Interpolatory model reduction of large-scale dynamical systems

Thanos Antoulas

Rice University and Jacobs University

e-mail: aca@rice.edu
URL: www.ece.rice.edu/˜aca

Comson Autumn School in Model Reduction
Terschelling, 21 - 25 September 2009

Collaborators

- Chris Beattie
- Serkan Gugercin
- Sanda Lefteriu
- Cosmin Ionita
- Roxana Ionutiu
- Joost Rommes
- Andrew Mayo

Outline

Outline

Problem setting

Linear dynamical systems are principally characterized through their input-output map $\mathcal{S} : \mathbf{u} \mapsto \mathbf{y}$, via a state-space realization given as:

$$\mathcal{S} : \begin{cases} \mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \end{cases} \quad \text{with } \mathbf{x}(0) = \mathbf{0}, \tag{1}$$

where $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times m}$, $\mathbf{C} \in \mathbb{R}^{p\times n}$ and $\mathbf{D} \in \mathbb{R}^{p\times m}$, are constant matrices. $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{u}(t) \in \mathbb{R}^m$ and $\mathbf{y}(t) \in \mathbb{R}^p$ are, respectively, an internal variable (the state if $\mathbf{E}$ is non-singular), the input and the output of the system $\mathcal{S}$.

The goal of model reduction is to produce a surrogate dynamical system with I/O map, $\mathcal{S}_r : \mathbf{u} \mapsto \mathbf{y}_r$, $r \ll n$, described in state-space form as:

$$\mathcal{S}_r \quad \begin{cases} \mathbf{E}_r\dot{\mathbf{x}}_r(t) = \mathbf{A}_r\mathbf{x}_r(t) + \mathbf{B}_r\mathbf{u}(t) \\ \mathbf{y}_r(t) = \mathbf{C}_r\mathbf{x}_r(t) + \mathbf{D}_r\mathbf{u}(t) \end{cases} \quad \text{with } \mathbf{x}_r(0) = \mathbf{0}, \tag{2}$$

where $\mathbf{A}_r, \mathbf{E}_r \in \mathbb{R}^{r\times r}$, $\mathbf{B}_r \in \mathbb{R}^{r\times m}$, $\mathbf{C}_r \in \mathbb{R}^{p\times r}$ and $\mathbf{D}_r \in \mathbb{R}^{p\times m}$.

A successful reduced-order model should meet the following criteria:

Goals for Reduced Order Models

1. The reduced input-output map $\mathcal{S}_r$ should be uniformly close to $\mathcal{S}$ in an appropriate sense. That is, for the same inputs, $\mathbf{u}(t)$, the difference between full and reduced system outputs, $\mathbf{y} - \mathbf{y}_r$, should be *small*.
2. Critical system structure should be preserved, e.g. passivity, Hamiltonian structure, subsystem interconnectivity, or second-order structure.
3. Strategies for obtaining the reduced system should lead to robust, numerically stable algorithms and furthermore require minimal application-specific tuning with little to none expert intervention. They should be robust and largely automatic to allow the broadest level of flexibility and applicability in complex multiphysics setting.

We define transfer functions as

$$\mathbf{H}(s) = \mathbf{C}\left(s\mathbf{E} - \mathbf{A}\right)^{-1}\mathbf{B} + \mathbf{D}, \quad \mathbf{H}_r(s) = \mathbf{C}_r\left(s\mathbf{E}_r - \mathbf{A}_r\right)^{-1}\mathbf{B}_r + \mathbf{D}_r$$

so that

$$\widehat{\mathbf{y}}(s) - \widehat{\mathbf{y}}_r(s) = \left[\mathbf{H}(s) - \mathbf{H}_r(s)\right]\widehat{\mathbf{u}}(s)$$

The general interpolation framework

- Interpolation is a very simple approach that is used for the general approximation of complex functions.
- The accuracy of the resulting approximations and its connection with strategic placement of interpolating points has been studied in many broad contexts – e.g., for interpolation of meromorphic functions by polynomials or rational functions, this work is tied in closely with potential theory and classical complex analysis.
- Our overarching goal is to produce a reduced order transfer function, $\mathbf{H}_r(s)$, that approximates with high fidelity a very large order transfer function, $\mathbf{H}(s)$: $\mathbf{H}_r(s) \approx \mathbf{H}(s)$. Interpolation is our primary vehicle. At its most elementary level it can be viewed as selecting a set of points $\{\sigma_i\}_{i=1}^r \subset \mathbb{C}$ and then seeking a reduced order transfer function, $\mathbf{H}_r(s)$, such that $\mathbf{H}_r(\sigma_i) = \mathbf{H}(\sigma_i)$ for $i = 1, \ldots, r$.
- This is a good starting place for SISO systems but turns out to be overly restrictive for MIMO systems, since the condition $\mathbf{H}_r(\sigma_i) = \mathbf{H}(\sigma_i)$ in effect imposes $m \cdot p$ scalar conditions at each interpolation point. It is more advantageous to consider interpolation conditions that are imposed in specified directions: *tangential interpolation*.

Problem 1: Model Reduction given state space system data.

Given a full-order model $\mathbf{E}$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ and given

left interpolation points: $\{\mu_i\}_{i=1}^{q} \subset \mathbb{C}$, with corresponding *left tangent directions*: $\{\tilde{c}_i\}_{i=1}^{q} \subset \mathbb{C}^p$,

and

right interpolation points: $\{\sigma_i\}_{i=1}^{r} \subset \mathbb{C}$ with corresponding *right tangent directions*: $\{\tilde{b}_i\}_{i=1}^{r} \subset \mathbb{C}^m$.

Find a reduced-order model $\mathbf{E}_r$, $\mathbf{A}_r$, $\mathbf{B}_r$, $\mathbf{C}_r$, and $\mathbf{D}_r$ such that the associated transfer function, $\mathbf{H}_r(s)$, is a *tangential interpolant* to $\mathbf{H}(s)$:

$$\tilde{c}_i^T \mathbf{H}_r(\mu_i) = \tilde{c}_i^T \mathbf{H}(\mu_i) \quad \text{for } i = 1, \cdots, q, \qquad \text{and} \qquad \mathbf{H}_r(\sigma_j)\tilde{b}_j = \mathbf{H}(\sigma_j)\tilde{b}_j, \quad \text{for } j = 1, \cdots, r, \tag{3}$$

Remark. This is a remarkably flexible framework within which to consider model reduction when one considers the significance of the interpolation data. Note that if $\tilde{y} = \mathbf{H}(\sigma)\tilde{b}$ then $e^{\sigma t}\tilde{y}$ is precisely the response of the full order system to a pure input given by $\mathbf{u}(t) = e^{\sigma t}\tilde{b}$, so the tangential interpolation conditions that characterize $\mathbf{H}_r(s)$ could (at least in principle) be obtained from *measured input-output data* drawn directly from observations on the original system. Similarly, if the *dual* dynamical system were driven by an input given by $e^{\mu t}\tilde{c}$ producing an output $e^{\mu t}\tilde{z}$ then $\tilde{c}^T\mathbf{H}(\mu) = \tilde{z}^T$.

This creates the following alternative problem.

Problem 2: Model reduction given input-output data.

Given a set of input-output response measurements on a system specified by

left driving frequencies: $\{\mu_i\}_{i=1}^q \subset \mathbb{C}$, using *left input directions*: $\{\tilde{c}_i\}_{i=1}^q \subset \mathbb{C}^p$, producing *left responses*: $\{\tilde{z}_i\}_{i=1}^q \subset \mathbb{C}^m$,

and

right driving frequencies: $\{\sigma_i\}_{i=1}^r \subset \mathbb{C}$ using *right input directions*: $\{\tilde{b}_i\}_{i=1}^r \subset \mathbb{C}^m$ producing *right responses*: $\{\tilde{y}_i\}_{i=1}^r \subset \mathbb{C}^p$

Find a system model by specifying (reduced) system matrices $\mathbf{E}_r$, $\mathbf{A}_r$, $\mathbf{B}_r$, $\mathbf{C}_r$, $\mathbf{D}_r$ such that $\mathbf{H}_r(s)$, is a *tangential interpolant* to the given data:

$$\tilde{c}_i^T \mathbf{H}_r(\mu_i) = \tilde{z}_i^T \text{ for } i = 1, \cdots, q, \quad \text{and} \quad \mathbf{H}_r(\sigma_j)\tilde{b}_j = \tilde{y}_j, \text{ for } j = 1, \cdots, r. \tag{4}$$

Interpolation points and tangent directions are determined (typically) by the availability of experimental data.

For both problems, it is necessary to have a computationally stable method for constructing the (reduced) system matrices $\mathbf{E}_r$, $\mathbf{A}_r$, $\mathbf{B}_r$, $\mathbf{C}_r$, and $\mathbf{D}_r$ that produces an associated transfer function, $\mathbf{H}_r(s)$, satisfying the interpolation conditions.

Model Reduction via Projection

Most model reduction methods proceed with some variation of a **Petrov-Galerkin projective approximation** to construct a reduced-order model $\mathcal{S}_r$.

$$\text{Find } \mathbf{x}(t) \text{ contained in } \mathbb{C}^n \text{ such that}$$
$$\mathbf{E}\dot{\mathbf{x}}(t) - \mathbf{A}\mathbf{x}(t) - \mathbf{B}\,\mathbf{u}(t) \quad \perp \quad \mathbb{C}^n \text{ (i.e., } = 0).$$
$$\text{Then the associated output is } \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t).$$

We choose an r-dimensional trial subspace, the **right modeling subspace**, $\mathcal{V}_r \subset \mathbb{C}^n$, and an r-dimensional test subspace, the **left modeling subspace**, $\mathcal{W}_r \subset \mathbb{C}^n$:

$$\text{Find } \mathbf{v}(t) \text{ contained in } \mathcal{V}_r \text{ such that}$$
$$\mathbf{E}\dot{\mathbf{v}}(t) - \mathbf{A}\mathbf{v}(t) - \mathbf{B}\,\mathbf{u}(t) \quad \perp \quad \mathcal{W}_r. \tag{5}$$
$$\text{Then the associated output is } \mathbf{y}_r(t) = \mathbf{C}\mathbf{v}(t) + \mathbf{D}\mathbf{u}(t).$$

Let $\text{Ran}(\mathbf{R})$ denote the range of a matrix $\mathbf{R}$. Let $\mathbf{V}_r \in \mathbb{C}^{n\times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n\times r}$ be matrices defined so that $\mathcal{V}_r = \text{Ran}(\mathbf{V}_r)$ and $\mathcal{W}_r = \text{Ran}(\overline{\mathbf{W}_r})$. We can represent the reduced system trajectories as $\mathbf{v}(t) = \mathbf{V}_r\mathbf{x}_r(t)$ with $\mathbf{x}_r(t) \in \mathbb{C}^r$ for each t and the Petrov-Galerkin approximation (5) can be rewritten as

$$\mathbf{W}_r^T\left(\mathbf{E}\mathbf{V}_r\dot{\mathbf{x}}_r(t) - \mathbf{A}\mathbf{V}_r\mathbf{x}_r(t) - \mathbf{B}\,\mathbf{u}(t)\right) = \mathbf{0} \quad \text{and} \quad \mathbf{y}_r(t) = \mathbf{C}\mathbf{V}_r\mathbf{x}_r(t) + \mathbf{D}\mathbf{u}(t),$$

leading to the reduced order state-space representation (2) with

$$\mathbf{E}_r = \mathbf{W}_r^T\mathbf{E}\mathbf{V}_r, \quad \mathbf{A}_r = \mathbf{W}_r^T\mathbf{A}\mathbf{V}_r, \quad \mathbf{B}_r = \mathbf{W}_r^T\mathbf{B}, \quad \mathbf{C}_r = \mathbf{C}\mathbf{V}_r \quad \text{and} \quad \mathbf{D}_r = \mathbf{D}. \tag{6}$$

We choose the modeling subspaces to enforce interpolation.

Interpolatory Projections

We seek $\mathbf{H}_r(s)$ so that

$$\left.\begin{array}{ll} \mathbf{H}(\sigma_i)\mathrm{b}_i = \mathbf{H}_r(\sigma_i)\mathrm{b}_i, & \text{for } \; i = 1, \cdots, r, \\ \mathrm{c}_j^T \mathbf{H}(\mu_j) = \mathrm{c}_j^T \mathbf{H}_r(\mu_j), & \text{for } \; j = 1, \cdots, r, \end{array}\right\} \tag{7}$$

The goal is to interpolate $\mathbf{H}(s)$ without ever computing the quantities to be matched since these numbers are numerically ill-conditioned. This is achieved by employing the projection framework.

Theorem: solution of rational tangential interpolation by projection.

Let σ, $\mu \in \mathbb{C}$ be such that $s\,\mathbf{E} - \mathbf{A}$ and $s\,\mathbf{E}_r - \mathbf{A}_r$ are invertible for $s = \sigma, \mu$. Also let $\mathbf{V}_r, \mathbf{W}_r \in \mathbb{C}^{n \times r}$ in (6) have full-rank. If $\mathrm{b} \in \mathbb{C}^m$ and $\mathrm{c} \in \mathbb{C}^\ell$ are fixed vectors then

(a) if $(\sigma\,\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{B}\mathrm{b} \in \mathrm{Ran}(\mathbf{V}_r)$, then $\mathbf{H}(\sigma)\mathrm{b} = \mathbf{H}_r(\sigma)\mathrm{b}$;

(b) if $\left(\mathrm{c}^T\mathbf{C}\,(\mu\,\mathbf{E} - \mathbf{A})^{-1}\right)^T \in \mathrm{Ran}(\mathbf{W}_r)$, then $\mathrm{c}^T\mathbf{H}(\mu) = \mathrm{c}^T\mathbf{H}_r(\mu)$; and

(c) if both (a) and (b) hold, and $\sigma = \mu$, then $\mathrm{c}^T\mathbf{H}'(\sigma)\mathrm{b} = \mathrm{c}^T\mathbf{H}_r'(\sigma)\mathrm{b}$ as well.

Proof 1.1 *Define*

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

Both $\mathcal{P}_r(z)$ and $\mathcal{Q}_r(z)$ are analytic matrix-valued functions in neighborhoods of $z = \sigma$ and $z = \mu$. It is easy to verify that both $\mathcal{P}_r(z)$ and $\mathcal{Q}_r(z)$ are projectors, i.e. $\mathcal{P}_r^2(z) = \mathcal{P}_r(z)$ and $\mathcal{Q}_r^2(z) = \mathcal{Q}_r(z)$. Moreover, for all z in a neighborhood of σ, $\mathcal{V}_r = Ran(\mathcal{P}_r(z)) = Ker(\mathbf{I} - \mathcal{P}_r(z))$ and $\mathcal{W}_r^\perp = Ker(\mathcal{Q}_r(z)) = Ran(\mathbf{I} - \mathcal{Q}_r(z))$ where $Ran(\mathbf{R})$ denotes the kernel of a matrix $\mathbf{R}$. First observe that

$$\mathbf{H}(z) - \mathbf{H}_r(z) = \mathbf{C}(z\mathbf{E} - \mathbf{A})^{-1}(\mathbf{I} - \mathcal{Q}_r(z))(z\mathbf{E} - \mathbf{A})\Big(\mathbf{I} - \mathcal{P}_r(z)\Big)(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$$

Evaluating this expression at $z = \sigma$ and postmultiplying by b *yields the first assertion; evaluating (1.1) at $z = \mu$ and premultiplying by c^T yields the second. Note that*

$$((\sigma + \varepsilon)\mathbf{E} - \mathbf{A})^{-1} = (\sigma\mathbf{E} - \mathbf{A})^{-1} - \varepsilon(\sigma\mathbf{E} - \mathbf{A})^{-1}\mathbf{E}(\sigma\mathbf{E} - \mathbf{A})^{-1} + \mathcal{O}(\varepsilon^2)$$
$$((\sigma + \varepsilon)\mathbf{E}_r - \mathbf{A}_r)^{-1} = (\sigma\mathbf{E}_r - \mathbf{A}_r)^{-1} - \varepsilon(\sigma\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{E}_r(\sigma\mathbf{E}_r - \mathbf{A}_r)^{-1} + \mathcal{O}(\varepsilon^2)$$

so evaluating (1.1) at $z = \sigma + \varepsilon$, premultiplying by c^T, and postmultiplying by b *under the hypotheses of the third assertion yields*

$$\mathrm{c}^T\mathbf{H}(\sigma + \varepsilon)\mathrm{b} - \mathrm{c}^T\mathbf{H}_r(\sigma + \varepsilon)\mathrm{b} = \mathcal{O}(\varepsilon^2).$$

Since $\mathrm{c}^T\mathbf{H}(\sigma)\mathrm{b} = \mathrm{c}^T\mathbf{H}_r(\sigma)\mathrm{b}$,

$$\frac{1}{\varepsilon}\Big(\mathrm{c}^T\mathbf{H}(\sigma + \varepsilon)\mathrm{b} - \mathrm{c}^T\mathbf{H}(\sigma)\mathrm{b}\Big) - \frac{1}{\varepsilon}\Big(\mathrm{c}^T\mathbf{H}_r(\sigma + \varepsilon)\mathrm{b} - \mathrm{c}^T\mathbf{H}_r(\sigma)\mathrm{b}\Big) \rightarrow 0,$$

as $\varepsilon \rightarrow 0$, which proves the third assertion.

i

Thus given a set of distinct right points (shifts), a set of distinct left points (shifts), left-tangential and right tangential directions, the solution of the tangential rational interpolation problem is straightforward. Simply construct $\mathbf{V}_r$ and $\mathbf{W}_r$.

