

Model Reduction by Rational Interpolation

Christopher Beattie and Serkan Gugercin

Abstract. The last two decades have seen major developments in interpolatory methods for model reduction of large-scale linear dynamical systems. Advances of note include the ability to produce (locally) optimal reduced models at modest cost; refined methods for deriving interpolatory reduced models directly from input/output measurements; and extensions for the reduction of parametrized systems. This chapter offers a survey of interpolatory model reduction methods starting from basic principles and ranging up through recent developments that include weighted model reduction and structure-preserving methods based on generalized coprime representations. Our discussion is supported by an assortment of numerical examples.

Mathematics Subject Classification (2010). 41A05, 93A15, 93C05, 37M99.

Keywords. Rational interpolation, model reduction, \mathcal{H}_2 approximation, parametric systems, generalized coprime factorization, weighted model reduction, Loewner framework

SUBMITTED FOR PUBLICATION IN FEBRUARY 2014 .

1. Introduction

Numerous techniques exist for model reduction of large-scale dynamical systems among them, Proper Orthogonal Decomposition (see Chapter ??), Balanced Truncation (see Chapter ??), and Interpolatory Methods, to be discussed both here and in Chapter ?. Interpolatory model reduction methods include methods referred to as Rational Krylov Methods, but should be viewed as distinct for reasons we describe later. Over the past two decades, major progress has been made in interpolation-based model reduction approaches and as a result, these methods have emerged as one of the leading choices for reducing large scale dynamical systems.

This chapter gives a survey of projection-based interpolatory methods for model reduction. Section 2 introduces the model reduction problem setting that we consider. In Section 3, we give general projection results for

This work was supported in part by the National Science Foundation under contract DMS-1217156.

interpolatory model reduction followed by a discussion of \mathcal{H}_2 optimal model reduction by interpolation which is presented in Section 4. Up through Section 4, we assume that the original system to be reduced is in a standard first-order state-space form. Beginning in Section 5, we discuss how interpolatory methods can be extended with ease to much more general settings that include systems with delays and systems with polynomial structure. For the most part, we assume that the internal dynamics of the full order system are specified, accessible, and manifested either in a known state-space generalized coprime representation. This assumption will be relaxed in Section 6 (and also in Chapter ??), where a data-driven framework for interpolatory methods is introduced, useful for situations with no direct access to internal dynamics and where only input/output measurements are available. Finally in Section 7, we show how to use interpolatory methods for reducing parametric dynamical systems.

The methods we discuss in this chapter have been applied with great success to very large-scale dynamical systems. Motivated by brevity, we do not present such examples here, preferring instead to illustrate the ideas with simple and approachable (albeit more academic) examples that may better reveal details of the process. For those with a hunger for more, we refer to the original papers where large-scale case studies are presented.

2. Model Reduction via Projection

In this section, we introduce the basic concepts of projection-based model reduction. We also discuss the main error measures with which the approximation error will be quantified.

2.1. The problem setting

We consider linear dynamical systems represented in state-space form as:

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

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. In (2.1), $\mathbf{x}(t) \in \mathbb{R}^n$ is the *internal variable*, or the *state variable* if \mathbf{E} is non-singular. The length, n , of \mathbf{x} is called the *dimension* of the underlying dynamical system. $\mathbf{u}(t) \in \mathbb{R}^m$ and $\mathbf{y}(t) \in \mathbb{R}^p$ are, respectively, the *inputs* and *outputs* of the system. Dynamical systems with $m = p = 1$ will be called SISO systems (*single-input (and) single-output*) while all other cases will be grouped together and referred to as MIMO systems (*multi-input (or) multi-output*).

For cases where the dimension n is large, e.g., $n \geq 10^5, 10^6$, the simulation and control of the system can lead to a huge computational burden; especially when the system must be resimulated over and over again, say, using different input selections, $\mathbf{u}(t)$. Our goal is to replace (2.1) with a simpler

reduced model having the form

$$\begin{aligned} \mathbf{E}_r \dot{\mathbf{x}}_r(t) &= \mathbf{A}_r \mathbf{x}_r(t) + \mathbf{B}_r \mathbf{u}(t) \\ \mathbf{y}_r(t) &= \mathbf{C}_r \mathbf{x}_r(t) + \mathbf{D}_r \mathbf{u}(t) \end{aligned} \quad \text{with } \mathbf{x}_r(0) = \mathbf{0}, \quad (2.2)$$

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 $\mathbf{D}_r \in \mathbb{R}^{p \times m}$ with $r \ll n$, and such that over a wide range of system inputs, the corresponding outputs of the reduced system, $\mathbf{y}_r(t)$, will be good approximations to the corresponding true outputs, $\mathbf{y}(t)$, with respect to an appropriate, physically relevant norm.

2.2. Transfer Function and Error Measures

For the linear dynamical systems considered here, the frequency domain representation is a powerful tool to quantify the model reduction error. Indeed, in this setting, error analysis in the frequency domain relates directly to error analysis in the time domain.

Let $\hat{\mathbf{y}}(s)$, $\hat{\mathbf{y}}_r(s)$, and $\hat{\mathbf{u}}(s)$ denote the Laplace transforms of $\mathbf{y}(t)$, $\mathbf{y}_r(t)$ and $\mathbf{u}(t)$, respectively. Taking the Laplace transforms of (2.1) and (2.2) yields

$$\hat{\mathbf{y}}(s) = \left(\mathbf{C} (s\mathbf{E} - \mathbf{A})^{-1} \mathbf{B} + \mathbf{D} \right) \hat{\mathbf{u}}(s), \quad (2.3)$$

$$\hat{\mathbf{y}}_r(s) = \left(\mathbf{C}_r (s\mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r + \mathbf{D}_r \right) \hat{\mathbf{u}}(s). \quad (2.4)$$

The mapping from $\hat{\mathbf{u}}(s)$ to $\hat{\mathbf{y}}(s)$ is called the *transfer function*. Likewise, the mapping from $\hat{\mathbf{u}}(s)$ to $\hat{\mathbf{y}}_r(s)$ is the *transfer function of the reduced model*. We denote them by $\mathbf{H}(s)$ and $\mathbf{H}_r(s)$, respectively:

$$\mathbf{H}(s) = \mathbf{C} (s\mathbf{E} - \mathbf{A})^{-1} \mathbf{B} + \mathbf{D}, \quad \text{and} \quad (2.5)$$

$$\mathbf{H}_r(s) = \mathbf{C}_r (s\mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r + \mathbf{D}_r. \quad (2.6)$$

$\mathbf{H}(s)$ is a $p \times m$ matrix-valued degree- n rational function in s , and $\mathbf{H}_r(s)$ is a $p \times m$ matrix-valued degree- r rational function in s . Thus, the model reduction problem could be viewed as a rational approximation problem in the complex domain. This perspective of model reduction is emphasized in Section 3.

2.3. Petrov-Galerkin Projections

Most model reduction methods can be formulated with the aid of either Petrov-Galerkin or Galerkin projections. Even though the original internal variable, $\mathbf{x}(t)$, evolves in a (large) n -dimensional space, it is often the case that it hews rather closely to some (typically unknown) r dimensional subspace. Let $\mathbf{V} \in \mathbb{R}^{n \times r}$ be a basis for this subspace, which is as yet undetermined. The original state may be approximated as $\mathbf{x}(t) \approx \mathbf{V}\mathbf{x}_r(t)$ for some $\mathbf{x}_r(t) \in \mathbb{R}^r$ and this expression may be used to represent the reduced model dynamics. Plug the approximation, $\mathbf{x}(t) \approx \mathbf{V}\mathbf{x}_r(t)$, into (2.1) to obtain a residual

$$\mathbf{R}(\mathbf{x}_r(t)) = \mathbf{E}\mathbf{V}\dot{\mathbf{x}}_r(t) - \mathbf{A}\mathbf{V}\mathbf{x}_r(t) - \mathbf{B}\mathbf{u}(t) \quad (2.7)$$

and the approximate output

$$\mathbf{y}_r(t) = \mathbf{C}\mathbf{V}\mathbf{x}_r(t) + \mathbf{D}\mathbf{u}(t). \quad (2.8)$$

The reduced state trajectory, $\mathbf{x}_r(t)$, is determined by enforcing a Petrov-Galerkin orthogonality condition on the residual $\mathbf{R}(t)$ in (2.7): we pick another r dimensional subspace with a basis $\mathbf{W} \in \mathbb{R}^{n \times r}$ and impose the Petrov-Galerkin condition:

$$\mathbf{W}^T \mathbf{R}(t) = \mathbf{W}^T (\widetilde{\mathbf{E}} \mathbf{V} \dot{\mathbf{x}}_r(t) - \mathbf{A} \mathbf{V} \mathbf{x}_r(t) - \mathbf{B} \mathbf{u}(t)) = \mathbf{0}.$$

This leads to a reduced model as in (2.2),

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

with reduced model quantities defined as

$$\mathbf{E}_r = \mathbf{W}^T \mathbf{E} \mathbf{V}, \quad \mathbf{A}_r = \mathbf{W}^T \mathbf{A} \mathbf{V}, \quad \mathbf{B}_r = \mathbf{W}^T \mathbf{B}, \quad \mathbf{C}_r = \mathbf{C} \mathbf{V}, \quad \text{and} \quad \mathbf{D}_r = \mathbf{D}. \quad (2.9)$$

A critical observation to make here is that the reduced model does not depend on the specific basis selection made for \mathbf{V} and \mathbf{W} , only on the subspaces themselves. To see this, let $\widetilde{\mathbf{V}} = \mathbf{V} \mathbf{T}_1$ and $\widetilde{\mathbf{W}} = \mathbf{W} \mathbf{T}_2$ where $\mathbf{T}_1 \in \mathbb{R}^{r \times r}$ and $\mathbf{T}_2 \in \mathbb{R}^{r \times r}$ are nonsingular matrices corresponding to a change-of-basis. This leads to a change in reduced model quantities

$$\widetilde{\mathbf{E}}_r = \mathbf{T}_2^T \mathbf{E}_r \mathbf{T}_1, \quad \widetilde{\mathbf{A}}_r = \mathbf{T}_2^T \mathbf{A}_r \mathbf{T}_1, \quad \widetilde{\mathbf{B}}_r = \mathbf{T}_2^T \mathbf{B}_r, \quad \widetilde{\mathbf{C}}_r = \mathbf{C}_r \mathbf{T}_1, \quad \text{and} \quad \widetilde{\mathbf{D}}_r = \mathbf{D}_r,$$

where \mathbf{E}_r , \mathbf{A}_r , \mathbf{B}_r , \mathbf{C}_r , and \mathbf{D}_r are as defined in (2.9). A straightforward comparison of the transfer functions reveals

$$\begin{aligned} \widetilde{\mathbf{H}}_r(s) &= \widetilde{\mathbf{C}}_r \left(s \widetilde{\mathbf{E}}_r - \widetilde{\mathbf{A}}_r \right)^{-1} \widetilde{\mathbf{B}}_r + \widetilde{\mathbf{D}}_r \\ &= \mathbf{C}_r \mathbf{T}_1 \left(s \mathbf{T}_2^T \mathbf{E}_r \mathbf{T}_1 - \mathbf{T}_2^T \mathbf{A}_r \mathbf{T}_1 \right)^{-1} \mathbf{T}_2^T \mathbf{B}_r + \mathbf{D}_r \\ &= \mathbf{C}_r (s \mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r + \mathbf{D}_r = \mathbf{H}_r(s). \end{aligned}$$

2.4. Error Measures

As in any approximation problem, error measures are necessary to quantify the approximation error appropriately. For linear dynamical systems, error analysis is best presented in the frequency domain, yet directly relates to time-domain error, $\mathbf{y}(t) - \mathbf{y}_r(t)$. It follows from (2.3) and (2.4) that

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

Thus, the closeness of $\widehat{\mathbf{y}}_r(s)$ to $\widehat{\mathbf{y}}(s)$ is directly related to the closeness of $\mathbf{H}_r(s)$ to $\mathbf{H}(s)$. The \mathcal{H}_2 and \mathcal{H}_∞ norms are the most common measures of closeness for transfer functions:

2.4.1. The \mathcal{H}_∞ Norm. Let $\mathbf{H}(s)$ be the transfer function of a stable dynamical system. The \mathcal{H}_∞ norm is defined as

$$\|\mathbf{H}\|_{\mathcal{H}_\infty} = \sup_{\omega \in \mathbb{R}} \|\mathbf{H}(j\omega)\|_2, \quad (2.10)$$

where $\|\mathbf{M}\|_2$ denotes the spectral (Euclidean-induced) norm of the complex matrix \mathbf{M} . When \mathbf{E} is nonsingular, all eigenvalues of the matrix pencil $\lambda \mathbf{E} - \mathbf{A}$ must lie in the left-half plane. When \mathbf{E} is singular, we assume additionally that 0 is not a defective eigenvalue of \mathbf{E} . This guarantees that $\mathbf{H}(s)$ remains bounded as $s \rightarrow \infty$.

The importance of the \mathcal{H}_∞ -norm stems from it being the $(L_2$ - L_2)-induced operator norm of an underlying convolution operator mapping the system inputs,

\mathbf{u} , to system outputs, \mathbf{y} : $\|\mathbf{H}\|_{\mathcal{H}_\infty} = \sup_{\mathbf{u} \in L_2} \frac{\|\mathbf{y}\|_{L_2}}{\|\mathbf{u}\|_{L_2}}$. where $\|\mathbf{z}\|_{L_2} = \sqrt{\int_0^\infty \|\mathbf{z}(t)\|_2^2 dt}$. With respect to model reduction error, one directly obtains

$$\|\mathbf{y} - \mathbf{y}_r\|_{L_2} \leq \|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_\infty} \|\mathbf{u}\|_{L_2}.$$

If one wishes to produce reduced models that generate outputs, $\mathbf{y}_r(t)$, that are always close (with respect to the L_2 norm) to the corresponding true outputs, $\mathbf{y}(t)$, uniformly so over all L_2 -bounded inputs, $\mathbf{u}(t)$, then one should apply a model reduction technique that produces small \mathcal{H}_∞ error.

