MAX-PLANCK-GESELLSCHAFT

Max Planck Institute Magdeburg
Preprints

Christopher Beattie Peter Benner

$\mathcal{H}_2$-Optimality Conditions for Structured Dynamical Systems

MAX–PLANCK–INSTITUT
FÜR DYNAMIK KOMPLEXER
TECHNISCHER SYSTEME
MAGDEBURG

MPIMD/14-18 **October, 2014**

Imprint:

Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg

Publisher:
Max Planck Institute for
Dynamics of Complex Technical Systems

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

http://www.mpi-magdeburg.mpg.de/preprints/

Abstract

Dynamical systems often have structural features that incorporate underlying physics and conservation laws that reflect basic properties of phenomena of interest. Reduced models for these dynamical systems that do not share such key structural features, even if they otherwise have high fidelity, may produce responses that are "unphysical" and as a result may be unsuitable for use as dependable surrogates.

We seek systems that have structure characterized as either port-Hamiltonian or second-order (or both), and that, within the latitude allowed by those constraints, is also a best possible approximation to the original system as discerned by the $\mathcal{H}_2$ error measure. In this work, we develop necessary optimality conditions that must be satisfied by such reduced systems.

Keywords: structured model reduction, $\mathcal{H}_2$-optimal model reduction, port-Hamiltonian systems, second-order systems

Author's addresses:

Christopher Beattie
Department of Mathematics
Virginia Tech
Blacksburg, VA 24061-0123, USA

beattie@vt.edu

Peter Benner,
Computational Methods in Systems and Control Theory,
Max Planck Institute for Dynamics of Complex Technical Systems,
Sandtorstraße 1, 39106 Magdeburg, Germany

benner@mpi-magdeburg.mpg.de

1 Introduction

Dynamical systems are a basic framework in modeling and control of many physical phenomena that are of interest in science and industry. Examples include chatter suppression in high precision machine tools, protein folding and molecular dynamics, and control of micro-electro-mechanical systems (MEMS). Direct numerical simulation is one of few available means to examine these systems and others like it in order to accurate predict or control their behavior. Persistent needs for greater accuracy lead to inclusion of greater detail in the model and potential coupling with other systems that may themselves operate in different time and spatial scales; this can produce computational tasks that make unmanageably large demands on resources. Efficient model utilization is a necessary component of simulations in such large-scale settings.

A key observation that leads one to consider the potential for inexpensive, high-fidelity system surrogates is that often the internal states of the original system model evolve along trajectories that do not fully occupy the state space but hew rather more closely to some subspace of substantially lower dimension. The system behaves nearly as if it had very many fewer internal degrees-of-freedom. A natural goal then is to replace the original system with a lower dimensional dynamical system having as much of the same input/output response characteristics as the original system as possible. The resulting reduced-order model could then be used as a surrogate replacing the original system model as a component in a larger simulation [1, 5, 6, 7].

Dynamical systems often have structural features that encode underlying physics and conservation laws. Reduced models that do not share such key structural features with the original system, even if they otherwise have high fidelity, may produce responses that are "unphysical" and as a result may be unsuitable for use as dependable surrogates for the original system. The structural features that we focus on here are port-Hamiltonian and second-order systems.

The goal here is to derive necessary conditions that must be satisfied by reduced-order models with particular structure in order to achieve, at least locally, error minimization. Following $\mathcal{H}_2$-norm minimization ideas successful for (unstructured) first-order linear systems [13], we generalize these concepts to dynamical systems constrained to have port-Hamiltonian or second-order structure.

We provide in §2 background information about $\mathcal{H}_2$-error measures and associated (unstructured) necessary conditions for optimality. We discuss particular system structures of interest in §3; $\mathcal{H}_2$-optimal port-Hamiltonian approximations are considered in §4; §5 considers necessary conditions for best $\mathcal{H}_2$-optimal approximations among second-order modally damped systems; finally, in §6 structural constraints are combined and we consider systems that are both port-Hamiltonian and second order.

2 Setting and Background

Let $\mathcal{H}_2^{m\times p}$ denote the set of $m \times p$ matrix-valued functions, $\mathfrak{H}(s)$, with components, $h_{ij}(s)$, that are analytic for s in the open right half plane, $Re(s) > 0$, and such that for each fixed $Re(s) = x > 0$, $h_{ij}(x + \imath y)$ is square integrable as a function of $y \in (-\infty, \infty)$ in such a way that

$$\sup_{x>0} \int_{-\infty}^{\infty} |h_{ij}(x + \imath y)|^2 \, dy < \infty.$$

$\mathcal{H}_2^{m\times p}$ is a Hilbert space. Indeed, if $\mathfrak{G}(s)$ and $\mathfrak{H}(s)$ are $\mathcal{H}_2^{m\times p}$-functions then the $\mathcal{H}_2^{m\times p}$-inner product can be defined as

$$\langle \mathfrak{G},\ \mathfrak{H}\rangle_{\mathcal{H}_2} \stackrel{\text{def}}{=} \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathsf{trace}\left(\overline{\mathfrak{G}(\imath\omega)}\,\mathfrak{H}(\imath\omega)^T\right) d\omega \qquad (1)$$

with an associated norm defined as

$$\|\mathfrak{H}\|_{\mathcal{H}_2} \stackrel{\text{def}}{=} \left(\frac{1}{2\pi}\int_{-\infty}^{+\infty} \|\mathfrak{H}(\imath\omega)\|_F^2 \, d\omega\right)^{1/2}. \qquad (2)$$

Here $\|\mathbf{M}\|_F = \sqrt{\langle \mathbf{M},\ \mathbf{M}\rangle_F}$ and $\langle \mathbf{M},\ \mathbf{N}\rangle_F = \mathsf{trace}\left(\overline{\mathbf{M}}\,\mathbf{N}^T\right)$ denote the *Frobenius norm* and *Frobenius inner product*, respectively. Notice that if $\mathfrak{G}(s)$ and $\mathfrak{H}(s)$ represent real dynamical systems then $\langle \mathfrak{G},\ \mathfrak{H}\rangle_{\mathcal{H}_2} = \langle \mathfrak{H},\ \mathfrak{G}\rangle_{\mathcal{H}_2}$ and $\langle \mathfrak{G},\ \mathfrak{H}\rangle_{\mathcal{H}_2}$ itself must be real.

There is a substantial body of literature addressing the problem of optimal $\mathcal{H}_2$ model reduction for general linear time-invariant systems. Such systems can be characterized through standard first-order state-space realizations of the form, $\mathfrak{H}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$. The optimal $\mathcal{H}_2$ model reduction problem may then be posed as seeking a system $\hat{\mathfrak{H}}_r$ of order r which solves:

$$\min_{\mathfrak{H}_r \text{ is stable}} \|\mathfrak{H} - \mathfrak{H}_r\|_{\mathcal{H}_2} \qquad (3)$$

— see, for example, [32, 25, 8, 15, 19, 16, 31, 18, 33, 13, 29]. Necessary conditions for a reduced order model to be an $\mathcal{H}_2$-optimal approximation are built typically from the following lemma:

Lemma 1. *Suppose that* $\{\widehat{\mathbf{H}}_r^{(\varepsilon)}\}_{\varepsilon>0} \subset \mathcal{Q}_r$ *is a family of dynamical systems parameterized by* $\varepsilon > 0$ *such that* $\|\widehat{\mathbf{H}}_r - \widehat{\mathbf{H}}_r^{(\varepsilon)}\|_{\mathcal{H}_2} = \mathcal{O}(\varepsilon)$ *as* $\varepsilon \to 0$. *Then* $\mathcal{H}_2$*-optimality of any* $\widehat{\mathbf{H}}_r$ *solving (3) implies that as* $\varepsilon \to 0$,

$$\left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\ \frac{\widehat{\mathbf{H}}_r - \widehat{\mathbf{H}}_r^{(\varepsilon)}}{\|\widehat{\mathbf{H}}_r - \widehat{\mathbf{H}}_r^{(\varepsilon)}\|_{\mathcal{H}_2}} \right\rangle_{\mathcal{H}_2} \longrightarrow 0$$

First-order necessary conditions for $\hat{\mathfrak{H}}_r$ to solve (3) (and more generally to be a local minimizer of the $\mathcal{H}_2$ error) were formulated by Wilson [31] and Hyland and Bernstein [16] in terms of a pair coupled Lyapunov equations. When $\mathfrak{H}(s)$ is not necessarily known in the customary first-order form, Meier and Luenberger [19] provided interpolation-based first-order conditions for $\mathcal{H}_2$ optimality, at least for single-input single-output systems. Gugercin *et al.* [13] extended these interpolation-based conditions to the multi-input/multi-output case. For convenience, we state them here: Assume that $\hat{\mathfrak{H}}_r$ is a local minimizer of $\|\mathfrak{H} - \mathfrak{H}_r\|_{\mathcal{H}_2}$ among those (stable) transfer functions having r distinct poles. The residue of $\hat{\mathfrak{H}}_r$ at each pole, $\widetilde{\lambda}_i$, is matrix-valued and has rank one, $\mathsf{res}[\hat{\mathfrak{H}}_r(s), \widetilde{\lambda}_i] = \hat{\mathbf{c}}_i \hat{\mathbf{b}}_i^*$. Then for $i = 1, \ldots, r$,

$$\begin{aligned} \mathfrak{H}(-\widetilde{\lambda}_i)\hat{\mathbf{b}}_i &= \mathfrak{H}_r(-\widetilde{\lambda}_i)\hat{\mathbf{b}}_i, \\ \hat{\mathbf{c}}_i^*\mathfrak{H}(-\widetilde{\lambda}_i) &= \hat{\mathbf{c}}_i^*\mathfrak{H}_r(-\widetilde{\lambda}_i), \end{aligned} \quad \text{and} \quad \hat{\mathbf{c}}_i^*\mathfrak{H}'(-\widetilde{\lambda}_i)\hat{\mathbf{b}}_i = \hat{\mathbf{c}}_i^*\mathfrak{H}_r'(-\widetilde{\lambda}_i)\hat{\mathbf{b}}_i. \tag{4}$$

Thus, first-order necessary conditions for $\mathcal{H}_2$-optimality without further structural constraints require *tangential interpolation* at mirror images of the reduced system poles, $\widetilde{\lambda}_i$, reflected across the imaginary axis. Significantly, these conditions do not require any particular realization for the original system, $\mathfrak{H}(s)$ to be known although the optimal reduced system, $\hat{\mathfrak{H}}_r$, is typically delivered in standard first order form, $\hat{\mathfrak{H}}_r = \mathbf{C}_r(s\mathbf{I} - \mathbf{A}_r)^{-1}\mathbf{B}_r$. We describe below some approaches that have been developed to accomplish this, using only evaluations of $\mathfrak{H}(s)$; these approaches are not dependent on any particular realization for $\mathfrak{H}(s)$ being available.

3 Structured Systems

Dynamical systems may have additional discernible structure that reflects underlying physics and conservation laws. Port-based network modeling [12] takes advantage of a common situation where the system under study is decomposable into subsystems that are interconnected through pairs of dynamic quantities whose pairwise product gives the power exchanged among subsystems. This approach is especially useful for multi-physics systems, where subsystems may be associated with different categories of physical phenomena (e.g, mechanical, electrical, or hydraulic). This leads one to consider *port-Hamiltonian* system representations (see [12, 27]) which encode structural features related to the manner in which energy is distributed within and across subsystems.

Although greater generality is possible, it suffices for our purposes to consider port-Hamiltonian systems that are linear time invariant $\mathcal{H}_2$ dynamical systems. Finite-dimensional systems of this sort have realizations of the form

$$\begin{aligned} \dot{\mathbf{x}} &= (\mathbf{J} - \mathbf{R})\mathbf{Q}\mathbf{x} + \mathbf{B}\mathbf{u} \\ \mathbf{y} &= \mathbf{B}^T\mathbf{Q}\mathbf{x} \end{aligned} \tag{5}$$

where

1. $\mathbf{J} = -\mathbf{J}^T$ is skew-symmetric,
2. $\mathbf{R} = \mathbf{R}^T$ is symmetric positive-semidefinite, and

3. $\mathbf{Q}$ is symmetric positive-definite.

A key feature of the class of port-Hamiltonian systems is that it is closed under power-conserving interconnection, that is, if an array of port-Hamiltonian systems are connected together in a way that preserves the integrity of the shared quantities, the resulting aggregate system is also port-Hamiltonian, and hence, passive. Thus, it may be important to substitute a portHamiltonian subsystem with a low order surrogate system that is also port-Hamiltonian. There has been earlier work along these lines on structure-preserving reduction of port-Hamiltonian systems, notably [14, 21, 22, 23, 28].

Let $\mathcal{PH}(r)$ denote the set of all port-Hamiltonian systems with state-space dimension r. The model reduction problem that we pursue may then be posed as seeking a port-Hamiltonian system $\hat{\boldsymbol{\mathcal{H}}}_r \in \mathcal{PH}(r)$ which solves:

$$\min_{\substack{\boldsymbol{\mathcal{H}}_r \text{ is stable} \\ \boldsymbol{\mathcal{H}}_r \in \mathcal{PH}(r)}} \|\boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{H}}_r\|_{\mathcal{H}_2} \tag{6}$$

The reduced system, $\hat{\boldsymbol{\mathcal{H}}}_r \in \mathcal{PH}(r)$ may be written as

$$\begin{aligned} \dot{\mathbf{x}}_r &= (\mathbf{J}_r - \mathbf{R}_r)\mathbf{Q}_r\mathbf{x}_r + \mathbf{B}_r\mathbf{u} \\ \mathbf{y}_r &= \mathbf{B}_r^T\mathbf{Q}_r\mathbf{x}_r \end{aligned} \tag{7}$$

where (similar to the full order system),

1. $\mathbf{J}_r = -\mathbf{J}_r^T$ is skew-symmetric,
2. $\mathbf{R}_r = \mathbf{R}_r^T$ is symmetric positive-semidefinite, and
3. $\mathbf{Q}_r$ is symmetric positive-definite.