Interpolatory projections

$$\mathbf{V}_r = \left[(\sigma_1 \mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathrm{b}_1,\ \cdots,\ (\sigma_r \mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathrm{b}_r\right], \tag{8}$$

$$\mathbf{W}_r^T = \begin{bmatrix} \mathrm{c}_1^T \mathbf{C}(\mu_1 \mathbf{E} - \mathbf{A})^{-1} \\ \vdots \\ \mathrm{c}_r^T \mathbf{C}(\mu_r \mathbf{E} - \mathbf{A})^{-1} \end{bmatrix}. \tag{9}$$

Interpolatory projections with derivative constraints

For completeness:

Interpolation with derivative constraints

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

(a) if $\left((\sigma\,\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{E}\right)^{j-1}(\sigma\,\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{B}\mathrm{b} \in \mathrm{Ran}(\mathbf{V}_r)$ for $j = 1, ., N$

then $\mathbf{H}^{(\ell)}(\sigma)\mathrm{b} = \mathbf{H}_r^{(\ell)}(\sigma)\mathrm{b}$ *for* $\ell = 0, 1, \ldots, N-1$

(b) if $\left((\mu\,\mathbf{E} - \mathbf{A})^{-T}\,\mathbf{E}^T\right)^{j-1}(\mu\,\mathbf{E} - \mathbf{A})^{-T}\,\mathbf{C}^T\mathrm{c} \in \mathrm{Ran}(\mathbf{W}_r)$ for $j = 1, ., M$,

then $\mathrm{c}^T\mathbf{H}^{(\ell)}(\mu) = \mathrm{c}^T\mathbf{H}_r^{(\ell)}(\mu)\mathrm{b}$ *for* $\ell = 0, 1, \ldots, M-1$;

(c) if both (a) and (b) hold, and if $\sigma = \mu$, then $\mathrm{c}^T\mathbf{H}^{(\ell)}(\sigma)\mathrm{b} = \mathrm{c}^T\mathbf{H}_r^{(\ell)}(\sigma)\mathrm{b}$,

for $\ell = 1, \ldots, M + N + 1$

The case $\mathbf{D}_r \neq \mathbf{D}$

- The Petrov-Galerkin framework leads to $\mathbf{D}_r = \mathbf{D}$.
- Sometimes it is useful to have the flexibility of choosing $\mathbf{D}_r \neq \mathbf{D}$, while still satisfying the interpolation constraints. For instance if we wish to minimize the maximum mismatch over the imaginary axis.

The $\mathbf{D}_r$ term

Given $\mathbf{H}(s)$, $2r$ distinct points, $\{\mu_i\}_{i=1}^r \cup \{\sigma_j\}_{j=1}^r$, together with $2r$ nontrivial vectors, $\{\mathrm{c}_i\}_{i=1}^r \subset \mathbb{C}^p$ and $\{\mathrm{b}_j\}_{j=1}^r \subset \mathbb{C}^m$, let $\mathbf{V}_r \in \mathbb{C}^{n\times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n\times r}$ be as before. Define $\mathbf{F}$ and $\mathbf{G}$ as

$$\mathbf{F} = [\mathrm{b}_1,\, \mathrm{b}_2,\, ...,\, \mathrm{b}_r\,] \quad \text{and} \quad \mathbf{G} = [\mathrm{c}_1,\, \mathrm{c}_2,\, \ldots,\, \mathrm{c}_r]^T$$

For any $\mathbf{D}_r \in \mathbb{C}^{p\times m}$, define

$$\mathbf{E}_r = \mathbf{W}_r^T\mathbf{E}\mathbf{V}_r,\quad \mathbf{A}_r = \mathbf{W}_r^T\mathbf{A}\mathbf{V}_r + \mathbf{G}\mathbf{D}_r\mathbf{F},\quad \mathbf{B}_r = \mathbf{W}_r^T\mathbf{B} - \mathbf{G}\mathbf{D}_r,\quad \mathbf{C}_r = \mathbf{C}\mathbf{V}_r^T - \mathbf{D}_r\mathbf{F}.$$

Then the reduced-order model satisfies the interpolation constraints.

Measures of Performance: the $\mathcal{H}_\infty$ and $\mathcal{H}_2$ norms

The $\mathcal{H}_\infty$ norm. We want the output error $\mathbf{y}(t) - \mathbf{y}_r(t)$ to be small in a root mean square sense: $\int_0^\infty \|\mathbf{y}(t) - \mathbf{y}_r(t)\|_2^2 \, dt$ to be small, uniformly over all inputs, $\mathbf{u}(t)$, having bounded "energy," $\int_0^\infty \|\mathbf{u}(t)\|_2^2 \, dt \leq 1$. There holds $\widehat{\mathbf{y}}(s) - \widehat{\mathbf{y}}_r(s) = [\mathbf{H}(s) - \mathbf{H}_r(s)]\, \widehat{\mathbf{u}}(s)$, so by the Parseval relation:

$$
\begin{aligned}
\int_0^\infty \|\mathbf{y}(t) - \mathbf{y}_r(t)\|_2^2 \, dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \|\, \widehat{\mathbf{y}}(\imath\omega) - \widehat{\mathbf{y}}_r(\imath\omega)\|_2^2 \; d\omega \\
&\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_2^2 \; \|\widehat{\mathbf{u}}(\imath\omega)\|_2^2 \, d\omega \\
&\leq \max_\omega \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_2^2 \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \|\widehat{\mathbf{u}}(\imath\omega)\|_2^2 \, d\omega \right)^{1/2} \\
&\leq \max_\omega \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_2^2 \left(\int_0^\infty \|\mathbf{u}(t)\|_2^2 \, dt \right)^{1/2} \leq \max_\omega \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_2^2 = \|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_\infty}^2
\end{aligned}
$$

Thus the $\|\mathbf{H}\|_{\mathcal{H}_\infty}$ norm is the L^2 induced operator norm of the associated system mapping, $\mathcal{S} : \mathbf{u} \mapsto \mathbf{y}$. If $\mathbf{E}$ is singular, we must require in addition that 0 is a nondefective eigenvalue of $\mathbf{E}$ so that $\mathbf{H}(s)$ is bounded at infinity.

The $\mathcal{H}_2$ norm

The $\mathcal{H}_2$ norm is:

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

where now $\|\mathbf{R}\|_{\mathrm{F}}^2$ is the Frobenius norm of $\mathbf{R}$. If $\mathbf{E}$ is singular then 0 must be a nondefective eigenvalue of $\mathbf{E}$ and that $\lim_{s\to\infty} \mathbf{H}(s) = 0$, for the $\mathcal{H}_2$ norm of the system to be finite.
If we want $\max_{t>0} \|\mathbf{y}(t) - \mathbf{y}_r(t)\|_\infty$ to be small, over all inputs, $\mathbf{u}(t)$ with $\int_0^\infty \|\mathbf{u}(t)\|_2^2 \, dt \leq 1$:

$$\begin{aligned}
\max_{t>0} \|\mathbf{y}(t) - \mathbf{y}_r(t)\|_\infty &= \max_{t>0} \left\| \frac{1}{2\pi} \int_{-\infty}^{\infty} (\widehat{\mathbf{y}}(\imath\omega) - \widehat{\mathbf{y}}_r(\imath\omega))\, e^{\imath\omega t} \, d\omega \right\|_\infty \\
&\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \|\widehat{\mathbf{y}}(\imath\omega) - \widehat{\mathbf{y}}_r(\imath\omega)\|_\infty \, d\omega \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_F \, \|\widehat{\mathbf{u}}(\imath\omega)\|_2 \, d\omega \\
&\leq \left(\int_{-\infty}^{\infty} \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_F^2 \, d\omega \right)^{1/2} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \|\widehat{\mathbf{u}}(\imath\omega)\|_2^2 \, d\omega \right)^{1/2} \\
&\leq \left(\int_{-\infty}^{\infty} \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_F^2 \, d\omega \right)^{1/2} \left(\int_0^{\infty} \|\mathbf{u}(t)\|_2^2 \, dt \right)^{1/2} \\
&\leq \left(\int_{-\infty}^{+\infty} \|\mathbf{H}(\imath\omega) - \mathbf{H}_r(\imath\omega)\|_F^2 \, d\omega \right)^{1/2} = \|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2}
\end{aligned}$$

In the single-input single-output case, the above relation holds with equality sign, because the $\mathcal{H}_2$ norm is equal to the $(2, \infty)$ induced norm of the convolution operator.

$\mathcal{H}_2$ is the set of matrix-valued functions, $\mathbf{G}(z)$, with components that are analytic for z in the open right half plane, $Re(z) > 0$, and such that for each fixed $Re(z) = x > 0$, $\mathbf{G}(x + \imath y)$ is square integrable in the sense that

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

$\mathcal{H}_2$ is a Hilbert space. Transfer functions associated with stable finite dimensional dynamical systems are elements of $\mathcal{H}_2$. If $\mathbf{G}(s)$ and $\mathbf{H}(s)$ are transfer functions associated with stable dynamical systems the $\mathcal{H}_2$-inner product can be defined as

$$\langle \mathbf{G},\ \mathbf{H} \rangle_{\mathcal{H}_2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{Tr}\left(\overline{\mathbf{G}(\imath\omega)} \mathbf{H}(\imath\omega)^T \right) \, d\omega = \int_{-\infty}^{\infty} \mathrm{Tr}\left(\overline{\mathbf{G}}(-\imath\omega) \mathbf{H}(\imath\omega)^T \right) \, d\omega,$$

with a norm defined as

$$\|\mathbf{G}\|_{\mathcal{H}_2} = \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} \|\mathbf{G}(\imath\omega)\|_F^2 \, d\omega \right)^{1/2}. \tag{10}$$

where $\mathrm{Tr}(\mathbf{M})$ and $\|\mathbf{M}\|_F$ denote the trace and Frobenius norm of $\mathbf{M}$, respectively.

There are two characterizations of this inner product.

First characterization

Suppose $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\widetilde{\mathbf{A}} \in \mathbb{R}^{\tilde{n}\times\tilde{n}}$ are stable and given $\mathbf{B}$, $\mathbf{C} \in \mathbb{R}^{m\times n}$ and $\widetilde{\mathbf{B}}$, $\widetilde{\mathbf{C}} \in \mathbb{R}^{\tilde{m}\times\tilde{n}}$ define associated transfer functions,

$$\mathbf{G}(s) = \mathbf{C}^T(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{B} \quad \text{and} \quad \mathbf{H}(s) = \widetilde{\mathbf{C}}^T(s\mathbf{I}-\widetilde{\mathbf{A}})^{-1}\widetilde{\mathbf{B}}.$$

The inner product $\langle \mathbf{G},\ \mathbf{H}\rangle_{\mathcal{H}_2}$ is associated with solutions to Sylvester equations:

$$\text{If } \mathbf{P} \text{ solves } \mathbf{A}\mathbf{P} + \mathbf{P}\widetilde{\mathbf{A}}^T + \mathbf{B}\widetilde{\mathbf{B}}^T = 0 \text{ then } \langle \mathbf{G},\ \mathbf{H}\rangle_{\mathcal{H}_2} = \operatorname{trace}\left(\mathbf{C}\mathbf{P}\widetilde{\mathbf{C}}^T\right)$$
$$\text{If } \mathbf{Q} \text{ solves } \mathbf{Q}\mathbf{A} + \widetilde{\mathbf{A}}^T\mathbf{Q} + \widetilde{\mathbf{C}}^T\mathbf{C} = 0 \text{ then } \langle \mathbf{G},\ \mathbf{H}\rangle_{\mathcal{H}_2} = \operatorname{trace}\left(\widetilde{\mathbf{B}}^T\mathbf{Q}\mathbf{B}\right)$$

Note that if $\mathbf{A} = \widetilde{\mathbf{A}}$, $\mathbf{B} = \widetilde{\mathbf{B}}$, and $\mathbf{C} = \widetilde{\mathbf{C}}$ then $\mathbf{P}$ is the "reachability Gramian" of $\mathbf{G}(s)$ and $\mathbf{Q}$ is the "observability Gramian" of $\mathbf{G}(s)$. Then

$$\|\mathbf{G}\|^2_{\mathcal{H}_2} = \operatorname{trace}\left(\mathbf{C}\mathbf{P}\mathbf{C}^T\right) = \operatorname{trace}\left(\mathbf{B}^T\mathbf{Q}\mathbf{B}\right)$$

Recently, we obtained a new expression for $\|G\|_{\mathcal{H}_2}$ based on the poles and residues of the transfer function $G(s)$ that complements the widely known alternative expression above. Let λ be a simple pole of $f(s)$, and the residue is nontrivial:

$\mathrm{res}[f(s), \lambda] = \lim_{s\to\lambda}(s-\lambda)f(s) \neq 0$. For matrix-valued functions $\mathbf{F}(s)$, if λ is simple $\mathrm{res}[\mathbf{F}(s), \lambda] = \lim_{s\to\lambda}(s-\lambda)f(s)$ has rank 1. We define the *order* or *dimension* of $\mathbf{F}(s)$ by $\dim \mathbf{F} = \sum_\lambda \mathrm{rank}\,(\mathrm{res}[\mathbf{F}(s), \lambda])$ where the sum is taken over all poles λ. In this case, we can represent $\mathbf{F}(s)$ as

$$\mathbf{F}(s) = \sum_{i=1}^{\dim \mathbf{F}} \frac{1}{s-\lambda_i} \mathrm{c}_i \mathrm{b}_i^T,$$

where λ_i are indexed according to multiplicity as indicated by $\mathit{rank}(\mathrm{res}[\mathbf{F}(s), \lambda_i])$ times.

Second characterization

Suppose that $\mathbf{G}(s)$ and $\mathbf{H}(s)$ are stable (poles contained in the open left halfplane) and suppose that $\mathbf{H}(s)$ has poles at $\mu_1, \mu_2, \ldots \mu_m$. Then

$$\langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2} = \sum_{k=1}^{m} \mathrm{res}[\mathrm{Tr}\left(\overline{\mathbf{G}}(-s)\mathbf{H}(s)^T\right), \mu_k].$$

In particular, if $\mathbf{H}(s)$ has only simple or semi-simple poles at $\mu_1, \mu_2, \ldots \mu_m$ and $m = \dim \mathbf{H}$ then $\mathbf{H}(s) = \sum_{i=1}^{m} \frac{1}{s-\mu_i} \mathrm{c}_i \mathrm{b}_i^T$ and

$$\langle \mathbf{G}, \mathbf{H} \rangle_{\mathcal{H}_2} = \sum_{k=1}^{m} \mathrm{c}_k^T \overline{\mathbf{G}}(-\mu_k) \mathrm{b}_k$$

and

$$\|\mathbf{H}\|_{\mathcal{H}_2} = \left(\sum_{k=1}^{m} \mathrm{c}_k^T \overline{\mathbf{H}}(-\mu_k) \mathrm{b}_k \right)^{1/2}.$$

Proof. Notice that the function $\mathrm{Tr}\left(\overline{\mathbf{G}}(-s)\mathbf{H}(s)^T\right)$ has singularities in the left half plane only at $\mu_1, \mu_2, \ldots \mu_m$. For any $R > 0$, define the semicircular contour in the left halfplane:

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

Γ_R bounds a region that for sufficiently large R contains all the system poles of $\mathbf{H}(s)$ and so, by the residue theorem

$$\begin{aligned}\langle \mathbf{G},\ \mathbf{H}\rangle_{\mathcal{H}_2} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{Tr}\left(\overline{\mathbf{G}}(-\imath\omega)\mathbf{H}(\imath\omega)^T\right)\, d\omega \\ &= \lim_{R\to\infty} \frac{1}{2\pi\imath}\int_{\Gamma_R} \mathrm{Tr}\left(\overline{\mathbf{G}}(-s)\mathbf{H}(s)^T\right)\, ds \\ &= \sum_{k=1}^{m} \mathrm{res}[\mathrm{Tr}\left(\overline{\mathbf{G}}(-s)\mathbf{H}(s)^T\right), \mu_k].\end{aligned}$$

The remaining assertions follow from the definition.