2.4.2. The \mathcal{H}_2 Norm. Let $\mathbf{H}(s)$ be the transfer function of a stable dynamical system. Then, the \mathcal{H}_2 norm is defined as

$$\|\mathbf{H}\|_{\mathcal{H}_2} := \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \|\mathbf{H}(i\omega)\|_{\mathbb{F}}^2 \right)^{1/2}, \quad (2.11)$$

where $\|\mathbf{M}\|_{\mathbb{F}}$ denotes the Frobenius norm of a complex matrix \mathbf{M} .

When \mathbf{E} is nonsingular, all eigenvalues of the matrix pencil $\lambda\mathbf{E} - \mathbf{A}$ must lie in the left-half plane. When \mathbf{E} is singular, we assume additionally that 0 is not a defective eigenvalue of \mathbf{E} .

When \mathbf{E} is nonsingular, we require that all the eigenvalues of the matrix pencil $\lambda\mathbf{E} - \mathbf{A}$ lie in the left-half plane and $\mathbf{D} = \mathbf{0}$ for the \mathcal{H}_2 norm to be finite. When \mathbf{E} is singular, we assume in addition that 0 is not a defective eigenvalue of \mathbf{E} and that $\lim_{s \rightarrow \infty} \mathbf{H}(s) = \mathbf{0}$. The \mathcal{H}_2 norm bears a direct relationship to the time domain norm of $\mathbf{y}(t)$:

$$\|\mathbf{y}\|_{L_\infty} = \sup_{t > 0} \|\mathbf{y}(t)\|_\infty \leq \|\mathbf{H}\|_{\mathcal{H}_2} \|\mathbf{u}\|_{L_2},$$

and this bound is best possible for MISO systems ($p = 1$), SIMO systems ($m = 1$), and SISO systems ($m = p = 1$), reflecting the fact that the \mathcal{H}_2 norm is simply the (L_2 - L_∞)-induced norm of the underlying convolution operator in these cases. With respect to model reduction error, we have in general,

$$\|\mathbf{y} - \mathbf{y}_r\|_{L_\infty} \leq \|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2} \|\mathbf{u}\|_{L_2}.$$

So, if one wishes to produce reduced models that generate outputs, $\mathbf{y}_r(t)$, that are uniformly and instantaneously close to the corresponding true outputs, $\mathbf{y}(t)$, uniformly so over all L_2 -bounded inputs, $\mathbf{u}(t)$, then one should apply a model reduction technique that produces small \mathcal{H}_2 error. Note that the \mathcal{H}_2 error may be made small even for original systems \mathbf{H} with $\lim_{s \rightarrow \infty} \mathbf{H}(s) \neq \mathbf{0}$ with an appropriate choice of \mathbf{D}_r and \mathbf{E}_r ; $\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2}$ may be small even if $\|\mathbf{H}\|_{\mathcal{H}_2}$ is unboundedly large. This is discussed in more detail in Section 3.2.1.

3. Model Reduction by Interpolation

In this section, we present the fundamental projection theorems used in interpolatory model reduction and illustrate their use with some simple examples.

3.1. Tangential Interpolation Problem

One easily observes that the model reduction problem for linear, time invariant dynamical systems can be formulated as a rational approximation problem: Given a degree- n rational function, $\mathbf{H}(s)$ (the full model), find a degree- r rational function, $\mathbf{H}_r(s)$ (the reduced model), that approximates $\mathbf{H}(s)$ accurately with respect to either the \mathcal{H}_∞ or \mathcal{H}_2 norm. Interpolation is a commonly applied tool for function approximation; typically, effective polynomial interpolants are easy to calculate. Here, we will develop straightforward methods for obtaining rational interpolants, however, the precise notion of “interpolation” that will be used must be clarified. Since $\mathbf{H}(s)$ is a $p \times m$ matrix valued rational function, the immediate extension of point-wise interpolation to matrix-valued functions suggests that we attempt to enforce conditions such as $\mathbf{H}(s_0) = \mathbf{H}_r(s_0)$ at each interpolation point $s_0 \in \mathbb{C}$. But viewed element-wise, this would require in effect, $p \times m$ interpolation conditions at every interpolation point. For systems with even modestly large number of input and outputs dimensions m and p , this will lead to a large number of interpolation conditions requiring as a result, quite a large reduced order r . Thus, for MIMO systems, instead of this notion of “full matrix interpolation”, we require only that the interpolating matrix function match the original only along certain directions, “tangential interpolation.” We will show later that this relaxed notion of interpolation is adequate to characterize necessary conditions for optimal approximation in the \mathcal{H}_2 norm.

Tangential interpolation involves choosing interpolation directions in addition to interpolation points. We separate the interpolation points and directions into two categories: “left” and “right.” We say that $\mathbf{H}_r(s)$ is a right-tangential interpolant to $\mathbf{H}(s)$ at $s = \sigma_i$ along the right tangent direction $\mathbf{r}_i \in \mathbb{C}^m$ if

$$\mathbf{H}(\sigma_i)\mathbf{r}_i = \mathbf{H}_r(\sigma_i)\mathbf{r}_i.$$

Similarly, we say that $\mathbf{H}_r(s)$ is a left-tangential interpolant to $\mathbf{H}(s)$ at $s = \mu_i$ along the left tangent direction $\mathbf{l}_i \in \mathbb{C}^p$ if

$$\mathbf{l}_i^T \mathbf{H}(\mu_i) = \mathbf{l}_i^T \mathbf{H}_r(\mu_i).$$

Our model reduction task can now be formulated as tangential interpolation as follows: Given a set of r right interpolation points $\{\sigma_i\}_{i=1}^r \in \mathbb{C}$, r left interpolation points $\{\mu_i\}_{i=1}^r \in \mathbb{C}$, r right-tangential directions $\{\mathbf{r}_i\}_{i=1}^r \in \mathbb{C}^m$, and r left-tangential directions $\{\mathbf{l}_i\}_{i=1}^r \in \mathbb{C}^p$, find a degree- r reduced transfer function $\mathbf{H}_r(s)$ so that

$$\begin{aligned} \mathbf{H}(\sigma_i)\mathbf{r}_i &= \mathbf{H}_r(\sigma_i)\mathbf{r}_i, \\ \mathbf{l}_i^T \mathbf{H}(\mu_i) &= \mathbf{l}_i^T \mathbf{H}_r(\mu_i), \end{aligned} \quad \text{for } i = 1, \dots, r. \quad (3.1)$$

We say that $\mathbf{H}_r(s)$ is a bitangential Hermite interpolant to $\mathbf{H}(s)$ at $s = \sigma_i$ along the right tangent direction $\mathbf{r}_i \in \mathbb{C}^m$ and the left tangent direction $\mathbf{l}_i \in \mathbb{C}^p$, if

$$\mathbf{l}_i^T \mathbf{H}'(\sigma_i)\mathbf{r}_i = \mathbf{l}_i^T \mathbf{H}'_r(\sigma_i)\mathbf{r}_i.$$

where $'$ denotes differentiation with respect to s . Therefore, in addition to (3.1), we may also require $\mathbf{H}_r(s)$ to satisfy

$$\mathbf{l}_i^T \mathbf{H}'(\sigma_i)\mathbf{r}_i = \mathbf{l}_i^T \mathbf{H}'_r(\sigma_i)\mathbf{r}_i, \quad \text{for } i = 1, \dots, r. \quad (3.2)$$

In Section 4, we will show how to choose interpolation points and tangent directions to produce optimal approximation with respect to the \mathcal{H}_2 norm.

3.2. Petrov-Galerkin Projections for Tangential Interpolation

Our goal here is to appropriately pick the model reduction bases \mathbf{V} and \mathbf{W} so that the reduced model obtained by a Petrov-Galerkin projection as in (2.9) satisfies the tangential interpolation conditions (3.1) and (3.2). We first present the projection theorem followed by a historical perspective:

Theorem 3.1. *Given the transfer function, $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$, let $\mathbf{H}_r(s)$ denote a reduced transfer function obtained by projection as in (2.9) using the model reduction bases \mathbf{V} and \mathbf{W} . For interpolation points $\sigma, \mu \in \mathbb{C}$, suppose that $\sigma\mathbf{E} - \mathbf{A}$ and $\mu\mathbf{E} - \mathbf{A}$ be invertible. Let $\mathbf{r} \in \mathbb{C}^m$ and $\mathbf{l} \in \mathbb{C}^\ell$ be designated (nontrivial) tangential directions. Then,*

(a) if
$$(\sigma\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathbf{r} \in \text{Ran}(\mathbf{V}), \quad (3.3)$$

then

$$\mathbf{H}(\sigma)\mathbf{r} = \mathbf{H}_r(\sigma)\mathbf{r}; \quad (3.4)$$

(b) if

$$\left(\mathbf{l}^T\mathbf{C}(\mu\mathbf{E} - \mathbf{A})^{-1}\right)^T \in \text{Ran}(\mathbf{W}), \quad (3.5)$$

then

$$\mathbf{l}^T\mathbf{H}(\mu) = \mathbf{l}^T\mathbf{H}_r(\mu); \quad (3.6)$$

(c) if both (3.3) and (3.5) hold, and $\sigma = \mu$, then

$$\mathbf{l}^T\mathbf{H}'(\sigma)\mathbf{r} = \mathbf{l}^T\mathbf{H}'_r(\sigma)\mathbf{r} \quad (3.7)$$

as well.

Theorem 3.1 illustrates that imposing either a left or right tangential interpolation condition requires adding only one vector either to the left or right model reduction bases. For the case of repeated left and right interpolation points, the bitangential Hermite condition is satisfied for free, in the sense that no additional vectors must be included in the model reduction bases. Notice that the values that are interpolated are never explicitly computed; this is a significant advantage of the Petrov-Galerkin projection framework as used in interpolatory model reduction.

A projection framework for interpolatory model reduction was introduced by Skelton *et al.* in [38, 92, 93]. This approach was put into a robust numerical framework by Grimme [52], who employed the rational Krylov subspace method of Ruhe [80]. The tangential interpolation framework of Theorem 3.1 was developed by Gallivan *et al.* [51]. For SISO systems, the model reduction bases, \mathbf{V} and \mathbf{W} , produced from Theorem 3.1 become rational Krylov subspaces; and so interpolatory model reduction is sometimes referred to as *rational Krylov methods*. However, the connection to rational Krylov subspaces is lost for general MIMO systems, unless all tangent directions are the same. So, we prefer the simpler descriptive label, *interpolatory methods*. Indeed, for the more general systems that we consider in Section 5, the direct extensions we develop for interpolatory methods have no connection to rational Krylov subspaces even in the SISO case. Another term that has been in use when referring to interpolatory model reduction methods is *Moment Matching Methods*. The k th moment of $\mathbf{H}(s)$ around σ is the k th derivative of $\mathbf{H}(s)$ evaluated at $s = \sigma$. In the SISO case, the reduced transfer function obtained via rational interpolation will match those moments - in effect, generalized Hermite interpolation. The notion of moment matching for MIMO systems with respect to tangent directions is not so clearly stated, however. See [10, 43, 50, 12, 48, 9, 26] and the references therein for other related work on model reduction by interpolation.

A simple example. Consider the following linear dynamical system with $n = 3$, $m = p = 2$, $\mathbf{E} = \mathbf{I}_3$, $\mathbf{D} = \mathbf{0}$,

$$\mathbf{A} = \begin{bmatrix} -6 & -11 & -6 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}.$$

The transfer function of this dynamical system can be computed as:

$$\mathbf{H}(s) = \frac{1}{s^3 + 6s^2 + 11s + 6} \begin{bmatrix} 10 & s^2 - 10s + 1 \\ -s^2 - 5s + 6 & -18s - 6 \end{bmatrix}.$$

Let $\sigma_1 = \mu_1 = 0$ be the left and right interpolation points together with tangent directions

$$\mathbf{r}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{l}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

Using Theorem 3.1, we compute the interpolatory model reduction bases:

$$\mathbf{V} = (\sigma_1 \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{r}_1 = \begin{bmatrix} -2 \\ -1 \\ 4 \end{bmatrix} \quad \text{and}$$

$$\mathbf{W} = (\sigma_1 \mathbf{E} - \mathbf{A})^{-T} \mathbf{C}^T \mathbf{l}_1 = \begin{bmatrix} 0.5 \\ -1 \\ 6.5 \end{bmatrix}.$$

Then, the Petrov-Galerkin projection in (2.9) leads to the reduced model quantities

$$\mathbf{E}_r = 26, \quad \mathbf{A}_r = -5, \quad \mathbf{B}_r = \begin{bmatrix} 6 & -0.5 \end{bmatrix}, \quad \mathbf{C}_r = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad \mathbf{D}_r = \mathbf{0},$$

and consequently to the reduced model transfer function

$$\mathbf{H}_r(s) = \frac{1}{26s + 5} \begin{bmatrix} 12 & -1 \\ -6 & 0.5 \end{bmatrix}.$$

Now that we have $\mathbf{H}_r(s)$, we can check the interpolation conditions explicitly:

$$\begin{aligned} \mathbf{H}(\sigma_1) \mathbf{r}_1 &= \mathbf{H}_r(\sigma_1) \mathbf{r}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad \checkmark \\ \mathbf{l}_1^T \mathbf{H}(\sigma_1) &= \mathbf{l}_1^T \mathbf{H}_r(\sigma_1) = \begin{bmatrix} 6 & -0.5 \end{bmatrix}, \quad \checkmark \\ \mathbf{l}_1^T \mathbf{H}'(\sigma_1) \mathbf{r}_1 &= \mathbf{l}_1^T \mathbf{H}'_r(\sigma_1) \mathbf{r}_1 = -26. \quad \checkmark \end{aligned}$$

Notice that $\mathbf{H}_r(s)$ does not fully interpolate $\mathbf{H}(s)$ at $s = \sigma_1 = 0$:

$$\mathbf{H}(\sigma_1) = \begin{bmatrix} 5/3 & 1/6 \\ 1 & -1 \end{bmatrix} \neq \begin{bmatrix} 2.4 & -0.2 \\ -1.2 & 0.1 \end{bmatrix} = \mathbf{H}_r(\sigma_1).$$