Another category of dynamical system that we consider arises in the modeling of forced vibration of an n degree–of–freedom mechanical structure:

$$\begin{aligned} \mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{D}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) &= \mathbf{B}_1\dot{\mathbf{u}}(t) + \mathbf{B}_0\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}_1\dot{\mathbf{x}}(t) + \mathbf{C}_0\mathbf{x}(t), \end{aligned} \tag{8}$$

where $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{K} \in \mathbb{R}^{n\times n}$ are positive (semi)-definite symmetric matrices describing, respectively, mass distribution, energy dissipation, and stiffness distribution throughout the structure. The input $\mathbf{u}(t) \in \mathbb{R}^m$ is a time-dependent force or displacement applied along degrees-of-freedom specified in $\mathbf{B}_0, \mathbf{B}_1 \in \mathbb{R}^{n\times m}$ and $\mathbf{y}(t) \in \mathbb{R}^p$ is a vector of output measurements defined through observation matrices $\mathbf{C}_0, \mathbf{C}_1 \in \mathbb{R}^p$. The transfer function $\boldsymbol{\mathcal{H}}(s)$ from $\mathbf{u}(t)$ to $\mathbf{y}(t)$ is given by

$$\boldsymbol{\mathcal{H}}(s) = (s\mathbf{C}_1 + \mathbf{C}_0)(s^2\mathbf{M} + s\mathbf{D} + \mathbf{K})^{-1}(s\mathbf{B}_1 + \mathbf{B}_0).$$

Second order systems of the form (8) arise naturally in the analysis of other phenomena apart from structural vibration, such as the response of electrical circuits and micro-electro-mechanical systems; see [11, 24, 3, 10, 30, 9, 17, 2], and references therein. Note that second-order systems of the form (8) cannot be converted to first-order form in a straightforward way when $\mathbf{B}_1$ is nontrivial. The goal is to generate, for some $r \ll n$, an r^{th} order reduced second-order system of the form

$$\mathbf{M}_r\ddot{\mathbf{x}}_r(t) + \mathbf{D}_r\dot{\mathbf{x}}_r(t) + \mathbf{K}_r\mathbf{x}_r(t) = \mathbf{B}_{1,r}\dot{\mathbf{u}}(t) + \mathbf{B}_{0,r}\mathbf{u}(t)$$
$$\mathbf{y}_r(t) = \mathbf{C}_{1,r}\dot{\mathbf{x}}_r(t) + \mathbf{C}_{0,r}\mathbf{x}_r(t),$$

where $\mathbf{M}_r, \mathbf{D}_r, \mathbf{K}_r \in \mathbb{R}^{r\times r}$ are positive (semi)-definite symmetric matrices, $\mathbf{B}_{0,r}$, $\mathbf{B}_{1,r} \in \mathbb{R}^{r\times m}$, and $\mathbf{C}_{0,r}$, $\mathbf{C}_{1,r} \in \mathbb{R}^{p\times r}$ are chosen in such a way so that $\mathbf{y}_r(t)$ approximates $\mathbf{y}(t)$ over a wide range of inputs, $\mathbf{u}(t)$.

One classical approach to model reduction of second-order systems is to the apply standard model reduction techniques to a first-order realization of the system. However, even in cases where a first-order realization is straightforward to obtain, performing the reduction on the first-order realization has several disadvantages: it destroys the original second-order system structure and the physical meaning of the states. Moreover, once the reduction is performed in the first-order framework, it will not always be possible to convert this back to a corresponding second-order system of the form (8), see [20]. Even when this is possible, one typically cannot guarantee that structural properties such as positive definite symmetric reduced mass, damping, and stiffness matrices, will be retained. Keeping the original structure is crucial both to preserve physical meaning of the states and to retain physically significant properties such as stability and passivity. Due to these considerations, we assert a better strategy will be to reduce directly in the second-order setting, constructing first a projecting subspace $\boldsymbol{\mathcal{V}}_r \in \mathbb{R}^{n\times r}$ and then defining the reduced coefficient matrices:

$$\begin{gathered}\mathbf{M}_r = \boldsymbol{\mathcal{V}}_r^T\mathbf{M}\boldsymbol{\mathcal{V}}_r,\ \mathbf{D}_r = \boldsymbol{\mathcal{V}}_r^T\mathbf{D}\boldsymbol{\mathcal{V}}_r,\ \mathbf{K}_r = \boldsymbol{\mathcal{V}}_r^T\mathbf{K}\boldsymbol{\mathcal{V}}_r, \\ \mathbf{B}_{1,r} = \boldsymbol{\mathcal{V}}_r^T\mathbf{B}_1,\ \mathbf{B}_{0,r} = \boldsymbol{\mathcal{V}}_r^T\mathbf{B}_0,\ \mathbf{C}_{0,r} = \mathbf{C}_0\boldsymbol{\mathcal{V}}_r,\ \text{and}\ \mathbf{C}_{1,r} = \mathbf{C}_1\boldsymbol{\mathcal{V}}_r.\end{gathered} \tag{9}$$

There have been significant efforts in this direction. Building on the earlier work of [26], Bai and Su [4] introduced *"second-order" Krylov subspaces* and showed how to obtain a reduced-order system directly in a second-order framework as in (9) while still satisfying interpolation conditions at selected points. Their method does not treat cases in which $\mathbf{C}_1$ in (8) is nontrivial, i.e., when the velocities are observed. Another second-order structure preserving interpolation-based reduction technique was introduced by Chahlaoui *et al.* [9], though this approach also requires a first-order state-space realization and so is not applicable to cases where $\mathbf{B}_1$ is nontrivial.

Let $\mathcal{Q}(r)$ denote the set of all second-order systems with state-space dimension r. The second model reduction problem that we pursue may then be posed as seeking a a second-order system $\hat{\boldsymbol{\mathcal{H}}}_r \in \mathcal{PH}(r)$ which solves:

$$\min_{\substack{\boldsymbol{\mathcal{H}}_r \text{ is stable} \\ \boldsymbol{\mathcal{H}}_r \in \mathcal{Q}(r)}} \|\boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{H}}_r\|_{\mathcal{H}_2} \tag{10}$$

4 $\mathcal{H}_2$-optimal Port-Hamiltonian Approximations

Consider the first problem of finding a $\mathbf{G}_r \in \mathcal{PH}(r)$ that solves :

$$\|\mathbf{G}-\mathbf{G}_r\|_{\mathcal{H}_2} = \min_{\widetilde{\mathbf{G}}_r \in \mathcal{PH}(r)} \left\|\mathbf{G}-\widetilde{\mathbf{G}}_r\right\|_{\mathcal{H}_2}. \tag{11}$$

That is, $\mathbf{G}_r$ is an $\mathcal{H}_2$-*optimal* reduced order port-Hamiltonian approximation of order r to (5).

We will denote the *Loewner matrix* associated with a transfer function $\mathbf{H}$ on tangential interpolation data $\mathcal{S} = \{\{\sigma_i\}_1^r,\ \{\mathbf{w}_i\}_1^r,\ \{\mathbf{v}_i\}_1^r\}$ as

$$(\mathbb{L}[\mathbf{H},\mathcal{S}])_{i,j} := \begin{cases} \dfrac{\mathbf{w}_i^T\mathbf{H}(\sigma_i)\mathbf{v}_j - \mathbf{w}_i^T\mathbf{H}(\sigma_j)\mathbf{v}_j}{\sigma_i-\sigma_j} & \text{if } \ i \neq j \\ \mathbf{w}_i^T\mathbf{H}'(\sigma_i)\mathbf{v}_i & \text{if } \ i = j \end{cases}$$

Theorem 1. *Suppose that $\mathbf{G}_r(s)$ is a solution to (11) with a reduced dissipation matrix $\mathbf{R}_r$ that is positive definite. Suppose further that $\mathbf{G}_r(s)$ has r distinct poles and is represented as $\mathbf{G}_r(s) = \sum_{i=1}^r \frac{1}{s-\lambda_i}\mathsf{c}_i\mathsf{d}_i^T$. Then*

$$\mathbb{L}[\mathbf{G},\mathcal{S}] = \mathbb{L}[\mathbf{G}_r,\mathcal{S}].$$

where $\mathcal{S}$ here denotes derived interpolation data: $\mathcal{S} = \{\{-\lambda_i\}_1^r,\ \{\mathsf{c}_i\}_1^r,\ \{\mathsf{d}_i\}_1^r\}$.

Proof: Perform a state-space transformation on (7) to obtain a reduced system realization in *scaled energy coordinates* (transforming $\mathbf{Q}_r \to \widetilde{\mathbf{Q}}_r = \mathbf{I}$):

$$\begin{aligned} \dot{\tilde{\mathbf{x}}}_r &= (\widetilde{\mathbf{J}}_r - \widetilde{\mathbf{R}}_r)\tilde{\mathbf{x}}_r + \widetilde{\mathbf{B}}_r\mathbf{u} \\ \mathbf{y}_r &= \widetilde{\mathbf{B}}_r^T\tilde{\mathbf{x}}_r \end{aligned} \tag{12}$$

The poles of $\mathbf{G}_r(s)$ are the eigenvalues of $\widetilde{\mathbf{A}}_r = \widetilde{\mathbf{J}}_r - \widetilde{\mathbf{R}}_r$, so for some invertible matrix $\mathbf{X}_r$: $\widetilde{\mathbf{A}}_r\mathbf{X}_r = \mathbf{X}_r\Lambda_r$ where $\Lambda_r = \mathsf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$ lists the poles of $\mathbf{G}_r(s)$. Given the structure of (12), writing $\widetilde{\mathbf{A}}_r = \widetilde{\mathbf{J}}_r - \widetilde{\mathbf{R}}_r$, notice that perturbations of $\widetilde{\mathbf{A}}_r \longrightarrow \widetilde{\mathbf{A}}_r^\varepsilon$ will remain port-Hamiltonian if and only if the numerical range of $\widetilde{\mathbf{A}}_r^\varepsilon$ remains in the closed left half-plane. Indeed, if such perturbations $\widetilde{\mathbf{A}}_r \longrightarrow \widetilde{\mathbf{A}}_r^\varepsilon$ depend continuously with respect to ε and we define $\widetilde{\mathbf{A}}_r^\varepsilon = \frac{1}{2}(\widetilde{\mathbf{A}}_r^\varepsilon - \widetilde{\mathbf{A}}_r^{*\varepsilon}) + \frac{1}{2}(\widetilde{\mathbf{A}}_r^\varepsilon + \widetilde{\mathbf{A}}_r^{*\varepsilon}) = \mathbf{J}_r^\varepsilon - \mathbf{R}_r^\varepsilon$ with skew Hermitian $\mathbf{J}_r^\varepsilon$ and Hermitian $\mathbf{R}_r^\varepsilon$, then $\mathbf{R}_r^\varepsilon$ is a perturbation of the (unperturbed) $\widetilde{\mathbf{R}}_r$, and since $\widetilde{\mathbf{R}}_r$ is assumed to be positive definite, $\mathbf{R}_r^\varepsilon$ will also be positive definite for all sufficiently small ε.

Define directions

$$[\mathsf{c}_1, \mathsf{c}_2, \ldots, \mathsf{c}_r] = \widetilde{\mathbf{B}}_r^T\mathbf{X}_r \quad \text{and} \quad \begin{bmatrix} \mathsf{d}_1^T \\ \mathsf{d}_2^T \\ \vdots \\ \mathsf{d}_r^T \end{bmatrix} = \mathbf{X}_r^{-1}\widetilde{\mathbf{B}}_r. \tag{13}$$

Write the reduced transfer function as

$$\mathbf{G}_r(s) = \widetilde{\mathbf{B}}_r^T \left(s\mathbf{I} - \widetilde{\mathbf{A}}_r\right)^{-1} \widetilde{\mathbf{B}}_r = \sum_{i=1}^{r} \frac{1}{s - \lambda_i} \mathsf{c}_i \mathsf{d}_i^T \tag{14}$$

Pick an index $1 \leq i \leq r$ and a vector $\mathbf{v} \in \mathbb{C}^r$ such that $v_i = \mathbf{e}_i^T \mathbf{v} = 0$, and then for $\varepsilon > 0$, consider perturbations to $\mathbf{G}_r(s)$ of the form:

$$\widehat{\mathbf{G}}_r(s) = \widetilde{\mathbf{B}}_r^T \mathbf{X}_r \left(s\mathbf{I} - \Lambda_r - \varepsilon e^{-\imath\theta} \mathbf{e}_i \mathbf{v}^T\right)^{-1} \mathbf{X}_r^{-1} \widetilde{\mathbf{B}}_r, \tag{15}$$

where θ is to be determined later.

Since the numerical range of $\widetilde{\mathbf{A}}_r$ lies in the open left halfplane, for all ε sufficiently small, $\widetilde{\mathbf{A}}_r + \varepsilon\, e^{-\imath\theta}\, \mathbf{X}_r \mathbf{e}_i \mathbf{v}^T \mathbf{X}_r^{-1}$ will have numerical range in the open left halfplane as well and can be decomposed into skew-symmetric/symmetric form, $\mathbf{J}_r - \mathbf{R}_r$ with $\mathbf{R}_r$ positive-definite. Thus, $\widehat{\mathbf{G}}_r(s)$ will have a port-Hamiltonian realization for all ε sufficiently small.