$\mathcal{H}_2$ norm of error system

Given a full-order real system, $\mathbf{H}(s)$ and a reduced model, $\mathbf{H}_r(s)$, having the form $\mathbf{H}_r(s) = \sum_{i=1}^{r} \frac{1}{s-\hat{\lambda}_i}\mathrm{c}_i\mathrm{b}_i{}^T$ ($\mathbf{H}_r$ has simple poles at $\hat{\lambda}_1, \hat{\lambda}_2, \ldots \hat{\lambda}_r$ and rank-1 residues $\mathrm{c}_1\mathrm{b}_1^T, \ldots, \mathrm{c}_r\mathrm{b}_r^T$.), the $\mathcal{H}_2$ norm of the error system is given by

$$\|\mathbf{H}-\mathbf{H}_r\|_{\mathcal{H}_2}^2 = \|\mathbf{H}\|_{\mathcal{H}_2}^2 - 2\sum_{k=1}^{r} \mathrm{c}_k^T\mathbf{H}(-\hat{\lambda}_k)\mathrm{b}_k + \sum_{k,\ell=1}^{r} \frac{\mathrm{c}_k^T\mathrm{c}_\ell\,\mathrm{b}_\ell^T\mathrm{b}_k}{-\hat{\lambda}_k-\hat{\lambda}_\ell}$$

Outline

Interpolatory Optimal $\mathcal{H}_2$ Approximation

Given a full-order system $\mathbf{H}(s)$, the optimal $\mathcal{H}_2$ model reduction problem seeks to find a reduced-order model $\mathbf{H}_r(s)$ that minimized the $\mathcal{H}_2$ error; i.e.

$$\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2} = \min_{\substack{\dim(\widetilde{\mathbf{H}}_r) = r \\ \widetilde{\mathbf{H}}_r : \text{ stable}}} \|\mathbf{H} - \widetilde{\mathbf{H}}_r\|_{\mathcal{H}_2}.$$

The optimization problem is **nonconvex**. The common approach is to find reduced order models that satisfy first-order necessary optimality conditions.

Condition for optimal $\mathcal{H}_2$ reduced model

Suppose $\mathbf{H}(s)$ is a real stable dynamical system and that $\mathbf{H}_r(s) = \sum_{i=1}^{r} \frac{1}{s-\hat{\lambda}_i} \mathrm{c}_i \mathrm{b}_i^T$ is a real dynamical system that is the best stable r^{th} order approximation of $\mathbf{H}$ with respect to the $\mathcal{H}_2$ norm. ($\mathbf{H}_r$ has simple poles at $\hat{\lambda}_1, \hat{\lambda}_2, \ldots \hat{\lambda}_r$ and rank-1 residues $\mathrm{c}_1\mathrm{b}_1^T, \ldots, \mathrm{c}_r\mathrm{b}_r^T$.) Then

$$\textbf{(a)}\ \mathbf{H}(-\hat{\lambda}_k)\mathrm{b}_k = \mathbf{H}_r(-\hat{\lambda}_k)\mathrm{b}_k, \ \ \textbf{(b)}\ \mathrm{c}_k^T\mathbf{H}(-\hat{\lambda}_k) = \mathrm{c}_k^T\mathbf{H}_r(-\hat{\lambda}_k),$$

$$\textbf{(c)}\ \mathrm{c}_k^T\mathbf{H}'(-\hat{\lambda}_k)\mathrm{b}_k = \mathrm{c}_k^T\mathbf{H}_r'(-\hat{\lambda}_k)\mathrm{b}_k, \quad \text{for} \quad k = 1, \cdots, r.$$

Proof. Suppose $\widetilde{\mathbf{H}}_r(s)$ is a transfer function associated with a stable r-th order dynamical system. Then

$$\begin{aligned} \|\mathbf{H}-\mathbf{H}_r\|_{\mathcal{H}_2}^2 \leq \; & \|\mathbf{H}-\widetilde{\mathbf{H}}_r\|_{\mathcal{H}_2}^2 = \|\mathbf{H}-\mathbf{H}_r+\mathbf{H}_r-\widetilde{\mathbf{H}}_r\|_{\mathcal{H}_2}^2 \\ & = \|\mathbf{H}-\mathbf{H}_r\|_{\mathcal{H}_2}^2 + 2\,\Re e\,\langle \mathbf{H}-\mathbf{H}_r, \mathbf{H}_r-\widetilde{\mathbf{H}}_r\rangle_{\mathcal{H}_2} + \|\mathbf{H}_r-\widetilde{\mathbf{H}}_r\|_{\mathcal{H}_2}^2 \\ \text{so that} \quad & 0 \leq 2\,\Re e\,\langle \mathbf{H}-\mathbf{H}_r, \mathbf{H}_r-\widetilde{\mathbf{H}}_r\rangle_{\mathcal{H}_2} + \|\mathbf{H}_r-\widetilde{\mathbf{H}}_r\|_{\mathcal{H}_2}^2 \end{aligned}$$

By making judicious choices in how $\widetilde{\mathbf{H}}_r$ is made to differ from $\mathbf{H}_r$, we arrive at tangential interpolation conditions. Toward this end, pick an arbitrary unit vector $\boldsymbol{\xi} \in \mathbb{C}^m$, $\varepsilon > 0$, and for some ℓ, define $\theta = \pi - \arg \boldsymbol{\xi}^T \left(\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)\right) \mathrm{b}_\ell$, and so that

$$\mathbf{H}_r(s) - \widetilde{\mathbf{H}}_r(s) = \frac{\varepsilon e^{\imath\theta}}{s-\hat{\lambda}_\ell}\boldsymbol{\xi}\mathrm{b}_\ell^T$$

and hence

$$\langle \mathbf{H}-\mathbf{H}_r, \mathbf{H}_r-\widetilde{\mathbf{H}}_r\rangle_{\mathcal{H}_2} = -\varepsilon\,|\boldsymbol{\xi}^T\left(\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)\right)\mathrm{b}_\ell|.$$

Now we have

$$0 \leq |\boldsymbol{\xi}^T\left(\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)\right)\mathrm{b}_\ell| \leq \varepsilon\frac{\|\mathrm{b}_\ell\|_2^2}{-2\Re e(\hat{\lambda}_\ell)}$$

which by taking ε small implies first that

$$\boldsymbol{\xi}^T\left(\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)\right)\mathrm{b}_\ell = 0$$

but then since $\boldsymbol{\xi}$ was chosen arbitrarily, we must have that

$$\left(\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)\right)\mathrm{b}_\ell = 0.$$

A similar argument yields (b).
If (c) did not hold then $c_\ell^T \mathbf{H}'(-\hat{\lambda}_\ell) b_\ell \neq c_\ell^T \mathbf{H}_r'(-\hat{\lambda}_\ell) b_\ell$ and we may pick $0 < \varepsilon < |\Re e(\hat{\lambda}_\ell)|$ and $\theta = -\arg c_\ell^T \left(\mathbf{H}'(-\hat{\lambda}_\ell) - \mathbf{H}_r'(-\hat{\lambda}_\ell)\right) b_\ell$ in such a way that $\mu = \hat{\lambda}_\ell + \varepsilon e^{\imath\theta}$ does not coincide with any reduced order poles $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_r$. Note that $\Re e(\mu) < 0$. Define $\widetilde{\mathbf{H}}_r(s)$ so that

$$\mathbf{H}_r(s) - \widetilde{\mathbf{H}}_r(s) = \left(\frac{1}{s - \hat{\lambda}_\ell} - \frac{1}{s - \mu}\right) c_\ell b_\ell^T.$$

$\widetilde{\mathbf{H}}_r(s)$ has the same poles and residues as $\mathbf{H}_r(s)$ aside from μ which replaces $\hat{\lambda}_\ell$ as a pole in $\widetilde{\mathbf{H}}_r(s)$ with no change in the associated residue. Because of conditions (a-b), we calculate

$$\begin{aligned}\langle \mathbf{H} - \mathbf{H}_r, \mathbf{H}_r - \widetilde{\mathbf{H}}_r \rangle_{\mathcal{H}_2} &= c_\ell^T \left(\mathbf{H}(-\hat{\lambda}_\ell) - \mathbf{H}_r(-\hat{\lambda}_\ell)\right) b_\ell - c_\ell^T \left(\mathbf{H}(-\mu) - \mathbf{H}_r(-\mu)\right) b_\ell \\ &= -\, c_\ell^T \left(\mathbf{H}(-\mu) - \mathbf{H}_r(-\mu)\right) b_\ell\end{aligned}$$

Then we get

$$0 \leq -2\Re e(c_\ell^T \left(\mathbf{H}(-\mu) - \mathbf{H}_r(-\mu)\right) b_\ell) - \frac{|\mu - \hat{\lambda}_\ell|^2}{|\overline{\mu} + \hat{\lambda}_\ell|^2} \frac{\Re e(\mu + \hat{\lambda}_\ell)}{2\Re e(\hat{\lambda}_\ell)\Re e(\mu)} \|c_\ell\|_2^2 \, \|b_\ell\|_2^2$$

Now, easy manipulations yield first a resolvent identity

$$(-\mu\mathbf{E} - \mathbf{A})^{-1} = (-\hat{\lambda}_\ell\mathbf{E} - \mathbf{A})^{-1} + (\mu - \hat{\lambda}_\ell)(-\hat{\lambda}_\ell\mathbf{E} - \mathbf{A})^{-1}\mathbf{E}(-\mu\mathbf{E} - \mathbf{A})^{-1}$$

and then resubstituting,

$$\begin{aligned}(-\mu\mathbf{E} - \mathbf{A})^{-1} &= (-\hat{\lambda}_\ell\mathbf{E} - \mathbf{A})^{-1} + (\mu - \hat{\lambda}_\ell)(-\hat{\lambda}_\ell\mathbf{E} - \mathbf{A})^{-1}\mathbf{E}(-\hat{\lambda}_\ell\mathbf{E} - \mathbf{A})^{-1} \\ &+ (\mu - \hat{\lambda}_\ell)^2(-\hat{\lambda}_\ell\mathbf{E} - \mathbf{A})^{-1}\mathbf{E}(-\mu\mathbf{E} - \mathbf{A})^{-1}\mathbf{E}(-\hat{\lambda}_\ell\mathbf{E} - \mathbf{A})^{-1}\end{aligned}$$

Premultiplying by **C** and postmultiplying by **B**, implies

$$\begin{aligned}\mathbf{H}(-\mu) &= \mathbf{H}(-\hat{\lambda}_\ell) + (\mu - \hat{\lambda}_\ell)\mathbf{H}'(-\hat{\lambda}_\ell) \\ &\quad + (\mu - \hat{\lambda}_\ell)^2 \mathbf{C}(-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1}\mathbf{E}(-\mu\mathbf{E} - \mathbf{A})^{-1}\mathbf{E}(-\hat{\lambda}_\ell \mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\end{aligned}$$

and analogous arguments yield

$$\begin{aligned}\mathbf{H}_r(-\mu) &= \mathbf{H}_r(-\hat{\lambda}_\ell) + (\mu - \hat{\lambda}_\ell)\mathbf{H}_r'(-\hat{\lambda}_\ell) \\ &\quad + (\mu - \hat{\lambda}_\ell)^2 \mathbf{C}_r(-\hat{\lambda}_\ell \mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{E}_r(-\mu\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{E}_r(-\hat{\lambda}_\ell \mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r\end{aligned}$$

Using these expressions yields

$$0 \leq -2\varepsilon |\mathsf{c}_\ell^T \left(\mathbf{H}'(-\hat{\lambda}_\ell) - \mathbf{H}_r'(-\hat{\lambda}_\ell)\right) \mathsf{b}_\ell| + \mathcal{O}(\varepsilon^2).$$

Letting $\varepsilon \longrightarrow 0$ forces a contradiction unless $|\mathsf{c}_\ell^T \left(\mathbf{H}'(-\hat{\lambda}_\ell) - \mathbf{H}_r'(-\hat{\lambda}_\ell)\right) \mathsf{b}_\ell| = 0$. ∎

Remarks.

(a) One consequence of this Theorem is that if $\mathbf{H}_r(s)$ interpolates a real system $\mathbf{H}(s)$ at the *mirror images* its own poles (i.e., at the poles of $\mathbf{H}_r(s)$ reflected across the imaginary axis), then $\mathbf{H}_r(s)$ is guaranteed to be an *optimal* approximation of $\mathbf{H}(s)$ relative to the $\mathcal{H}_2$ norm among all reduced order systems having the same reduced system poles $\{\mu_i\}_{i=1}^r$.

(b) The set of stable r^{th} order dynamical systems is not convex and so, the original problem allows for multiple minimizers. A reduced order system, $\mathbf{H}_r$, is a *local minimizer*, if for all $\varepsilon > 0$ sufficiently small,

$$\|G - G_r\|_{\mathcal{H}_2} \leq \|G - \widetilde{G}_r^{(\varepsilon)}\|_{\mathcal{H}_2},$$

for all stable dynamical systems, $\widetilde{G}_r^{(\varepsilon)}$ with $\dim(\widetilde{G}_r^{(\varepsilon)}) = r$ and

$$\|G_r - \widetilde{G}_r^{(\varepsilon)}\|_{\mathcal{H}_2} \leq C\,\varepsilon,$$

C is a constant that may depend on the particular family $\widetilde{G}_r^{(\varepsilon)}$ considered.

(c) As a practical matter, the global minimizers are difficult to obtain with certainty; current approaches favor seeking reduced order models that satisfy a local (first-order) necessary condition for optimality. Even though such strategies do not guarantee global minimizers, they often produce effective reduced order models.

A numerical algorithm for optimal $\mathcal{H}_2$ model reduction

The *Iterative Rational Krylov Algorithm* (**IRKA**) resolves this problem.

Let $\mathbf{Y}^*$ and $\mathbf{X}$ denote the left and right eigenvectors for $\lambda \mathbf{E}_r - \mathbf{A}_r$ so that $\mathbf{Y}^* \mathbf{A}_r \mathbf{X} = \mathrm{diag}(\widetilde{\lambda}_i)$ and $\mathbf{Y}^* \mathbf{E}_r \mathbf{X} = \mathbf{I}_r$.

Denote the columns of $\mathbf{C}_r \mathbf{X}$ as $\tilde{c}_i$ and the rows of $\mathbf{Y}^* \mathbf{B}_r$ as $\tilde{b}_i^T$.

Then in **IRKA** interpolation points used in the next step are $\lambda(\mathbf{A}_r, \mathbf{E}_r)$, of the pencil $\lambda \mathbf{E}_r - \mathbf{A}_r$ in the current step. The tangent directions are corrected using the residues of the previous reduced-ordered model until the optimality condition is satisfied.

A numerical algorithm for optimal $\mathcal{H}_2$ model reduction

Optimal $\mathcal{H}_2$ MIMO tangetial interpolation alorithm

1. Make an initial r-fold shift selection: $\{\sigma_1, \ldots, \sigma_r\}$ that is closed under conjugation and initial tangent directions $\hat{b}_1, \ldots, \hat{b}_r$ and $\hat{c}_1, \ldots, \hat{c}_r$, also closed under conjugation.
2. $\mathbf{V}_r = \left[(\sigma_1\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\hat{b}_1 \ \cdots \ (\sigma_r\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\hat{b}_r \right]$
   $\mathbf{W}_r = \left[(\sigma_1\,\mathbf{E} - \mathbf{A}^T)^{-1}\mathbf{C}^T\hat{c}_1 \ \cdots \ (\sigma_r\,\mathbf{E} - \mathbf{A}^T)^{-1}\mathbf{C}^T\hat{c}_1 \right]$.
3. while (not converged)
   1. $\mathbf{A}_r = \mathbf{W}_r^T\mathbf{A}\mathbf{V}_r$, $\mathbf{E}_r = \mathbf{W}_r^T\mathbf{E}\mathbf{V}_r$, $\mathbf{B}_r = \mathbf{W}_r^T\mathbf{B}$, and $\mathbf{C}_r = \mathbf{C}\mathbf{V}_r$
   2. Compute $\mathbf{Y}^*\mathbf{A}_r\mathbf{X} = diag(\tilde{\lambda}_i)$ and $\mathbf{Y}^*\mathbf{E}_r\mathbf{X} = \mathbf{I}_r$ where $\mathbf{Y}^*$ and $\mathbf{X}$ are the left and right eigenvectors of $\lambda\mathbf{E}_r - \mathbf{A}_r$.
   3. $\sigma_i \longleftarrow -\lambda_i(\mathbf{A}_r, \mathbf{E}_r)$ for $i = 1, \ldots, r$, $\hat{b}_i^* \longleftarrow \mathbf{e}_i^T\mathbf{Y}^*\mathbf{B}_r$ and $\hat{c}_i \longleftarrow \mathbf{C}_r\mathbf{X}\mathbf{e}_i$.
   4. $\mathbf{V}_r = \left[(\sigma_1\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\hat{b}_1 \ \cdots \ (\sigma_r\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\hat{b}_r \right]$
   5. $\mathbf{W}_r = \left[(\sigma_1\,\mathbf{E} - \mathbf{A}^T)^{-1}\mathbf{C}^T\hat{c}_1 \ \cdots \ (\sigma_r\,\mathbf{E} - \mathbf{A}^T)^{-1}\mathbf{C}^T\hat{c}_1 \right]$.
4. $\mathbf{A}_r = \mathbf{W}_r^T\mathbf{A}\mathbf{V}_r$, $\mathbf{E}_r = \mathbf{W}_r^T\mathbf{E}\mathbf{V}_r$, $\mathbf{B}_r = \mathbf{W}_r^T\mathbf{B}$, $\mathbf{C}_r = \mathbf{C}\mathbf{V}_r$

A numerical algorithm for optimal $\mathcal{H}_2$ model reduction

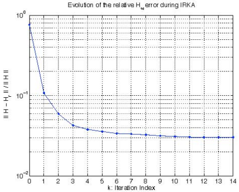
• The main computational cost of this algorithm involves solving $2r$ linear systems to generate $\mathbf{V}_r$ and $\mathbf{W}_r$. Computing the eigenvectors $\mathbf{Y}$ and $\mathbf{X}$, and the eigenvalues of the reduced pencil $\lambda\mathbf{E}_r - \mathbf{A}_r$ are cheap since the dimension r is small.