To enforce full matrix interpolation, we need to modify the construction of the model reduction bases and remove the tangential vectors. Denote the new bases for full matrix interpolation by \mathbf{V}_m and \mathbf{W}_m :

$$\mathbf{V}_m = (\sigma_1 \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \\ 5/3 & 7/6 \end{bmatrix}$$

$$\mathbf{W}_m = (\sigma_1 \mathbf{E} - \mathbf{A})^{-T} \mathbf{C}^T = \begin{bmatrix} 1/6 & 0 \\ 0 & -1 \\ 11/6 & 1 \end{bmatrix}.$$

Note that even with a single interpolation point, in the case of full matrix interpolation, the reduction spaces has dimension 2 leading a degree-2 reduced model with transfer function

$$\mathbf{H}_{rm}(s) = \begin{bmatrix} 10 & 1 \\ 6 & -6 \end{bmatrix} \left(s \begin{bmatrix} 110 & 71 \\ 96 & 42 \end{bmatrix} - \begin{bmatrix} -60 & -6 \\ -36 & 36 \end{bmatrix} \right)^{-1} \begin{bmatrix} 10 & 1 \\ 6 & -6 \end{bmatrix}$$

This new reduced model fully interpolates $\mathbf{H}(s)$ and $\mathbf{H}'(s)$ at $s = \sigma_1 = 0$:

$$\begin{aligned} \mathbf{H}(\sigma_1) &= \mathbf{H}_{rm}(\sigma_1) = \begin{bmatrix} 5/3 & 1/6 \\ 1 & -1 \end{bmatrix} \\ \mathbf{H}'(\sigma_1) &= \mathbf{H}'_{rm}(\sigma_1) = \begin{bmatrix} -55/18 & -71/36 \\ -8/3 & -7/6 \end{bmatrix}. \end{aligned}$$

This simple example illustrates the fundamental difference between tangential and full interpolation. In the case of tangential interpolation, each interpolation condition only adds one degree of freedom to the reduced dimension; however in the case of full interpolation, a single (full matrix) interpolation condition will generically add m or p dimensions to the reduced model. We will see in §4 that optimality requires only tangential interpolation. \square

The computations in the previous example can be extended easily to the case of r interpolation points: Given $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$, r right interpolation points $\{\sigma_i\}_{i=1}^r$ and right directions $\{\mathbf{r}_k\}_{k=1}^r \in \mathbb{C}^m$, r left interpolation points $\{\mu_j\}_{j=1}^r$ and left directions $\{\mathbf{l}_k\}_{k=1}^r \in \mathbb{C}^p$, construct

$$\mathbf{V} = [(\sigma_1\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathbf{r}_1, \dots, (\sigma_r\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathbf{r}_r] \text{ and} \quad (3.8)$$

$$\mathbf{W} = [(\mu_1\mathbf{E} - \mathbf{A})^{-T}\mathbf{C}^T\mathbf{l}_1, \dots, (\mu_r\mathbf{E} - \mathbf{A})^{-T}\mathbf{C}^T\mathbf{l}_r]. \quad (3.9)$$

Then, $\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r$ constructed by a Petrov-Galerkin projection as in (2.9) satisfies the Lagrange tangential interpolation conditions (3.1) and the bitangential Hermite interpolation conditions (3.2), if in addition $\sigma_i = \mu_i$ (provided that $\sigma_i\mathbf{E}_r - \mathbf{A}_r$ and $\mu_i\mathbf{E}_r - \mathbf{A}_r$ are nonsingular for each $i = 1, \dots, r$).

Theorem 3.1 can be extended readily to include higher-order Hermite interpolation:

Theorem 3.2. *Given a full order model with transfer function*

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

let $\mathbf{H}_r(s)$ denote a reduced transfer function obtained by projection as in (2.9), using model reduction bases, \mathbf{V} and \mathbf{W} . Let $\mathbf{H}^{(k)}(\sigma)$ denote the k th derivative of $\mathbf{H}(s)$ with respect to s evaluated at $s = \sigma$. For interpolation points $\sigma, \mu \in \mathbb{C}$, suppose $\sigma\mathbf{E} - \mathbf{A}$ and $\mu\mathbf{E} - \mathbf{A}$ are invertible, and that $\mathbf{r} \in \mathbb{C}^m$ and $\mathbf{l} \in \mathbb{C}^p$ are given (nontrivial) tangent directions. Then,

(a) if

$$((\sigma\mathbf{E} - \mathbf{A})^{-1}\mathbf{E})^{j-1}(\sigma\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}\mathbf{r} \in \text{Ran}(\mathbf{V}), \text{ for } j = 1, \dots, N \quad (3.10)$$

then

$$\mathbf{H}^{(k)}(\sigma)\mathbf{r} = \mathbf{H}_r^{(k)}(\sigma)\mathbf{r} \text{ for } k = 0, 1, \dots, N-1; \quad (3.11)$$

(b) if

$$\left((\mu\mathbf{E} - \mathbf{A})^{-T}\mathbf{E}^T \right)^{j-1} (\mu\mathbf{E} - \mathbf{A})^{-T}\mathbf{C}^T\mathbf{l} \in \text{Ran}(\mathbf{W}) \text{ for } j = 1, \dots, M, \quad (3.12)$$

then

$$\mathbf{I}^T \mathbf{H}^{(k)}(\mu) = \mathbf{I}^T \mathbf{H}_r^{(k)}(\mu) \text{ for } k = 0, 1, \dots, M-1; \quad (3.13)$$

(c) if $\sigma = \mu$ and both (3.10) and (3.12) hold, then

$$\mathbf{I}^T \mathbf{H}^{(k)}(\sigma) \mathbf{r} = \mathbf{I}^T \mathbf{H}_r^{(k)}(\sigma) \mathbf{r} \text{ for } k = 0, \dots, M+N-1 \quad (3.14)$$

as well.

The main cost in interpolatory model reduction originates from the need to solve large-scale (typically sparse) shifted linear systems. There is no need to solve large-scale Lyapunov or Riccati equations, giving interpolatory methods a computational advantage over competing methods. The discussion here assumes that these linear systems are solved by direct methods (e.g., Gauss elimination). However, for systems with millions of degrees of freedom, one would prefer to incorporate iterative solution strategies to construct the model reduction bases, \mathbf{V} and \mathbf{W} . We refer to [19, 90, 16] for detailed analyses of the effects of iterative solves on interpolatory model reduction and to [19, 4, 2, 24, 3] for development of effective iterative solves in the context of interpolatory model reduction.

3.2.1. Rational interpolants with $\mathbf{D}_r \neq \mathbf{D}$. So far, we have assumed that $\mathbf{D}_r = \mathbf{D}$. This is the logical choice if one is interested in minimizing the \mathcal{H}_2 norm of the error system. For the case of ordinary differential equations where \mathbf{E} is nonsingular, choosing $\mathbf{D}_r \neq \mathbf{D}$ will lead to unbounded \mathcal{H}_2 error norm. However, if, instead, one is interested in \mathcal{H}_∞ error, then flexibility in choosing \mathbf{D}_r will be necessary as the optimal \mathcal{H}_∞ approximation will have $\mathbf{D}_r \neq \mathbf{D}$ (see, e.g., [20, 47]).

Another case that may require choosing $\mathbf{D}_r \neq \mathbf{D}$ is the case of an index 1 system of differential algebraic equations (DAEs). In our setting, this means that the \mathbf{E} matrix in (2.1) has a non-defective eigenvalue at 0. (Interpolatory projection methods for DAEs is considered in detail in Section 3.3 below.) In this case, $\lim_{s \rightarrow \infty} \mathbf{H}(s) \neq \mathbf{D}$, so for $\mathbf{H}_r(s)$ to match $\mathbf{H}(s)$ asymptotically well at high frequencies, we require

$$\mathbf{D}_r = \lim_{s \rightarrow \infty} (\mathbf{H}(s) - \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r).$$

Since \mathbf{E}_r will be generically nonsingular (assuming $r < \text{rank}(\mathbf{E})$), setting $\mathbf{D}_r = \lim_{s \rightarrow \infty} \mathbf{H}(s)$ will guarantee $\lim_{s \rightarrow \infty} \mathbf{H}(s) = \lim_{s \rightarrow \infty} \mathbf{H}_r(s)$.

The next theorem shows how one may construct reduced-models with $\mathbf{D}_r \neq \mathbf{D}$ without losing interpolation properties. Without loss of generality, we assume $\mathbf{D} = \mathbf{0}$, i.e.

$$\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}. \quad (3.15)$$

In the general case with $\mathbf{D} \neq \mathbf{0}$, one simply need to replace \mathbf{D}_r with $\mathbf{D}_r - \mathbf{D}$. The result below was first given in [72] and later generalized in [17].

Theorem 3.3. *Given are a full-order model with transfer function $\mathbf{H}(s)$ as in (3.15), r distinct left-interpolation points $\{\mu_i\}_{i=1}^r$ together with r left tangent directions $\{\mathbf{l}_i\}_{i=1}^r \subset \mathbb{C}^p$, and, r distinct right-interpolation points $\{\sigma_j\}_{j=1}^r$, together with r right tangent directions $\{\mathbf{r}_j\}_{j=1}^r \subset \mathbb{C}^m$. Let the model reduction bases $\mathbf{V}_r \in \mathbb{C}^{n \times r}$ and $\mathbf{W}_r \in \mathbb{C}^{n \times r}$ be constructed as in (3.8) and (3.9), respectively. Define $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{l}}$ as*

$$\tilde{\mathbf{r}} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_r] \quad \text{and} \quad \tilde{\mathbf{l}}^T = [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_r]^T$$

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

$$\begin{aligned} \mathbf{E}_r &= \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r, & \mathbf{A}_r &= \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r + \widetilde{\mathbf{I}}^T \mathbf{D}_r \widetilde{\mathbf{r}}, \\ \mathbf{B}_r &= \mathbf{W}_r^T \mathbf{B} - \widetilde{\mathbf{I}}^T \mathbf{D}_r, & \text{and } \mathbf{C}_r &= \mathbf{C} \mathbf{V}_r - \mathbf{D}_r \widetilde{\mathbf{r}} \end{aligned} \quad (3.16)$$

Then the reduced-order model $\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r + \mathbf{D}_r$ satisfies

$$\mathbf{H}(\sigma_i)\mathbf{r}_i = \mathbf{H}_r(\sigma_i)\mathbf{r}_i \quad \text{and} \quad \mathbf{l}_i^T \mathbf{H}(\mu_i) = \mathbf{l}_i^T \mathbf{H}_r(\mu_i) \quad \text{for } i = 1, \dots, r.$$

Theorem 3.3 shows how to construct a rational tangential interpolant with an arbitrary \mathbf{D}_r term. This \mathbf{D}_r can be chosen to satisfy specific design goals.

3.3. Interpolatory Projections for Differential Algebraic Systems

The interpolation conditions in Theorems 3.1 and 3.2 are valid regardless of whether or not the matrix \mathbf{E} is singular, as long as $s\mathbf{E} - \mathbf{A}$ and $s\mathbf{E}_r - \mathbf{A}_r$ are invertible matrices for $s = \sigma, \mu$. When \mathbf{E} is nonsingular, the underlying model is a system of ordinary differential equations (ODE); when \mathbf{E} is singular, the underlying model is a system of differential algebraic equations (DAE)). Thus, from a pure interpolation perspective, the distinction does not make a difference. However, from the perspective of error measures, there is a crucial difference.

A crucial difference between a DAE system and an ODE system is that the transfer function of a DAE system could contain a polynomial part that may grow unboundedly as $s \rightarrow \infty$. In the case of ODE systems, the polynomial part is simply the constant feed-forward term, \mathbf{D} .

Let $\mathbf{H}(s)$ be the transfer function of a DAE system. We decompose $\mathbf{H}(s)$ as:

$$\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} = \mathbf{G}(s) + \mathbf{P}(s), \quad (3.17)$$

where $\mathbf{G}(s)$ is the strictly proper rational part, i.e., $\lim_{s \rightarrow \infty} \mathbf{G}(s) = 0$ and $\mathbf{P}(s)$ is the polynomial part. Now, assume that the Petrov-Galerkin projection is applied to $\mathbf{H}(s)$ as in (2.9). Then, even though \mathbf{E} is singular, the reduced matrix $\mathbf{E}_r = \mathbf{W}^T \mathbf{E} \mathbf{V}$ will generically be a nonsingular matrix for $r \leq \text{rank}(\mathbf{E})$. This means that unlike $\mathbf{H}(s)$, which contains a polynomial part $\mathbf{P}(s)$, the reduced model will correspond to an ODE and the polynomial part of the reduced transfer function $\mathbf{H}_r(s)$ will be \mathbf{D} . Decompose $\mathbf{H}_r(s)$ in a similar way

$$\mathbf{H}_r(s) = \mathbf{G}_r(s) + \mathbf{D},$$

where $\mathbf{G}_r = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r$ is strictly proper. Then, the error transfer function

$$\mathbf{H}(s) - \mathbf{H}_r(s) = (\mathbf{G}(s) - \mathbf{G}_r(s)) + (\mathbf{P}(s) - \mathbf{D})$$

has a polynomial part $\mathbf{P}(s) - \mathbf{D}$. Even when $\mathbf{P}(s)$ is a polynomial of degree 1, the difference $\mathbf{P}(s) - \mathbf{D}$ will grow without bound as $s \rightarrow \infty$, leading to unbounded \mathcal{H}_∞ and \mathcal{H}_2 error norms. Even when $\mathbf{P}(s)$ is a constant polynomial (i.e. degree 0), unless $\mathbf{P}(s) = \mathbf{D}$, this will still lead to unbounded \mathcal{H}_2 error. The only way to guarantee bounded error norms is to make sure that the reduced transfer function $\mathbf{H}_r(s)$ has exactly the same polynomial part as $\mathbf{H}(s)$, i.e. $\mathbf{H}_r(s) = \mathbf{G}_r(s) + \mathbf{P}(s)$ so that the error function is simply $\mathbf{H}(s) - \mathbf{H}_r(s) = \mathbf{G}(s) - \mathbf{G}_r(s)$ having only a null polynomial component. Based on these observations, [58, 90] discusses how to modify the interpolatory projection bases \mathbf{V} and \mathbf{W} in order to achieve this goal. As expected, the left and right deflating subspaces of the pencil $\lambda\mathbf{E} - \mathbf{A}$ corresponding to finite and infinite eigenvalues play a crucial role:

Theorem 3.4. *Suppose the transfer function $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} = \mathbf{G}(s) + \mathbf{P}(s)$ is associated with a DAE, where $\mathbf{G}(s)$ and $\mathbf{P}(s)$ are, respectively, the strictly proper and the polynomial parts of $\mathbf{H}(s)$. Let \mathbb{P}_l and \mathbb{P}_r be the spectral projectors onto the left and right deflating subspaces of the pencil $\lambda\mathbf{E} - \mathbf{A}$ corresponding to the finite eigenvalues. Also, let the columns of \mathbf{W}_∞ and \mathbf{V}_∞ span the left and right deflating subspaces of $\lambda\mathbf{E} - \mathbf{A}$ corresponding to the eigenvalue at infinity. For interpolation points $\sigma, \mu \in \mathbb{C}$, suppose $\sigma\mathbf{E} - \mathbf{A}$ and $\mu\mathbf{E} - \mathbf{A}$ are invertible and $\mathbf{r} \in \mathbb{C}^m$ and $\mathbf{l} \in \mathbb{C}^\ell$ are given (nontrivial) tangent directions. Suppose further that $\mathbf{H}_r(s)$ is the reduced transfer function obtained by projection as in (2.9) using the model reduction bases \mathbf{V} and \mathbf{W} . Construct \mathbf{V}_f and \mathbf{W}_f so that*

$$((\sigma\mathbf{E} - \mathbf{A})^{-1}\mathbf{E})^{j-1}(\sigma\mathbf{E} - \mathbf{A})^{-1}\mathbb{P}_l\mathbf{B}\mathbf{r} \in \text{Ran}(\mathbf{V}_f), \quad (3.18)$$

for $j = 1, \dots, N$, and

$$((\mu\mathbf{E} - \mathbf{A})^{-T}\mathbf{E}^T)^{j-1}(\mu\mathbf{E} - \mathbf{A})^{-T}\mathbb{P}_r^T\mathbf{C}^T\mathbf{l} \in \text{Ran}(\mathbf{W}_f) \quad (3.19)$$

for $j = 1, \dots, M$. Define \mathbf{W} and \mathbf{V} using $\mathbf{W} = [\mathbf{W}_f, \mathbf{W}_\infty]$ and $\mathbf{V} = [\mathbf{V}_f, \mathbf{V}_\infty]$. Then, $\mathbf{H}_r(s) = \mathbf{G}_r(s) + \mathbf{P}_r(s)$ satisfies $\mathbf{P}_r(s) = \mathbf{P}(s)$ together with (3.11) and (3.13). If, in addition, $\sigma = \mu$, then (3.14) holds as well.

Theorem 3.4, taken from [58], shows how to apply projection-based tangential interpolation to DAEs in the most general case where the index of the DAE and the interpolation points are arbitrary. By appropriately incorporating the deflating projectors \mathbb{P}_r and \mathbb{P}_l in the model reduction bases, the polynomial part of $\mathbf{H}(s)$ is exactly matched as desired while simultaneously enforcing interpolation. For the special case of DAEs with proper transfer functions and interpolation around $s = \infty$, a solution has been given in [28]. For descriptor systems of index 1, [10] offered a solution that uses an appropriately chosen \mathbf{D}_r term.

Remark 3.5. A fundamental difficulty in the reduction of DAEs is the need to compute deflating projectors \mathbb{P}_r and \mathbb{P}_l . For large-scale DAEs, construction of \mathbb{P}_r and \mathbb{P}_l is at best very costly, if even feasible. However, for the cases of semi-explicit descriptor systems of index-1 and Stokes-type descriptor systems of index 2, it is possible to apply interpolatory projections without forming \mathbb{P}_r and \mathbb{P}_l explicitly [58, 90]. Thus, no greater effort is required to produce reduced models than for the case of ODEs. The Stokes-type descriptor systems of index 2 were also studied in [63] in a balanced truncation setting. Recently, [1] extended the work of [58] to address the reduction of index-3 DAEs without forming projectors explicitly. For some structured problems arising in circuit simulation, multibody systems or computational fluid dynamics, these projectors can be constructed without much computational effort [82]. We choose to omit these details from the present discussion but refer the interested reader to the original sources.

4. Interpolatory Projections for \mathcal{H}_2 Optimal Approximation

When interpolation points and tangent directions are specified, Section 3 presents an approach that one may follow in order to construct a reduced model satisfying the desired (tangential) conditions. Notably, this development does not suggest a strategy for choosing interpolation points and tangent directions that lead to high-fidelity reduced models. In this section, we approach this issue by developing

interpolatory conditions that are necessary for optimal approximation with respect to the \mathcal{H}_2 norm.

4.1. Interpolatory \mathcal{H}_2 -optimality Conditions

Consider the following optimization problem: Given a full-order system, $\mathbf{H}(s)$, find a reduced model, $\mathbf{H}_r(s)$ that minimizes the \mathcal{H}_2 error; i.e.,

$$\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2} = \min_{\dim(\tilde{\mathbf{H}}_r)=r} \left\| \mathbf{H} - \tilde{\mathbf{H}}_r \right\|_{\mathcal{H}_2}. \quad (4.1)$$

As we pointed out in Section 2.4.2, small \mathcal{H}_2 error induces small time domain error $\|\mathbf{y} - \mathbf{y}_r\|_{L_\infty}$, so attempting to *minimize* the \mathcal{H}_2 error is a worthy goal.

The \mathcal{H}_2 optimization problem (4.1) is nonconvex ; finding a global minimizer is typically infeasible. A common approach used instead involves finding *locally* optimal reduced models that satisfy first-order necessary conditions for optimality. The (simpler) problem of finding locally optimal \mathcal{H}_2 reduced models has been studied extensively. Optimality conditions have been formulated either in terms of Lyapunov and Sylvester equations [89, 64, 61, 96, 91, 81, 27] or in terms of rational (tangential) interpolation conditions [73, 54, 55, 57, 84, 35, 67, 21, 22, 18, 66, 76]. [57] showed the equivalence between the Lyapunov/Sylvester equation conditions and the interpolation framework that we describe here.

We will first assume that \mathbf{E} is nonsingular, so that $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$ will correspond to a system of ODEs and $\lim_{s \rightarrow \infty} \mathbf{H}(s) = \mathbf{D}$. In order to have a bounded \mathcal{H}_2 error norm, $\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2}$, it is necessary that $\mathbf{D}_r = \mathbf{D}$. Therefore, without loss of generality, one may take $\mathbf{D}_r = \mathbf{D} = \mathbf{0}$.

For MIMO systems, interpolatory first-order conditions for \mathcal{H}_2 optimality are best understood from the pole-residue expansion for $\mathbf{H}_r(s)$. We write $\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r$ in the following way:

$$\mathbf{H}_r(s) = \sum_{i=1}^r \frac{\mathbf{l}_i \mathbf{r}_i^T}{s - \lambda_i} \quad (4.2)$$

where we have assumed that the $\lambda_i s$ are distinct. We refer to $\mathbf{l}_i \in \mathbb{C}^p$ and $\mathbf{r}_i \in \mathbb{C}^m$ in (4.2) respectively as left/right residue directions associated with the pole λ_i of $\mathbf{H}_r(s)$; $\mathbf{l}_i \mathbf{r}_i^T$ is the (matrix) residue of $\mathbf{H}_r(s)$ at $s = \lambda_i$. The pole-residue expansion in (4.2) can be computed effectively by computing a generalized eigenvalue decomposition for the matrix pencil $\lambda \mathbf{E}_r - \mathbf{A}_r$, which is a trivial computation for the small to modest orders of r typically encountered. Note that finding such a representation for the full model $\mathbf{H}(s)$ will generally be infeasible.

Theorem 4.1. *Let $\mathbf{H}_r(s)$ in (4.2) be the best r th order rational approximation of $\mathbf{H}(s)$ with respect to the \mathcal{H}_2 norm. Then,*

$$\mathbf{H}(-\lambda_k) \mathbf{r}_k = \mathbf{H}_r(-\lambda_k) \mathbf{r}_k, \quad (4.3a)$$

$$\mathbf{l}_k^T \mathbf{H}(-\lambda_k) = \mathbf{l}_k^T \mathbf{H}_r(-\lambda_k), \text{ and} \quad (4.3b)$$

$$\mathbf{l}_k^T \mathbf{H}'(-\lambda_k) \mathbf{r}_k = \mathbf{l}_k^T \mathbf{H}'_r(-\lambda_k) \mathbf{r}_k \quad (4.3c)$$

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

In particular, any optimal \mathcal{H}_2 approximation $\mathbf{H}_r(s)$ must be a bitangential Hermite interpolant to $\mathbf{H}(s)$, and this theorem directly connects optimal model reduction to interpolation. The optimal interpolation points and tangent directions are derived from the pole-residue representation of $\mathbf{H}_r(s)$: The optimal interpolation points are the mirror images of the poles of $\mathbf{H}_r(s)$ reflected across the imaginary

axis, and the optimal tangent directions are the residue directions associated with that pole.

Interpolatory conditions for SISO systems were initially introduced by Meier and Luenberger [73]. However, until recently effective numerical algorithms to find reduced systems that satisfy these conditions were lacking, especially for large-scale settings. Gugercin *et al.* in [54, 55] introduced such an algorithm, called the *Iterative Rational Krylov Algorithm* (IRKA). In practice, IRKA has significantly expanded the utility of optimal \mathcal{H}_2 model reduction. The optimality conditions for MIMO systems as presented in Theorem 4.1 were developed in [57, 35, 84], and led to an analogous algorithm for IRKA in the MIMO case. This is the main focus of Section 4.2. Recall that we have assumed that $\mathbf{H}_r(s)$ has distinct (reduced) poles: $\lambda_1, \dots, \lambda_r$. Optimality conditions for cases when $\mathbf{H}_r(s)$ has repeated poles are derived in [83].

4.2. IRKA for optimal \mathcal{H}_2 approximation

Theorem 4.1 gives optimality conditions that depend on the poles and residues of a reduced order system, a locally \mathcal{H}_2 -optimal reduced system, that is unknown *a priori* and is yet to be determined. IRKA utilizes the construction of Theorem 3.1 to force interpolation at the mirror images of successive sets of reduced poles, iteratively correcting the reduced model until the optimality conditions of Theorem 4.1 hold. The method proceeds as follows: Given some initial interpolation points $\{\sigma_i\}_{i=1}^r$ and directions $\{\mathbf{r}_i\}_{i=1}^r$ and $\{\mathbf{l}_i\}_{i=1}^r$, construct \mathbf{V} and \mathbf{W} as in (3.8) and (3.9), respectively, and construct an intermediate reduced model $\mathbf{H}_r(s)$ using (2.9). Then, compute the pole-residue decomposition of $\mathbf{H}_r(s)$,

$$\mathbf{H}_r(s) = \sum_{i=1}^r \frac{\widehat{\mathbf{l}}_i \widehat{\mathbf{r}}_i^T}{s - \lambda_i}$$

(by solving a small $r \times r$ generalized eigenvalue problem). For $\mathbf{H}_r(s)$ to satisfy the first-order necessary conditions, we need $\sigma_i = -\lambda_i$, $\mathbf{r}_i = \widehat{\mathbf{r}}_i$, and $\mathbf{l}_i = \widehat{\mathbf{l}}_i$, for $i = 1, \dots, r$. Therefore, set

$$\sigma_i \leftarrow -\lambda_i, \quad \mathbf{r}_i \leftarrow \widehat{\mathbf{r}}_i, \quad \text{and} \quad \mathbf{l}_i \leftarrow \widehat{\mathbf{l}}_i, \quad \text{for} \quad i = 1, \dots, r$$

as the next interpolation data and iterate until convergence is reached. A brief sketch of IRKA is given below:

Algorithm 4.1. MIMO \mathcal{H}_2 -Optimal Tangential Interpolation (“IRKA”)

1. Make an initial r -fold shift selection: $\{\sigma_1, \dots, \sigma_r\}$ that is closed under conjugation (i.e., $\{\sigma_1, \dots, \sigma_r\} \equiv \{\bar{\sigma}_1, \dots, \bar{\sigma}_r\}$ viewed as sets) and initial tangent directions $\mathbf{r}_1, \dots, \mathbf{r}_r$ and $\mathbf{l}_1, \dots, \mathbf{l}_r$, also closed under conjugation.
2. $\mathbf{V}_r = [(\sigma_1 \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{r}_1 \ \cdots \ (\sigma_r \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{r}_r]$
3. $\mathbf{W}_r = [(\sigma_1 \mathbf{E} - \mathbf{A}^T)^{-1} \mathbf{C}^T \mathbf{l}_1 \ \cdots \ (\sigma_r \mathbf{E} - \mathbf{A}^T)^{-1} \mathbf{C}^T \mathbf{l}_1]$
4. while (not converged)
 - (a) $\mathbf{A}_r = \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r$, $\mathbf{E}_r = \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r$, $\mathbf{B}_r = \mathbf{W}_r^T \mathbf{B}$, and $\mathbf{C}_r = \mathbf{C} \mathbf{V}_r$
 - (b) Compute a pole-residue expansion of $\mathbf{H}_r(s)$:

$$\mathbf{H}_r(s) = \mathbf{C}_r (s \mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{B}_r = \sum_{i=1}^r \frac{\hat{\mathbf{l}}_i \hat{\mathbf{r}}_i^T}{s - \lambda_i}$$
 - (c) $\sigma_i \leftarrow -\lambda_i$, $\mathbf{r}_i \leftarrow \hat{\mathbf{r}}_i$, and $\mathbf{l}_i \leftarrow \hat{\mathbf{l}}_i$, for $i = 1, \dots, r$
 - (d) $\mathbf{V}_r = [(\sigma_1 \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{r}_1 \ \cdots \ (\sigma_r \mathbf{E} - \mathbf{A})^{-1} \mathbf{B} \mathbf{r}_r]$
 - (e) $\mathbf{W}_r = [(\sigma_1 \mathbf{E} - \mathbf{A}^T)^{-1} \mathbf{C}^T \mathbf{l}_1 \ \cdots \ (\sigma_r \mathbf{E} - \mathbf{A}^T)^{-1} \mathbf{C}^T \mathbf{l}_1]$
5. $\mathbf{A}_r = \mathbf{W}_r^T \mathbf{A} \mathbf{V}_r$, $\mathbf{E}_r = \mathbf{W}_r^T \mathbf{E} \mathbf{V}_r$, $\mathbf{B}_r = \mathbf{W}_r^T \mathbf{B}$, $\mathbf{C}_r = \mathbf{C} \mathbf{V}_r$

Upon convergence, the reduced model, $\mathbf{H}_r(s)$, satisfies the interpolatory first-order necessary conditions (4.3) for \mathcal{H}_2 optimality by construction. Convergence is generally observed to be rapid; though it slows as input/output orders grow. Convergence may be guaranteed *a priori* in some circumstances [46]; yet there are known cases where convergence may fail [57, 46]. When convergence occurs, the resulting reduced model is guaranteed to be a local \mathcal{H}_2 -minimizer since the local maxima of the \mathcal{H}_2 minimization problem are known to be repellent [66]. Overall in practice, IRKA has seen significant success in computing high fidelity (locally) optimal reduced models and has been successfully applied in large-scale settings to find \mathcal{H}_2 -optimal reduced models for systems with hundreds of thousands of state variables; for example, see [65] for application in cellular neurophysiology, [30] for energy efficient building design in order to produce accurate compact models for the indoor-air environment, [57] for optimal cooling for steel profiles. Moreover, [23] has extended IRKA to the reduction of *bilinear* dynamical systems, a special class of weakly nonlinear dynamical systems.