Now consider,

$$(s\mathbf{I} - \Lambda_r)^{-1} - \left(s\mathbf{I} - \Lambda_r - \varepsilon \mathbf{e}_i \mathbf{v}^T\right)^{-1} = -\varepsilon\, e^{-\imath\theta}\, (s\mathbf{I} - \Lambda_r)^{-1} \mathbf{e}_i \mathbf{v}^T (s\mathbf{I} - \Lambda_r)^{-1}.$$

Thus,

$$\begin{aligned} \mathbf{G}_r(s) - \widehat{\mathbf{G}}_r(s) &= -\,\varepsilon\, e^{-\imath\theta}\, \widetilde{\mathbf{B}}_r^T \mathbf{X}_r (s\mathbf{I} - \Lambda_r)^{-1} \mathbf{e}_i \mathbf{v}^T (s\mathbf{I} - \Lambda_r)^{-1} \mathbf{X}_r^{-1} \widetilde{\mathbf{B}}_r \\ &= \sum_{j \neq i} \frac{-\varepsilon\, e^{-\imath\theta}\, v_j}{(s - \lambda_i)(s - \lambda_j)} \mathsf{c}_i \mathsf{d}_j^T \end{aligned}$$

Suppose that $\langle \mathbf{G} - \mathbf{G}_r,\ \sum_{j \neq i} \frac{v_j}{(s-\lambda_i)(s-\lambda_j)} \mathsf{c}_i \mathsf{d}_j^T \rangle_{\mathcal{H}_2} \neq 0$ and define the θ in (15) as

$$\theta = \mathsf{arg} \langle \mathbf{G} - \mathbf{G}_r,\ \sum_{j \neq i} \frac{v_j}{(s - \lambda_i)(s - \lambda_j)} \mathsf{c}_i \mathsf{d}_j^T \rangle_{\mathcal{H}_2}.$$

$\mathcal{H}_2$-optimality of $\mathbf{G}_r(s)$ requires that

$$\|\mathbf{G}_r - \mathbf{G}_r\|_{\mathcal{H}_2} \leq \|\mathbf{G} - \widehat{\mathbf{G}}_r\|_{\mathcal{H}_2}$$

for all rth-order port-Hamiltonian systems, $\widehat{\mathbf{G}}_r(s)$, in a neighborhood of $\mathbf{G}_r(s)$. This in turn, implies

$$0 \leq 2\,\mathsf{Re}\,\langle \mathbf{G} - \mathbf{G}_r,\ \mathbf{G}_r - \widehat{\mathbf{G}}_r \rangle_{\mathcal{H}_2} + \|\mathbf{G}_r - \widehat{\mathbf{G}}_r\|_{\mathcal{H}_2}^2 \tag{16}$$

or

$$\mathsf{Re}\left(e^{-\imath\theta} \langle \mathbf{G} - \mathbf{G}_r,\ \sum_{j \neq i} \frac{v_j}{(s - \lambda_i)(s - \lambda_j)} \mathsf{c}_i \mathsf{d}_j^T \rangle_{\mathcal{H}_2} \right) \leq \frac{\varepsilon}{2} \left\| \sum_{j \neq i} \frac{v_j}{(s - \lambda_i)(s - \lambda_j)} \mathsf{c}_i \mathsf{d}_j^T \right\|_{\mathcal{H}_2}^2$$

or

$$\left|\langle \mathbf{G}-\mathbf{G}_r,\ \sum_{j\neq i}\frac{v_j}{(s-\lambda_i)(s-\lambda_j)}\mathsf{c}_i\mathsf{d}_j^T\rangle_{\mathcal{H}_2}\right| \leq \frac{\varepsilon}{2}\left\|\sum_{j\neq i}\frac{v_j}{(s-\lambda_i)(s-\lambda_j)}\mathsf{c}_i\mathsf{d}_j^T\right\|_{\mathcal{H}_2}^2 .$$

The left-hand side is independent of ε, which can be taken arbitrarily small. Thus we have

$$\begin{aligned}
\langle \mathbf{G}-\mathbf{G}_r,\ \sum_{j\neq i}\frac{v_j}{(s-\lambda_i)(s-\lambda_j)}\mathsf{c}_i\mathsf{d}_j^T\rangle_{\mathcal{H}_2} &=0\\
\sum_{j\neq i}\frac{v_j}{\lambda_i-\lambda_j}\langle \mathbf{G}-\mathbf{G}_r,\ \left(\frac{1}{s-\lambda_i}-\frac{1}{s-\lambda_j}\right)\mathsf{c}_i\mathsf{d}_j^T\rangle_{\mathcal{H}_2} &=0\\
\sum_{j\neq i}\frac{v_j}{\lambda_i-\lambda_j}\left[\mathsf{c}_i^T\mathbf{G}(-\lambda_i)\mathsf{d}_j-\mathsf{c}_i^T\mathbf{G}_r(-\lambda_i)\mathsf{d}_j-\left(\mathsf{c}_i^T\mathbf{G}(-\lambda_j)\mathsf{d}_j-\mathsf{c}_i^T\mathbf{G}_r(-\lambda_j)\mathsf{d}_j\right)\right] &=0\\
\sum_{j\neq i}v_j\left[\frac{\mathsf{c}_i^T\mathbf{G}(-\lambda_i)\mathsf{d}_j-\mathsf{c}_i^T\mathbf{G}(-\lambda_j)\mathsf{d}_j}{(-\lambda_i)-(-\lambda_j)}-\frac{\mathsf{c}_i^T\mathbf{G}_r(-\lambda_i)\mathsf{d}_j-\mathsf{c}_i^T\mathbf{G}_r(-\lambda_j)\mathsf{d}_j}{(-\lambda_i)-(-\lambda_j)}\right] &=0.
\end{aligned}$$

Since v_j was chosen arbitrarily, the conclusion holds for the case $i \neq j$.

Now for the case $i = j$, consider perturbations to $\mathbf{G}_r(s)$ of the form:

$$\widehat{\mathbf{G}}_r(s) = \widetilde{\mathbf{B}}_r^T\mathbf{X}_r\left(s\mathbf{I}-\Lambda_r-\varepsilon e^{-\imath\theta}\mathbf{e}_i\mathbf{e}_i^T\right)^{-1}\mathbf{X}_r^{-1}\widetilde{\mathbf{B}}_r, \tag{17}$$

where $\varepsilon > 0$ and θ is (as before) to be determined later. $\widehat{\mathbf{G}}_r(s)$ will have a port-Hamiltonian realization for all ε sufficiently small. Then

$$\mathbf{G}_r(s)-\widehat{\mathbf{G}}_r(s) = \left(\frac{1}{s-\lambda_i}-\frac{1}{s-(\lambda_i+\varepsilon\, e^{-\imath\theta})}\right)\mathsf{c}_i\mathsf{d}_i^T$$

With some calculation, we find:

$$\|\mathbf{G}_r-\widehat{\mathbf{G}}_r\|_{\mathcal{H}_2}^2 = \frac{\varepsilon^2\|\mathsf{c}_i\|^2\|\mathsf{d}_i\|^2}{4(-\mathsf{Re}\,\lambda_i)^3}+\mathcal{O}(\varepsilon^3)$$

$\mathcal{H}_2$-optimality of $\mathbf{G}_r(s)$ again requires that

$$0 \leq 2\,\mathsf{Re}\,\langle \mathbf{G}-\mathbf{G}_r,\ \mathbf{G}_r-\widehat{\mathbf{G}}_r\rangle_{\mathcal{H}_2}+\|\mathbf{G}_r-\widehat{\mathbf{G}}_r\|_{\mathcal{H}_2}^2.$$

Defining for the moment, $\Delta(s) = \mathsf{c}_i^T(\mathbf{G}(s)-\mathbf{G}_r(s))\mathsf{d}_i$, this means

$$\begin{aligned}
0 \leq 2\,\mathsf{Re}\left[\Delta(-\lambda_i)-\Delta(-\lambda_i-\varepsilon\, e^{-\imath\theta})\right]+\frac{\varepsilon^2}{4(-\mathsf{Re}\,\lambda_i)^3}+\mathcal{O}(\varepsilon^3) \quad \text{or}\\
0 \leq 2\,\mathsf{Re}\left(\varepsilon\, e^{-\imath\theta}\Delta'(-\lambda_i)+\mathcal{O}(\varepsilon^2)\right)+\frac{\varepsilon^2}{4(-\mathsf{Re}\,\lambda_i)^3}+\mathcal{O}(\varepsilon^3) \quad \text{or}\\
0 \leq \mathsf{Re}\left(e^{-\imath\theta}\Delta'(-\lambda_i)\right)+\mathcal{O}(\varepsilon)
\end{aligned}$$

If $\Delta'(-\lambda_i) \neq 0$ this leads to a contradiction if we pick $\theta = \mathsf{arg}(\Delta'(-\lambda_i))-\pi$. Thus, $\Delta'(-\lambda_i) = \mathsf{c}_i^T\mathbf{G}'(-\lambda_i)\mathsf{d}_i-\mathsf{c}_i^T\mathbf{G}_r'(-\lambda_i)\mathsf{d}_i = 0$ giving the conclusion for the case $i = j$. $\square$.

Theorem 2. *Suppose $\mathbf{G}_r$ is a solution to (11) and has simple poles. Using the notation of Theorem 1, define left and right residuals as:*

$$\mathbf{R} = [\mathbf{G}(-\lambda_1)\mathsf{d}_1 - \mathbf{G}_r(-\lambda_1)\mathsf{d}_1,\, \ldots,\, \mathbf{G}(-\lambda_r)\mathsf{d}_r - \mathbf{G}_r(-\lambda_r)\mathsf{d}_r]$$
$$\mathbf{L} = \left[\mathbf{G}(-\lambda_1)^T\mathsf{c}_1 - \mathbf{G}_r(-\lambda_1)^T\mathsf{c}_1,\, \ldots,\, \mathbf{G}(-\lambda_r)^T\mathsf{c}_r - \mathbf{G}_r(-\lambda_r)^T\mathsf{c}_r\right]$$

Then

$$\mathbf{R}\begin{bmatrix} \mathsf{c}_1^T \\ \vdots \\ \mathsf{c}_r^T \end{bmatrix} + \mathbf{L}\begin{bmatrix} \mathsf{d}_1^T \\ \vdots \\ \mathsf{d}_r^T \end{bmatrix} = \mathbf{0} \tag{18}$$

Proof: Proceed as in Theorem 1 with $\widetilde{\mathbf{A}}_r = \widetilde{\mathbf{J}}_r - \widetilde{\mathbf{R}}_r$ and $\mathbf{X}_r$ such that $\widetilde{\mathbf{A}}_r\mathbf{X}_r = \mathbf{X}_r\Lambda_r$ and $\Lambda_r = \mathsf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$ lists the poles of $\mathbf{G}_r(s)$. Residue directions are defined as in (13).

We consider perturbations to $\mathbf{G}_r(s)$ of the form:

$$\begin{aligned} \widehat{\mathbf{G}}_r(s) &= \left(\widetilde{\mathbf{B}}_r + \delta\widetilde{\mathbf{B}}_r\right)^T \left(s\mathbf{I} - \widetilde{\mathbf{A}}_r\right)^{-1} \left(\widetilde{\mathbf{B}}_r + \delta\widetilde{\mathbf{B}}_r\right) \\ &= \left(\widetilde{\mathbf{B}}_r + \delta\widetilde{\mathbf{B}}_r\right)^T \mathbf{X}_r\left(s\mathbf{I} - \Lambda_r\right)^{-1}\mathbf{X}_r^{-1}\left(\widetilde{\mathbf{B}}_r + \delta\widetilde{\mathbf{B}}_r\right) \end{aligned} \tag{19}$$

which evidently has a port-Hamiltonian realization for all choices of $\delta\widetilde{\mathbf{B}}_r$.

Observe that

$$\begin{aligned} \mathbf{G}_r(s) - \widehat{\mathbf{G}}_r(s) = &-\widetilde{\mathbf{B}}_r^T\mathbf{X}_r\left(s\mathbf{I} - \Lambda_r\right)^{-1}\mathbf{X}_r^{-1}\,\delta\widetilde{\mathbf{B}}_r \\ &-\delta\widetilde{\mathbf{B}}_r^T\mathbf{X}_r\left(s\mathbf{I} - \Lambda_r\right)^{-1}\mathbf{X}_r^{-1}\,\widetilde{\mathbf{B}}_r \\ &-\delta\widetilde{\mathbf{B}}_r^T\mathbf{X}_r\left(s\mathbf{I} - \Lambda_r\right)^{-1}\mathbf{X}_r^{-1}\,\delta\widetilde{\mathbf{B}}_r \\ = &-\sum_{i=1}^{r}\frac{1}{s-\lambda_i}\mathsf{c}_i\,\delta\mathsf{d}_i^T - \sum_{i=1}^{r}\frac{1}{s-\lambda_i}\,\delta\mathsf{c}_i\,\mathsf{d}_i^T - \sum_{i=1}^{r}\frac{1}{s-\lambda_i}\,\delta\mathsf{c}_i\,\delta\mathsf{d}_i^T, \end{aligned}$$

where we have defined

$$[\delta\mathsf{c}_1, \delta\mathsf{c}_2, \ldots, \delta\mathsf{c}_r] = \delta\widetilde{\mathbf{B}}_r^T\mathbf{X}_r \quad \text{and} \quad \begin{bmatrix} \delta\mathsf{d}_1^T \\ \delta\mathsf{d}_2^T \\ \vdots \\ \delta\mathsf{d}_r^T \end{bmatrix} = \mathbf{X}_r^{-1}\,\delta\widetilde{\mathbf{B}}_r.$$

Notice that

$$\mathbf{X}_r^T\mathbf{X}_r\begin{bmatrix} \delta\mathsf{d}_1^T \\ \delta\mathsf{d}_2^T \\ \vdots \\ \delta\mathsf{d}_r^T \end{bmatrix} = \begin{bmatrix} \delta\mathsf{c}_1^T \\ \delta\mathsf{c}_2^T \\ \vdots \\ \delta\mathsf{c}_r^T \end{bmatrix}.$$

Since the poles of $\mathbf{G}_r$ are closed under conjugation, there is a permutation matrix, $\mathbf{\Pi}$, such that $\mathbf{\Pi}\mathbf{X}_r^T = \mathbf{X}_r^*$ and $\mathbf{\Pi}^2 = \mathbf{I}$. Suppose $\mathbf{u} = [\upsilon_1, \upsilon_2, \ldots, \upsilon_r]^T$ is an eigenvector of the positive definite matrix $\mathbf{X}_r^*\mathbf{X}_r$: $\mathbf{X}_r^*\mathbf{X}_r\mathbf{u} = \mu\mathbf{u}$. Then $\mu > 0$, $(\mathbf{\Pi}\mathbf{X}_r^*\mathbf{X}_r\mathbf{\Pi})\,\mathbf{\Pi}\mathbf{u} = \mu\,\mathbf{\Pi}\mathbf{u}$, and $(\mathbf{\Pi}\mathbf{X}_r^*\mathbf{X}_r\mathbf{\Pi})\,\overline{\mathbf{u}} = \overline{\mathbf{X}_r^*\mathbf{X}_r}\,\overline{\mathbf{u}} = \mu\,\overline{\mathbf{u}}$. So,

$$\mathbf{X}_r^T\mathbf{X}_r\mathbf{u} = \mathbf{\Pi}\mathbf{X}_r^*\mathbf{X}_r\mathbf{u} = \mu\mathbf{\Pi}\mathbf{u} = \mu\overline{\mathbf{u}}.$$

Pick an index $1 \le \ell \le r$, $\varepsilon > 0$, and a $\theta \in [0, 2\pi]$. Then choose the perturbation so that

$$\begin{bmatrix} \delta\mathsf{d}_1^T \\ \delta\mathsf{d}_2^T \\ \vdots \\ \delta\mathsf{d}_r^T \end{bmatrix} = \varepsilon\, e^{-\imath\theta}\,\mathbf{u}\,\mathbf{e}_\ell^T \quad \Rightarrow \quad \begin{bmatrix} \delta\mathsf{c}_1^T \\ \delta\mathsf{c}_2^T \\ \vdots \\ \delta\mathsf{c}_r^T \end{bmatrix} = \varepsilon\, e^{-\imath\theta}\,\mu\,\overline{\mathbf{u}}\,\mathbf{e}_\ell^T$$

so that, in particular, $\delta\mathsf{d}_k = \varepsilon\, e^{-\imath\theta}\,\upsilon_k\,\mathbf{e}_\ell$ and $\delta\mathsf{c}_k = \varepsilon\, e^{-\imath\theta}\,\mu\,\overline{\upsilon_k}\,\mathbf{e}_\ell$.

