'\left(-\overline{\lambda_{ij}}\right)\right) b_i = c_i^*\left(\sum_{j=-\infty}^{\infty} \overline{\phi_{ij}} \widehat{H}'\left(-\overline{\lambda_{ij}}\right)\right) b_i.$$

7 Conclusion

We have developed interpolatory $\mathcal{H}_2$-optimality conditions for approximating non-parametric structured dynamical systems, namely for second-order, pH, and time-delay systems. We have shown that bitangential Hermite interpolation is the common unifying framework across all these different settings. In this paper, we have mainly focused on the theoretical analysis of deriving the optimality conditions. A natural future direction would be to develop (iterative) numerical methods, such as IRKA, that directly use the interpolation conditions discussed here as opposed to the gradient-based iterative optimization methods employed in the literature.

A Proofs

Proof of Lemma 3.1. By direct calculation,

$$\operatorname{Res}(f_1, c) = \lim_{z\to c}(z - c)f_1(z) = \lim_{z\to c} \frac{z - c}{h(z)} \cdot g(z) = \frac{g(c)}{h'(c)},$$

which proves the first result in (3.1). Then, following similarly for f_2 (but with much more tedious calculations), we obtain

$$\begin{aligned}
&\operatorname{Res}(f_2, c)\\
&= \lim_{z\to c} \frac{\mathrm{d}}{\mathrm{d}z}\Big((z-c)^2 f_2(z)\Big) = \lim_{z\to c} \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{(z-c)^2 g(z)}{h(z)^2}\right)\\
&= \lim_{z\to c} \frac{\Big(2(z-c)g(z) + (z-c)^2 g'(z)\Big)h(z)^2 - 2(z-c)^2 g(z)h(z)h'(z)}{h(z)^4}\\
&= \lim_{z\to c} \frac{2(z-c)g(z)h(z) + (z-c)^2 g'(z)h(z) - 2(z-c)^2 g(z)h'(z)}{h(z)^3}\\
&= \lim_{z\to c} \frac{z-c}{h(z)} \cdot \frac{2g(z)h(z) + (z-c)g'(z)h(z) - 2(z-c)g(z)h'(z)}{h(z)^2}\\
&= \frac{1}{h'(c)} \lim_{z\to c} \frac{3g'(z)h(z) + (z-c)g''(z)h(z) - (z-c)g'(z)h'(z) - 2(z-c)g(z)h''(z)}{2h(z)h'(z)}\\
&= \frac{1}{h'(c)}\left(\frac{3g'(c)}{2h'(c)} + \lim_{z\to c} \frac{z-c}{h(z)} \cdot \frac{g''(z)h(z) - g'(z)h'(z) - 2g(z)h''(z)}{2h'(z)}\right)\\
&= \frac{1}{h'(c)}\left(\frac{3g'(c)}{2h'(c)} - \frac{1}{h'(c)} \cdot \frac{g'(c)h'(c) + 2g(c)h''(c)}{2h'(c)}\right)\\
&= \frac{g'(c)}{h'(c)^2} - \frac{g(c)h''(c)}{h'(c)^3},
\end{aligned}$$

where we used L'Hôpital's rule in the fifth equality. □

References

[ABG20] A. C. Antoulas, C. A. Beattie, and S. Güğercin. *Interpolatory methods for model reduction*. Computational Science and Engineering 21. SIAM, Philadelphia, PA, 2020. doi:10.1137/1.9781611976083.

[AW23] Q. Aumann and S. W. R. Werner. Adaptive choice of near-optimal expansion points for interpolation-based structure-preserving model reduction. arXiv preprint 2305.10806, 2023. doi:10.48550/arXiv.2305.10806.

[Bai02] Z. Bai. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics*, 43(1–2):9–44, 2002. doi:10.1016/S0168-9274(02)00116-2.

[BB14] C. A. Beattie and P. Benner. $\mathcal{H}_2$-optimality conditions for structured dynamical systems. Preprint MPIMD/14-18, Max Planck Institute Magdeburg, 2014. URL: https://csc.mpi-magdeburg.mpg.de/preprints/2014/MPIMD14-18.pdf.

[BFSZ16] T. Bonin, H. Faßbender, A. Soppa, and M. Zaeh. A fully adaptive rational global Arnoldi method for the model-order reduction of second-order MIMO systems with proportional damping. *Mathematics and Computers in Simulation*, 122:1–19, 2016. doi:10.1016/j.matcom.2015.08.017.

[BG05] C. A. Beattie and S. Gugercin. Krylov-based model reduction of second-order systems with proportional damping. *44th IEEE Conference on Decision and Control / European Control Conference. CDC-ECC '05.*, pages 2278–2283, December 2005. doi:10.1109/CDC.2005.1582501.

[BG09] C. A. Beattie and S. Gugercin. Interpolatory projection methods for structure-preserving model reduction. *Syst. Control Lett.*, 58(3):225–232, 2009. doi:10.1016/j.sysconle.2008.10.016.

[BG12] C. A. Beattie and S. Gugercin. Realization-independent $\mathcal{H}_2$-approximation. In *51st IEEE Conference on Decision and Control (CDC)*, pages 4953–4958, 2012. doi:10.1109/CDC.2012.6426344.

[BS05] Z. Bai and Y. Su. Dimension reduction of second order dynamical systems via a second-order Arnoldi method. *SIAM J. Sci. Comput.*, 5:1692–1709, 2005. doi:10.1137/040605552.

[BU22] T. Breiten and B. Unger. Passivity preserving model reduction via spectral factorization. *Automatica*, 142:110368, 2022. doi:10.1016/j.automatica.2022.110368.

[CGM15] R. Cepeda-Gomez and W. Michiels. Some special cases in the stability analysis of multi-dimensional time-delay systems using the matrix Lambert W function. *Automatica*, 53:339–345, 2015. doi:10.1016/j.automatica.2015.01.016.