Our analysis so far has assumed that \mathbf{E} is a nonsingular matrix. Interpolatory optimal \mathcal{H}_2 model reduction for the case of singular \mathbf{E} , i.e., for systems of DAEs, has been developed in [58] and IRKA has been extended to DAEs. Similar to the ODE case where we require $\mathbf{D} = \mathbf{D}_r$, the DAE case requires that the polynomial part of $\mathbf{H}_r(s)$ match that of $\mathbf{H}(s)$ exactly and the strictly proper part of $\mathbf{H}_r(s)$ be an optimal tangential interpolant to the strictly proper part of $\mathbf{H}(s)$. For details, we refer the reader to [58].

4.3. Interpolatory Weighted- \mathcal{H}_2 Model Reduction

The error measures we have considered thus far give the same weight to all frequencies equally and they are global in nature to the extent that degradation in fidelity is penalized in the same way throughout the full frequency spectrum. However, some applications require that certain frequencies be weighted more than others. For example, certain dynamical systems, such as mechanical systems or electrical circuits, might operate only in certain frequency bands and retaining fidelity outside this frequency band carries no value. This leads to the problem of weighted model reduction. We formulate it here in terms of a weighted- \mathcal{H}_2 norm.

Let $\mathbf{W}(s)$ be an input weighting function, a “shaping filter.” We will assume that $\mathbf{W}(s)$ is a rational function itself in the form

$$\mathbf{W}(s) = \mathbf{C}_w (s\mathbf{I} - \mathbf{A}_w)^{-1} \mathbf{B}_w + \mathbf{D}_w = \sum_{k=1}^{n_w} \frac{\mathbf{e}_k \mathbf{f}_k^T}{s - \gamma_k} + \mathbf{D}_w, \quad (4.4)$$

where n_w denotes the dimension of $\mathbf{W}(s)$. Then, given the full-model $\mathbf{H}(s)$ and the weight $\mathbf{W}(s)$, define the weighted \mathcal{H}_2 error norm as:

$$\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2(W)} \stackrel{\text{def}}{=} \|(\mathbf{H} - \mathbf{H}_r) \cdot \mathbf{W}\|_{\mathcal{H}_2}. \quad (4.5)$$

In addition to the input weighting $\mathbf{W}(s)$, one may also define a filter for output weighting. For simplicity of presentation, we focus here on one-sided weighting only. The goal is to find a reduced model $\mathbf{H}_r(s)$ that minimizes the weighted error (4.5):

$$\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2(W)} = \min_{\dim(\tilde{\mathbf{H}}_r)=r} \|\mathbf{H} - \tilde{\mathbf{H}}_r\|_{\mathcal{H}_2(W)}. \quad (4.6)$$

Weighted- \mathcal{H}_2 model reduction has been considered in [61] and [81] using a framework that uses Riccati and/or Lyapunov equations. A numerically more efficient, interpolation-based approach was introduced in [6] for the SISO case. This initial interpolatory framework was significantly extended and placed on a more rigorous theoretical footing (which allowed straightforward extension to MIMO systems) in [32] where the equivalence of the Riccati and interpolation-based frameworks were also proved. Our presentation below follows [32]. We use the notation $\mathbf{H} \in \mathcal{H}_2$ to indicate that $\mathbf{H}(s)$ is a stable dynamical system with $\|\mathbf{H}\|_{\mathcal{H}_2} < \infty$. We use the notation $\mathbf{H} \in \mathcal{H}_2(W)$ analogously.

The interpolatory framework for the weighted- \mathcal{H}_2 problem is best understood by defining a new linear transformation ([32, 6])

$$\mathfrak{F}[\mathbf{H}](s) = \mathbf{H}(s)\mathbf{W}(s)\mathbf{W}(-s)^T + \sum_{k=1}^{n_w} \mathbf{H}(-\gamma_k)\mathbf{W}(-\gamma_k) \frac{\mathbf{f}_k \mathbf{e}_k^T}{s + \gamma_k}, \quad (4.7)$$

where $\mathbf{H} \in \mathcal{H}_2(W)$, and \mathbf{f}_k and \mathbf{e}_k are as defined in (4.4). $\mathfrak{F}[\mathbf{H}](s)$ is a bounded linear transformation from $\mathcal{H}_2(W)$ to \mathcal{H}_2 [32]. A state-space representation for $\mathfrak{F}[\mathbf{H}](s)$ is given by

$$\begin{aligned} \mathfrak{F}[\mathbf{H}](s) &= \mathbf{C}_{\mathfrak{F}}(s\mathbf{I} - \mathcal{A}_{\mathfrak{F}})^{-1} \mathcal{B}_{\mathfrak{F}} \\ &= \underbrace{\begin{bmatrix} \mathbf{C} & \mathbf{D}\mathbf{C}_w \end{bmatrix}}_{\mathbf{C}_{\mathfrak{F}}} \left(s\mathbf{I} - \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{B}\mathbf{C}_w \\ \mathbf{0} & \mathbf{A}_w \end{bmatrix}}_{\mathcal{A}_{\mathfrak{F}}} \right)^{-1} \underbrace{\begin{bmatrix} \mathbf{Z}\mathbf{C}_w^T + \mathbf{B}\mathbf{D}_w\mathbf{D}_w^T \\ \mathbf{P}_w\mathbf{C}_w^T + \mathbf{B}_w\mathbf{D}_w^T \end{bmatrix}}_{\mathcal{B}_{\mathfrak{F}}}, \end{aligned} \quad (4.8)$$

where \mathbf{P}_w and \mathbf{Z} solve, respectively,

$$\mathbf{A}_w\mathbf{P}_w + \mathbf{P}_w\mathbf{A}_w^T + \mathbf{B}_w\mathbf{B}_w^T = \mathbf{0} \quad \text{and} \quad (4.9)$$

$$\mathbf{A}\mathbf{Z} + \mathbf{Z}\mathbf{A}^T + \mathbf{B}(\mathbf{C}_w\mathbf{P}_w + \mathbf{D}_w\mathbf{B}_w^T) = \mathbf{0}. \quad (4.10)$$

For $\mathbf{H}, \mathbf{H}_r \in \mathcal{H}_2(W)$, denote the impulse responses corresponding to $\mathfrak{F}[\mathbf{H}](s)$ and $\mathfrak{F}[\mathbf{H}_r](s)$, respectively by $\mathbf{F}(t)$ and $\mathbf{F}_r(t)$, so that $\mathfrak{F}[\mathbf{H}] = \mathcal{L}\{\mathbf{F}\}$ and $\mathfrak{F}[\mathbf{H}_r] = \mathcal{L}\{\mathbf{F}_r\}$, where $\mathcal{L}\{\cdot\}$ denotes the Laplace transform.

Theorem 4.2. *Given the input weighting $\mathbf{W}(s)$, let $\mathbf{H}_r \in \mathcal{H}_2(W)$ be the best order- r rational approximation to \mathbf{H} in the weighted- \mathcal{H}_2 norm. Suppose that \mathbf{H}_r has the form*

$$\mathbf{H}_r(s) = \mathbf{C}_r (s\mathbf{I} - \mathbf{A}_r)^{-1} \mathbf{B}_r + \mathbf{D}_r = \sum_{k=1}^{n_r} \frac{\mathbf{l}_k \mathbf{r}_k^T}{s - \lambda_k} + \mathbf{D}_r. \quad (4.11)$$

where $\lambda_1, \dots, \lambda_r$ are assumed to be distinct. Then

$$\mathfrak{F}[\mathbf{H}](-\lambda_k) \mathbf{r}_k = \mathfrak{F}[\mathbf{H}_r](-\lambda_k) \mathbf{r}_k \quad (4.12a)$$

$$\mathbf{l}_k^T \mathfrak{F}[\mathbf{H}](-\lambda_k) = \mathbf{l}_k^T \mathfrak{F}[\mathbf{H}_r](-\lambda_k), \text{ and} \quad (4.12b)$$

$$\mathbf{l}_k^T \mathfrak{F}'[\mathbf{H}](-\lambda_k) \mathbf{r}_k = \mathbf{l}_k^T \mathfrak{F}'[\mathbf{H}_r](-\lambda_k) \mathbf{r}_k. \quad (4.12c)$$

$$\mathbf{F}(0) \mathbf{n} = \mathbf{F}_r(0) \mathbf{n}. \quad (4.12d)$$

for $k = 1, 2, \dots, r$ and for all $\mathbf{n} \in \text{Ker}(\mathbf{D}_w^T)$ where $\mathfrak{F}'[\cdot](s) = \frac{d}{ds} \mathfrak{F}[\cdot](s)$.

Bitangential Hermite interpolation once again appears as a necessary condition for optimality. However, unlike the unweighted \mathcal{H}_2 case, the interpolation conditions need to be satisfied by the maps $\mathfrak{F}[\mathbf{H}](s)$ and $\mathfrak{F}[\mathbf{H}_r](s)$ as opposed to $\mathbf{H}(s)$ and $\mathbf{H}_r(s)$. For $\mathbf{W}(s) = \mathbf{I}$, (4.12a)-(4.12c) simplify to (4.3a)-(4.3c) and (4.12d) is automatically satisfied since $\text{Ker}(\mathbf{D}_w^T) = \{\mathbf{0}\}$ when $\mathbf{W}(s) = \mathbf{I}$.

Guided by how IRKA is employed to satisfy the \mathcal{H}_2 optimality conditions (4.3a)-(4.3c), one might consider simply applying IRKA to the state-space representation of $\mathfrak{F}[\mathbf{H}](s)$ given in (4.8) to satisfy the weighted interpolation conditions (4.12a)-(4.12c). Unfortunately, this solves a different problem: If IRKA is applied to $\mathfrak{F}[\mathbf{H}](s)$ directly, then one obtains a reduced model $\mathbf{H}_r(s)$ that interpolates $\mathfrak{F}[\mathbf{H}](s)$. That is, instead of (4.12a), one obtains instead $\mathfrak{F}[\mathbf{H}](-\lambda_k) \mathbf{r}_k = \mathbf{H}_r(-\lambda_k) \mathbf{r}_k$ which is clearly not appropriate. The need to preserve the structure in the maps $\mathfrak{F}[\mathbf{H}](s)$ and $\mathfrak{F}[\mathbf{H}_r](s)$ while satisfying interpolatory conditions makes the development of an IRKA-like algorithm for the weighted- \mathcal{H}_2 problem quite nontrivial. Breiten *et al* in [32] proposed an algorithm, called ‘‘Nearly Optimal Weighted Interpolation’’ (NOWI), that *nearly satisfies* the interpolatory optimality conditions (4.12a)-(4.12c) while preserving the structure in $\mathfrak{F}[\mathbf{H}](s)$ and $\mathfrak{F}[\mathbf{H}_r](s)$. The deviation from exact interpolation is quantified explicitly. Even though NOWI proves itself to be a very effective numerical algorithm in many circumstances, the development of an algorithm that satisfies (4.12a)-(4.12c) exactly remains an important future goal.

4.4. Descent Algorithms for \mathcal{H}_2 Model Reduction

At its core, IRKA is a fixed point iteration. Excepting the special case of symmetric state-space systems,¹ where convergence is guaranteed, convergence of IRKA for general systems is not guaranteed, see [57, 46], although superior performance is commonly observed. More significantly, IRKA is not a descent algorithm; that is, the \mathcal{H}_2 error might fluctuate during intermediate steps and premature termination of the algorithm could result (at least in principle) in a worse approximation than what was provided for initialization. To address these issues, Beattie and Gugercin [22] developed an \mathcal{H}_2 descent algorithm that reduces the \mathcal{H}_2 error at each step of the iteration and assures global convergence to a local minimum.

¹ $\mathbf{E} = \mathbf{E}^T$ is positive definite, $\mathbf{A} = \mathbf{A}^T$ is negative definite, and $\mathbf{B} = \mathbf{C}^T$

The key to their approach is the following representation of the \mathcal{H}_2 error norm for MIMO systems [22]:

Theorem 4.3. *Given a full-order model $\mathbf{H}(s)$, let $\mathbf{H}_r(s)$ have the form in (4.2), i.e.,*

$$\mathbf{H}_r(s) = \sum_{i=1}^r \frac{\mathbf{l}_i \mathbf{r}_i^T}{s - \lambda_i}.$$

Then, the \mathcal{H}_2 norm of the error system is given by

$$\|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2}^2 = \|\mathbf{H}\|_{\mathcal{H}_2}^2 - 2 \sum_{k=1}^r \mathbf{l}_k^T \mathbf{H}(-\lambda_k) \mathbf{r}_k + \sum_{k,\ell=1}^r \frac{\mathbf{l}_k^T \mathbf{l}_\ell \mathbf{r}_\ell^T \mathbf{r}_k}{-\lambda_k - \lambda_\ell}. \quad (4.13)$$

For SISO systems, Krajewski *et al.* [66] developed and proved a similar expression, which was later rediscovered in [56, 53, 9].