We may directly compute

$$\begin{aligned}
\langle \mathbf{G} - \mathbf{G}_r,\ \mathbf{G}_r - \widehat{\mathbf{G}}_r\rangle_{\mathcal{H}_2} &= -\sum_{i=1}^r \mathsf{c}_i^T\,(\mathbf{G}(-\lambda_i) - \mathbf{G}_r(-\lambda_i))\ \delta\mathsf{d}_i \\
-\sum_{i=1}^r \delta\mathsf{c}_i^T\ (\mathbf{G}(-\lambda_i) - \mathbf{G}_r(-\lambda_i))\,\mathsf{d}_i &- \sum_{i=1}^r \delta\mathsf{c}_i^T\ (\mathbf{G}(-\lambda_i) - \mathbf{G}_r(-\lambda_i))\ \delta\mathsf{d}_i \\
&= -\,\varepsilon\, e^{-\imath\theta}\sum_{i=1}^r \upsilon_i\mathsf{c}_i^T\,(\mathbf{G}(-\lambda_i) - \mathbf{G}_r(-\lambda_i))\ \mathbf{e}_\ell \\
&\quad -\,\varepsilon\, e^{-\imath\theta}\sum_{i=1}^r \mu\,\overline{\upsilon_i}\,\mathbf{e}_\ell^T\ (\mathbf{G}(-\lambda_i) - \mathbf{G}_r(-\lambda_i))\,\mathsf{d}_i \\
&\quad -\,\varepsilon^2\, e^{-\imath 2\theta}\sum_{i=1}^r \mu\,|\upsilon_i|^2\,\mathbf{e}_\ell^T\ (\mathbf{G}(-\lambda_i) - \mathbf{G}_r(-\lambda_i))\ \mathbf{e}_\ell \\
&= -\,\varepsilon\, e^{-\imath\theta}\mathbf{e}_\ell^T\,(\mu\mathbf{R}\overline{\mathbf{u}} + \mathbf{L}\mathbf{u}) + \mathcal{O}(\varepsilon^2)
\end{aligned}$$

and

$$\begin{aligned}
\|\mathbf{G}_r - \widehat{\mathbf{G}}_r\|_{\mathcal{H}_2}^2 &= \varepsilon^2 \sum_{i,j} \frac{-1}{\lambda_i + \overline{\lambda_j}} \left[\sum_{k\neq\ell} \upsilon_i\mathsf{c}_i(k)\,\overline{\upsilon_j\mathsf{c}_j(k)} \right. \\
&\quad \left. +\mu(\upsilon_i\mathsf{c}_i(\ell) + \overline{\upsilon_i}\mathsf{d}_i(\ell))\overline{(\upsilon_j\mathsf{c}_j(\ell) + \overline{\upsilon_j}\mathsf{d}_j(\ell))} + \mu^2\sum_{k\neq\ell} \overline{\upsilon_i}\mathsf{d}_i(k)\,\overline{\overline{\upsilon_j}\mathsf{d}_j(k)} \right] + \mathcal{O}(\varepsilon^3) \\
&= \varepsilon^2 \underbrace{\left(\sum_{k\neq\ell} \mathbf{a}_k^*\mathbf{M}\mathbf{a}_k + \mu(\mathbf{a}_\ell + \mathbf{b}_\ell)^*\mathbf{M}(\mathbf{a}_\ell + \mathbf{b}_\ell) + \mu^2\sum_{k\neq\ell} \mathbf{b}_k^*\mathbf{M}\mathbf{b}_k \right)}_{M>0} + \mathcal{O}(\varepsilon^3) \\
&= M\,\varepsilon^2 + \mathcal{O}(\varepsilon^3),
\end{aligned}$$

where $\mathbf{a}_k = \upsilon_k \begin{Bmatrix} \mathsf{c}_1(k) \\ \mathsf{c}_2(k) \\ \vdots \\ \mathsf{c}_r(k) \end{Bmatrix}$, $\mathbf{b}_k = \overline{\upsilon_k} \begin{Bmatrix} \mathsf{d}_1(k) \\ \mathsf{d}_2(k) \\ \vdots \\ \mathsf{d}_r(k) \end{Bmatrix}$, and $\mathbf{M} = \left[\frac{-1}{\lambda_i + \overline{\lambda_j}}\right]$.

Note that $\mathbf{M}$ is a (positive definite) Cauchy matrix.

Choose $\theta = \arg\left(\mathbf{e}_\ell^T\left(\mu\mathbf{R}\overline{\mathbf{u}} + \mathbf{L}\mathbf{u}\right)\right)$ so that $e^{-\imath\theta}\mathbf{e}_\ell^T\left(\mu\mathbf{R}\overline{\mathbf{u}} + \mathbf{L}\mathbf{u}\right) = \left|\mathbf{e}_\ell^T\left(\mu\mathbf{R}\overline{\mathbf{u}} + \mathbf{L}\mathbf{u}\right)\right|$. Since $\mathbf{G}_r$ is $\mathcal{H}_2$-optimal, (16) holds and so,

$$\left|\mathbf{e}_\ell^T\left(\mu\mathbf{R}\overline{\mathbf{u}} + \mathbf{L}\mathbf{u}\right)\right| \leq M\,\varepsilon + \mathcal{O}(\varepsilon^2).$$

Thus, $\mathbf{e}_\ell^T\left(\mu\mathbf{R}\overline{\mathbf{u}} + \mathbf{L}\mathbf{u}\right) = 0$ and since this is true for each index $\ell = 1, \ldots, r$, we have $\mu\mathbf{R}\overline{\mathbf{u}} + \mathbf{L}\mathbf{u} = \mathbf{0}$. The same argument can be made using each eigenvalue/eigenvector pair of $\mathbf{X}_r^*\mathbf{X}_r$ to construct a perturbation, which can then be summarized together as

$$\mathbf{R}\overline{\mathbf{U}}\mathsf{diag}(\mu_1, \ldots, \mu_r) + \mathbf{L}\mathbf{U} = \mathbf{0},$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r]$ is the matrix of eigenvectors for $\mathbf{X}_r^*\mathbf{X}_r$. Indeed, we have

$$\mathbf{X}_r^T\mathbf{X}_r\,\mathbf{U} = \mathbf{\Pi}\,\mathbf{X}_r^*\mathbf{X}_r\,\mathbf{U} = \mathbf{\Pi}\,\mathbf{U}\,\mathsf{diag}(\mu_1, \ldots, \mu_r) = \overline{\mathbf{U}}\mathsf{diag}(\mu_1, \ldots, \mu_r),$$

so we have in turn

$$\mathbf{R}\mathbf{X}_r^T\mathbf{X}_r\mathbf{U} + \mathbf{L}\mathbf{U} = \mathbf{0} \;\Rightarrow\; \left(\mathbf{R}\mathbf{X}_r^T + \mathbf{L}\mathbf{X}_r^{-1}\right)\mathbf{X}_r\mathbf{U} = \mathbf{0} \;\Rightarrow\; \mathbf{R}\mathbf{X}_r^T + \mathbf{L}\mathbf{X}_r^{-1} = \mathbf{0}.$$

Postmultiplying by $\widetilde{\mathbf{B}}_r$ and using (13) yields the conclusion. $\square$

5 Optimal Second-order Modally-damped Systems

Let $\mathcal{Q}_r \subset \mathcal{H}_2^{m\times p}$ be defined as the set of $m \times p$ matrix-valued transfer functions associated with second-order, modally damped dynamical systems:

$$\mathcal{Q}_r = \left\{ \mathbf{C}_r(s^2\mathbf{M}_r + s\mathbf{D}_r + \mathbf{K}_r)^{-1}\mathbf{B}_r \;\middle|\; \begin{array}{l} \mathbf{M}_r, \mathbf{D}_r, \mathbf{K}_r \in \mathbb{R}^{r\times r} \text{ SPD}, \\ \mathbf{B}_r \in \mathbb{R}^{r\times p}, \mathbf{C}_r \in \mathbb{R}^{m\times r}, \\ \text{and } \mathbf{D}_r\mathbf{M}_r^{-1}\mathbf{K}_r = \mathbf{K}_r\mathbf{M}_r^{-1}\mathbf{D}_r \end{array} \right\}$$

(SPD means Symmetric Positive Definite).

Suppose $\widehat{\mathbf{H}}_r(s) = \widehat{\mathbf{C}}_r(s^2\widehat{\mathbf{M}}_r + s\widehat{\mathbf{D}}_r + \widehat{\mathbf{K}}_r)^{-1}\widehat{\mathbf{B}}_r \in \mathcal{Q}_r$ solves

$$\left\|\mathbf{H} - \widehat{\mathbf{H}}_r\right\|_{\mathcal{H}_2} = \min_{\mathbf{H}_r \in \mathcal{Q}_r} \|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2}. \tag{20}$$

Let $\widehat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_r] \in \mathbb{R}^{r\times r}$ be a matrix of eigenvectors solving the generalized eigenvalue problem $\widehat{\mathbf{K}}_r\hat{\mathbf{x}}_i = \omega_i^2\widehat{\mathbf{M}}_r\hat{\mathbf{x}}_i$, represented collectively as $\widehat{\mathbf{K}}_r\widehat{\mathbf{X}} = \widehat{\mathbf{M}}_r\widehat{\mathbf{X}}\mathbf{\Omega}^2$ for

$\mathbf{\Omega}^2 = \mathrm{diag}(\omega_1^2, \omega_2^2, \ldots, \omega_r^2)$. Assume without loss of generality that eigenvectors are normalized so that

$$\widehat{\mathbf{X}}^T \widehat{\mathbf{K}}_r \widehat{\mathbf{X}} = \mathbf{\Omega} \quad \text{and} \quad \widehat{\mathbf{X}}^T \widehat{\mathbf{M}}_r \widehat{\mathbf{X}} = \mathbf{\Omega}^{-1}.$$

Since $\widehat{\mathbf{D}}_r \widehat{\mathbf{M}}_r^{-1} \widehat{\mathbf{K}}_r = \widehat{\mathbf{K}}_r \widehat{\mathbf{M}}_r^{-1} \widehat{\mathbf{D}}_r$, evidently $\widehat{\mathbf{M}}_r^{-1} \widehat{\mathbf{D}}_r$ commutes with $\widehat{\mathbf{M}}_r^{-1} \widehat{\mathbf{K}}_r$ and so may be simultaneously diagonalized, implying for an appropriate choice of positive scalars (damping ratios), $\{\xi_1, \xi_2, \ldots, \xi_r\}$,

$$\widehat{\mathbf{D}}_r \widehat{\mathbf{X}} = \widehat{\mathbf{M}}_r \widehat{\mathbf{X}} \, (2\mathbf{\Xi}) \quad \text{with} \quad \mathbf{\Xi} = \mathrm{diag}(\xi_1, \xi_2, \ldots, \xi_r).$$

Thus,

$$\begin{aligned}
\widehat{\mathbf{H}}_r(s) =& \widehat{\mathbf{C}}_r (s^2 \widehat{\mathbf{M}}_r + s \widehat{\mathbf{D}}_r + \widehat{\mathbf{K}}_r)^{-1} \widehat{\mathbf{B}}_r \\
=& \widehat{\mathbf{C}}_r (s^2 \widehat{\mathbf{X}}^{-T} \mathbf{\Omega}^{-1} \widehat{\mathbf{X}}^{-1} + 2\, s\, \widehat{\mathbf{X}}^{-T} \mathbf{\Xi} \widehat{\mathbf{X}}^{-1} + \widehat{\mathbf{X}}^{-T} \mathbf{\Omega} \widehat{\mathbf{X}}^{-1})^{-1} \widehat{\mathbf{B}}_r \\
=& \widehat{\mathbf{C}}_r \widehat{\mathbf{X}} (s^2 \mathbf{\Omega}^{-1} + 2\, s\, \mathbf{\Xi} + \mathbf{\Omega})^{-1} \widehat{\mathbf{X}}^T \widehat{\mathbf{B}}_r \\
=& \sum_{k=1}^{r} \frac{\omega_k \, \phi_k \; \mathsf{c}_k \, \mathsf{b}_k^T}{s^2 + 2\, \xi_k \, \omega_k \, s + \omega_k^2} = \sum_{k=1}^{r} \frac{\omega_k \, \phi_k \; \mathsf{c}_k \, \mathsf{b}_k^T}{(s - \lambda_k^+)(s - \lambda_k^-)};
\end{aligned}$$

the vector residues are defined so that

$$\widehat{\mathbf{C}}_r \widehat{\mathbf{X}} \cdot \widehat{\mathbf{X}}^T \widehat{\mathbf{B}}_r = [\mathsf{c}_1, \mathsf{c}_2, \ldots, \mathsf{c}_r] \; \mathrm{diag}(\phi_1, \phi_2, \ldots, \phi_r) [\mathsf{b}_1, \mathsf{b}_2, \ldots, \mathsf{b}_r]^T;$$

where scale factors, $\phi_k \geq 0$, are introduced so that $\|\mathsf{c}_k\| = \|\mathsf{b}_k\| = 1$.