• **IRKA** has been remarkably successful in producing high fidelity reduced-order approximations; it has been successfully applied to finding $\mathcal{H}_2$-optimal reduced models for systems of order $n > 160,000$.

Numerical Results for IRKA

• This problem arises during a cooling process in a rolling mill and is modeled as boundary control of a two dimensional heat equation. A finite element discretization results in a descriptor system state-dimension $n = 79,841$, i.e., $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{79841 \times 79841}$, $\mathbf{B} \in \mathbb{R}^{79841 \times 7}$, $\mathbf{C} \in \mathbb{R}^{6 \times 79841}$.

• Using IRKA, we reduce the order of the full-order system, $\mathbf{H}(s)$, to $r = 20$ to obtain our $\mathcal{H}_2$ optimal reduced model, $\mathbf{H}_{\rm IRKA}(s)$. The figure depicts how the relative $\mathcal{H}_\infty$ error, $\dfrac{\|\mathbf{H} - \mathbf{H}_{\rm IRKA}\|_{\mathcal{H}_\infty}}{\|\mathbf{H}\|_{\mathcal{H}_\infty}}$ evolves through out the iteration



• The important observation is that starting from an initial relative error values close to 1, the method automatically, corrects both the interpolation points and tangential directions, and reaches an optimal solution with an error value around 3×10^{-2}.

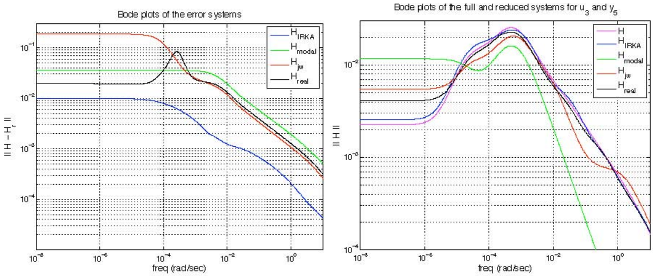
We compare our approach with other commonly used MIMO model reduction techniques.

1. Modal Approximation: We reduce the order $r = 20$ using 20 dominant modes of $\mathbf{H}(s)$. The reduced model is denoted by $\mathbf{H}_{\rm modal}$
2. Interpolation points on the imaginary axis: Based on the Bode plot of $\mathbf{H}(s)$, we have placed interpolation points on the imaginary axis where $\|\mathbf{H}(\jmath\omega)\|$ is dominant. This was a very common approach of choosing the interpolation before the optimal shift selection strategy have been developed. The reduced model is denoted by $\mathbf{H}_{\jmath\omega}$.
3. 20 real interpolation points are chosen in the mirror images of the poles of $\mathbf{H}(s)$. This selection is a good initialization for IRKA in the SISO case. The reduced model is denoted by $\mathbf{H}_{\rm real}$.

	$\mathbf{H}_{\rm IRKA}$	$\mathbf{H}_{\rm modal}$	$\mathbf{H}_{\jmath\omega}$	$\mathbf{H}_{\rm real}$
Relative $\mathcal{H}_\infty$ error	3.03×10^{-2}	1.03×10^{-1}	5.42×10^{-1}	2.47×10^{-1}

In order to be able to find reasonable interpolation points and directions for $\mathbf{H}_{\jmath\omega}$ and $\mathbf{H}_{\rm real}$, we needed to run several experiments. They either resulted in unstable systems or very poor performance results. Here we are presenting the best selection we were able to find.

We initiate IRKA it once, randomly in this case, and the algorithm automatically find the optimal points and directions. *There is no need for an ad hoc search.*
Note from Figure 14 that the initial guess for IRKA has a higher error than all the other methods. However, even after only two steps of the iteration long before convergence, the IRKA iterate has already a smaller error norm than all other three approaches. Note that $\mathbf{H}(s)$ has 7 inputs and 6 outputs; hence there are 42 input-output channels. $\mathbf{H}_{\rm IRKA}$ with order $r = 20$, less than the total number of input-output channels, is able to replicate these behaviors with a relative accuracy of order 10^{-2}. Even though IRKA is an $\mathcal{H}_2$-based approach, superior $\mathcal{H}_\infty$ performance is also observed and is not surprising. It is an efficient general purpose $\mathcal{H}_2$ and $\mathcal{H}_\infty$ model reduction method, not only $\mathcal{H}_2$ optimal.



Left pane: depicts the Bode Plots of the error systems. It is clear that $\mathbf{H}_{\mathrm{IRKA}}$ outperforms the rest of the methods. To give an example how the reduced models match the full-order model for a specific input/output channel, we show in the right pane the Bode plots for the transfer function between the third input $\mathbf{u}_3$ and the fifth output $\mathbf{y}_5$.

Since IRKA yielded a much lower error value, we have checked what lowest order model from IRKA would yield similar error norms as the other approaches. We have found that IRKA for order $r = 2$ yields a relative $\mathcal{H}_\infty$ error of 2.26×10^{-1}; already better than 20^{th} order $\mathbf{H}_{\jmath\omega}$ and $\mathbf{H}_{\mathrm{real}}$. For $r = 6$, for example, IRKA yielded a relative $\mathcal{H}_\infty$ error of 1.64×10^{-1}, a number close to the one obtained by 20^{th} order $\mathbf{H}_{\mathrm{modal}}$.

Outline

Interpolatory Passivity Preserving Model Reduction

A system is *passive* if

$$\mathcal{R}e \int_{-\infty}^{t} \mathbf{u}(\tau)^* \mathbf{y}(\tau) \mathrm{d}\tau \geq 0,$$

for all $t \in \mathbb{R}$ and for all $\mathbf{u} \in \mathcal{L}_2(\mathbb{R})$. A rational (square) matrix function $\mathbf{H}(s)$ is *positive real* if:

1. $\mathbf{H}(s)$, is analytic for $\mathcal{R}e(s) > 0$,
2. $\overline{\mathbf{H}(s)} = \mathbf{H}(\bar{s})$ for all $s \in \mathbb{C}$, and
3. the Hermitian part of $\mathbf{H}(s)$, i.e., $\mathbf{H}(s) + \mathbf{H}^T(\bar{s})$, is positive semidefinite for all $\mathcal{R}e(s) \geq 0$.

• Dynamical systems are passive if and only if the associated transfer function $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$, is positive real.
• A consequence of positive realness is the existence of a *spectral factorization*: a matrix function $\mathbf{\Phi}(s)$ for which $\mathbf{H}(s) + \mathbf{H}^T(-s) = \mathbf{\Phi}(s)\mathbf{\Phi}^T(-s)$, and the poles as well as the (finite) zeros of $\mathbf{\Phi}(s)$ are all stable. The *spectral zeros* are λ, for which $\mathbf{\Phi}(\lambda)$ (and hence $\mathbf{H}(\lambda) + \mathbf{H}^T(-\lambda)$) loses rank. Assume for simplicity that λ is a spectral zero with multiplicity one (so that nullity$(\mathbf{\Phi}(\lambda)) = 1$). Then there is a right spectral zero direction, $\mathbf{z}$, such that $(\mathbf{H}(\lambda) + \mathbf{H}^T(-\lambda))\mathbf{z} = \mathbf{0}$. Evidently, if $(\lambda, \mathbf{z})$ is a right spectral zero pair for the system represented by $\mathbf{H}(s)$, then $(-\overline{\lambda}, \mathbf{z}^*)$ is a left spectral zero pair: $\mathbf{z}^*(\mathbf{H}(-\overline{\lambda}) + \mathbf{H}^T(\overline{\lambda})) = \mathbf{0}$. The key result that is:

Passivity preserving tangential interpolation

Suppose the dynamical system given in (1) and represented by the transfer function $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$ is stable and passive. Suppose that for some index $r \geq 1$, $\lambda_1, \cdots, \lambda_r$ are stable spectral zeros of $\mathbf{H}$ with corresponding right spectral zero directions $\mathbf{z}_1, \ldots, \mathbf{z}_r$.

If a reduced order system $\mathbf{H}_r(s)$ tangentially interpolates $\mathbf{H}(s)$ with $\sigma_i = \lambda_i$, $\mathrm{b}_i = \mathbf{z}_i$, $\mu_i = -\overline{\lambda_i}$, and $\mathrm{c}_i^T = \mathbf{z}^*$ for $i = 1, \ldots, r$, then $\mathbf{H}_r(s)$ is stable and passive.

The computation of the spectral zeros of the system can be formulated as a structured eigenvalue problem. Let

$$\mathcal{H} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & -\mathbf{A}^T & -\mathbf{C}^T \\ \mathbf{C} & \mathbf{B}^T & \mathbf{D}+\mathbf{D}^T \end{bmatrix}, \ \mathcal{E} = \begin{bmatrix} \mathbf{E} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

The *spectral zeros* of the system are the generalized eigenvalues of the pencil: $\mathcal{H}\mathbf{x}_i = \lambda_i \, \mathcal{E}\mathbf{x}_i$. To see this, partition the eigenvector $\mathbf{x}_i$ into components $\mathbf{v}_i$, $\overline{\mathbf{w}_i}$, and $\mathbf{z}_i$ such that

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & -\mathbf{A}^T & -\mathbf{C}^T \\ \mathbf{C} & \mathbf{B}^T & \mathbf{D}+\mathbf{D}^T \end{bmatrix} \begin{bmatrix} \mathbf{v}_i \\ \overline{\mathbf{w}_i} \\ \mathbf{z}_i \end{bmatrix} = \lambda_i \begin{bmatrix} \mathbf{E} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_i \\ \overline{\mathbf{w}_i} \\ \mathbf{z}_i \end{bmatrix},$$

Then

$$\mathbf{v}_i = (\lambda_i \mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathbf{z}_i, \ \ \mathbf{w}_i^T = \mathbf{z}_i^* \mathbf{C}(-\overline{\lambda_i}\mathbf{E} - \mathbf{A})^{-1} \ \text{ and } \ (\mathbf{H}(\lambda_i) + \mathbf{H}^T(-\lambda_i))\mathbf{z}_i = \mathbf{0}.$$

Thus, λ_i are spectral zeros of $\mathbf{H}(s)$ associated with the right spectral zero directions, $\mathbf{z}_i$, for $i = 1, \cdots, r$. Furthermore, the right and left conditions are determined by the remaining two components of $\mathbf{x}_i$: $\mathbf{V}_r = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r]$ and $\mathbf{W}_r = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r]$.
Since $\mathcal{H}$ and $\mathcal{E}$ are real, the eigenvalues of $\mathcal{H}\mathbf{x} = \lambda \, \mathcal{E}\mathbf{x}$ occur in complex conjugate pairs. Thus, the spectral zeros, associated spectral zero directions, and bases for the left and right modeling subspaces can be obtained by means of the above Hamiltonian eigenvalue problem.

The question remains of which r spectral zeros to choose. The concept of *dominance* arising in modal approximation is useful in distinguishing effective choices of spectral zero sets.
Assume for simplicity that $\mathbf{D} + \mathbf{D}^T$ is invertible, take $\Delta = (\mathbf{D} + \mathbf{D}^T)^{-1}$ and define

$$\mathcal{B} = \begin{bmatrix} \mathbf{B} \\ -\mathbf{C}^T \\ \mathbf{0} \end{bmatrix} \Delta, \quad \mathcal{C} = -\Delta[\mathbf{C}, \quad \mathbf{B}^T, \quad \mathbf{0}],$$

It can be checked that

$$\mathbf{G}(s) \overset{\Delta}{=} [\mathbf{H}(s) + \mathbf{H}^T(-s)]^{-1} = \Delta + \mathcal{C}(s\mathcal{E} - \mathcal{H})^{-1}\mathcal{B}.$$

Let the partial fraction expansion of $\mathbf{G}(s)$ be

$$\mathbf{G}(s) = \sum_{j=1}^{2n} \frac{\mathbf{R}_j}{s - \lambda_j}, \quad \text{with} \quad \mathbf{R}_j = \frac{1}{\mathbf{y}_j^* \mathcal{E} \mathbf{x}_j} \mathcal{C} \mathbf{x}_j \mathbf{y}_j^* \mathcal{B},$$

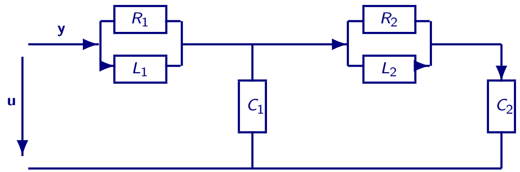
where λ_i are the spectral zeros of the original system (poles of the associated Hamiltonian system) and $\mathbf{R}_j$ are the *residues*. The left and right eigenvectors of $\mathbf{y}_j$, $\mathbf{x}_j$, are computed from $\mathcal{H}\mathbf{x}_j = \lambda_j\, \mathcal{E}\mathbf{x}_j$ and $\mathbf{y}_j^* \mathcal{H} = \lambda_j\, \mathbf{y}_j^* \mathcal{E}$.
A spectral zero λ_i is *dominant* over another spectral zero λ_j, if

$$\frac{\|\mathbf{R}_i\|_2}{|\Re e(\lambda_i)|} > \frac{\|\mathbf{R}_j\|_2}{|\Re e(\lambda_j)|}.$$

To efficiently compute the r *most dominant* spectral zeros of a dynamical system represented by $\mathbf{H}(s)$, the algorithm SADPA (Subspace Accelerated Dominant Pole Algorithm) is used.

An example

Consider the following RLC circuit:



Using the voltages across C_1, C_2, and the currents through L_1, L_2, as state variables, $\mathbf{x}_i$, $= 1, 2, 3, 4$, respectively, we end up with equations of the form $\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$, $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)$, where

$$\mathbf{E} = \begin{pmatrix} C_1 & 0 & -G_1 L_1 & G_2 L_2 \\ 0 & C_2 & 0 & -G_2 L_2 \\ 0 & 0 & L_1 & 0 \\ 0 & 0 & 0 & L_2 \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{C} = [-G_1, \ \ 0, \ \ 1, \ \ 0], \quad \mathbf{D} = G_1,$$

$G_i = \frac{1}{R_i}$, $i = 1, 2$, are the corresponding conductances. By construction, the system is passive and it is easy to see that its transfer function has a zero at $s = 0$. Hence the system has a double spectral zero at $s = 0$. According to the definition of dominance mentioned above, among all finite spectral zeros, those on the imaginary axis are dominant. Hence we will compute a second order reduced system by using the the eigenpairs of $(\mathcal{H},\ \mathcal{E})$, corresponding to the double zero eigenvalue. The Hamiltonian pair is:

$$\mathcal{H} = \begin{pmatrix} 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & G_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ -G_1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 2G_1 \end{pmatrix},$$

and

$$\mathcal{E} = \begin{pmatrix} C_1 & 0 & -G_1L_1 & G_2L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_2 & 0 & -G_2L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & L_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & C_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & C_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -G_1L_1 & 0 & L_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & G_2L_2 & -G_2L_2 & 0 & L_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

It follows that although the algebraic multiplicity of this eigenvalue is two, its geometric multiplicity is only one. Hence we need a Jordan chain of eigenvectors $\mathbf{x}_1$, $\mathbf{x}_2$, corresponding to this eigenvalue. In particular, $\mathbf{x}_1$ satisfies $\mathcal{H}\mathbf{x}_1 = 0$, while $\mathbf{x}_2$, satisfies $\mathcal{H}\mathbf{x}_2 = \mathcal{E}\mathbf{x}_1$. These eigenvectors are:

$$[\mathbf{x}_1, \quad \mathbf{x}_2] = \left(\begin{array}{c|c} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & \frac{C_2}{C_1+C_2} \\ \hline -1 & 0 \\ -1 & 0 \\ -G_1 & -1 \\ 0 & \frac{-C_2}{C_1+C_2} \\ \hline 1 & 0 \end{array}\right).$$

Thus the projection is defined by means of $\mathbf{V}_r = [\mathbf{x}_1(1:4,:), \quad \mathbf{x}_2(1:4,:)]$ and $\mathbf{W}_r = -[\mathbf{x}_1(5:8,:), \quad \mathbf{x}_2(5:8,:)]$, that is

$$\mathbf{V}_r = \left(\begin{array}{c|c} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & \frac{C_2}{C_1+C_2} \end{array}\right), \quad \mathbf{W}_r = \left(\begin{array}{c|c} 1 & 0 \\ 1 & 0 \\ G_1 & 1 \\ 0 & \frac{C_2}{C_1+C_2} \end{array}\right).$$

Therefore by (6) the reduced quantities are:

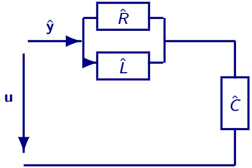
$$\mathbf{E}_r = \mathbf{W}_r^*\mathbf{E}\mathbf{V}_r = \begin{pmatrix} C_1+C_2 & 0 \\ 0 & L_1 + \frac{{C_2}^2}{(C_1+C_2)^2}L_2 \end{pmatrix}, \mathbf{A}_r = \mathbf{W}_r^*\mathbf{A}\mathbf{V}_r = \begin{pmatrix} -G_1 & 1 \\ -1 & 0 \end{pmatrix},$$

$$\mathbf{B}_r = \mathbf{W}_r^*\mathbf{B} = \begin{pmatrix} G_1 \\ 1 \end{pmatrix}, \quad \mathbf{C}_r = \mathbf{C}\mathbf{V}_r = \begin{pmatrix} -G_1 & 1 \end{pmatrix}, \quad \mathbf{D}_r = \mathbf{D}.$$

The corresponding transfer function $\mathbf{H}_r(s) = \mathbf{D} + \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r$, can be expressed as follows:

$$\mathbf{H}_r^{-1}(s) = \frac{1}{s(C_1 + C_2)} + \frac{1}{G_1 + \frac{\kappa}{s}} \quad \text{where} \quad \kappa = \frac{L_1 L_2 (C_1 + C_2)^2}{L_1 {C_2}^2 + L_2 (C_1 + C_2)^2}.$$

From this expression or from the state space matrices, we can read-off an RLC realization. The reduced order circuit contains namely, a capacitor $\hat{C} = C_1 + C_2$, an inductor $\hat{L} = \frac{1}{\kappa} = L_1 + \frac{{C_2}^2}{(C_1+C_2)^2} L_2$, and a resistor of value $\hat{R} = R_1$. Thereby the capacitor is in series with a parallel connection of the inductor and the resistor as shown below.