If one considers $\mathbf{H}_r(s) = \sum_{i=1}^r \frac{\mathbf{l}_i \mathbf{r}_i^T}{s - \lambda_i}$ in the pole-residue form, the variables defining the reduced model are the residue directions \mathbf{l}_i , \mathbf{r}_i and the poles λ_i , for $i = 1, \dots, r$. The formula (4.13) expresses the error in terms of these variables. Thus, one can compute the gradient and Hessian of the error with respect to unknowns and construct globally convergent descent (optimization) algorithms. Gradient and Hessian expressions were derived in [22]. For brevity of presentation, we include only the gradient expressions here.

Theorem 4.4. *Given the full-order model $\mathbf{H}(s)$ and the reduced model $\mathbf{H}_r(s)$ as in (4.2), define*

$$\mathcal{J} \stackrel{\text{def}}{=} \|\mathbf{H} - \mathbf{H}_r\|_{\mathcal{H}_2}^2.$$

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

$$\frac{\partial \mathcal{J}}{\partial \lambda_i} = -2 \mathbf{l}_i^T (\mathbf{H}'_r(-\lambda_i) - \mathbf{H}'(-\lambda_i)) \mathbf{r}_i. \quad (4.14)$$

Moreover, the gradient of \mathcal{J} with respect to residue directions listed as

$$\{\mathbf{r}, \mathbf{l}\} = [\mathbf{r}_1^T, \mathbf{l}_1^T, \mathbf{r}_2^T, \mathbf{l}_2^T, \dots, \mathbf{r}_r^T, \mathbf{l}_r^T]^T,$$

is given by $\nabla_{\{\mathbf{r}, \mathbf{l}\}} \mathcal{J}$, a vector of length $r(m+p)$, partitioned into r vectors of length $m+p$ as

$$(\nabla_{\{\mathbf{r}, \mathbf{l}\}} \mathcal{J})_\ell = \begin{pmatrix} 2 (\mathbf{l}_\ell^T \mathbf{H}_r(-\lambda_\ell) - \mathbf{l}_\ell^T \mathbf{H}(-\lambda_\ell))^T \\ 2 (\mathbf{H}_r(-\lambda_\ell) \mathbf{r}_\ell - \mathbf{H}(-\lambda_\ell) \mathbf{r}_\ell) \end{pmatrix} \quad (4.15)$$

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

One may observe that setting the gradient expression in (4.14) and (4.15) to zero leads immediately to the interpolatory optimality conditions (4.3). Having gradient and Hessian expressions at hand, one may then develop a globally convergent descent algorithm for \mathcal{H}_2 optimal reduction as done in [22] where the optimization algorithm was put in a trust-region framework. Unlike in IRKA, the intermediate reduced models are not interpolatory. However, upon convergence, they satisfy the interpolatory optimality conditions.

In a recent paper, for SISO systems $H(s) = \mathbf{c}^T (s\mathbf{E} - \mathbf{A})^{-1} \mathbf{b}$ where $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$ are length- n vectors, Panzer *et al* [76] applied a descent-type algorithm successively. Instead of designing a degree- r rational function directly, [76] first constructs a SISO degree-2 rational function $H_r(s) = \mathbf{c}_r^T (s\mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{b}_r$ where $\mathbf{A}_r, \mathbf{E}_r \in \mathbb{R}^{2 \times 2}$ and $\mathbf{b}_r, \mathbf{c}_r^T \in \mathbb{R}^2$ by a descent method where only Lagrange optimality conditions

are enforced (without Hermite conditions). Then, the error transfer function is decomposed in a multiplicative form

$$H - H_r(s) = \underbrace{\left(\mathbf{c}^T (s\mathbf{E} - \mathbf{A})^{-1} \mathbf{b}_\perp \right)}_{H_\perp(s)} \left(\mathbf{c}_r^T (s\mathbf{E}_r - \mathbf{A}_r)^{-1} \mathbf{b}_r \right)$$

where $\mathbf{b}_\perp = \mathbf{b} - \mathbf{E}\mathbf{V}(\mathbf{W}^T\mathbf{E}\mathbf{V})^{-1}\mathbf{b}_r$ and the method proceeds by constructing another degree-2 approximation to $H_\perp(s)$ in a descent framework, once more only enforcing the Lagrange optimality conditions. At the end, all the intermediate degree-2 approximants are put together in a special way to form the final reduced model of degree- r . For details, we refer the reader to [76]. The final reduced model will not generally satisfy the full set of interpolatory \mathcal{H}_2 optimality conditions - only the Lagrange conditions are satisfied. Moreover, the incremental approach means optimization over a smaller set; thus for a given r , optimization directly over a degree- r rational function using the gradient and Hessian expressions in [22] will lead to a smaller model reduction error than an incremental search. However since [76] performs the optimization over a very small number of variables in each step, this approach can provide some numerical advantages.

Druskin *et al* in [39] and [40] suggest alternative greedy-type algorithms for interpolatory model reduction. Instead of constructing r interpolation points (and directions at every step) as done in IRKA or in the descent framework of [22], [39] and [40] start instead with an interpolation point and corresponding tangent directions. Then, a greedy search on the residual determines the next set of interpolation data. Since the greedy search is not done in a descent setting, this is not a descent method and at the end optimality conditions will not be satisfied typically. Nonetheless, the final reduced model is still an interpolatory method. Even though the resulting reduced models will not generally be as accurate as those obtained by IRKA, the methods of [39] and [40] provide satisfactory approximants at relatively low cost.

Descent-type algorithms have been extended to the weighted- \mathcal{H}_2 norm minimization as well; for details see [77, 86, 31].

5. Interpolatory Model Reduction for Generalized Coprime Framework

So far, we have assumed that the original transfer function has a generic first-order state-space representation: $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}$. This representation is quite general and a wide range of linear dynamical systems can be converted to this form, at least in principle. However, problem formulations often lead to somewhat different structures that reflect the underlying physics or other important system features. One may wish to retain such structural features and conversion to standard first-order representations often obfuscates these features and may lead then to “unphysical” reduced models. Neutral and delay differential equations present another interesting class of dynamical systems that are generically of infinite order, so they do not accept a standard first-order representation using a finite dimensional state space. In this section, we follow the discussion of Beattie and Gugercin [17], and show how interpolatory methods can be used to preserve relevant system structure in reduced models, often avoiding entirely the need to convert the system to an equivalent first-order state-space representation.

5.1. Polynomial and Delay Systems

A common example of a situation where conversion to the standard state-space form is possible but may not be prudent is the case of constant coefficient ordinary differential equations of order two or more, with dynamics given by

$$\begin{aligned} \mathbf{A}_0 \frac{d^\ell \mathbf{x}}{dt^\ell} + \mathbf{A}_1 \frac{d^{\ell-1} \mathbf{x}}{dt^{\ell-1}} + \cdots + \mathbf{A}_\ell \mathbf{x}(t) &= \mathbf{B} \mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{C}_1 \frac{d^{\ell-1} \mathbf{x}}{dt^{\ell-1}} + \mathbf{C}_2 \frac{d^{\ell-2} \mathbf{x}}{dt^{\ell-2}} + \cdots + \mathbf{C}_\ell \mathbf{x}(t), \end{aligned} \quad (5.1)$$

where $\mathbf{A}_i \in \mathbb{R}^{n \times n}$, for $i = 0, \dots, \ell$, $\mathbf{B} \in \mathbb{R}^{n \times m}$ and $\mathbf{C}_i \in \mathbb{R}^{p \times n}$ for $i = 1, \dots, \ell$. By defining a state vector $\mathbf{q} = [\mathbf{x}^T, \dot{\mathbf{x}}^T, \ddot{\mathbf{x}}^T, \dots, (\mathbf{x}^{(\ell-1)})^T]^T$, one may easily convert (5.1) into an equivalent first-order form (2.1). However, this has two major disadvantages:

1. By forming the vector $\mathbf{q}(t)$, the physical meaning of the state vector $\mathbf{x}(t)$ is lost in the model reduction state since the model reduction process will mix the physical quantities such as displacement and velocity; the reduced state loses its physical significance.
2. The dimension of the aggregate state $\mathbf{q}(t)$ is $\ell \times n$. As a consequence, conversion to first-order form has made the model reduction problem numerically much harder. For example, if reduction is approached via interpolation, we now need to solve linear systems of size $(\ell n) \times (\ell n)$.

Therefore, it is desirable to perform model reduction in the original state-space associated with the original representation (5.1); we wish to preserve the structure of (5.1) in the reduced model and produce a reduced model of the form

$$\begin{aligned} \mathbf{A}_{0,r} \frac{d^\ell \mathbf{x}_r}{dt^\ell} + \mathbf{A}_{1,r} \frac{d^{\ell-1} \mathbf{x}_r}{dt^{\ell-1}} + \cdots + \mathbf{A}_{\ell,r} \mathbf{x}_r(t) &= \mathbf{B}_r \mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}_{1,r} \frac{d^{\ell-1} \mathbf{x}_r}{dt^{\ell-1}} + \mathbf{C}_{2,r} \frac{d^{\ell-2} \mathbf{x}_r}{dt^{\ell-2}} + \cdots + \mathbf{C}_{\ell,r} \mathbf{x}_r(t) \end{aligned} \quad (5.2)$$

where $\mathbf{A}_{i,r} \in \mathbb{R}^{r \times r}$, for $i = 0, \dots, \ell$, $\mathbf{B}_r \in \mathbb{R}^{r \times m}$ and $\mathbf{C}_{i,r} \in \mathbb{R}^{p \times r}$ for $i = 1, \dots, \ell$.

Another example where the structure of a dynamical system presents an obstacle to reduction using methods that depend on availability of a standard first-order form is the class of delay differential equations. Consider a linear dynamical system with an internal delay, given in state space form as:

$$\mathbf{E} \dot{\mathbf{x}}(t) = \mathbf{A}_0 \mathbf{x}(t) + \mathbf{A}_1 \mathbf{x}(t - \tau) + \mathbf{B} \mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{C} \mathbf{x}(t) \quad (5.3)$$

with $\tau > 0$, and $\mathbf{E}, \mathbf{A}_0, \mathbf{A}_1 \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$ and $\mathbf{C} \in \mathbb{R}^{p \times n}$. The system in (5.3) is not associated with a rational transfer function due to the delay term; it is intrinsically of infinite order. Preserving the delay structure in the reduced model is crucial for accurate representation and so, we seek a reduced model of the form

$$\mathbf{E}_r \dot{\mathbf{x}}_r(t) = \mathbf{A}_{0,r} \mathbf{x}_r(t) + \mathbf{A}_{1,r} \mathbf{x}_r(t - \tau) + \mathbf{B}_r \mathbf{u}(t), \quad \mathbf{y}_r(t) = \mathbf{C}_r \mathbf{x}_r(t) \quad (5.4)$$

with $\tau > 0$, and $\mathbf{E}_r, \mathbf{A}_{0,r}, \mathbf{A}_{1,r} \in \mathbb{R}^{r \times r}$, $\mathbf{B}_r \in \mathbb{R}^{r \times m}$ and $\mathbf{C}_r \in \mathbb{R}^{p \times r}$. We want to perform this reduction step without the need for approximating the delay term with an additional rational approximation.

5.2. Generalized Coprime Representation

Examples such as these lead us to consider transfer functions having the following *Generalized Coprime Representation*:

$$\mathcal{H}(s) = \mathcal{C}(s) \mathcal{K}(s)^{-1} \mathcal{B}(s) + \mathcal{D} \quad (5.5)$$

where \mathcal{D} is a constant $p \times m$ matrix, 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 halfplane. Note that both (5.1) and (5.3) fit this framework: for the polynomial system (5.2), we obtain

$$\mathcal{H}(s) = \underbrace{\left(\sum_{i=1}^{\ell} s^{\ell-i} \mathbf{C}_i \right)}_{\mathcal{C}(s)} \underbrace{\left(\sum_{i=0}^{\ell} s^{\ell-i} \mathbf{A}_i \right)^{-1}}_{\mathcal{K}(s)^{-1}} \underbrace{\mathbf{B}}_{\mathcal{B}(s)}, \quad (5.6)$$

and for the delay system (5.3), we have

$$\mathcal{H}(s) = \underbrace{\mathbf{C}}_{\mathcal{C}(s)} \underbrace{(s \mathbf{E} - \mathbf{A}_0 - e^{-\tau s} \mathbf{A}_1)^{-1}}_{\mathcal{K}(s)^{-1}} \underbrace{\mathbf{B}}_{\mathcal{B}(s)}. \quad (5.7)$$

Our model reduction goals are the same: Construct a reduced (generalized coprime) transfer function that tangentially interpolates the original one. To this end, we choose two model reduction bases $\mathcal{V} \in \mathbb{R}^{n \times r}$ and $\mathcal{W} \in \mathbb{R}^{n \times r}$ as before. This leads to a reduced transfer function

$$\mathcal{H}_r(s) = \mathcal{C}_r(s) \mathcal{K}_r(s)^{-1} \mathcal{B}_r(s) + \mathcal{D}_r \quad (5.8)$$

where $\mathcal{C}_r(s) \in \mathbb{C}^{p \times r}$, $\mathcal{B}_r(s) \in \mathbb{C}^{r \times m}$, $\mathcal{D}_r \in \mathbb{C}^{p \times m}$ and $\mathcal{K}_r(s) \in \mathbb{C}^{r \times r}$ are obtained by Petrov-Galerkin projection:

$$\begin{aligned} \mathcal{K}_r(s) &= \mathcal{W}^T \mathcal{K}(s) \mathcal{V}, & \mathcal{B}_r(s) &= \mathcal{W}^T \mathcal{B}(s), \\ \mathcal{C}_r(s) &= \mathcal{C}(s) \mathcal{V}, & \text{and } \mathcal{D}_r &= \mathcal{D}. \end{aligned} \quad (5.9)$$

5.3. Interpolatory Projections for Generalized Coprime Factors

Interpolatory projections for generalized coprime representations of transfer functions were introduced in [17]. We follow the notation in [17] and use $\mathcal{D}_\sigma^\ell f$ to denote the ℓ^{th} derivative of the univariate function $f(s)$ evaluated at $s = \sigma$, with the usual convention that $\mathcal{D}_\sigma^0 f = f(\sigma)$.