For reference note that:

$$\frac{1}{(s - \lambda^+)(s - \lambda^-)} = \frac{1}{\lambda^+ - \lambda^-} \left(\frac{1}{s - \lambda^+} - \frac{1}{s - \lambda^-} \right) \tag{21}$$

$$\begin{aligned}
\frac{1}{(s - \lambda^+)^2 (s - \lambda^-)^2} =& \frac{-2}{(\lambda^+ - \lambda^-)^3} \left(\frac{1}{s - \lambda^+} - \frac{1}{s - \lambda^-} \right) \\
&+ \frac{1}{(\lambda^+ - \lambda^-)^2} \left(\frac{1}{(s - \lambda^+)^2} + \frac{1}{(s - \lambda^-)^2} \right)
\end{aligned} \tag{22}$$

$$\frac{s}{(s - \lambda^+)(s - \lambda^-)} = \frac{1}{\lambda^+ - \lambda^-} \left(\frac{\lambda^+}{s - \lambda^+} - \frac{\lambda^-}{s - \lambda^-} \right) \tag{23}$$

$$\begin{aligned}
\frac{s}{(s - \lambda^+)^2 (s - \lambda^-)^2} =& - \frac{\lambda^+ + \lambda^-}{(\lambda^+ - \lambda^-)^3} \left(\frac{1}{s - \lambda^+} - \frac{1}{s - \lambda^-} \right) \\
&+ \frac{1}{(\lambda^+ - \lambda^-)^2} \left(\frac{\lambda^+}{(s - \lambda^+)^2} + \frac{\lambda^-}{(s - \lambda^-)^2} \right)
\end{aligned} \tag{24}$$

$$\frac{s^2}{(s-\lambda^+)^2(s-\lambda^-)^2} = \frac{1}{(\lambda^+-\lambda^-)^2}\left(\left(\frac{\lambda^-}{s-\lambda^-}\right)^2+\left(\frac{\lambda^+}{s-\lambda^+}\right)^2\right) - \frac{2\ \lambda^-\ \lambda^+}{(\lambda^+-\lambda^-)^3}\left(\frac{1}{s-\lambda^+}-\frac{1}{s-\lambda^-}\right) \tag{25}$$

Defining the auxiliary quantity $\rho_k = \xi_k - \sqrt{\xi_k^2-1}$, the poles are

$$\lambda_k^+ = -\xi_k\omega_k + \omega_k\sqrt{\xi_k^2-1} = -\frac{\omega_k}{\rho_k} \quad \text{and}$$
$$\lambda_k^- = -\xi_k\omega_k - \omega_k\sqrt{\xi_k^2-1} = -\omega_k\ \rho_k.$$

where $\omega_k > 0$ (as before) and either ρ_k is real with $\rho_k \in (0,1]$ or ρ_k is complex with $\rho_k = e^{-\imath\theta}$ for $\theta \in (0, \frac{\pi}{2})$.

Lemma 2. *Suppose* $\mathbf{G}(s) \in \mathcal{H}_2^{m\times p}$*; choose* $\mathsf{c} \in \mathbb{R}^m$, $\mathsf{b} \in \mathbb{R}^p$*; and let*

$$\lambda^\pm = -\xi\omega \pm \omega\sqrt{\xi^2-1} \textit{ for } \xi,\, \omega > 0 \textit{ and } \xi \neq 1.$$

Define

$$\mathbf{\Delta}_1(s) = \frac{\mathsf{c}\,\mathsf{b}^T}{(s-\lambda^+)(s-\lambda^-)} \in \mathcal{H}_2^{m\times p} \textit{ and } \mathbf{\Delta}_2(s) = \frac{\mathsf{c}\,\mathsf{b}^T}{(s-\lambda^+)^2(s-\lambda^-)^2} \in \mathcal{H}_2^{m\times p}.$$

Then

$$\|\mathbf{\Delta}_1\|_{\mathcal{H}_2} = \frac{\|\mathsf{c}\|\cdot\|\mathsf{b}\|}{2\omega\sqrt{\xi\omega}} \quad \textit{and} \quad \langle \mathbf{G},\, \mathbf{\Delta}_1\rangle_{\mathcal{H}_2} = \frac{\mathsf{c}^T\overline{\mathbf{G}}(-\lambda^+)\mathsf{b} - \mathsf{c}^T\overline{\mathbf{G}}(-\lambda^-)\mathsf{b}}{\lambda^+-\lambda^-}.$$

$$\|s\mathbf{\Delta}_1\|_{\mathcal{H}_2} = \frac{\|\mathsf{c}\|\cdot\|\mathsf{b}\|}{2\omega\sqrt{\xi\omega}} \quad \textit{and} \quad \langle \mathbf{G},\, s\mathbf{\Delta}_1\rangle_{\mathcal{H}_2} = \frac{\mathsf{c}^T\overline{\mathbf{G}}(-\lambda^+)\mathsf{b} - \mathsf{c}^T\overline{\mathbf{G}}(-\lambda^-)\mathsf{b}}{\lambda^+-\lambda^-}.$$

$$\frac{s^2}{(s-\lambda^+)^2(s-\lambda^-)^2} = \frac{1}{(\lambda^+-\lambda^-)^2}\left(\frac{(\lambda^+)^2}{(s-\lambda^+)^2}+\frac{(\lambda^-)^2}{(s-\lambda^-)^2}\right) - \frac{2\ \lambda^-\ \lambda^+}{(\lambda^+-\lambda^-)^3}\left(\frac{1}{s-\lambda^+}-\frac{1}{s-\lambda^-}\right) \tag{26}$$

$$\langle \mathbf{G},\, \mathbf{\Delta}_2\rangle_{\mathcal{H}_2} = \frac{-1}{(\lambda^+-\lambda^-)^2}\left(\mathsf{c}^T\overline{\mathbf{G}}'(-\lambda^+)\mathsf{b} + \mathsf{c}^T\overline{\mathbf{G}}'(-\lambda^-)\mathsf{b}\right) - \frac{2}{(\lambda^+-\lambda^-)^3}\left(\mathsf{c}^T\overline{\mathbf{G}}(-\lambda^+)\mathsf{b} - \mathsf{c}^T\overline{\mathbf{G}}(-\lambda^-)\mathsf{b}\right).$$

PROOF: Notice that the function $\mathsf{trace}\big(\overline{\mathbf{G}}(-s)\boldsymbol{\Delta}_m(s)^T\big)$ has singularities in the left half plane only at λ^+ and λ^-. For any $R > 0$, define the semicircular contour in the left halfplane:

$$\Gamma_R = \{z\,|z = \imath\omega \text{ with } \omega \in [-R, R]\} \cup \left\{z \,\middle|\, z = R\,e^{\imath\theta} \text{ with } \theta \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]\right\}.$$

Γ_R bounds a region that for sufficiently large R contains λ^+ and λ^- and so, by the residue theorem (for $m = 1, 2$)

$$\begin{aligned}\langle \mathbf{G},\ \boldsymbol{\Delta}_m\rangle_{\mathcal{H}_2} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathsf{trace}\big(\overline{\mathbf{G}}(-\imath\omega)\boldsymbol{\Delta}_m(\imath\omega)^T\big)\ d\omega \\ &= \lim_{R\to\infty}\frac{1}{2\pi\imath}\int_{\Gamma_R} \mathsf{trace}\big(\overline{\mathbf{G}}(-s)\boldsymbol{\Delta}_m(s)^T\big)\ ds \\ &= \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s-\lambda^+)^m(s-\lambda^-)^m}, \lambda^+\right] + \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s-\lambda^+)^m(s-\lambda^-)^m}, \lambda^-\right].\end{aligned}$$

For $\boldsymbol{\Delta}_1$, consider first the case $\xi \neq 1$, so that $\lambda^- \neq \lambda^+$. One directly calculates:

$$\begin{aligned}\langle \mathbf{G},\ \boldsymbol{\Delta}_1\rangle_{\mathcal{H}_2} &= \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s-\lambda^+)(s-\lambda^-)}, \lambda^+\right] + \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s-\lambda^+)(s-\lambda^-)}, \lambda^-\right] \\ &= \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{\lambda^+-\lambda^-}\left(\frac{1}{s-\lambda^+} - \frac{1}{s-\lambda^-}\right), \lambda^+\right] \\ &\qquad + \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{\lambda^+-\lambda^-}\left(\frac{1}{s-\lambda^+} - \frac{1}{s-\lambda^-}\right), \lambda^-\right] \\ &= \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{\lambda^+-\lambda^-}\left(\frac{1}{s-\lambda^+}\right), \lambda^+\right] - \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{\lambda^+-\lambda^-}\left(\frac{1}{s-\lambda^-}\right), \lambda^-\right] \\ &= \frac{\mathsf{c}^T\overline{\mathbf{G}}(-\lambda^+)\mathsf{b} - \mathsf{c}^T\overline{\mathbf{G}}(-\lambda^-)\mathsf{b}}{\lambda^+-\lambda^-}.\end{aligned}$$

Setting $\mathbf{G} = \boldsymbol{\Delta}_1$ in this expression, we find

$$\mathsf{c}^T\overline{\boldsymbol{\Delta}_1}(-\lambda^+)\mathsf{b} = \frac{\mathsf{c}^T\mathsf{c}\ \mathsf{b}^T\mathsf{b}}{(-\lambda^+-\overline{\lambda^+})(-\lambda^+-\overline{\lambda^-})} = \frac{\|\mathsf{c}\|^2\ \|\mathsf{b}\|^2}{-4\xi\omega\cdot\lambda^+}$$

$$\mathsf{c}^T\overline{\boldsymbol{\Delta}_1}(-\lambda^-)\mathsf{b} = \frac{\mathsf{c}^T\mathsf{c}\ \mathsf{b}^T\mathsf{b}}{(-\lambda^--\overline{\lambda^+})(-\lambda^--\overline{\lambda^-})} = \frac{\|\mathsf{c}\|^2\ \|\mathsf{b}\|^2}{-4\xi\omega\cdot\lambda^-}$$

Writing $s^2 + 2\xi\omega s + \omega^2 = (s-\lambda^+)(s-\lambda^-)$, the last equality in each case comes from noticing that regardless of whether $\lambda^\pm$ is real or a conjugate pair, in either case we have that

$$(-\lambda^+ - \overline{\lambda^+})(-\lambda^+ - \overline{\lambda^-}) = -4\xi\omega\cdot\lambda^+ \text{ and}$$

$$(-\lambda^- - \overline{\lambda^+})(-\lambda^- - \overline{\lambda^-}) = -4\xi\omega\cdot\lambda^-.$$

So we may calculate,

$$\|\boldsymbol{\Delta}_1\|_{\mathcal{H}_2}^2 = \left\langle \boldsymbol{\Delta}_1,\ \frac{\mathsf{c}\,\mathsf{b}^T}{(s-\lambda^+)(s-\lambda^-)} \right\rangle_{\mathcal{H}_2} = \frac{\mathsf{c}^T\overline{\boldsymbol{\Delta}_1}(-\lambda^+)\mathsf{b} - \mathsf{c}^T\overline{\boldsymbol{\Delta}_1}(-\lambda^-)\mathsf{b}}{\lambda^+ - \lambda^-}$$
$$= \frac{-\|\mathsf{c}\|^2\ \|\mathsf{b}\|^2}{4\xi\omega(\lambda^+ - \lambda^-)}\left(\frac{1}{\lambda^+} - \frac{1}{\lambda^-}\right) = \frac{\|\mathsf{c}\|^2\ \|\mathsf{b}\|^2}{4\xi\omega(\lambda^+\lambda^-)} = \frac{\|\mathsf{c}\|^2\ \|\mathsf{b}\|^2}{4\xi\omega^3}$$

Now for the case that $\xi = 1$, $\lambda^+ = \lambda^- = -\omega$ and we have first

$$\langle \mathbf{G},\ \boldsymbol{\Delta}_1\rangle_{\mathcal{H}_2} = \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s+\omega)^2}, -\omega\right] = -\mathsf{c}^T\overline{\mathbf{G}}'(\omega)\mathsf{b}.$$

and likewise,

$$\|\boldsymbol{\Delta}_1\|_{\mathcal{H}_2}^2 = \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\boldsymbol{\Delta}_1}(-s)\mathsf{b}}{(s+\omega)^2}, -\omega\right] = -\mathsf{c}^T\overline{\boldsymbol{\Delta}_1}'(\omega)\mathsf{b} = \frac{\|\mathsf{c}\|^2\ \|\mathsf{b}\|^2}{4\omega^3}.$$

Pleasantly, both expressions match the general cases above as $\xi \to 1$.

$$\langle \mathbf{G},\ s\boldsymbol{\Delta}_1\rangle_{\mathcal{H}_2} = \mathsf{res}\left[\frac{s\,\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s-\lambda^+)(s-\lambda^-)}, \lambda^+\right] + \mathsf{res}\left[\frac{s\,\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s-\lambda^+)(s-\lambda^-)}, \lambda^-\right]$$

$$= \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{\lambda^+ - \lambda^-}\left(\frac{\lambda^+}{s-\lambda^+} - \frac{\lambda^-}{s-\lambda^-}\right), \lambda^+\right]$$
$$+ \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{\lambda^+ - \lambda^-}\left(\frac{\lambda^+}{s-\lambda^+} - \frac{\lambda^-}{s-\lambda^-}\right), \lambda^-\right]$$

$$= \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{\lambda^+ - \lambda^-}\left(\frac{\lambda^+}{s-\lambda^+}\right), \lambda^+\right] - \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{\lambda^+ - \lambda^-}\left(\frac{\lambda^-}{s-\lambda^-}\right), \lambda^-\right]$$

$$= \mathsf{c}^T\overline{\mathbf{G}}(-\lambda^+)\mathsf{b} - \mathsf{c}^T\overline{\mathbf{G}}(-\lambda^-)\mathsf{b}.$$