Hence in this particular case, after reduction besides passivity, the structure (topology) of the circuit using the spectral zero reduction method, is preserved.

Outline

Model reduction for co-prime factorization models

There exist linear dynamical systems whose description takes a from different from the standard one.

Example: the dynamic response of a viscoelastic body. Vibrations of an isotropic incompressible viscoelastic solid are described as

$$\partial_{tt}\mathbf{w}(x,t) - \eta\,\Delta\mathbf{w}(x,t) - \int_0^t \rho(t-\tau)\,\Delta\mathbf{w}(x,\tau)\,d\tau + \nabla\varpi(x,t) = \mathbf{b}(x)\cdot\mathbf{u}(t),$$

$$\nabla\cdot\mathbf{w}(x,t) = 0 \quad \text{determining} \quad \mathbf{y}(t) = [\mathbf{w}(\hat{x}_1,t), \ldots \mathbf{w}(\hat{x}_p,t)]^T,$$

where

$\mathbf{w}(x,t)$ is the displacement field
$\varpi(x,t)$ is the associated pressure field
$\nabla\cdot\mathbf{w} = 0$ represents the incompressibility constraint
$\rho(\tau)$ is a known "relaxation function" satisfying $\rho(\tau) \geq 0$ and $\int_0^\infty \rho(\tau)\,d\tau < \infty$
$\eta > 0$ is a known constant associated with the "initial elastic response"
$\mathbf{b}(x)\cdot\mathbf{u}(t) = \sum_{i=1}^m b_i(x)\,u_i(t)$ is a superposition of the m inputs
Displacements at $\hat{x}_1, \ldots, \hat{x}_p$ are the outputs.

Semidiscretization with respect to space produces a large order linear dynamical system of the form:

$$\mathbf{M}\,\ddot{\mathbf{x}}(t) + \eta\,\mathbf{K}\,\mathbf{x}(t) + \int_0^t \rho(t-\tau)\,\mathbf{K}\,\mathbf{x}(\tau)\,d\tau + \mathbf{D}\,\boldsymbol{\varpi}(t) = \mathbf{B}\,\mathbf{u}(t),$$

$$\mathbf{D}^T\,\mathbf{x}(t) = \mathbf{0}, \qquad \text{which determines} \quad \mathbf{y}(t) = \mathbf{C}\,\mathbf{x}(t). \tag{11}$$

where

$\mathbf{x} \in \mathbb{R}^{n_1}$ is the discretization of the displacement field $\mathbf{w}$
$\boldsymbol{\varpi} \in \mathbb{R}^{n_2}$ is the discretization of the pressure field ϖ
$\mathbf{M}$ and $\mathbf{K}$ are $n_1 \times n_1$ real, symmetric, positive-definite matrices
The state dimension is $n = n_1 + n_2$.

Applying Laplace transform yields

$$\widehat{\mathbf{y}}(s) = \begin{bmatrix} \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} s^2\mathbf{M} + (\widehat{\rho}(s) + \eta)\,\mathbf{K} & \mathbf{D} \\ \mathbf{D}^T & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix} \widehat{\mathbf{u}}(s) \;=\; \mathcal{H}(s)\,\widehat{\mathbf{u}}(s),$$

defining the transfer function, $\mathcal{H}(s)$. This is a *descriptor* system described by differential-algebraic equations with a hereditary damping. A reformulation into the standard form is usually not feasible. Hence the question becomes how one can obtain a reduced-order model in this setting.
An effective reduced model should take into account the structure associated with the effect of distributed material properties. Therefore, we consider reduced models of the form

$$\mathbf{M}_r\,\ddot{\mathbf{x}}_r(t) \;+\; \eta\,\mathbf{K}_r\,\mathbf{x}_r(t) \;+\; \int_0^t \rho(t-\tau)\,\mathbf{K}_r\,\mathbf{x}_r(\tau)\,d\tau + \mathbf{D}_r\,\boldsymbol{\varpi}_r = \mathbf{B}_r\,\mathbf{u}(t)$$

$$\mathbf{D}_r^T\,\mathbf{x}_r(t) = \mathbf{0} \qquad \text{which determines} \quad \mathbf{y}_r(t) = \mathbf{C}_r\,\mathbf{x}_r(t) \tag{12}$$

where

$\mathbf{M}_r$ and $\mathbf{K}_r$ are $r_1 \times r_1$ real, symmetric, positive-definite matrices
the reduced state space dimension is $r = r_1 + r_2$.

We define matrices of trial vectors $\mathbf{U}_r \in \mathbb{R}^{n_1 \times r_1}$ and $\mathbf{Z}_r \in \mathbb{R}^{n_2 \times r_2}$; use the ansatz $\mathbf{x}(t) \approx \mathbf{U}_r\mathbf{x}_r(t)$ and $\boldsymbol{\varpi}(t) \approx \mathbf{Z}_r\boldsymbol{\varpi}_r(t)$; and force Galerkin condition on the reduced state-space trajectories to obtain the reduced coefficient matrices as

$$\mathbf{M}_r = \mathbf{U}_r^T\mathbf{M}\mathbf{U}_r, \quad \mathbf{K}_r = \mathbf{U}_r^T\mathbf{K}\mathbf{U}_r, \quad \mathbf{D}_r = \mathbf{U}_r^T\mathbf{D}\mathbf{Z}_r, \quad \mathbf{B}_r = \mathbf{U}_r^T\mathbf{B} \;\text{ and }\; \mathbf{C}_r = \mathbf{C}\mathbf{U}_r.$$

This construction prevents mixing displacement state variables and pressure state variables. Also both symmetry and positive-definiteness are preserved automatically. Now, in addition to preserving the structure we want to choose $\mathbf{U}_r$ and $\mathbf{Z}_r$ so that the reduced model $\mathbf{H}_r(s)$ to interpolate $\mathbf{H}(s)$ as before.

Coprime factorization systems

We now consider MIMO systems with the following structure in the Laplace transform domain:

$$\mathcal{K}(s)\,\widehat{\mathbf{v}}(s) = \mathcal{B}(s)\widehat{\mathbf{u}}(s) \quad \text{then} \quad \widehat{\mathbf{y}}(s) = \mathcal{C}(s)\,\widehat{\mathbf{v}}(s) + \mathbf{D}\,\widehat{\mathbf{u}}(s)$$

$$\text{with the transfer function} \quad \mathbf{H}(s) = \mathcal{C}(s)\mathcal{K}(s)^{-1}\mathcal{B}(s) + \mathbf{D},$$

where $\mathbf{D} \in \mathbb{R}^{p\times m}$, both $\mathcal{C}(s) \in \mathbb{C}^{p\times n}$ and $\mathcal{B}(s) \in \mathbb{C}^{n\times m}$ are analytic in the right half plane; and $\mathcal{K}(s) \in \mathbb{C}^{n\times n}$ is analytic and full rank throughout the right half plane. Hence, the goal is to find a reduced transfer function using a Petrov-Galerkin projection:

$$\mathbf{H}_r(s) = \mathcal{C}_r(s)\mathcal{K}_r(s)^{-1}\mathcal{B}_r(s) + \mathbf{D}_r$$

where $\mathcal{C}_r(s) = \mathcal{C}(s)\mathbf{V}_r \in \mathbb{C}^{p\times r}$, $\mathcal{B}_r(s) = \mathbf{W}_r^T\mathcal{B}(s) \in \mathbb{C}^{r\times m}$,
$\mathcal{K}_r(s) = \mathbf{W}_r^T\mathcal{K}(s)\mathbf{V}_r \in \mathbb{C}^{n\times n}$, and $\mathbf{W}_r, \mathbf{V}_r \in \mathbb{C}^{n\times r}$.

For simplicity we only consider $\mathbf{D} = \mathbf{D}_r$, which is equivalent to $\mathbf{D} = \mathbf{D}_r = \mathbf{0}$.

Coprime factorization systems

Model reduction by generalized interpolation

Suppose that $\mathcal{B}(s)$, $\mathcal{C}(s)$, and $\mathcal{K}(s)$ are analytic at $\sigma \in \mathbb{C}$ and $\mu \in \mathbb{C}$. Also let $\mathcal{K}(\sigma)$, $\mathcal{K}(\mu)$, $\mathcal{K}_r(\sigma) = \mathbf{W}_r^T \mathcal{K}(\sigma) \mathbf{V}_r$, and $\mathcal{K}_r(\mu) = \mathbf{W}_r^T \mathcal{K}(\mu) \mathbf{V}_r$ have full rank. Let nonnegative integers M and N be given as well as vectors $\mathrm{b} \in \mathbb{R}^m$ and $\mathrm{c} \in \mathbb{R}^p$.

(a) If $\mathcal{D}_\sigma^i [\mathcal{K}(s)^{-1}\mathcal{B}(s)]\mathrm{b} \in \mathrm{Ran}(\mathbf{V}_r)$ for $i = 0, \ldots, N$
then $\mathbf{H}^{(\ell)}(\sigma)\mathrm{b} = \mathbf{H}_r^{(\ell)}(\sigma)\mathrm{b}$ for $\ell = 0, \ldots, N$.

(b) If $\left(\mathrm{c}^T \mathcal{D}_\mu^j [\mathcal{C}(s)\mathcal{K}(s)^{-1}]\right)^T \in \mathrm{Ran}(\mathbf{W}_r)$ for $j = 0, \ldots, M$
then $\mathrm{c}^T \mathbf{H}^{(\ell)}(\mu) = \mathrm{c}^T \mathbf{H}_r^{(\ell)}(\mu)$ for $\ell = 0, \ldots, M$.

(c) If both (a) and (b) hold and if $\sigma = \mu$,
then $\mathrm{c}^T \mathbf{H}^{(\ell)}(\sigma)\mathrm{b} = \mathrm{c}^T \mathbf{H}_r^{(\ell)}(\sigma)\mathrm{b}$ for $\ell = 0, \ldots, M + N + 1$.

Coprime factorization systems

For $\mathbf{H}(s) = \mathcal{C}(s)\mathcal{K}(s)^{-1}\mathcal{B}(s)$, let r interpolation points $\{\sigma_i\}_{i=1}^r$, the left-directions $\{\mathrm{c}_i\}_{i=1}^r$ and the right-directions $\{\mathrm{b}_i\}_{i=1}^r$ be given. Construct,

$$\begin{aligned}\mathbf{V}_r &= \left[\mathcal{K}(\sigma_1)^{-1}\mathcal{B}(\sigma_1)\mathrm{b}_1, \ldots, \mathcal{K}(\sigma_r)^{-1}\mathcal{B}(\sigma_r)\mathrm{b}_r\right], \\ \mathbf{W}_r &= \left[\mathcal{K}(\sigma_1)^{-T}\mathcal{C}(\sigma_1)^T\mathrm{c}_1, \ldots, \mathcal{K}(\sigma_r)^{-T}\mathcal{C}(\sigma_r)^T\mathrm{c}_r\right]^T.\end{aligned}$$

Then, $\mathbf{H}_r(s) = \mathcal{C}_r(s)\mathcal{K}_r(s)^{-1}\mathcal{B}_r(s)$ satisfies the bi-tangential interpolation conditions, i.e. $\mathbf{H}(\sigma_i)\mathrm{b}_i = \mathbf{H}_r(\sigma_i)\mathrm{b}_i$, $\mathrm{c}_i^T\mathbf{H}(\sigma_i) = \mathrm{c}_i^T\mathbf{H}_r(\sigma_i)$ and $\mathrm{c}_i^T\mathbf{H}'(\sigma_i)\mathrm{b}_i = \mathrm{c}_i^T\mathbf{H}_r'(\sigma_i)\mathrm{b}_i$ for $i = 1, \ldots, r$.

This result extends interpolatory model reduction to a more general framework of $\mathbf{H}(s) = \mathcal{C}(s)\mathcal{K}(s)^{-1}\mathcal{B}(s) + \mathbf{D}$.

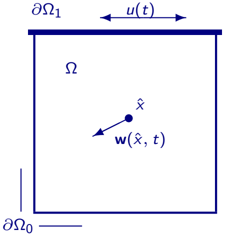
A numerical example: driven cavity flow in two dimensions I

Consider a square domain $\Omega = [0,1]^2$ (below) representing a volume filled with a viscoelastic material with boundary separated into a top edge ("lid"), $\partial\Omega_1$, and the complement, $\partial\Omega_0$ (bottom, left, and right edges). The material in the cavity is excited through shearing forces on the material caused by transverse displacement of the lid, $u(t)$. We are interested in the displacement response of the material, $\mathbf{w}(\hat{x}, t)$, at the center of Ω, i.e. $\hat{x} = (0.5, 0.5)$.

$$\partial_{tt}\mathbf{w}(x,t) - \eta_0\,\Delta\mathbf{w}(x,t) - \eta_1\partial_t \int_0^t \frac{\Delta\mathbf{w}(x,\tau)}{(t-\tau)^\alpha}\,d\tau \;+\; \nabla\varpi(x,t) = 0 \;\text{ for } x \in \Omega$$

$$\nabla\cdot\mathbf{w}(x,t) = 0 \text{ for } x \in \Omega \quad \text{and} \quad \begin{array}{l}\mathbf{w}(x,t) = 0 \text{ for } x \in \partial\Omega_0 \\ \mathbf{w}(x,t) = u(t) \text{ for } x \in \partial\Omega_1\end{array}$$

which determines the output $\mathbf{y}(t) = \mathbf{w}(\hat{x}, t)$,



We take $\eta_0 = 1.05 \times 10^6$, $\eta_1 = 2.44 \times 10^5$, and $\alpha = 0.519$, corresponding to experimentally derived values for the polymer butyl B252. The transfer function of the discretized system is given by $\mathbf{H}(s) = \mathcal{C}(s)\mathcal{K}(s)^{-1}\mathcal{B}(s)$ where

$$\mathcal{K}(s) = \begin{bmatrix} s^2\mathbf{M} + \widehat{\rho}(s)\mathbf{K} & \mathbf{D} \\ \mathbf{D}^T & \mathbf{0} \end{bmatrix}, \quad \mathcal{C}(s) = [\mathbf{C} \quad \mathbf{0}], \text{ and } \mathcal{B}(s) = \begin{bmatrix} s^2\mathbf{m} + \widehat{\rho(s)}\mathbf{k} \\ \mathbf{0} \end{bmatrix}.$$