[CGVV05] V. Chahlaoui, K. A. Gallivan, A. Vandendorpe, and P. Van Dooren. Model reduction of second-order systems. In P. Benner, V. Mehrmann, and D. C. Sorensen, editors, *Dimension Reduction of Large-Scale Systems*, volume 45 of *Lecture Notes in Computational Science and Engineering*, pages 149–172, Berlin/Heidelberg, 2005. Springer-Verlag. doi:10.1007/3-540-27909-1_6.

[DGA10] P. Van Dooren, K. A. Gallivan, and P.-A. Absil. $\mathcal{H}_2$-optimal model reduction with higher-order poles. *SIAM J. Matrix Anal. Appl.*, 31(5):2738–2753, 2010. doi:10.1137/080731591.

[Fri14] E. Fridman. *Introduction to Time-Delay Systems*. Springer International Publishing, Switzerland, 2014. doi:10.1007/978-3-319-09393-2.

[GAB08] 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(2):609–638, 2008. doi:10.1137/060666123.

[GEMM19] M. A. Gomez, A. V. Egorov, S. Mondié, and W. Michiels. Optimization of the $\mathcal{H}_2$ norm for single-delay systems, with application to control design and model approximation. *IEEE Trans. Autom. Control*, 64(2):804–811, 2019. doi:10.1109/TAC.2018.2836019.

[GPBvdS12] 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(9):1963–1974, 2012. doi:10.1016/j.automatica.2012.05.052.

[JZ12] B. Jacob and H. J. Zwart. *Linear port-Hamiltonian systems on infinite-dimensional spaces*, volume 223. Springer, Basel, 2012. doi:10.1007/978-3-0348-0399-1.

[MG23a] P. Mlinarić and S. Gugercin. $\mathcal{L}_2$-optimal reduced-order modeling using parameter-separable forms. *SIAM J. Sci. Comput.*, 45(2):A554–A578, 2023. doi:10.1137/22M1500678.

[MG23b] P. Mlinarić and S. Gugercin. A unifying framework for interpolatory $\mathcal{L}_2$-optimal reduced-order modeling. *SIAM J. Numer. Anal.*, 61(5):2133–2156, 2023. doi:10.1137/22M1516920.

[ML67] L. Meier and D. Luenberger. Approximation of linear constant systems. *IEEE Trans. Autom. Control*, 12(5):585–588, 1967. doi:10.1109/TAC.1967.1098680.

[ML20] Tim Moser and Boris Lohmann. A new Riemannian framework for efficient $\mathcal{H}_2$-optimal model reduction of port-Hamiltonian systems. In *59th IEEE Conference on Decision and Control (CDC)*, pages 5043–5049, 2020. doi:10.1109/CDC42340.2020.9304134.

[MSMV22] T. Moser, P. Schwerdtner, V. Mehrmann, and M. Voigt. Structure-preserving model order reduction for index two port-Hamiltonian descriptor systems. arXiv preprint 2206.03942, 2022. doi:10.48550/arXiv.2206.03942.

[MU22] V. Mehrmann and B. Unger. Control of port-Hamiltonian differential-algebraic systems and applications. arXiv preprint 2201.06590, 2022. doi:10.48550/arXiv.2201.06590.

[PGB+16] I. Pontes Duff, S. Gugercin, C. Beattie, C. Poussot-Vassal, and C. Seren. $\mathcal{H}_2$-optimality conditions for reduced time-delay systems of dimensions one. *IFAC-PapersOnLine*, 49(10):7–12, 2016. 13th IFAC on Time Delay Systems TDS 2019. doi:10.1016/j.ifacol.2016.07.464.

[Pon23] I. Pontes Duff. personal communication, 2023.

[RS08] T. Reis and T. Stykel. Balanced truncation model reduction of second-order systems. *Math. Comput. Model. Dyn. Syst.*, 14(5):391–406, 2008. doi:10.1080/13873950701844170.

[SGB16] K. Sinani, S. Gugercin, and C. Beattie. A structure-preserving model reduction algorithm for dynamical systems with nonlinear frequency dependence. *IFAC-PapersOnLine*, 49(9):56–61, 2016. 6th IFAC Symposium on System Structure and Control SSSC 2016. doi:10.1016/j.ifacol.2016.07.492.

[SMMV22] P. Schwerdtner, T. Moser, V. Mehrmann, and M. Voigt. Structure-preserving model order reduction for index one port-Hamiltonian descriptor systems. arXiv preprint 2206.01608, 2022. doi:10.48550/arXiv.2206.01608.

[SSW19] J. Saak, D. Siebelts, and S. W. R. Werner. A comparison of second-order model order reduction methods for an artificial fishtail. *at-Automatisierungstechnik*, 67(8):648–667, 2019. doi:10.1515/auto-2019-0027.

[vdSJ14] A. van der Schaft and D. Jeltsema. Port-Hamiltonian systems theory: An introductory overview. *Foundations and Trends in Systems and Control*, 1(2-3):173–378, 2014. doi:10.1561/2600000002.

[Vui14] P. Vuillemin. *Frequency-limited model approximation of large-scale dynamical models*. PhD thesis, Université de Toulouse, 2014. URL: https://hal.science/tel-01092051/.

[Wer21] S. W. R. Werner. *Structure-Preserving Model Reduction for Mechanical Systems*. Dissertation, Otto-von-Guericke-Universität, Magdeburg, Germany, 2021. doi:10.25673/38617.

[Wya12] S. Wyatt. *Issues in Interpolatory Model Reduction: Inexact Solves, Second-order Systems and DAEs*. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA, May 2012. URL: http://hdl.handle.net/10919/27668.