Now take $\mathbf{\Delta}_2 = \dfrac{\mathsf{c}\,\mathsf{b}^T}{(s-\lambda^+)^2(s-\lambda^-)^2}$, consider $\xi \neq 1$, and evaluate

$$\langle \mathbf{G},\, \mathbf{\Delta}_2\rangle_{\mathcal{H}_2} = \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s-\lambda^+)^2(s-\lambda^-)^2}, \lambda^+\right] + \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s-\lambda^+)^2(s-\lambda^-)^2}, \lambda^-\right]$$

$$= \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(\lambda^+-\lambda^-)^2}\frac{1}{(s-\lambda^+)^2}, \lambda^+\right] - \mathsf{res}\left[\frac{2\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(\lambda^+-\lambda^-)^3}\frac{1}{s-\lambda^+}, \lambda^+\right]$$
$$+\mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(\lambda^+-\lambda^-)^2}\frac{1}{(s-\lambda^-)^2}, \lambda^-\right] + \mathsf{res}\left[\frac{2\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(\lambda^+-\lambda^-)^3}\frac{1}{s-\lambda^-}, \lambda^-\right]$$

$$= -\frac{\mathsf{c}^T\overline{\mathbf{G}}'(-\lambda^+)\mathsf{b}}{(\lambda^+-\lambda^-)^2} - 2\frac{\mathsf{c}^T\overline{\mathbf{G}}(-\lambda^+)\mathsf{b}}{(\lambda^+-\lambda^-)^3} - \frac{\mathsf{c}^T\overline{\mathbf{G}}'(-\lambda^-)\mathsf{b}}{(\lambda^+-\lambda^-)^2} + 2\frac{\mathsf{c}^T\overline{\mathbf{G}}(-\lambda^-)\mathsf{b}}{(\lambda^+-\lambda^-)^3}$$

$$= -\frac{\mathsf{c}^T\overline{\mathbf{G}}'(-\lambda^+)\mathsf{b} + \mathsf{c}^T\overline{\mathbf{G}}'(-\lambda^-)\mathsf{b}}{(\lambda^+-\lambda^-)^2} - 2\frac{\mathsf{c}^T\overline{\mathbf{G}}(-\lambda^+)\mathsf{b} - \mathsf{c}^T\overline{\mathbf{G}}(-\lambda^-)\mathsf{b}}{(\lambda^+-\lambda^-)^3} \tag{27}$$

Set $\mathbf{G} = \mathbf{\Delta}_2$ in this expression to find,

$$\|\mathbf{\Delta}_2\|^2_{\mathcal{H}_2} = \frac{-1}{(\lambda^+-\lambda^-)^2}\Bigg(\mathsf{c}^T\overline{\mathbf{\Delta}_2}'(-\lambda^+)\mathsf{b} + \mathsf{c}^T\overline{\mathbf{\Delta}_2}'(-\lambda^-)\mathsf{b}$$
$$+ 2\left(\frac{\mathsf{c}^T\overline{\mathbf{\Delta}_2}(-\lambda^+)\mathsf{b} - \mathsf{c}^T\overline{\mathbf{\Delta}_2}(-\lambda^-)\mathsf{b}}{\lambda^+-\lambda^-}\right)\Bigg) = \frac{1+4\xi^2}{32\xi^3\omega^7}\|\mathsf{c}\|^2\|\mathsf{b}\|^2$$

Finally,

$$\langle \mathbf{G},\, s\mathbf{\Delta}_2\rangle_{\mathcal{H}_2} = \mathsf{res}\left[\frac{s\,\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s-\lambda^+)^2(s-\lambda^-)^2}, \lambda^+\right] + \mathsf{res}\left[\frac{s\,\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(s-\lambda^+)^2(s-\lambda^-)^2}, \lambda^-\right]$$

$$= \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(\lambda^+-\lambda^-)^2}\left(\frac{\lambda^+}{(s-\lambda^+)^2} + \frac{\lambda^-}{(s-\lambda^-)^2}\right), \lambda^+\right]$$
$$- \mathsf{res}\left[\frac{\lambda^++\lambda^-}{(\lambda^+-\lambda^-)^3}\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}\left(\frac{1}{s-\lambda^+} - \frac{1}{s-\lambda^-}\right), \lambda^+\right]$$
$$+ \mathsf{res}\left[\frac{\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}}{(\lambda^+-\lambda^-)^2}\left(\frac{\lambda^+}{(s-\lambda^+)^2} + \frac{\lambda^-}{(s-\lambda^-)^2}\right), \lambda^-\right]$$
$$- \mathsf{res}\left[\frac{\lambda^++\lambda^-}{(\lambda^+-\lambda^-)^3}\mathsf{c}^T\overline{\mathbf{G}}(-s)\mathsf{b}\left(\frac{1}{s-\lambda^+} - \frac{1}{s-\lambda^-}\right), \lambda^-\right]$$

$$= -\frac{1}{(\lambda^+-\lambda^-)^2}\left(\lambda^+\mathsf{c}^T\overline{\mathbf{G}}'(-\lambda^+)\mathsf{b} + \lambda^-\mathsf{c}^T\overline{\mathbf{G}}'(-\lambda^-)\mathsf{b}\right) \tag{28}$$
$$- \frac{\lambda^++\lambda^-}{(\lambda^+-\lambda^-)^3}\left(\mathsf{c}^T\overline{\mathbf{G}}(-\lambda^+)\mathsf{b} - \mathsf{c}^T\overline{\mathbf{G}}(-\lambda^-)\mathsf{b}\right) \qquad \square$$

This lemma provides a useful collection of technical facts that aid in calculating necessary optimality conditions.

5.1 Perturbations in residues for second-order systems.

Necessary conditions typically are derived variationally by observing stationarity in error norms under various categories of perturbations. Pick $1 \leq k \leq r$, consider an arbitrary $\delta\mathsf{c}_k \in \mathbb{R}^m$ with $\|\delta\mathsf{c}_k\| = 1$, and define a perturbation of $\widehat{\mathbf{H}}_r$ that involves a small variation in the kth left vector residue in the direction of $\delta\mathsf{c}_k$:

$$\begin{aligned}\widehat{\mathbf{H}}_r^{(\varepsilon)} &= \widehat{\mathbf{H}}_r - \left(\frac{\omega_k\,\phi_k\,\mathsf{c}_k\,\mathsf{b}_k^T}{(s-\lambda_k^+)(s-\lambda_k^-)} - \frac{\omega_k\,\phi_k\,(\mathsf{c}_k + \varepsilon\delta\mathsf{c}_k)\,\mathsf{b}_k^T}{(s-\lambda_k^+)(s-\lambda_k^-)}\right)\\ &= \widehat{\mathbf{H}}_r + \varepsilon\frac{\omega_k\,\phi_k}{(s-\lambda_k^+)(s-\lambda_k^-)}\delta\mathsf{c}_k\,\mathsf{b}_k^T\end{aligned}$$

Optimality of $\widehat{\mathbf{H}}_r$ implies that as $\varepsilon \to 0$:

$$\left|\left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\ \varepsilon\frac{\omega_k\,\phi_k\ \delta\mathsf{c}_k\,\mathsf{b}_k^T}{(s-\lambda_k^+)(s-\lambda_k^-)}\right\rangle_{\mathcal{H}_2}\right| = \mathcal{O}(\varepsilon^2),$$

which implies in turn,

$$\begin{aligned}0 &= \left|\left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\ \frac{\delta\mathsf{c}_k\,\mathsf{b}_k^T}{(s-\lambda_k^+)(s-\lambda_k^-)}\right\rangle_{\mathcal{H}_2}\right|\\ &= \frac{1}{\lambda_k^+ - \lambda_k^-}\ \delta\mathsf{c}_k^T\left(\left(\mathbf{H}(-\lambda_k^+) - \widehat{\mathbf{H}}_r(-\lambda_k^+)\right)\mathsf{b}_k - \left(\mathbf{H}(-\lambda_k^-) - \widehat{\mathbf{H}}_r(-\lambda_k^-)\right)\mathsf{b}_k\right).\end{aligned}$$

Equivalently, since $\delta\mathsf{c}_k$ is arbitrary,

$$\boxed{\left(\mathbf{H}(-\lambda_k^+) - \mathbf{H}(-\lambda_k^-)\right)\mathsf{b}_k = \left(\widehat{\mathbf{H}}_r(-\lambda_k^+) - \widehat{\mathbf{H}}_r(-\lambda_k^-)\right)\mathsf{b}_k,} \tag{29}$$

for each $k = 1, \ldots, r$. Notice this is a "distributed" interpolation condition acting across pairs of reduced order poles. We may pursue a similar analysis using perturbations in b_k to find analogous left interpolation conditions:

$$\boxed{\mathsf{c}_k^T\left(\mathbf{H}(-\lambda_k^+) - \mathbf{H}(-\lambda_k^-)\right) = \mathsf{c}_k^T\left(\widehat{\mathbf{H}}_r(-\lambda_k^+) - \widehat{\mathbf{H}}_r(-\lambda_k^-)\right),} \tag{30}$$

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

5.2 Perturbations in poles for second-order systems.

Pick $1 \leq k \leq r$ and consider the effect of arbitrary (but small) perturbations in ξ_k and ω_k. Suppressing k-dependence in ξ_k and ω_k (likewise in $\lambda_k^{\pm}$ and associated residues, ϕ_k, c_k, and b_k):

$$\begin{aligned}
\widehat{\mathbf{H}}_r^{(\varepsilon)} &= \widehat{\mathbf{H}}_r - \left(\frac{\omega\,\phi\,\mathsf{c}\,\mathsf{b}^T}{s^2+2\xi\omega s+\omega^2} - \frac{(\omega+\delta\omega)\,\phi\,\mathsf{c}\,\mathsf{b}^T}{s^2+2(\xi+\delta\xi)(\omega+\delta\omega)s+(\omega+\delta\omega)^2}\right) \\
&= \widehat{\mathbf{H}}_r - \frac{\omega\,\phi\,\mathsf{c}\,\mathsf{b}^T}{s^2+2\xi\omega s+\omega^2}\left(1 - \frac{\left(1+\frac{\delta\omega}{\omega}\right)}{\left(\frac{s^2+2(\xi+\delta\xi)(\omega+\delta\omega)s+(\omega+\delta\omega)^2}{s^2+2\xi\omega s+\omega^2}\right)}\right) \\
&= \widehat{\mathbf{H}}_r - \frac{\omega\,\phi\,\mathsf{c}\,\mathsf{b}^T}{s^2+2\xi\omega s+\omega^2}\left(1 - \frac{\left(1+\frac{\delta\omega}{\omega}\right)}{1+\left(\frac{2(\xi\delta\omega+\omega\delta\xi)s+2\omega\delta\omega}{s^2+2\xi\omega s+\omega^2}\right)}\right) + \mathcal{O}\left((\delta\omega)^2+(\delta\xi)^2\right) \\
&= \widehat{\mathbf{H}}_r - \frac{\omega\,\phi\,\mathsf{c}\,\mathsf{b}^T}{s^2+2\xi\omega s+\omega^2}\left(1 - \left(1+\frac{\delta\omega}{\omega}\right)\left(1-\frac{2(\xi\delta\omega+\omega\delta\xi)s+2\omega\delta\omega}{s^2+2\xi\omega s+\omega^2}\right)\right) \\
&\qquad + \mathcal{O}\left((\delta\omega)^2+(\delta\xi)^2\right) \\
&= \widehat{\mathbf{H}}_r - \frac{\omega\,\phi\,\mathsf{c}\,\mathsf{b}^T}{s^2+2\xi\omega s+\omega^2}\left(1 - \left(1+\frac{\delta\omega}{\omega}\right)\left(1-\frac{2(\xi\delta\omega+\omega\delta\xi)s+2\omega\delta\omega}{s^2+2\xi\omega s+\omega^2}\right)\right) \\
&\qquad + \mathcal{O}\left((\delta\omega)^2+(\delta\xi)^2\right) \\
&= \widehat{\mathbf{H}}_r - \frac{\omega\,\phi\,\mathsf{c}\,\mathsf{b}^T}{s^2+2\xi\omega s+\omega^2}\left(-\frac{\delta\omega}{\omega}+\frac{2(\xi\delta\omega+\omega\delta\xi)s+2\omega\delta\omega}{s^2+2\xi\omega s+\omega^2}\right) + \mathcal{O}\left((\delta\omega)^2+(\delta\xi)^2\right) \\
&= \widehat{\mathbf{H}}_r + (\mathbf{\Delta}_1(s) - 2\,\omega^2\mathbf{\Delta}_2(s))\phi\;\delta\omega \;-2\,s\mathbf{\Delta}_2(s)\,\omega\,\phi\,(\xi\delta\omega+\omega\delta\xi) + \mathcal{O}\left((\delta\omega)^2+(\delta\xi)^2\right)
\end{aligned}$$

where $\delta\xi$, $\delta\omega$ are $\mathcal{O}(\varepsilon)$ as $\varepsilon \to 0$.