A numerical example: driven cavity flow in two dimensions II

C corresponds to measuring the horizontal and vertical displacement at $\hat{x} = (0.5, 0.5)$, **m** and **k** are the sum of the columns of the free-free mass and stiffness matrix associated with x-displacement degrees of freedom on the top lid boundary and $\widehat{\rho}(s) = \eta_0 + \eta_1 s^\alpha$; hence producing a frequency dependent input-state map due to the system input being a boundary *displacement* as opposed to a boundary *force*. Note the non-linear frequency dependency in the state-input map; however our procedure can be still applied to obtain an interpolatory reduced-order model of the same form as shown next: Given are the interpolation point $\sigma \in \mathbb{C}$ and direction $\mathrm{c} \in \mathbb{C}^p$. Since the input is a scaler, there is no need for a right tangential direction. For simplicity we would only force bitangential Hermite interpolating conditions, i.e.

$$\mathcal{H}_r(\sigma) = \mathcal{H}(\sigma),\ \mathrm{c}^T\mathcal{H}_r(\sigma) = \mathrm{c}^T\mathcal{H}(\sigma),\ \text{and } \mathrm{c}^T\mathcal{H}_r'(\sigma) = \mathrm{c}^T\mathcal{H}'(\sigma).$$

Following the theorem, we solve the following two linear systems of equations:

$$\begin{bmatrix} \mathbf{F}(\sigma) & \mathbf{D} \\ \mathbf{D}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{z}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{N}(\sigma) \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{F}(\sigma) & \mathbf{D} \\ \mathbf{D}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_2 \\ \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}^T\mathrm{c} \\ \mathbf{0} \end{bmatrix}.$$

where $\mathbf{F}(\sigma) = \sigma^2\mathbf{M} + \widehat{\rho}(\sigma)\mathbf{K}$ and $\mathbf{N}(\sigma) = s^2\mathbf{m} + \widehat{\rho}(\sigma)\mathbf{k}$. Define the matrices $\mathbf{U}_r = [\mathbf{u}_1,\ \mathbf{u}_2]$ and $\mathbf{Z}_r = [\mathbf{z}_1,\ \mathbf{z}_2]$. Then the reduced system matrices are $\mathbf{V}_r = \mathbf{W}_r = \mathbf{U}_r \oplus \mathbf{Z}_r$ and the bitangential interpolation conditions are satisfied.
Below, we compare three different models including the full-order model:

1. $\mathbf{H}_{\rm fine}$, using a fine mesh FEM discretization with 51,842 displacement degrees of freedom and 6,651 pressure degrees of freedom (mesh size $h = \frac{1}{80}$);
2. $\mathbf{H}_{\rm coarse}$, for a coarse mesh discretization with 29,282 displacement degrees of freedom and 3721 pressure degrees of freedom;
3. $\mathbf{H}_{30}$, a generalized interpolatory reduced order model as defined above with $r = 30$, corresponding to 30 reduced displacement degrees of freedom and 30 reduced pressure degrees of freedom satisfying the bitangential interpolation conditions.

A numerical example: driven cavity flow in two dimensions III

The resulting frequency response plots shown in Figure 1:

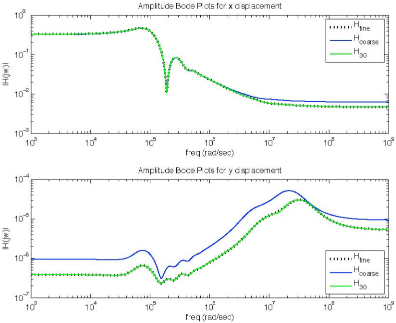


Figure: Bode Plots of $\mathbf{H}_{\text{fine}}$, $\mathbf{H}_{\text{coarse}}$ and reduced models $\mathbf{H}_{20}$ and $\mathbf{H}_{30}$

The reduced model is also a descriptor system with the same damping structure. $\mathbf{H}_{30}$ clearly outperforms $\mathbf{H}_{\text{coarse}}$.

Second-order Dynamical Systems

One important variation on the dynamical systems above arises in n degree–of–freedom mechanical (or other) structures:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{G}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = \mathbf{B}\mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t),$$

with the transfer function $\mathbf{H}(s) = \mathbf{C}(s^2\mathbf{M} + s\mathbf{G} + \mathbf{K})^{-1}\mathbf{B}$, where $\mathbf{M},\ \mathbf{G}$, and $\mathbf{K} \in \mathbb{R}^{n\times n}$ are positive (semi)-definite symmetric matrices describing, mass distribution, energy dissipation, and stiffness distribution.
The goal is to generate, for some $r \ll n$, an r^{th} order reduced second-order system of the form

$$\mathbf{M}_r\ddot{\mathbf{x}}_r(t) + \mathbf{G}_r\dot{\mathbf{x}}_r(t) + \mathbf{K}_r\mathbf{x}_r(t) = \mathbf{B}_r\dot{\mathbf{u}}(t), \qquad \mathbf{y}_r(t) = \mathbf{C}_r\mathbf{x}_r(t),$$

where $\mathbf{M}_r, \mathbf{G}_r, \mathbf{K}_r \in \mathbb{R}^{r\times r}$, are positive (semi)-definite symmetric matrices, $\mathbf{B} \in \mathbb{R}^{r\times m}$, and $\mathbf{C} \in \mathbb{R}^{p\times r}$. In order to preserve the symmetry and positive definiteness of $\mathbf{M}$, $\mathbf{G}$ and $\mathbf{K}$ in the course of model reduction, one-sided reduction is applied, i.e. one usually takes $\mathbf{W}_r = \mathbf{V}_r$; hence resulting in

$$\mathbf{M}_r = \mathbf{V}_r^T\mathbf{M}\mathbf{V}_r,\ \mathbf{G}_r = \mathbf{V}_r^T\mathbf{G}\mathbf{V}_r,\ \mathbf{K}_r = \mathbf{V}_r^T\mathbf{K}\mathbf{V}_r, \mathbf{B}_r = \mathbf{V}_r^T\mathbf{B},,\ \mathbf{C}_r = \mathbf{C}\mathbf{V}_r.$$

Algorithm 4.1. Second-order Tangential MIMO Order Reduction

Given interpolation points $\{\sigma_1, \cdots, \sigma_N\}$, directions $\{\mathrm{b}_1, \cdots, \mathrm{b}_N\}$, and interpolation orders $\{J_1, \cdots, J_N\}$ (with $r = \sum_{i=1}^{N} J_i$).

1. For $i = 1, \ldots, N$
   1. For each shift σ_i and tangent direction b_i, define $\mathcal{K}_0^{(i)} = \sigma_i^2 \mathbf{M} + \sigma_i \mathbf{G} + \mathbf{K}$, $\mathcal{K}_1^{(i)} = 2\sigma_i \mathbf{M} + \mathbf{G}$, and $\mathcal{K}_2 = \mathbf{M}$.
   2. Solve $\mathcal{K}_0^{(i)} \mathbf{f}_{i1} = \mathbf{B}\mathrm{b}_i$ and set $\mathbf{f}_{i0} = 0$.
   3. For $j = 2 : J_i$, solve $\mathcal{K}_0^{(i)} \mathbf{f}_{ij} = -\mathcal{K}_1^{(i)} \mathbf{f}_{i,j-1} - \mathcal{K}_2 \mathbf{f}_{i,j-2}$,
2. Take $\mathbf{V}_r = \left[\mathbf{f}_{11}, \mathbf{f}_{12}, \ldots, \mathbf{f}_{1J_1}, \mathbf{f}_{21}, \ldots, \mathbf{f}_{2J_2}, \ldots, \mathbf{f}_{N1}, \ldots, \mathbf{f}_{NJ_N}\right]$ and then $\mathbf{M}_r = \mathbf{V}_r^T \mathbf{M} \mathbf{V}_r$, $\mathbf{G}_r = \mathbf{V}_r^T \mathbf{G} \mathbf{V}_r$, $\mathbf{K}_r = \mathbf{V}_r^T \mathbf{K} \mathbf{V}_r$, $\mathbf{B}_r = \mathbf{V}_r^T \mathbf{B}$, and $\mathbf{C}_r = \mathbf{C}\mathbf{V}_r$.

Remark. The discussion above and the algorithm can be easily generalized to higher order constant coefficient ordinary differential equations as well where the system dynamics follow

$$\mathbf{A}_0 \frac{d^\ell \mathbf{x}}{dt^\ell} + \mathbf{A}_1 \frac{d^{\ell-1} \mathbf{x}}{dt^{\ell-1}} + \cdots + \mathbf{A}_\ell \mathbf{x}(t) = \mathbf{B}\mathbf{u}(t) \quad \text{and} \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t).$$

Model Reduction of Parametric Systems

In parameterized model reduction, the goal is to construct a high fidelity parametric reduced order models which recover the response of the original full order parametric sytem throughout the range of variation of interest of the parameters.

We consider a multi-input/multi-output linear dynamical system that is parameterized with q parameters $\mathrm{p} = [\mathrm{p}_1, \ldots, \mathrm{p}_q]$:

$$\mathbf{H}(s, \mathrm{p}) = \mathcal{C}(s, \mathrm{p})\mathcal{K}(s, \mathrm{p})^{-1}\mathrm{b}(s, \mathrm{p}),$$

with $\mathcal{K}(s, \mathrm{p}) \in \mathbb{C}^{n\times n}$ and $\mathcal{B}(s, \mathrm{p}) \in \mathbb{C}^{n\times m}$ and $\mathcal{C}(s, \mathrm{p}) \in \mathbb{C}^{p\times n}$. We assume that

$$\begin{aligned}
\mathcal{K}(s, \mathrm{p}) &= \mathcal{K}^{[0]}(s) + a_1(\mathrm{p})\,\mathcal{K}^{[1]}(s) + \ldots + a_\nu(\mathrm{p})\,\mathcal{K}^{[\nu]}(s)\\
\mathcal{B}(s, \mathrm{p}) &= \mathcal{B}^{[0]}(s) + b_1(\mathrm{p})\,\mathcal{B}^{[1]}(s) + \ldots + b_\nu(\mathrm{p})\,\mathcal{B}^{[\nu]}(s),\\
\mathcal{C}(s, \mathrm{p}) &= \mathcal{C}^{[0]}(s) + c_1(\mathrm{p})\,\mathcal{C}^{[1]}(s) + \ldots + c_\nu(\mathrm{p})\,\mathcal{C}^{[\nu]}(s).
\end{aligned}$$

where $a_1(\mathrm{p}), a_2(\mathrm{p}) \ldots, b_1(\mathrm{p}), \ldots, c_\nu(\mathrm{p})$ are scalar-valued parameter functions that could be linear or non-linear.

Our goal is to to generate, for some $r \ll n$, a reduced-order system with dimension r having the same parametric structure. Suppose matrices $\mathbf{V}_r \in \mathbb{C}^{n\times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n\times r}$ are specified and consider:

$$\mathbf{H}_r(s, \mathrm{p}) = \mathcal{C}_r(s, \mathrm{p})\mathcal{K}_r(s, \mathrm{p})^{-1}\mathcal{B}_r(s, \mathrm{p}),$$

with $\mathcal{K}_r(s, \mathrm{p}) = \mathbf{W}_r^T\mathcal{K}(s, \mathrm{p})\mathbf{V}_r$, $\mathcal{B}_r(s, \mathrm{p}) = \mathbf{W}_r^T\mathcal{B}(s, \mathrm{p})$, and $\mathcal{C}_r(s, \mathrm{p}) = \mathcal{C}(s, \mathrm{p})\mathbf{V}_r$.

We say that the reduced model has *the same parametric structure* if

$$\mathcal{K}_r(s,\mathsf{p}) = \left(\mathbf{W}_r^T\mathcal{K}^{[0]}(s)\mathbf{V}_r\right) + a_1(\mathsf{p})\left(\mathbf{W}_r^T\mathcal{K}^{[1]}(s)\mathbf{V}_r\right) + \ldots + a_\nu(\mathsf{p})\left(\mathbf{W}_r^T\mathcal{K}^{[\nu]}(s)\mathbf{V}_r\right)$$

$$\mathcal{B}_r(s,\mathsf{p}) = \left(\mathbf{W}_r^T\mathcal{B}^{[0]}(s)\right) + b_1(\mathsf{p})\left(\mathbf{W}_r^T\mathcal{B}^{[1]}(s)\right) + \ldots + b_\nu(\mathsf{p})\left(\mathbf{W}_r^T\mathcal{B}^{[\nu]}(s)\right),$$

$$\mathcal{C}_r(s,\mathsf{p}) = \left(\mathcal{C}^{[0]}(s)\mathbf{V}_r\right) + c_1(\mathsf{p})\left(\mathcal{C}^{[1]}(s)\mathbf{V}_r\right) + \ldots + c_\nu(\mathsf{p})\left(\mathcal{C}^{[\nu]}(s)\mathbf{V}_r\right).$$

with the *same* parameter functions $a_1(\mathsf{p}), \ldots, c_\nu(\mathsf{p})$, but with smaller coefficient matrices. The next result extends the previous results to the parameterized dynamical system setting:

Interpolatory parametrized model reduction

Suppose $\mathcal{K}(s,\mathsf{p})$, $\mathcal{B}(s,\mathsf{p})$, and $\mathcal{C}(s,\mathsf{p})$ are analytic with respect to s at $\sigma \in \mathbb{C}$ and $\mu \in \mathbb{C}$, and are continuously differentiable with respect to p in a neighborhood of $\hat{\mathsf{p}} = [\hat{\mathsf{p}}_1, \ldots, \hat{\mathsf{p}}_q]$. Suppose further that both $\mathcal{K}(\sigma,\hat{\mathsf{p}})$ and $\mathcal{K}(\mu,\hat{\mathsf{p}})$ are nonsingular and matrices $\mathbf{V}_r \in \mathbb{C}^{n\times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n\times r}$ are given such that both $\mathcal{K}_r(\sigma,\hat{\mathsf{p}}) = \mathbf{W}_r^T\mathcal{K}(\sigma,\hat{\mathsf{p}})\mathbf{V}_r$ and $\mathcal{K}_r(\mu,\hat{\mathsf{p}}) = \mathbf{W}_r^T\mathcal{K}(\mu,\hat{\mathsf{p}})\mathbf{V}_r$ are also nonsingular. For nontrivial tangential directions $\mathsf{b} \in \mathbb{C}^m$ and $\mathsf{c} \in \mathbb{C}^p$:

- **(a)** If $\mathcal{K}(\sigma,\hat{\mathsf{p}}))^{-1}\mathcal{B}(\sigma,\hat{\mathsf{p}}))\mathsf{b} \in \mathrm{Ran}(\mathbf{V}_r)$ then $\mathbf{H}(\sigma,\hat{\mathsf{p}}))\mathsf{b} = \mathbf{H}_r(\sigma,\hat{\mathsf{p}}))\mathsf{b}$
- **(b)** If $\left(\mathsf{c}^T\mathcal{C}(\mu,\hat{\mathsf{p}}))\mathcal{K}(\mu,\hat{\mathsf{p}}))^{-1}\right)^T \in \mathrm{Ran}(\mathbf{W}_r)$ then $\mathsf{c}^T\mathbf{H}(\mu,\hat{\mathsf{p}})) = \mathsf{c}^T\mathbf{H}_r(\mu,\hat{\mathsf{p}}))$
- **(c)** If both (a) and (b) hold and if $\sigma = \mu$, then

$$\nabla_{\mathsf{p}}\mathsf{c}^T\mathbf{H}(\sigma,\hat{\mathsf{p}})\mathsf{b} = \nabla_{\mathsf{p}}\mathsf{c}^T\mathbf{H}_r(\sigma,\hat{\mathsf{p}})\mathsf{b} \quad \text{and} \quad \mathsf{c}^T\mathbf{H}'(\sigma,\hat{\mathsf{p}})\mathsf{b} = \mathsf{c}^T\mathbf{H}_r'(\sigma,\hat{\mathsf{p}})\mathsf{b}$$

Outline

Model Reduction from Measurements

Consider a set of scalar points:

$$(s_i, \phi_i),\ i = 1, 2, \cdots, N, \quad s_i \neq s_j,\ i \neq j.$$

We seek a rational function $\mathbf{H}(s) = \frac{\mathbf{n}(s)}{\mathbf{d}(s)}$, such that $\mathbf{H}(s_i) = \phi_i$, $i = 1, 2, \cdots, N$, and $\mathbf{n}, \mathbf{d}$ are coprime polynomials.
A solution always exists, e.g. the *Lagrange interpolating polynomial*:

$$\phi_0(s) = \sum_{j=1}^{N} \frac{\prod_{i\neq j}(s - \phi_i)}{\prod_{i\neq j}(s_j - \phi_i)}.$$

Then all solutions can be expressed as

$$\phi(s) = \phi_0(s) + \rho(s) \prod_{i=1}^{N}(s - s_i), \quad \rho(s_i) \text{ is finite.}$$

Additional constraints for $\mathbf{H}$ may include *minimality*, *stability*, *bounded realness*.