Theorem 5.1. *Given the original model transfer function*

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

let let $\mathcal{H}_r(s)$ denote the reduced transfer function in (5.8) obtained by projection as in (5.9) using the model reduction bases \mathcal{V} and \mathcal{W} . For the interpolation points $\sigma, \mu \in \mathbb{C}$, 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 $\mathcal{K}(\sigma)$ and $\mathcal{K}(\mu)$ have full rank. Also, let $\mathbf{r} \in \mathbb{C}^m$ and $\mathbf{l} \in \mathbb{C}^\ell$ be nontrivial tangential direction vectors. The following implications hold:

(a) If

$$\mathcal{D}_\sigma^i [\mathcal{K}(s)^{-1} \mathcal{B}(s)] \mathbf{r} \in \text{Ran}(\mathcal{V}) \quad \text{for } i = 0, \dots, N, \quad (5.10)$$

then

$$\mathcal{H}^{(\ell)}(\sigma) \mathbf{r} = \mathcal{H}_r^{(\ell)}(\sigma) \mathbf{r}, \quad \text{for } \ell = 0, \dots, N. \quad (5.11)$$

(b) If

$$\left(\mathbf{l}^T \mathcal{D}_\mu^j [\mathcal{C}(s) \mathcal{K}(s)^{-1}] \right)^T \in \text{Ran}(\mathcal{W}) \quad \text{for } j = 0, \dots, M, \quad (5.12)$$

then

$$\mathbf{l}^T \mathcal{H}^{(\ell)}(\mu) = \mathbf{l}^T \mathcal{H}_r^{(\ell)}(\mu) \quad \text{for } \ell = 0, \dots, M. \quad (5.13)$$

(c) If both (5.10) and (5.12) hold and if $\sigma = \mu$, then

$$\mathbf{1}^T \mathcal{H}^{(\ell)}(\sigma) \mathbf{r} = \mathbf{1}^T \mathcal{H}_r^{(\ell)}(\sigma) \mathbf{r}, \quad \text{for } \ell = 0, \dots, M + N - 1. \quad (5.14)$$

assuming $\mathcal{K}_r(\sigma) = \mathbf{W}^T \mathcal{K}(\sigma) \mathbf{V}$, and $\mathcal{K}_r(\mu) = \mathbf{W}^T \mathcal{K}(\mu) \mathbf{V}$ have full rank.

Theorem 5.1 proves the power and flexibility of the interpolatory framework for model reduction. The earlier interpolation result, Theorem 3.2, directly extends to this much more general class of transfer function, requiring very similar subspace conditions. Moreover, this structure guarantees that the reduced transfer function will have a similar generalized coprime representation. Computational complexity is comparable; one need only solve $n \times n$ (often sparse) linear systems.

Recall the delay example (5.3). Assume that $r = 2$ interpolation points $\{\sigma_1, \sigma_2\}$ together with the right-directions $\{\mathbf{r}_1, \mathbf{r}_2\}$ and the left-directions $\{\mathbf{l}_1, \mathbf{l}_2\}$ are given. Based on Theorem 5.1, we construct $\mathbf{V} \in \mathbb{C}^{n \times 2}$ and $\mathbf{W} \in \mathbb{C}^{n \times 2}$ using

$$\mathbf{V} = \left[(\sigma_1 \mathbf{E} - \mathbf{A}_0 - e^{-\tau\sigma_1} \mathbf{A}_1)^{-1} \mathbf{B} \mathbf{r}_1 \quad (\sigma_2 \mathbf{E} - \mathbf{A}_0 - e^{-\tau\sigma_2} \mathbf{A}_1)^{-1} \mathbf{B} \mathbf{r}_2 \right]$$

and

$$\mathbf{W} = \left[(\sigma_1 \mathbf{E} - \mathbf{A}_0 - e^{-\tau\sigma_1} \mathbf{A}_1)^{-T} \mathbf{C}^T \mathbf{l}_1 \quad (\sigma_2 \mathbf{E} - \mathbf{A}_0 - e^{-\tau\sigma_2} \mathbf{A}_1)^{-T} \mathbf{C}^T \mathbf{l}_2 \right].$$

Notice, first of all, that structure is preserved: The reduced model of dimension-2 has the same internal delay structure

$$\begin{aligned} \mathbf{W}^T \mathbf{E} \mathbf{V} \dot{\mathbf{x}}_r(t) &= \mathbf{W}^T \mathbf{A}_0 \mathbf{V} \mathbf{x}_r(t) + \mathbf{W}^T \mathbf{A}_1 \mathbf{V} \mathbf{x}_r(t - \tau) + \mathbf{W}^T \mathbf{B} \mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C} \mathbf{V} \mathbf{x}_r(t) \end{aligned} \quad (5.15)$$

with a correspondingly structured transfer function

$$\mathcal{H}_r(s) = \mathbf{C} \mathbf{V} \left(s \mathbf{W}^T \mathbf{E} \mathbf{V} - \mathbf{W}^T \mathbf{A}_0 \mathbf{V} - e^{-\tau s} \mathbf{W}^T \mathbf{A}_1 \mathbf{V} \right)^{-1} \mathbf{W}^T \mathbf{B}.$$

Moreover, due to the interpolation-based construction of \mathbf{V} and \mathbf{W} , the reduced transfer function is a Hermite bitangential interpolant:

$$\mathcal{H}(\sigma_i) \mathbf{r}_i = \mathcal{H}_r(\sigma_i) \mathbf{r}_i, \quad \mathbf{l}_i^T \mathcal{H}(\sigma_i) = \mathbf{l}_i^T \mathcal{H}_r(\sigma_i) \quad \text{and} \quad \mathbf{l}_i^T \mathcal{H}'(\sigma_i) \mathbf{r}_i = \mathbf{l}_i^T \mathcal{H}'_r(\sigma_i) \mathbf{r}_i$$

for $i = 1, 2$. Note that the reduced transfer function fully incorporates the delay structure and *exactly* interpolates the original transfer function. This would not be true if the delay term $e^{-\tau s}$ had been approximated via a rational approximation, such as Padé approximation, as is commonly done while seeking to convert to a standard first-order form.

Remark 5.2. The construction of rational interpolants $\mathcal{H}_r(s) = \mathbf{C}_r(s) \mathcal{K}_r(s)^{-1} \mathbf{B}_r(s) + \mathcal{D}_r$ with $\mathcal{D}_r \neq \mathcal{D}$ can be achieved for generalized coprime representations similarly as described in Theorem 3.3 for standard first-order realizations. For details, we refer the reader to the original source [17].

6. Realization Independent Optimal \mathcal{H}_2 Approximation

We described IRKA in Section 4.2 and promoted it as an effective tool for constructing at modest cost locally optimal rational \mathcal{H}_2 approximations. One may observe, however, that the formulation of IRKA, as it appears in Algorithm 4.1, assumes that a first-order realization for $\mathbf{H}(s)$ is available: $\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B}$. As we found in Section 5, there are several important examples where the original transfer function, $\mathbf{H}(s)$, is not naturally represented in this way. In order to address these

situations among others, Beattie and Gugercin in [18] removed the need for *any* particular realization and extended applicability of IRKA to any evaluable \mathcal{H}_2 -transfer function. We focus on this extension of IRKA in the present section. Since $\mathbf{H}(s)$ is not required to have a first-order realization here, we will follow the notation of Section 5 and use $\mathfrak{H}(s)$ to denote the original transfer function, however we do not require even that the original system have a generalized coprime representation. Since the reduced model is still a rational function of order r in the first-order form, we continue to use $\mathbf{H}_r(s)$ to denote the reduced model transfer function.

There are two main observations behind the methodology offered in [18]. The first observation is based on the first-order \mathcal{H}_2 optimality conditions (4.3) in Theorem 4.1. Recall that Theorem 4.1 does not put any restrictions on $\mathfrak{H}(s)$; the only assumption is that the approximant $\mathbf{H}_r(s)$ is a rational function; thus the theorem and the bitangential Hermite optimality conditions apply equally if $\mathfrak{H}(s)$ were to have the form, for example, $\mathfrak{H}(s) = \mathbf{C}(s^2\mathbf{M} + s\mathbf{G} + \mathbf{K})^{-1}\mathbf{B}$. A second observation is related to how IRKA constructs the solution: For the current set of interpolation points and tangent directions, IRKA constructs a bitangential Hermite interpolant and updates the interpolation data. Thus, the key issue becomes, given a set of interpolation data, how shall one construct a rational approximant $\mathbf{H}_r(s)$ that is a Hermite bitangential interpolant to $\mathfrak{H}(s)$ (which may not be presented as a first-order state-space model). The Loewner interpolatory framework introduced by Mayo and Antoulas [72] (discussed in detail in Chapter ??) is the right tool.

For the balance of this section, we assume that both $\mathfrak{H}(s)$ and its derivative, $\mathfrak{H}'(s)$, are only accessible through the evaluation, $s \mapsto (\mathfrak{H}(s), \mathfrak{H}'(s))$. No particular system realizations are assumed. Suppose we are given interpolation points $\{\sigma_1, \dots, \sigma_r\}$ with the corresponding tangential directions $\{\mathbf{r}_1, \dots, \mathbf{r}_r\}$ and $\{\mathbf{l}_1, \dots, \mathbf{l}_r\}$. We want to construct a degree- r rational approximant $\mathbf{H}_r(s)$ that is a bitangential Hermite interpolant to $\mathfrak{H}(s)$:

$$\mathfrak{H}(\sigma_k)\mathbf{r}_k = \mathbf{H}_r(\sigma_k)\mathbf{r}_k, \quad (6.1a)$$

$$\mathbf{r}_k^T \mathfrak{H}(\sigma_k) = \mathbf{r}_k^T \mathbf{H}_r(\sigma_k), \text{ and} \quad (6.1b)$$

$$\mathbf{r}_k^T \mathfrak{H}'(\sigma_k)\mathbf{r}_k = \mathbf{r}_k^T \mathbf{H}'_r(\sigma_k)\mathbf{r}_k \quad (6.1c)$$

for $k = 1, 2, \dots, r$. As seen in Chapter ??, the framework of [72] allows one to achieve this goal requiring only the evaluation $\mathfrak{H}(s)$ and $\mathfrak{H}'(s)$ at σ_k without any constraint on the structure of $\mathfrak{H}(s)$: Simply construct the matrices \mathbf{E}_r , \mathbf{A}_r , \mathbf{B}_r and \mathbf{C}_r using

$$(\mathbf{E}_r)_{i,j} := \begin{cases} -\frac{\mathbf{l}_i^T (\mathfrak{H}(\sigma_i) - \mathfrak{H}(\sigma_j)) \mathbf{r}_j}{\sigma_i - \sigma_j} & \text{if } i \neq j \\ -\mathbf{l}_i^T \mathfrak{H}'(\sigma_i) \mathbf{r}_i & \text{if } i = j \end{cases}, \quad (6.2)$$

$$(\mathbf{A}_r)_{i,j} := \begin{cases} -\frac{\mathbf{l}_i^T (\sigma_i \mathfrak{H}(\sigma_i) - \sigma_j \mathfrak{H}(\sigma_j)) \mathbf{r}_j}{\sigma_i - \sigma_j} & \text{if } i \neq j \\ -\mathbf{l}_i^T [s\mathfrak{H}(s)]' |_{s=\sigma_i} \mathbf{r}_i & \text{if } i = j \end{cases}, \quad (6.3)$$

and

$$\mathbf{C}_r = [\mathfrak{H}(\sigma_1)\mathbf{r}_1, \dots, \mathfrak{H}(\sigma_r)\mathbf{r}_r], \quad \mathbf{B}_r = \begin{bmatrix} \mathbf{l}_1^T \mathfrak{H}(\sigma_1) \\ \vdots \\ \mathbf{l}_r^T \mathfrak{H}(\sigma_r) \end{bmatrix}. \quad (6.4)$$

Then $\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r$ satisfies (6.1).

To use IRKA for \mathcal{H}_2 approximation without any structural constraints on $\mathcal{H}(s)$, one need only replace the projection-based construction of the intermediate Hermite interpolant with a Loewner-based construction. This is exactly what [18] introduced, leading to the following realization-independent optimal \mathcal{H}_2 approximation methodology:

Algorithm 6.1. TF-IRKA: IRKA using transfer function evaluations

1. *Make an initial r -fold shift selection: $\{\sigma_1, \dots, \sigma_r\}$ that is closed under conjugation (i.e., $\{\sigma_1, \dots, \sigma_r\} \equiv \{\bar{\sigma}_1, \dots, \bar{\sigma}_r\}$ viewed as sets) and initial tangent directions $\mathbf{r}_1, \dots, \mathbf{r}_r$ and $\mathbf{l}_1, \dots, \mathbf{l}_r$, also closed under conjugation.*
2. *while (not converged)*
 - (a) *Construct $\mathbf{E}_r, \mathbf{A}_r, \mathbf{C}_r$ and \mathbf{B}_r as in (6.2)-(6.4).*
 - (b) *Compute a pole-residue expansion of $\mathbf{H}_r(s)$:*

$$\mathbf{H}_r(s) = \mathbf{C}_r(s\mathbf{E}_r - \mathbf{A}_r)^{-1}\mathbf{B}_r = \sum_{i=1}^r \frac{\mathbf{l}_i \mathbf{r}_i^T}{s - \lambda_i}$$

- (c) $\sigma_i \leftarrow -\lambda_i, \quad \mathbf{r}_i \leftarrow \hat{\mathbf{r}}_i, \quad \text{and} \quad \mathbf{l}_i \leftarrow \hat{\mathbf{l}}_i, \quad \text{for} \quad i = 1, \dots, r$
3. *Construct $\mathbf{E}_r, \mathbf{A}_r, \mathbf{C}_r$ and \mathbf{B}_r as in (6.2)-(6.4).*

As for the original formulation of IRKA, upon convergence the rational approximant resulting from Algorithm 6.1 will satisfy the first-order necessary conditions (4.3) for \mathcal{H}_2 optimality.