We consider two regimes of perturbations. First, for small ε suppose that $\delta\omega = \mathcal{O}(\varepsilon)$ varies arbitrarily and that $\delta\xi$ covaries with $\delta\omega$ in such a way that $\dfrac{\delta\xi}{\xi} = -\dfrac{\delta\omega}{\omega}$. Then,

$$\begin{aligned}
\left\langle \mathbf{H}-\widehat{\mathbf{H}}_r,\, \widehat{\mathbf{H}}_r - \widehat{\mathbf{H}}_r^{(\varepsilon)}\right\rangle &= -\left\langle \mathbf{H}-\widehat{\mathbf{H}}_r,\, \mathbf{\Delta}_1\right\rangle \phi\,\delta\omega \\
&\quad + \left\langle \mathbf{H}-\widehat{\mathbf{H}}_r,\, \mathbf{\Delta}_2\right\rangle 2\,\omega^2\phi\;\delta\omega \;+ \mathcal{O}\left((\delta\omega)^2+(\delta\xi)^2\right)
\end{aligned} \tag{31}$$

Referring to Lemma 2, observe that either of the necessary conditions (29) or (30) implies immediately,

$$\left\langle \mathbf{H}-\widehat{\mathbf{H}}_r,\, \mathbf{\Delta}_1\right\rangle = 0.$$

Now, Theorem 1 together with (31) implies

$$\begin{aligned}
0 &= \left\langle \mathbf{H}-\widehat{\mathbf{H}}_r,\, \mathbf{\Delta}_2\right\rangle \\
&= -\frac{1}{(\lambda^+-\lambda^-)^2}\left(\mathsf{c}^T\left(\mathbf{H}'(-\lambda^+)-\widehat{\mathbf{H}}_r'(-\lambda^+)\right)\mathsf{b} + \mathsf{c}^T\left(\mathbf{H}'(-\lambda^-)-\widehat{\mathbf{H}}_r'(-\lambda^-)\right)\mathsf{b}\right) \\
&\quad - \frac{2}{(\lambda^+-\lambda^-)^3}\left(\mathsf{c}^T\left(\mathbf{H}(-\lambda^+)-\widehat{\mathbf{H}}_r(-\lambda^+)\right)\mathsf{b} - \mathsf{c}^T\left(\mathbf{H}(-\lambda^-)-\widehat{\mathbf{H}}_r(-\lambda^-)\right)\mathsf{b}\right) \\
&= -\frac{1}{(\lambda^+-\lambda^-)^2}\left(\mathsf{c}^T\left(\mathbf{H}'(-\lambda^+)-\widehat{\mathbf{H}}_r'(-\lambda^+)\right)\mathsf{b} + \mathsf{c}^T\left(\mathbf{H}'(-\lambda^-)-\widehat{\mathbf{H}}_r'(-\lambda^-)\right)\mathsf{b}\right)
\end{aligned} \tag{32}$$

Now, suppose that $\delta\xi = \mathcal{O}(\varepsilon)$ varies arbitrarily as $\varepsilon \to 0$ and let $\delta\omega = 0$. Then,

$$\left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, \widehat{\mathbf{H}}_r - \widehat{\mathbf{H}}_r^{(\varepsilon)} \right\rangle = -2 \left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, s\boldsymbol{\Delta}_2 \right\rangle \phi\,\omega^2 \delta\xi + \mathcal{O}\left((\delta\omega)^2 + (\delta\xi)^2\right) \tag{33}$$

Theorem 1 together with (33) as $\varepsilon \to 0$ implies

$$\begin{aligned} 0 &= \left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, s\boldsymbol{\Delta}_2 \right\rangle \\ &= -\frac{1}{(\lambda^+ - \lambda^-)^2} \left(\lambda^+ \mathsf{c}^T \left(\mathbf{H}'(-\lambda^+) - \widehat{\mathbf{H}}_r'(-\lambda^+) \right) \mathsf{b} + \lambda^- \mathsf{c}^T \left(\mathbf{H}'(-\lambda^-) - \widehat{\mathbf{H}}_r'(-\lambda^-) \right) \mathsf{b} \right) \\ &\qquad - \frac{\lambda^+ + \lambda^-}{(\lambda^+ - \lambda^-)^3} \left(\mathsf{c}^T \left(\mathbf{H}(-\lambda^+) - \widehat{\mathbf{H}}_r(-\lambda^+) \right) \mathsf{b} - \mathsf{c}^T \left(\mathbf{H}(-\lambda^-) - \widehat{\mathbf{H}}_r(-\lambda^-) \right) \mathsf{b} \right) \\ &= -\frac{1}{(\lambda^+ - \lambda^-)^2} \left(\lambda^+ \mathsf{c}^T \left(\mathbf{H}'(-\lambda^+) - \widehat{\mathbf{H}}_r'(-\lambda^+) \right) \mathsf{b} + \lambda^- \mathsf{c}^T \left(\mathbf{H}'(-\lambda^-) - \widehat{\mathbf{H}}_r'(-\lambda^-) \right) \mathsf{b} \right) \end{aligned} \tag{34}$$

The conditions (32) and (34) can be summarized as:

$$\begin{bmatrix} 1 & 1 \\ \lambda^+ & \lambda^- \end{bmatrix} \begin{pmatrix} \mathsf{c}^T \left(\mathbf{H}'(-\lambda^+) - \widehat{\mathbf{H}}_r'(-\lambda^+) \right) \mathsf{b} \\ \mathsf{c}^T \left(\mathbf{H}'(-\lambda^-) - \widehat{\mathbf{H}}_r'(-\lambda^-) \right) \mathsf{b} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Since $\lambda^+ \neq \lambda^-$, we find the bitangential condition: $\mathsf{c}^T \mathbf{H}'(-\lambda^\pm)\mathsf{b} = \mathsf{c}^T \widehat{\mathbf{H}}_r'(-\lambda^\pm)\mathsf{b}$.

Reinstating the k-dependence, we find the further necessary conditions

$$\boxed{\mathsf{c}_k^T \mathbf{H}'(-\lambda_k^+)\mathsf{b}_k = \mathsf{c}_k^T \widehat{\mathbf{H}}_r'(-\lambda_k^+)\mathsf{b}_k \quad \text{and} \quad \mathsf{c}_k^T \mathbf{H}'(-\lambda_k^-)\mathsf{b}_k = \mathsf{c}_k^T \widehat{\mathbf{H}}_r'(-\lambda_k^-)\mathsf{b}_k,} \tag{35}$$

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

6 Second-order Port-Hamiltonian Systems

We modify the set of second-order models that define $\mathcal{Q}_r \subset \mathcal{H}_2^{m \times p}$ so that they are in addition, *port-Hamiltonian.*

Introduce momentum degrees-of-freedom as $\mathbf{p}_r = \mathbf{M}_r \dot{\mathbf{x}}_r$. Then the model systems in $\mathcal{Q}_r$ may be rewritten as

$$\begin{pmatrix} \dot{\mathbf{x}}_r \\ \dot{\mathbf{p}}_r \end{pmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & -\mathbf{D}_r \end{bmatrix} \begin{bmatrix} \mathbf{K}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_r^{-1} \end{bmatrix} \begin{pmatrix} \mathbf{x}_r \\ \mathbf{p}_r \end{pmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_r \end{bmatrix} \mathbf{u} \tag{36}$$

$$\mathbf{y}_r = \begin{bmatrix} \mathbf{C}_r & \mathbf{0} \end{bmatrix} \begin{pmatrix} \mathbf{x}_r \\ \mathbf{p}_r \end{pmatrix} \tag{37}$$

Imposing port-Hamiltonian structure requires only changing the state-output map to

$$\mathbf{y}_r = \begin{bmatrix} \mathbf{0} & \mathbf{B}_r^T \end{bmatrix} \begin{bmatrix} \mathbf{K}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_r^{-1} \end{bmatrix} \begin{pmatrix} \mathbf{x}_r \\ \mathbf{p}_r \end{pmatrix} = \mathbf{B}_r^T \dot{\mathbf{x}}_r \tag{38}$$

Define the set of $p \times p$ matrix-valued transfer functions associated with second-order, modally damped dynamical systems with port-Hamiltonian structure:

$$\mathcal{P}_r = \left\{ s\mathbf{B}_r^T(s^2\mathbf{M}_r + s\mathbf{D}_r + \mathbf{K}_r)^{-1}\mathbf{B}_r \;\middle|\; \begin{array}{l} \mathbf{M}_r, \mathbf{D}_r, \mathbf{K}_r \in \mathbb{R}^{r\times r} \text{ SPD}, \\ \mathbf{B}_r \in \mathbb{R}^{r\times p}, \mathbf{C}_r \in \mathbb{R}^{m\times r}, \\ \text{and } \mathbf{D}_r\mathbf{M}_r^{-1}\mathbf{K}_r = \mathbf{K}_r\mathbf{M}_r^{-1}\mathbf{D}_r \end{array} \right\}$$

We follow a directly analogous path of development as before.

Suppose $\widehat{\mathbf{H}}_r(s) = s\widehat{\mathbf{B}}_r^T(s^2\widehat{\mathbf{M}}_r + s\widehat{\mathbf{D}}_r + \widehat{\mathbf{K}}_r)^{-1}\widehat{\mathbf{B}}_r \in \mathcal{P}_r$ solves

$$\left\| \mathbf{H} - \widehat{\mathbf{H}}_r \right\|_{\mathcal{H}_2} = \min_{\mathbf{H}_r \in \mathcal{P}_r} \| \mathbf{H} - \mathbf{H}_r \|_{\mathcal{H}_2}. \tag{39}$$

Using the same state-space transformation introduced before, we may write now

$$\widehat{\mathbf{H}}_r(s) = \sum_{k=1}^{r} \frac{\omega_k\,\phi_k\; s\; \mathsf{b}_k\,\mathsf{b}_k^T}{s^2 + 2\,\xi_k\,\omega_k\,s + \omega_k^2} = \sum_{k=1}^{r} \frac{\omega_k\,\phi_k\; s\; \mathsf{b}_k\,\mathsf{b}_k^T}{(s-\lambda_k^+)(s-\lambda_k^-)};$$

where scale factors, $\phi_k \geq 0$, are introduced so that $\|\mathsf{b}_k\| = 1$. Notice that $\widehat{\mathbf{H}}_r(s)^T = \widehat{\mathbf{H}}_r(s)$.

6.1 Perturbations in residues: second-order port-Hamiltonian case.

Pick $1 \leq k \leq r$, consider an arbitrary $\delta\mathsf{b}_k \in \mathbb{R}^m$ with $\|\delta\mathsf{b}_k\| = 1$, and define a perturbation of $\widehat{\mathbf{H}}_r$ that involves a small variation in the kth vector residue in the direction of $\delta\mathsf{b}_k$:

$$\begin{aligned} \widehat{\mathbf{H}}_r^{(\varepsilon)} &= \widehat{\mathbf{H}}_r - \left(\frac{\omega_k\,\phi_k\; s\; \mathsf{b}_k\,\mathsf{b}_k^T}{(s-\lambda_k^+)(s-\lambda_k^-)} - \frac{\omega_k\,\phi_k\, s\,(\mathsf{b}_k + \varepsilon\delta\mathsf{b}_k)\,(\mathsf{b}_k + \varepsilon\delta\mathsf{b}_k)^T}{(s-\lambda_k^+)(s-\lambda_k^-)} \right) \\ &= \widehat{\mathbf{H}}_r + \varepsilon \frac{\omega_k\,\phi_k\; s}{(s-\lambda_k^+)(s-\lambda_k^-)} \left(\delta\mathsf{b}_k\,\mathsf{b}_k^T + \mathsf{b}_k\,\delta\mathsf{b}_k^T \right) + \mathcal{O}(\varepsilon^2) \end{aligned}$$

Optimality of $\widehat{\mathbf{H}}_r$ implies that as $\varepsilon \to 0$:

$$\left| \left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\; \varepsilon \frac{\omega_k\,\phi_k\, s\, \left(\delta\mathsf{b}_k\,\mathsf{b}_k^T + \mathsf{b}_k\,\delta\mathsf{b}_k^T \right)}{(s-\lambda_k^+)(s-\lambda_k^-)} \right\rangle_{\mathcal{H}_2} \right| = \mathcal{O}(\varepsilon^2),$$

which implies in turn,

$$\begin{aligned} 0 &= \left| \left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\; \frac{s\, \left(\delta\mathsf{b}_k\,\mathsf{b}_k^T + \mathsf{b}_k\,\delta\mathsf{b}_k^T \right)}{(s-\lambda_k^+)(s-\lambda_k^-)} \right\rangle_{\mathcal{H}_2} \right| \\ &= \left(\delta\mathsf{b}_k^T \left(\mathbf{H}(-\lambda_k^+) - \widehat{\mathbf{H}}_r(-\lambda_k^+) \right) \mathsf{b}_k - \delta\mathsf{b}_k^T \left(\mathbf{H}(-\lambda_k^-) - \widehat{\mathbf{H}}_r(-\lambda_k^-) \right) \mathsf{b}_k \right) \\ &\quad + \left(\delta\mathsf{b}_k^T \left(\mathbf{H}(-\lambda_k^+)^T - \widehat{\mathbf{H}}_r(-\lambda_k^+)^T \right) \mathsf{b}_k - \delta\mathsf{b}_k^T \left(\mathbf{H}(-\lambda_k^-)^T - \widehat{\mathbf{H}}_r(-\lambda_k^-)^T \right) \mathsf{b}_k \right) \\ &= \delta\mathsf{b}_k^T \left(\left(\mathbf{H}(-\lambda_k^+) + \mathbf{H}(-\lambda_k^+)^T \right) \mathsf{b}_k - \left(\mathbf{H}(-\lambda_k^-) + \mathbf{H}(-\lambda_k^-)^T \right) \mathsf{b}_k \right. \\ &\quad \left. - 2 \left(\widehat{\mathbf{H}}_r(-\lambda_k^+) - \widehat{\mathbf{H}}_r(-\lambda_k^-) \right) \mathsf{b}_k \right) \end{aligned}$$

Denote the symmetric part of a $p \times p$ matrix $\mathbf{M}$ as $\mathsf{symm}(\mathbf{M}) = \frac{1}{2}(\mathbf{M} + \mathbf{M}^T)$. Since $\delta \mathsf{b}_k$ is arbitrary,

$$\boxed{\left(\mathsf{symm}\left[\mathbf{H}(-\lambda_k^+)\right] - \mathsf{symm}\left[\mathbf{H}(-\lambda_k^-)\right]\right) \mathsf{b}_k = \left(\widehat{\mathbf{H}}_r(-\lambda_k^+) - \widehat{\mathbf{H}}_r(-\lambda_k^-)\right) \mathsf{b}_k,} \tag{40}$$

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

6.2 Perturbations in poles: second-order port-Hamiltonian case.

Pick $1 \le k \le r$ and consider the effect of arbitrary (but small) perturbations in ξ_k and ω_k. Suppressing k-dependence in ξ_k and ω_k (likewise in $\lambda_k^\pm$ and associated residues, ϕ_k, , and b_k):

$$\begin{aligned}
\widehat{\mathbf{H}}_r^{(\varepsilon)} &= \widehat{\mathbf{H}}_r - \left(\frac{\omega\,\phi\, s\, \mathsf{b}\,\mathsf{b}^T}{s^2 + 2\xi\omega s + \omega^2} - \frac{(\omega + \delta\omega)\,\phi\, s\,\mathsf{b}\,\mathsf{b}^T}{s^2 + 2(\xi + \delta\xi)(\omega+\delta\omega)s + (\omega+\delta\omega)^2}\right) \\
&= \widehat{\mathbf{H}}_r - \frac{\omega\,\phi\, s\, \mathsf{b}\,\mathsf{b}^T}{s^2 + 2\xi\omega s + \omega^2}\left(1 - \frac{\left(1 + \frac{\delta\omega}{\omega}\right)}{1 + \left(\frac{2(\xi\delta\omega + \omega\delta\xi)s + 2\omega\delta\omega}{s^2+2\xi\omega s+\omega^2}\right)}\right) \\
&= \widehat{\mathbf{H}}_r - \frac{\omega\,\phi\, s\, \mathsf{b}\,\mathsf{b}^T}{s^2 + 2\xi\omega s + \omega^2}\left(1 - \left(1 + \frac{\delta\omega}{\omega}\right)\left(1 - \frac{2(\xi\delta\omega + \omega\delta\xi)s + 2\omega\delta\omega}{s^2 + 2\xi\omega s + \omega^2}\right)\right) + \mathcal{O}\left(\varepsilon^2\right) \\
&= \widehat{\mathbf{H}}_r - \frac{\omega\,\phi\, s\, \mathsf{b}\,\mathsf{b}^T}{s^2 + 2\xi\omega s + \omega^2}\left(-\frac{\delta\omega}{\omega} + \frac{2(\xi\delta\omega + \omega\delta\xi)s + 2\omega\delta\omega}{s^2 + 2\xi\omega s + \omega^2}\right) + \mathcal{O}\left(\varepsilon^2\right) \\
&= \widehat{\mathbf{H}}_r + \phi\,\delta\omega\,(s\,\boldsymbol{\Delta}_1(s) - 2\,\omega^2\, s\,\boldsymbol{\Delta}_2(s)) - 2\,\omega\,\phi\,(\xi\delta\omega + \omega\delta\xi)\, s^2 \boldsymbol{\Delta}_2(s) + \mathcal{O}\left(\varepsilon^2\right)
\end{aligned}$$

where we take $\mathsf{c} = \mathsf{b}$ in Lemma 2 and $\delta\xi$, $\delta\omega$ are $\mathcal{O}(\varepsilon)$ as $\varepsilon \to 0$.