A rational Lagrange-type formula

The data is now divided in disjoint sets: (σ_i, y_i), $i = 1, 2, \cdots, r$, (μ_j, z_j), $j = 1, 2, \cdots, q$, $k + q = N$. Consider:

$$\sum_{i=1}^{r} \gamma_i \frac{\phi(s) - y_i}{s - \sigma_i} = 0.$$

Then as long as $\gamma_i \neq 0$, there holds $\phi(\sigma_i) = y_i$, for $i = 1, \cdots, q$. If we make use of the freedom in satisfying the remaining interpolation conditions, we obtain the condition

The Loewner matrix

$$\mathbb{L}\mathbf{c} = 0 \quad \text{where} \quad \mathbb{L} = \underbrace{\begin{bmatrix} \frac{z_1 - y_1}{\mu_1 - \sigma_1} & \cdots & \frac{z_1 - y_r}{\mu_1 - \sigma_r} \\ \vdots & \ddots & \vdots \\ \frac{z_q - y_1}{\mu_q - \sigma_1} & \cdots & \frac{z_q - y_k}{\mu_q - \sigma_r} \end{bmatrix}}_{\textbf{Loewner matrix}} \in \mathbb{C}^{q \times r}, \quad \mathbf{c} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_r \end{bmatrix} \in \mathbb{C}^r.$$

The Loewner matrix

The main result affirms that the rank of $\mathbb{L}$ encodes the minimal degree of interpolants. Furthermore if an interpolant is provided in internal form: $\mathbf{H}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$, the Loewner matrix can be factorized as:

$$\mathbb{L} = -\underbrace{\begin{bmatrix} \mathbf{C}(\mu_1\mathbf{E} - \mathbf{A})^{-1} \\ \mathbf{C}(\mu_2\mathbf{E} - \mathbf{A})^{-1} \\ \vdots \\ \mathbf{C}(\mu_q\mathbf{E} - \mathbf{A})^{-1} \end{bmatrix}}_{\mathbf{W}_q^T} \underbrace{\begin{bmatrix} (\sigma_1\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} & \cdots & (\sigma_r\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} \end{bmatrix}}_{\mathbf{V}_r},$$

where $\mathbf{W}_q^T$, $\mathbf{V}_r$ were introduced earlier and represent generalized observability and reachability matrices.

This property parallels the corresponding property of the Hankel-matrix framework and thus indicate that the Loewner matrix is indeed the right tool. This is further enforced by the following fact. The problem of interpolation with multiplicities can also be solved by means of the Loewner matrix. In particular, if the value and a number of derivatives at a single point $(s_0; \phi_0, \phi_1, \cdots, \phi_{N-1})$, is provided, it turns out that the associated Loewner matrix has *Hankel structure*:

$$\mathbb{L} = \begin{bmatrix} \frac{\phi_1}{1!} & \frac{\phi_2}{2!} & \frac{\phi_3}{3!} & \frac{\phi_4}{4!} & \cdots \\ \frac{\phi_2}{2!} & \frac{\phi_3}{3!} & \frac{\phi_4}{4!} & \cdots & \\ \frac{\phi_3}{3!} & \frac{\phi_4}{4!} & & & \\ \frac{\phi_4}{4!} & \vdots & & \ddots & \\ \vdots & & & & \end{bmatrix}.$$

Thus the *Loewner matrix generalizes* the *Hankel matrix* when general interpolation replaces realization.

Remark. In this framework strict properness of interpolants is not required. Thus rational functions with polynomial part can be recovered from input-output data.

Model Reduction from Measurements

In many instances, input/output measurements replace an explicit model of a to-be-simulated system. In such cases it is of great interest to be able to efficiently construct models and reduced models from the available data.
An important tool is the S- or *scattering-parameter* system representation. The S-parameters represent a system as a black box. An advantage is that these parameters can be measured using VNAs (Vector Network Analyzers).

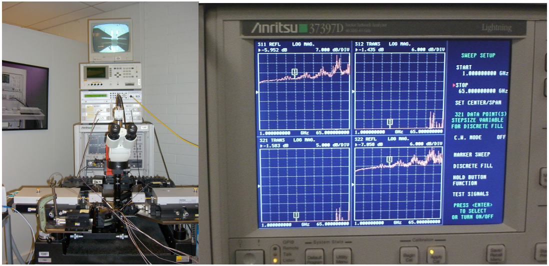


Figure: VNA (Vector Network Analyzer) and VNA screen showing the magnitude of the S-parameters for a 2 port.

Given a system in input/output representation: $\hat{\mathbf{y}}(s) = \mathbf{H}(s)\hat{\mathbf{u}}(s)$, the associated **S-parameter representation** is $\hat{\mathbf{y}}_s = \mathbf{S}(s)\mathbf{u}(s)$, where $\mathbf{S}(s) = [\mathbf{H}(s) + \mathbf{I}][\mathbf{H}(s) - \mathbf{I}]^{-1}$. Thereby $\hat{\mathbf{y}}_s = \frac{1}{2}(\hat{\mathbf{y}} + \hat{\mathbf{u}})$ are the *transmitted waves* and, $\hat{\mathbf{u}}_s = \frac{1}{2}(\hat{\mathbf{y}} - \hat{\mathbf{u}})$ are the *reflected waves*. Thus the *S-parameter measurements* $\mathbf{S}(j\omega_k)$, are samples of the frequency response of the S-parameter system representation.

The Loewner matrix pair and the construction of interpolants

Suppose we have observed response data as described in Problem 2. We are given r (right) driving frequencies: $\{\sigma_i\}_{i=1}^r \subset \mathbb{C}$ that use input directions $\{\tilde{b}_i\}_{i=1}^r \subset \mathbb{C}^m$, to produce system responses, $\{\tilde{y}_i\}_{i=1}^r \subset \mathbb{C}^p$ and q (left) driving frequencies: $\{\mu_i\}_{i=1}^q \subset \mathbb{C}$ that use dual (left) input directions $\{\tilde{c}_i\}_{i=1}^q \subset \mathbb{C}^p$, to produce dual (left) responses, $\{\tilde{z}_i\}_{i=1}^q \subset \mathbb{C}^m$. We assume that there is an underlying dynamical system as defined in (1) so that

$$\tilde{c}_i^T \mathbf{H}(\mu_i) = \tilde{z}_i^T \quad \text{for } i = 1, \cdots, q, \qquad \text{and} \qquad \mathbf{H}(\sigma_j)\tilde{b}_j = \tilde{y}_j, \quad \text{for } j = 1, \cdots, r.$$

yet we are given access only to the $r + q$ response observations listed above and have no other information about the underlying system $\mathbf{H}(s)$. Here we will sketch the solution of Problem 2. Towards this goal we will introduce the **Loewner matrix** pair in the **tangential interpolation** case:

$$\mathbb{L} = \begin{bmatrix} \dfrac{\tilde{z}_1^T\tilde{b}_1 - \tilde{c}_1^T\tilde{y}_1}{\mu_1 - \sigma_1} & \cdots & \dfrac{\tilde{z}_1^T\tilde{b}_r - \tilde{c}_1^T\tilde{y}_r}{\mu_1 - \sigma_r} \\ \vdots & \ddots & \vdots \\ \dfrac{\tilde{z}_q^T\tilde{b}_1 - \tilde{c}_q^T\tilde{y}_1}{\mu_q - \sigma_1} & \cdots & \dfrac{\tilde{z}_q^T\tilde{b}_r - \tilde{c}_q^T\tilde{y}_r}{\mu_q - \sigma_r} \end{bmatrix} \in \mathbb{C}^{q\times r}.$$

The Loewner matrix pair and the construction of interpolants

If we define matrices associated with the system observations as:

$$\widetilde{\mathsf{B}} = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \tilde{\mathsf{b}}_1 & \tilde{\mathsf{b}}_2 & \dots & \tilde{\mathsf{b}}_r \\ \vdots & \vdots & & \vdots \end{bmatrix} \quad \widetilde{\mathsf{Y}} = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \tilde{\mathsf{y}}_1 & \tilde{\mathsf{y}}_2 & \dots & \tilde{\mathsf{y}}_r \\ \vdots & \vdots & & \vdots \end{bmatrix}$$

$$\widetilde{\mathsf{Z}}^T = \begin{bmatrix} \dots & \tilde{\mathsf{z}}_1^T & \dots \\ \dots & \tilde{\mathsf{z}}_2^T & \dots \\ & \vdots & \\ \dots & \tilde{\mathsf{z}}_q^T & \dots \end{bmatrix} \quad \widetilde{\mathsf{C}}^T = \begin{bmatrix} \dots & \tilde{\mathsf{c}}_1^T & \dots \\ \dots & \tilde{\mathsf{c}}_2^T & \dots \\ & \vdots & \\ \dots & \tilde{\mathsf{c}}_q^T & \dots \end{bmatrix}$$

$\mathbb{L}$ satisfies the Sylvester equation

$$\mathbb{L}\boldsymbol{\Sigma} - M\mathbb{L} = \widetilde{\mathsf{Z}}^T\widetilde{\mathsf{B}} - \widetilde{\mathsf{C}}^T\widetilde{\mathsf{Y}},$$

where $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1,\ \sigma_2,\ \dots,\ \sigma_r) \in \mathbb{C}^{r\times r}$ and $M = \mathrm{diag}(\mu_1,\ \mu_2,\ \dots,\ \mu_q) \in \mathbb{C}^{q\times q}$ Suppose that state space data $(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, of minimal degree n are given such that $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E}-\mathbf{A})^{-1}\mathbf{B}+\mathbf{D}$. If the generalized eigenvalues of $(\mathbf{A}, \mathbf{E})$ are distinct from σ_i and μ_j, we define $\mathbf{V}_r$, so that its j^{th} column is: $(\sigma_j\mathbf{E}-\mathbf{A})^{-1}\mathbf{B}\tilde{\mathsf{b}}_j$, and $\mathbf{W}_q^T$ so that its i^{th} row is $\tilde{\mathsf{c}}_i^T\mathbf{C}(\mu_i\mathbf{E}-\mathbf{A})^{-1}$. It follows that

$$\mathbb{L} = -\mathbf{W}_q^T\mathbf{E}\mathbf{V}_r,$$

and we call $\mathbf{V}_r$, $\mathbf{W}_q^T$ **generalized tangential** reachability, observability matrices, respectively.

The Loewner matrix pair and the construction of interpolants

Next we introduce a new object which is pivotal in our approach. This is the *shifted Loewner matrix*, defined as follows:

$$\mathbb{M} = \begin{bmatrix} \frac{\mu_1 \tilde{z}_1^T \tilde{b}_1 - \sigma_1 \tilde{c}_1^T \tilde{y}_1}{\mu_1 - \sigma_1} & \cdots & \frac{\mu_1 \tilde{z}_1^T \tilde{b}_r - \sigma_r \tilde{c}_1^T \tilde{y}_r}{\mu_1 - \sigma_r} \\ \vdots & \ddots & \vdots \\ \frac{\mu_q \tilde{z}_q^T \tilde{b}_1 - \sigma_1 \tilde{c}_q^T \tilde{y}_1}{\mu_q - \sigma_1} & \cdots & \frac{\mu_q \tilde{z}_q^T \tilde{b}_r - \sigma_r \tilde{c}_q^T \tilde{y}_r}{\mu_q - \sigma_r} \end{bmatrix} \in \mathbb{C}^{q \times r}$$

$\mathbb{M}$ satisfies the Sylvester equation

$$\mathbb{M}\Sigma - M\mathbb{M} = M\widetilde{Z}^T\widetilde{B} - \widetilde{C}^T\widetilde{Y}\Sigma.$$

If an interpolant $\mathbf{H}(s)$ is associated with the interpolation data, the shifted Loewner matrix is the Loewner matrix associated to $s\mathbf{H}(s)$. If a state space representation is available, then like for the Loewner matrix, the shifted Loewner matrix can be factored as

$$\mathbb{M} = -\mathbf{W}_q^T \mathbf{A} \mathbf{V}_r.$$

It therefore becomes apparent that $\mathbb{L}$ contains information about $\mathbf{E}$ while $\mathbb{M}$ contains information about $\mathbf{A}$. These observations are formalized in one of the main results of this section which shows how straightforward the solution of the interpolation problem becomes, in the Loewner matrix framework.

Main result

Construction of interpolants in state space form

Assume that $r = q$ and that $\mu_i \neq \sigma_j$ for all $i, j = 1, \ldots, r$. Suppose that $\mathbb{M} - s\,\mathbb{L}$ is invertible for all $s \in \{\sigma_i\} \cup \{\mu_j\}$. Then, with

$$\mathbf{E}_r = -\mathbb{L}, \quad \mathbf{A}_r = -\mathbb{M}, \quad \mathbf{B}_r = \widetilde{\mathrm{Z}}^T, \quad \mathbf{C}_r = \widetilde{\mathrm{Y}}, \quad \mathbf{D}_r = 0,$$

$$\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r = \widetilde{\mathrm{Y}}(\mathbb{M} - s\,\mathbb{L})^{-1}\widetilde{\mathrm{Z}}^T$$

interpolates the data and furthermore is a minimal realization.

Next we will outline the proof of this important result which is straightforward and hence reveals the main attributes of this approach.

Proof 5.1 *Multiplying equation (**??**) by s and subtracting it from (**??**) we get*

$$(\mathbb{M}-s\mathbb{L})\boldsymbol{\Sigma}-M(\mathbb{M}-s\mathbb{L})=\widetilde{\mathrm{C}}^T\widetilde{\mathrm{Y}}(\boldsymbol{\Sigma}-s\mathbf{I})-(M-s\mathbf{I})\widetilde{\mathrm{Z}}^T\widetilde{\mathrm{B}}.$$

Multiplying this equation by $\mathbf{e}_i$ (the i^{th} unit vector) on the right and setting $s=\sigma_i$, we obtain

$$(\sigma_i\mathbf{I}-M)(\mathbb{M}-\sigma_i\mathbb{L})\mathbf{e}_i=(\sigma_i\mathbf{I}-M)\widetilde{\mathrm{Z}}^T\widetilde{\mathrm{b}}_i \quad\Rightarrow$$
$$(\sigma_i\mathbb{L}-\mathbb{M})\mathbf{e}_i=\widetilde{\mathrm{Z}}^T\widetilde{\mathrm{b}}_i \quad\Rightarrow\quad \widetilde{\mathrm{Y}}\mathbf{e}_i=\widetilde{\mathrm{Y}}(\sigma_i\mathbb{L}-\mathbb{M})^{-1}\widetilde{\mathrm{Z}}^T\mathrm{b}_i$$

Therefore $\mathbf{H}(\sigma_i)\tilde{\mathrm{b}}_i=\tilde{\mathrm{y}}_i$. This proves the right tangential interpolation property. To prove the left tangential interpolation property, we multiply the above equation by $\mathbf{e}_j^*$ (the transpose of the j^{th} unit vector) on the left and set $s=\mu_j$:

$$\mathbf{e}_j^*(\mathbb{M}-\mu_j\mathbb{L})(\boldsymbol{\Sigma}-\mu_j\mathbf{I})=\mathbf{e}_j^*\widetilde{\mathrm{C}}^T\widetilde{\mathrm{Y}}(\mu_j\mathbf{I}-\boldsymbol{\Sigma}) \quad\Rightarrow$$
$$\mathbf{e}_j^*(\mathbb{M}-\mu_j\mathbb{L})=\widetilde{\mathrm{c}}_j^T\widetilde{\mathrm{Y}} \quad\Rightarrow\quad \mathbf{e}_j^*\widetilde{\mathrm{Z}}^T=\widetilde{\mathrm{c}}_j^T\widetilde{\mathrm{Y}}(\mathbb{M}-\mu_j\,\mathbb{L})^{-1}\widetilde{\mathrm{Z}}^T.$$

Therefore $\tilde{\mathrm{c}}_j^T\mathbf{H}(\mu_j)=\tilde{\mathrm{z}}_j^T$, which completes the proof. ∎

The general case

If the assumption of the above theorem is not satisfied, one needs to project onto the column span and onto the row span of a linear combination of the two Loewner matrices. More precisely, let the following assumption be satisfied:

$$\mathrm{rank}\,(s\mathbb{L} - \mathbb{M}) = \mathrm{rank}\,\begin{pmatrix} \mathbb{L} & \mathbb{M} \end{pmatrix} = \mathrm{rank}\begin{pmatrix} \mathbb{L} \\ \mathbb{M} \end{pmatrix} = \rho, \text{ for all } s \in \{\sigma_i\} \cup \{\mu_j\}.$$

$\rho \leq q, r$ and can be either the *exact* or the *numerical* rank. The best tool for determining the rank of $s\mathbb{L} - \mathbb{M}$, is the SVD. Let

$$s\mathbb{L} - \mathbb{M} = \mathbf{Y}\Sigma\mathbf{X}^*,$$

for some choice of $s \in \{\sigma_i\} \cup \{\mu_j\}$ and consider a truncated SVD as $\mathbf{Y}_\rho \in \mathbb{C}^{q\times\rho}$, $\mathbf{X}_\rho \in \mathbb{C}^{r\times\rho}$.

Approximate tangetial interpolation

A realization $[\mathbf{E}_\rho, \mathbf{A}_\rho, \mathbf{B}_\rho, \mathbf{C}_\rho]$, of a minimal solution is given as follows:

$$\mathbf{E}_\rho = -\mathbf{Y}_\rho^*\mathbb{L}\mathbf{X}_\rho, \quad \mathbf{A}_\rho = -\mathbf{Y}_\rho^*\mathbb{M}\mathbf{X}_\rho, \quad \mathbf{B}_\rho = \mathbf{Y}_\rho^*\widetilde{\mathbb{Y}}, \quad \mathbf{C}_\rho = \widetilde{\mathbb{Z}}^T\mathbf{X}_\rho.$$

Depending on whether ρ is the exact or approximate rank, we obtain either an interpolant or an approximate interpolant of the data, respectively.

Loewner and Pick matrices

The positive real interpolation problem can be formulated as follows. Given triples $(\sigma_i, \tilde{b}_i, \tilde{y}_i)$, $i = 1, \cdots, q$, where σ_i are distinct complex numbers in the right-half of the complex plane, $\tilde{b}_i$ and $\tilde{y}_i$ are in $\mathbb{C}^r$, we seek a rational matrix function $\mathbf{H}(s)$ of size $r \times r$, such that $\mathbf{H}(\sigma_i)\tilde{b}_i = \tilde{y}_i$, $i = 1, \cdots, q$, and in addition $\mathbf{H}$ is positive real. This problem does not always have a solution. It is well known that the necessary and sufficient condition for its solution is that the associated *Pick* matrix

$$\mathbf{\Pi} = \begin{bmatrix} \frac{\tilde{y}_1^*\tilde{b}_1 + \tilde{b}_1^*\tilde{y}_1}{\overline{\sigma_1} + \sigma_1} & \cdots & \frac{\tilde{y}_1^*\tilde{b}_q + \tilde{b}_1^*\tilde{y}_q}{\overline{\sigma_1} + \sigma_q} \\ \vdots & \ddots & \vdots \\ \frac{\tilde{y}_q^*\tilde{b}_1 + \tilde{b}_q^*\tilde{y}_1}{\overline{\sigma_q} + \sigma_1} & \cdots & \frac{\tilde{y}_q^*\tilde{b}_q + \tilde{b}_q^*\tilde{y}_q}{\overline{\sigma_q} + \sigma_q} \end{bmatrix} \in \mathbb{C}^{q \times q},$$

be positive semi-definite, that is $\mathbf{\Pi} = \mathbf{\Pi}^* \geq \mathbf{0}$.

By comparing $\mathbf{\Pi}$ with the Loewner matrix $\mathbb{L}$, we conclude that if the right (column) array for the former is taken as $(\sigma_i, \tilde{b}_i, \tilde{y}_i)$, $i = 1, \cdots, q$, and the left (row) array as $(-\overline{\sigma_i}, \tilde{b}_i^*, -\tilde{y}_i^*)$, $i = 1, \cdots, q$, then

$$\mathbf{\Pi} = \mathbb{L}.$$

The left is then called the *mirror-image array*. Thus for this choice of interpolation data the Pick matrix is the same as the Loewner matrix. This shows the importance of the Loewner matrix as a tool for studying rational interpolation.

Remark. **(a)** The above considerations provide an algebraization of the positive real interpolation problem. If namely, $\mathbf{\Pi} \geq \mathbf{0}$, the minimal-degree rational functions which interpolate *simultaneously* the original array *and* its mirror image array, are automatically *positive real* and hence *stable* as well. The data in the model reduction problem, *automatically* satisfy this positive definiteness constraint, and therefore the reduced system is positive real.

(b) It readily follows that interpolants of the original and the mirror-image arrays constructed by means of the Loewner matrix, satisfy

$$[\mathbf{H}(\sigma_i) + \mathbf{H}^*(-\sigma_i)]\,\tilde{\mathrm{b}}_i = \mathbf{0}.$$

In general the zeros σ_i of $\mathbf{H}(s) + \mathbf{H}^*(-s)$ are called *spectral zeros*, and $\tilde{\mathrm{b}}_i$ are the corresponding (right) zero directions. Thus the construction of positive real interpolants by means of the Loewner (Pick) matrix, forces these interpolants to have the *given* interpolation points as *spectral zeros*.

Examples I

A simple low-order example. First we will illustrate the above results by means of a simple example. Consider a 2×2 rational function with minimal realization:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; \mathbf{E} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; \mathbf{B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \; \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Thus the transfer function is

$$\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} = \begin{bmatrix} s & 1 \\ 1 & \frac{1}{s} \end{bmatrix}.$$

Since $\mathrm{rank}\,\mathbf{E} = 2$, the McMillan degree of $\mathbf{H}$ is 2. Our goal is to recover this function through interpolation. The data will be chosen in two different ways.
First, we will choose *matrix data*, that is the values of the whole matrix are available at each interpolation point:

$$\sigma_1 = 1, \quad \sigma_2 = 1, \quad \sigma_3 = 2, \quad \sigma_4 = 2,$$

$$\tilde{b}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \tilde{b}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \tilde{b}_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \tilde{b}_4 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\tilde{y}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \tilde{y}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \tilde{y}_3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \tilde{y}_4 = \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}$$

$$\mu_1 = 1, \quad \mu_2 = 1, \quad \mu_3 = 2, \quad \mu_4 = 2,$$

$$\tilde{c}_1^T = \begin{pmatrix} 1, & 0 \end{pmatrix}, \quad \tilde{c}_2^T = \begin{pmatrix} 0, & 1 \end{pmatrix}, \quad \tilde{c}_3^T = \begin{pmatrix} 1, & 0 \end{pmatrix}, \quad \tilde{c}_4^T = \begin{pmatrix} 0, & 1 \end{pmatrix}$$

$$\tilde{z}_1^T = \begin{pmatrix} -1, & 1 \end{pmatrix}, \quad \tilde{z}_2^T = \begin{pmatrix} 1, & -1 \end{pmatrix}, \quad \tilde{z}_3^T = \begin{pmatrix} -2, & 1 \end{pmatrix}, \quad \tilde{z}_4^T = \begin{pmatrix} 1, & -\frac{1}{2} \end{pmatrix}$$

Examples II

The associated (block) Loewner and shifted Loewner matrices turn out to be:

$$\mathbb{L} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & \frac{1}{2} \\ 1 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{4} \end{pmatrix}, \quad \mathbb{M} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

Notice that the rank of both Loewner matrices is 2 while the rank of $x_i\mathbb{L} - \mathbb{M}$ is 3, for all x equal to a σ_i or μ_i. It can be readily verified that the column span of $\sigma_1\mathbb{L} - \mathbb{M} = \mathbb{L} - \mathbb{M}$ is the same as that of Π, where

$$\Pi = \begin{pmatrix} 1 & 1 & -2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Furthermore the row span of $\mathbb{L} - \mathbb{M}$ is the same as that of Π^* Thus

$$\hat{\mathbf{A}} = -\Pi^*\mathbb{M}\Pi = \begin{pmatrix} -2 & -3 & 1 \\ -1 & 0 & -4 \\ 1 & 0 & 4 \end{pmatrix}, \quad \hat{\mathbf{E}} = -\Pi^*\mathbb{L}\Pi = \begin{pmatrix} -2 & -2 & \frac{3}{2} \\ -2 & -4 & 4 \\ \frac{3}{2} & 4 & -\frac{17}{4} \end{pmatrix},$$

$$\hat{\mathbf{B}} = \Pi^*\widetilde{Z}^T = \begin{pmatrix} 0 & 0 \\ -3 & 2 \\ 3 & -\frac{5}{2} \end{pmatrix}, \quad \hat{\mathbf{C}} = \widetilde{Y}\Pi = \begin{pmatrix} 2 & 3 & -1 \\ 2 & 2 & -\frac{3}{2} \end{pmatrix},$$

Examples III

satisfy $\mathbf{H}(s) = \hat{\mathbf{C}}(s\hat{\mathbf{E}} - \hat{\mathbf{A}})\hat{\mathbf{B}}$, which shows that a (second) minimal realization of $\mathbf{H}$ has been obtained.

The *second* experiment involves tangential data, that is, at each interpolation point only values along certain directions are available.

$$\sigma_1 = 1, \qquad \sigma_2 = 2, \qquad \sigma_3 = 3,$$

$$\tilde{b}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \tilde{b}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \tilde{b}_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\tilde{y}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \tilde{y}_2 = \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}, \quad \tilde{y}_3 = \begin{pmatrix} 4 \\ \frac{4}{3} \end{pmatrix}$$

$$\mu_1 = -1, \qquad \mu_2 = -2, \qquad \mu_3 = -3,$$

$$\tilde{c}_1^T = \begin{pmatrix} 1, & 0 \end{pmatrix}, \quad \tilde{c}_2^T = \begin{pmatrix} 0, & 1 \end{pmatrix}, \quad \tilde{c}_3^T = \begin{pmatrix} 1, & 1 \end{pmatrix}$$

$$\tilde{z}_1^T = \begin{pmatrix} -1, & 1 \end{pmatrix}, \quad \tilde{z}_2^T = \begin{pmatrix} 1, & -\frac{1}{2} \end{pmatrix}, \quad \tilde{z}_3^T = \begin{pmatrix} -2, & \frac{2}{3} \end{pmatrix}.$$

Thus the associated Loewner and shifted Loewner matrices are:

$$\mathbb{L} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{1}{6} \\ 1 & \frac{1}{6} & \frac{10}{9} \end{bmatrix}, \quad \mathbb{M} = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

It readily follows that the conditions of theorem **??** are satified and hence the quadruple $(-\mathbb{M}, -\mathbb{L}, \widetilde{Z}^T, \widetilde{Y})$, provides a (third) minimal realization of the original rational function: $\mathbf{H}(s) = -\widetilde{Y}\,[s\mathbb{L} - \mathbb{M}]^{-1}\,\widetilde{Z}^T$.

Coupled mechanical system

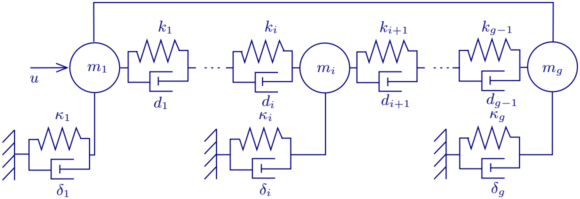


Figure: Constrained mechanical system

There are g masses in total; the i^{th} mass of weight m_i, is connected to the $(i+1)^{\text{st}}$ mass by a spring and a damper with constants k_i and d_i, respectively, and also to the ground by a spring and a damper with constants κ_i and δ_i, respectively. Additionally, the first mass is connected to the last one by a rigid bar (holonomic constraint) and it is influenced by the control $\mathbf{u}(t)$.

The vibration of this constrained system is described in generalized state space form as:

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

where $\mathbf{x}$ contains the positions and velocities of the masses,

$$\mathbf{E} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \; \mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{K} & \mathbf{D} & -\mathbf{G}^* \\ \mathbf{G} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \; \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_2 \\ \mathbf{0} \end{bmatrix}, \; \mathbf{C} = [\mathbf{C}_1, \;\; \mathbf{C}_2, \;\; \mathbf{C}_3];$$

furthermore $\mathbf{M}$ is the mass matrix ($g \times g$, diagonal, positive definite), $\mathbf{K}$ is the stiffness matrix ($g \times g$, tri-diagonal), $\mathbf{D}$ is the damping matrix ($g \times g$, tri-diagonal), $\mathbf{G} = [1, \;\; 0, \;\; \cdots \;, \; 0, \;\; -1]$, is the $1 \times g$ constraint matrix.

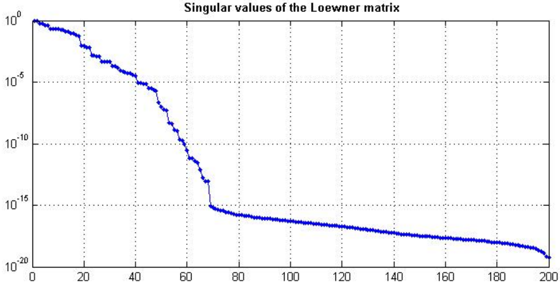


Figure: The singular values of the Loewner matrix

In this case balanced truncation methods for descriptor systems can be used to reduce this system. Here we will reduce this system by means of the Loewner framework.

Towards this goal, we compute 200 frequency response data, that is $\mathbf{H}(i\omega_i)$, where $\omega_i \in [-2, +2]$. Figure 4 shows the singular values of the Loewner matrix pair, which indicate that a system of order 20 will have an approximate error 10^{-3} (-60dB).

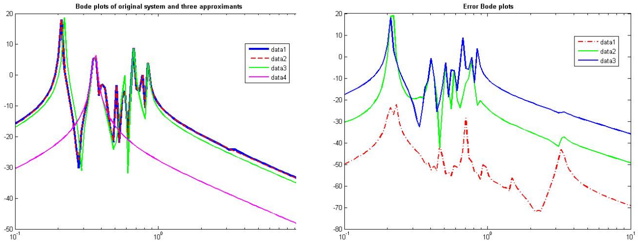


Figure: Left pane: Frequency responses of original system and approximants (orders 2, 10, 18). Right pane: Frequecy responses of error systems (orders 2,10,18)

This figure shows that (for the chosen values of the parameters) the frequency response has about 7 peaks. A second order approximant reproduces (approximately) the highest peak, a tenth order system reproduces (approximately) five peaks, while a system of order 18 provides a good approximation of the whole frequency response.

Four-pole band-pass filter I

In this case 1000 frequency response measurements are given, of the 2×2 S-parameters of a semi-conductor device which is meant to be a band-pass filter. There is no *a priori* model available. The range of frequencies is between 40 and 120 GHz; We will use the Loenwer matrix procedure applied to the S-parameters. This yields $\mathbb{L},\ \mathbb{M} \in \mathbb{C}^{2000 \times 2000}$.

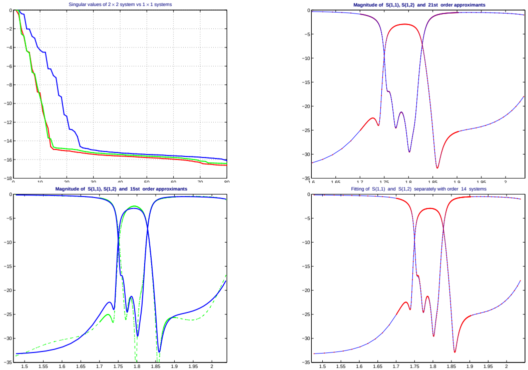


Figure: Upper row, left pane: The singular values of $x\mathbb{L} - \mathbb{M}$, for the 2-port and for two one-ports. Upper row, right pane: The $S(1,1)$ and $S(1,2)$ parameter data for a 17-th order model. Lower row: left pane: Fitting $S(1,1)$, $S(1,2)$ jointly with a 15th order approximant. Lower row, right pane: Fitting $S(1,1)$, $S(1,2)$ separately with 14th order approximants.

Four-pole band-pass filter II

In the upper left-hand plot of figure 6, the singular values of the Loewner matrix corresponing to the 2-port system (upper curve) is compared with the singular values of two one-port subsystems (lower curves). As the decay of all curves is fast, an approximant of order around 20, is expected to provide a good fit. Indeed, as the upper right-hand plot shows, a 21^{st} order approximant provides fits with error less than $-60dB$. For comparison the fit of a 15^{th} order model is shown in the lower left-hand plot. Sometimes in practical applications, the entries of the 2-port S-parameters are modeled separately. In our case 14^{th} order models are sufficient, but the McMillan degree os the 2-port is 28 or higher (depending on the symmetries involved, e.g. $S_{11} = S_{22}$, $S_{12} = S_{21}$).

Outline

References

- A.C. Antoulas, C.A. Beattie, and S. Gugercin, *Interpolatory model reduction od large-scale systems*, Report, September 2009.
- S. Lefteriu and A.C. Antoulas, *A new approach in modeling multip-port systems from frequency response data*, IEEE Trans. CAD (in press), 2009.
- R. Ionutiu, J. Rommes, and A.C. Antoulas, *Passivity preserving model reduction using spectral zero interpolation*, IEEE Trans. CAD, vol. 27, 2008.
- A. Bunse-Gerstner, D. Kubalinska, and D. Wilczek, $\mathcal{H}_2$ *optimal model reduction for large-scale discrete dynamical systems*, J. Computational and Applied Math, 2009.
- P. van Dooren, K. Gallivan, and P.-A. Absil, $\mathcal{H}_2$ *optimal model reduction of MIMO systems*, Applied Math. Letters, vol. 21, 2008.
- A.J. Mayo and A.C. Antoulas, *A framework for the solution of the generalized realization problem*, Linear Algebra and Its Applications, vol. 425, 2007.
- C.A. Beattie and S. Gugercin, *Interpolatory projection methods for structure preserving model reduction*, Systems and Control Letters, vol. 58, 2009.