We consider two regimes of perturbations. First, for small ε suppose that $\delta\omega = \mathcal{O}(\varepsilon)$ varies arbitrarily and that $\delta\xi$ covaries with $\delta\omega$ in such a way that $\dfrac{\delta\xi}{\xi} = -\dfrac{\delta\omega}{\omega}$. Then,

$$\begin{aligned}
\left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, \widehat{\mathbf{H}}_r - \widehat{\mathbf{H}}_r^{(\varepsilon)} \right\rangle &= -\phi\,\delta\omega\,\left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, s\,\boldsymbol{\Delta}_1 \right\rangle \\
&\quad + 2\,\omega^2 \phi\,\delta\omega\,\left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, s\,\boldsymbol{\Delta}_2 \right\rangle + \mathcal{O}\left(\varepsilon^2\right)
\end{aligned} \tag{41}$$

Referring to Lemma 2, observe that the necessary condition (40) implies immediately,

$$\left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, s\,\boldsymbol{\Delta}_1 \right\rangle = 0.$$

Now, Theorem 1 together with (41) implies Theorem 1 together with (33) as $\varepsilon \to 0$ implies

$$\begin{aligned}
0 &= \left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, s\boldsymbol{\Delta}_2 \right\rangle \\
&= -\frac{1}{(\lambda^+ - \lambda^-)^2}\left(\lambda^+ \mathsf{b}^T\left(\mathbf{H}'(-\lambda^+) - \widehat{\mathbf{H}}_r'(-\lambda^+)\right)\mathsf{b} + \lambda^- \mathsf{b}^T\left(\mathbf{H}'(-\lambda^-) - \widehat{\mathbf{H}}_r'(-\lambda^-)\right)\mathsf{b}\right) \\
&\qquad - \frac{\lambda^+ + \lambda^-}{(\lambda^+ - \lambda^-)^3}\left(\mathsf{b}^T\left(\mathbf{H}(-\lambda^+) - \widehat{\mathbf{H}}_r(-\lambda^+)\right)\mathsf{b} - \mathsf{b}^T\left(\mathbf{H}(-\lambda^-) - \widehat{\mathbf{H}}_r(-\lambda^-)\right)\mathsf{b}\right) \\
&= -\frac{1}{(\lambda^+ - \lambda^-)^2}\left(\lambda^+ \mathsf{b}^T\left(\mathbf{H}'(-\lambda^+) - \widehat{\mathbf{H}}_r'(-\lambda^+)\right)\mathsf{b} + \lambda^- \mathsf{b}^T\left(\mathbf{H}'(-\lambda^-) - \widehat{\mathbf{H}}_r'(-\lambda^-)\right)\mathsf{b}\right)
\end{aligned} \tag{42}$$

Now, suppose that $\delta\xi = \mathcal{O}(\varepsilon)$ varies arbitrarily as $\varepsilon \to 0$ and let $\delta\omega = 0$. Then,

$$\left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, \widehat{\mathbf{H}}_r - \widehat{\mathbf{H}}_r^{(\varepsilon)} \right\rangle = -2\,\phi\,\omega^2 \delta\xi \left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, s^2\,\mathbf{\Delta}_2 \right\rangle + \mathcal{O}\left(\varepsilon^2\right) \tag{43}$$

Theorem 1 together with (43) as $\varepsilon \to 0$ implies

$$\begin{aligned} 0 &= \left\langle \mathbf{H} - \widehat{\mathbf{H}}_r,\, s^2\,\mathbf{\Delta}_2 \right\rangle \\ &= \frac{1}{(\lambda^+ - \lambda^-)^2} \left((\lambda^+)^2 \mathsf{b}^T \left(\mathbf{H}'(-\lambda^+) - \widehat{\mathbf{H}}_r'(-\lambda^+) \right) \mathsf{b} + (\lambda^-)^2 \mathsf{b}^T \left(\mathbf{H}'(-\lambda^-) - \widehat{\mathbf{H}}_r'(-\lambda^-) \right) \mathsf{b} \right) \\ &\qquad - \frac{2\,\lambda^+\,\lambda^-}{(\lambda^+ - \lambda^-)^3} \left(\mathsf{b}^T \left(\mathbf{H}(-\lambda^+) - \widehat{\mathbf{H}}_r(-\lambda^+) \right) \mathsf{b} - \mathsf{b}^T \left(\mathbf{H}(-\lambda^-) - \widehat{\mathbf{H}}_r(-\lambda^-) \right) \mathsf{b} \right) \\ &= \frac{1}{(\lambda^+ - \lambda^-)^2} \left((\lambda^+)^2\, \mathsf{b}^T \left(\mathbf{H}'(-\lambda^+) - \widehat{\mathbf{H}}_r'(-\lambda^+) \right) \mathsf{b} \right. \\ &\qquad\qquad \left. +\, (\lambda^-)^2\, \mathsf{b}^T \left(\mathbf{H}'(-\lambda^-) - \widehat{\mathbf{H}}_r'(-\lambda^-) \right) \mathsf{b} \right) \end{aligned} \tag{44}$$

The conditions (41) and (44) can be summarized as:

$$\begin{bmatrix} \lambda^+ & \lambda^- \\ (\lambda^+)^2 & (\lambda^-)^2 \end{bmatrix} \begin{pmatrix} \mathsf{b}^T \left(\mathbf{H}'(-\lambda^+) - \widehat{\mathbf{H}}_r'(-\lambda^+) \right) \mathsf{b} \\ \mathsf{b}^T \left(\mathbf{H}'(-\lambda^-) - \widehat{\mathbf{H}}_r'(-\lambda^-) \right) \mathsf{b} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Since $\lambda^+ \neq \lambda^-$, we find the bitangential conditions: $\mathsf{b}^T \mathbf{H}'(-\lambda^\pm)\mathsf{b} = \mathsf{c}^T \widehat{\mathbf{H}}_r'(-\lambda^\pm)\mathsf{b}$.

Reinstating the k-dependence, we find the further necessary conditions

$$\boxed{\mathsf{b}_k^T \mathbf{H}'(-\lambda_k^+)\mathsf{b}_k = \mathsf{b}_k^T \widehat{\mathbf{H}}_r'(-\lambda_k^+)\mathsf{b}_k \quad \text{and} \quad \mathsf{b}_k^T \mathbf{H}'(-\lambda_k^-)\mathsf{b}_k = \mathsf{b}_k^T \widehat{\mathbf{H}}_r'(-\lambda_k^-)\mathsf{b}_k,} \tag{45}$$

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

7 Conclusions

We have derived necessary conditions that must be satisfied by the best $\mathcal{H}_2$-approximating reduced-order model that is also constrained to have particular structural features such as being port-Hamiltonian or second-order. Future work will address the construction of algorithms that can be used to find such systems systematically.

Acknowledgments

This research was funded by the *Deutsche Forschungsgemeinschaft (DFG)* as subproject A06 Model Order Reduction for Thermo-Elastic Assembly Group Models of the Collaborative Research Center/ Transregio 96 Thermo-Energetic Design of Machine Tools.

References

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

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

[3] Z. Bai, D. Bindel, J. Clark, J. Demmel, K.S.J. Pister, and N. Zhou. New numerical techniques and tools in SUGAR for 3D MEMS simulation. In *Technical Proceedings of the Fourth International Conference on Modeling and Simulation of Microsystems*, pages 31–34, 2000.

[4] Z. Bai and Y. Su. Dimension reduction of second order dynamical systems via a second-order Arnoldi method. *SIAM J. Sci. Comp*, 5:1692–1709, 2005.

[5] U. Baur, P. Benner, and L. Feng. Model order reduction for linear and nonlinear systems: a system-theoretic perspective. *Archives of Computational Methods in Engineering*, 2014 (electronic).

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

[7] P. Benner, V. Mehrmann, and D.C. Sorensen. *Dimension Reduction of Large-Scale Systems*, volume 45 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin/Heidelberg, Germany, 2005.

[8] A. Bryson. Second-order algorithm for optimal model order reduction. *J. Guidance, Control, Dynamics*, 13:887–892, 1990.

[9] V. Chahlaoui, K. A. Gallivan, A. Vandendorpe, and P. van Dooren. Model reduction of second-order system. 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. Springer-Verlag, Berlin/Heidelberg, Germany, 2005.

[10] J. V. Clark, N. Zhou, D. Bindel, L. Schenato, W. Wu, J. Demmel, and K. S. J. Pister. 3D MEMS simulation using modified nodal analysis. In *Proceedings of Microscale Systems: Mechanics and Measurements Symposium*, page 6875, 2000.

[11] R. R. Craig Jr. *Structral dynamics: An introduction to computer methods*. John Wiley & Sons, 1981.

[12] Vincent Duindam, Alessandro Macchelli, Stefano Stramigioli, and Herman Bruyninckx. *Modeling and control of complex physical systems*. Springer, 2009.

[13] S. Gugercin, A. C. Antoulas, and C. A. Beattie. $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM J. Matrix Anal. Appl.*, 30(2):609–638, 2008.

[14] S. Gugercin, R.V. Polyuga, C.A. Beattie, and A.J. van der Schaft. Interpolation-based $\mathcal{H}_2$ Model Reduction for port-Hamiltonian Systems. In *Proceedings of the Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Shanghai, PR China*, pages 5362–5369, 2009.

[15] Y. Halevi. Frequency weighted model reduction via optimal projection. *IEEE Trans. Automatic Control*, 37(10):1537–1542, 1992.

[16] D. Hyland and D. Bernstein. The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore. *IEEE Trans. Automatic Control*, 30(12):1201–1211, 1985.

[17] J.G. Korvink and E.B. Rudnyi. Oberwolfach benchmark collection. 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 311–315. Springer-Verlag, Berlin/Heidelberg, Germany, 2005.

[18] A. Lepschy, G. A. Mian, G. Pinato, and U. Viaro. Rational L^2 approximation: a non-gradient algorithm. In *Proceedings of the 30th IEEE Conference on Decision and Control*, pages 2321–2323, 1991.

[19] L. Meier III and D. Luenberger. Approximation of linear constant systems. *IEEE Trans. Automatic Control*, 12(5):585–588, 1967.

[20] D.G. Meyer and S. Srinivasan. Balancing and model reduction for second-order form linear systems. *Automatic Control, IEEE Transactions on*, 41(11):1632–1644, 1996.

[21] R.V. Polyuga and A.J. van der Schaft. Structure preserving model reduction of port-Hamiltonian systems by moment matching at infinity. *Automatica*, 46:665–672, 2010.

[22] R.V. Polyuga and A.J. van der Schaft. Structure preserving port-Hamiltonian model reduction of electrical circuits. In Peter Benner, Michael Hinze, and E. Jan W. ter Maten, editors, *Model Reduction for Circuit Simulation*, volume 74 of *Lecture Notes in Electrical Engineering*, pages 241–260. Springer-Verlag, Dordrecht, NL, 2011.

[23] R.V. Polyuga and A.J. van der Schaft. Model reduction of port-Hamiltonian systems as structured systems. *In Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems, Budapest, Hungary*, pages 1509–1513, 5–9 July, 2010.

[24] A. Preumont. *Vibration Control of Active Structures: An Introduction.* Springer, 2002.

[25] J. T. Spanos, M. H. Milman, and D. L. Mingori. A new algorithm for L^2 optimal model reduction. *Automatica (Journal of IFAC)*, 28(5):897–909, 1992.

[26] T.-J. Su and R.R. Craig Jr. Model reduction and control of flexible structures using Krylov vectors. *J. Guid. Control Dyn.*, 14:260–267, 1991.

[27] A. J. van der Schaft. L_2-Gain and Passivity Techniques in Nonlinear Control. *Lecture Notes in Control and Information Sciences*, 218, 2000.

[28] A.J. van der Schaft and R.V. Polyuga. Structure-preserving model reduction of complex physical systems. *In Proceedings of the Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Shanghai, P.R. China*, pages 4322–4327, December 16-18, 2009.

[29] P. Van Dooren, K. A. Gallivan, and P. A. Absil. $\mathcal{H}_2$-optimal model reduction of MIMO systems. *Applied Math. Letters*, 21(12):1267–1273, 2008.

[30] W. Weaver and P. Johnston. *Structural dynamics by finite elements.* Prentice Hall, Upper Saddle River, 1987.

[31] D. A. Wilson. Optimum solution of model-reduction problem. *Proc. Inst. Elec. Eng.*, 117(6):1161–1165, 1970.

[32] W. Y. Yan and J. Lam. An approximate approach to h^2 optimal model reduction. *IEEE Trans. Automatic Control*, 44(7):1341–1358, 1999.

[33] D. Zigic, L. T. Watson, and C. Beattie. Contragredient transformations applied to the optimal projection equations. *Linear Algebra Appl.*, 188:665–676, 1993.