6.1. An optimal rational approximation for a delay system

Consider the delay system given in (5.7), i.e.,

$$\mathcal{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A}_0 - e^{-\tau s}\mathbf{A}_1)^{-1}\mathbf{B}.$$

Following [17], we take $\mathbf{E} = \kappa \mathbf{I} + \mathbf{T}$, $\mathbf{A}_0 = \frac{3}{\tau}(\mathbf{T} - \kappa \mathbf{I})$, and $\mathbf{A}_1 = \frac{1}{\tau}(\mathbf{T} - \kappa \mathbf{I})$, for any $\kappa > 2$ and delay $\tau > 0$, where \mathbf{T} is an $n \times n$ matrix with ones on the first superdiagonal, on the first subdiagonal, at the $(1, 1)$ entry, and at the (n, n) entry. The remaining entries of \mathbf{T} are zero. We take the internal delay as $\tau = 0.1$ and a SISO system with $n = 1000$, i.e., $\mathbf{E}, \mathbf{A}_0, \mathbf{A}_1 \in \mathbb{R}^{1000 \times 1000}$, and $\mathbf{B}, \mathbf{C}^T \in \mathbb{R}^{1000 \times 1}$. Then, we use TF-IRKA as illustrated in Algorithm 6.1 to construct a degree $r = 20$ (locally) \mathcal{H}_2 optimal rational approximation $\mathbf{H}_r(s)$. TF-IRKA requires evaluating $\mathcal{H}(s)$ and $\mathcal{H}'(s)$. For this delay model, $\mathcal{H}'(s)$ is given by

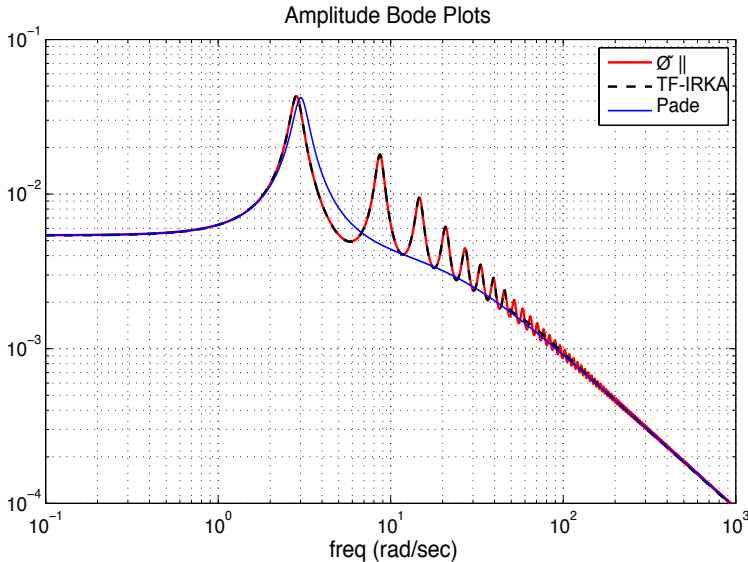
$$\mathcal{H}'(s) = -\mathbf{C}(s\mathbf{E} - \mathbf{A}_1 - e^{-\tau s}\mathbf{A}_2)^{-1}(\mathbf{E} + \tau e^{-\tau s}\mathbf{A}_2)(s\mathbf{E} - \mathbf{A}_1 - e^{-\tau s}\mathbf{A}_2)^{-1}\mathbf{B}.$$

Another approach to obtain a rational approximation for such a delay system would be to replace the exponent $e^{-\tau s}$ with a rational approximation and then reduce the resulting rational large-scale model with standard techniques. Here we will use the second order Padé approximation towards this goal where we replace $e^{-\tau s}$ by $\frac{12-6\tau s+\tau^2 s^2}{12+6\tau s+\tau^2 s^2}$, obtaining the large-scale approximate rational transfer function

$$\mathcal{H}^{[P_2]}(s) = (12\mathbf{C} + s6\tau\mathbf{C} + s^2\tau^2\mathbf{C})(\mathbf{N}s^3 + \hat{\mathbf{M}}s^2 + \hat{\mathbf{G}}s + \hat{\mathbf{K}})^{-1}\mathbf{B} \quad (6.5)$$

where $\mathbf{N} = \tau^2\mathbf{E}$, $\hat{\mathbf{M}} = 6\tau\mathbf{E} - \tau^2(\mathbf{A}_0 + \mathbf{A}_1)$, $\hat{\mathbf{G}} = 12\mathbf{E} + 6\tau(-\mathbf{A}_0 + \mathbf{A}_1)$, and $\hat{\mathbf{K}} = -12(\mathbf{A}_0 + \mathbf{A}_1)$. We used the notation $\mathcal{H}^{[P_2]}(s)$ to denote the resulting large-scale rational approximation due to the second-order Padé approximation. We note that the resulting approximation has a term s^3 and will result in an $N = 3000$ first-order model. Once an interpolatory model reduction technique is applied to

$\mathcal{H}^{[P_2]}(s)$, the reduced model will be an exact interpolant to $\mathcal{H}^{[P_2]}(s)$, but not to $\mathcal{H}(s)$. This starkly contrasts to TF-IRKA where the resulting rational approximation exactly interpolates the original delay model $\mathcal{H}(s)$. In Figure 6.1 below, we show the Amplitude Bode Plots of $\mathcal{H}(s)$ (denoted by “Full”), order $r = 20$ TF-IRKA approximant (denoted by “TF-IRKA”), and order $N = 3000$ Padé model (denoted by “Pade”). The figure illustrates clearly that the locally optimal \mathcal{H}_2 approximation due to TF-IRKA almost exactly replicates the original model. On the other hand, even the order $N = 3000$ Padé model $\mathcal{H}^{[P_2]}(s)$ is a very poor approximation.



7. Interpolatory Model Reduction of Parametric Systems

All dynamical systems considered here are *linear, time invariant* systems; the system properties are presumed to be *constant* at least with respect to time. Very often system properties will depend on external parameters and system dynamics vary as parameter values change. Parameters enter naturally into the models in various ways, representing changes in boundary conditions, material properties, system geometry, etc. Producing a new reduced model for every new set of parameter values could be very costly, so a natural goal is to generate parametrized reduced models that provide high-fidelity approximations throughout a wide range of parameter values. This is usually referred to as *parametric model reduction (PMR)*. It has found immediate applications in inverse problems [87, 49, 69, 41, 37], optimization [11, 7, 94, 95, 8], and design and control [70, 5, 36, 15, 45, 71, 62]. There are various approaches to parametric model reduction methods; see, e.g., [78, 79, 33, 75, 13, 60, 62, 85, 34] and the references therein. In this section, we focus on interpolatory methods. For a recent, detailed survey on parametric model reduction, we refer the reader to [25].

7.1. Parametric Structure

We consider MIMO transfer functions that are parametrized with ν parameters $\mathbf{p} = [p_1, \dots, p_\nu]$:

$$\mathcal{H}(s, \mathbf{p}) = \mathcal{C}(s, \mathbf{p})\mathcal{K}(s, \mathbf{p})^{-1}\mathcal{B}(s, \mathbf{p}) \quad (7.1)$$

with $\mathcal{K}(s, \mathbf{p}) \in \mathbb{C}^{n \times n}$ and $\mathcal{B}(s, \mathbf{p}) \in \mathbb{C}^{n \times m}$ and $\mathcal{C}(s, \mathbf{p}) \in \mathbb{C}^{p \times n}$. The standard case of parametric linear dynamical systems of the form

$$\mathbf{E}(\mathbf{p}) \dot{\mathbf{x}}(t; \mathbf{p}) = \mathbf{A}(\mathbf{p}) \mathbf{x}(t; \mathbf{p}) + \mathbf{B}(\mathbf{p}) \mathbf{u}(t), \quad \mathbf{y}(t; \mathbf{p}) = \mathbf{C}(\mathbf{p}) \mathbf{x}(t; \mathbf{p}) \quad (7.2)$$

then becomes a special case of the more general form (7.1) we consider here with $\mathcal{K}(s, \mathbf{p}) = s\mathbf{E}(\mathbf{p}) - \mathbf{A}(\mathbf{p})$, $\mathcal{B}(s, \mathbf{p}) = \mathbf{B}(\mathbf{p})$ and $\mathcal{C}(s, \mathbf{p}) = \mathbf{C}(\mathbf{p})$.

Even though the theoretical discussion applies to general parametric dependency, we assume an affine parametric form:

$$\begin{aligned} \mathcal{K}(s, \mathbf{p}) &= \mathcal{K}^{[0]}(s) + k_1(\mathbf{p})\mathcal{K}^{[1]}(s) + \dots + k_q(\mathbf{p})\mathcal{K}^{[\nu]}(s) \\ \mathcal{B}(s, \mathbf{p}) &= \mathcal{B}^{[0]}(s) + b_1(\mathbf{p})\mathcal{B}^{[1]}(s) + \dots + b_q(\mathbf{p})\mathcal{B}^{[\nu]}(s), \\ \mathcal{C}(s, \mathbf{p}) &= \mathcal{C}^{[0]}(s) + c_1(\mathbf{p})\mathcal{C}^{[1]}(s) + \dots + c_q(\mathbf{p})\mathcal{C}^{[\nu]}(s), \end{aligned} \quad (7.3)$$

where $\{k_i(\mathbf{p})\}$, $\{b_i(\mathbf{p})\}$, and $\{c_i(\mathbf{p})\}$ for $i = 1, \dots, q$ are scalar-valued nonlinear (or linear) parameter functions. Even though we have linear dynamics with respect to the state-variable, we allow nonlinear parametric dependency in the state-space representation.

Our reduction framework remains the same: Use a Petrov-Galerkin projection to construct, in this case, a reduced parametric model. Thus we will pick two model reduction bases $\mathcal{V} \in \mathbb{C}^{n \times r}$ and $\mathcal{W} \in \mathbb{C}^{n \times r}$ and obtain the reduced parametric model

$$\mathcal{H}_r(s, \mathbf{p}) = \mathcal{C}_r(s, \mathbf{p})\mathcal{K}_r(s, \mathbf{p})^{-1}\mathcal{B}_r(s, \mathbf{p}) \quad (7.4)$$

where we use a Petrov-Galerkin projection to obtain the reduced quantities $\mathcal{C}_r(s, \mathbf{p}) \in \mathbb{C}^{p \times r}$, $\mathcal{B}_r(s, \mathbf{p}) \in \mathbb{C}^{r \times m}$, and $\mathcal{K}_r(s, \mathbf{p}) \in \mathbb{C}^{r \times r}$, i.e.,

$$\mathcal{C}_r(s, \mathbf{p}) = \mathcal{C}(s, \mathbf{p})\mathcal{V}, \quad \mathcal{B}_r(s, \mathbf{p}) = \mathcal{W}^T\mathcal{B}(s, \mathbf{p}), \quad \mathcal{K}_r(s, \mathbf{p}) = \mathcal{W}^T\mathcal{K}(s, \mathbf{p})\mathcal{V}. \quad (7.5)$$

Applying (7.5) to the affine parametric structure (7.3) yields

$$\mathcal{K}_r(s, \mathbf{p}) = \mathcal{W}^T\mathcal{K}^{[0]}(s)\mathcal{V} + \sum_{i=1}^{\nu} k_i(\mathbf{p})\mathcal{W}^T\mathcal{K}^{[i]}(s)\mathcal{V}, \quad (7.6)$$

$$\mathcal{B}_r(s, \mathbf{p}) = \mathcal{W}^T\mathcal{B}^{[0]}(s) + \sum_{i=1}^{\nu} b_i(\mathbf{p})\mathcal{W}^T\mathcal{B}^{[i]}(s), \quad (7.7)$$

$$\mathcal{C}_r(s, \mathbf{p}) = \mathcal{C}^{[0]}(s)\mathcal{V} + \sum_{i=1}^{\nu} c_i(\mathbf{p})\mathcal{C}^{[i]}(s)\mathcal{V}.$$

The advantages are clear. The affine structure allows fast online evaluation of the reduced model: all the reduced order coefficients matrices can be precomputed and for a new parameter value, only the scalar nonlinear parametric coefficients need to be recomputed. No operation in the original dimension n is required.

7.2. Interpolatory Projections for Parametric Model Reduction

The question we want to answer in this section is how to choose \mathcal{W} and \mathcal{V} so that the reduced parametric model interpolates the original one. The main difference from the earlier cases is that we now have two variables with which to interpolate, namely the frequency variable $s \in \mathbb{C}$ and the parameter vector $\mathbf{p} \in \mathbb{R}^\nu$. Thus, we will require $\mathcal{H}_r(s, \mathbf{p})$ to (tangentially) interpolate $\mathcal{H}(s, \mathbf{p})$ at selected s and \mathbf{p}

values. In particular, this will require choosing both frequency interpolation points and parameter interpolation points.

Interpolatory parametric model reduction has been studied in various papers; see, e.g., [88, 59, 29, 36, 45, 68, 74, 42, 44]. These papers focus on matrix interpolation (as opposed to tangential interpolation) and in some cases are restricted to special cases where parametric dependence is allowed only within a subset of state-space matrices. See Chapter ?? for a careful of comparison of various parameterized model reduction strategies. Baur *et al.* [14] provide quite a general projection-based framework for approaching structure-preserving parametric model reduction via tangential interpolation. Our discussion below follows [14] closely. However, we note that instead of the standard first-order framework (7.2) that is considered there, we present results for more general *parametrized* generalized coprime representations as in (7.1). To keep the presentation concise, we only list the zeroth and first order interpolation conditions:

Theorem 7.1. *Given $\mathcal{H}(s, \mathbf{p}) = \mathcal{C}(s, \mathbf{p})\mathcal{K}(s, \mathbf{p})^{-1}\mathcal{B}(s, \mathbf{p})$, let $\mathcal{H}_r(s, \mathbf{p})$ denote the reduced transfer function in (7.4) obtained by projection as in (7.5) using the model reduction bases \mathcal{V} and \mathcal{W} . For the frequency interpolation points $\sigma, \mu \in \mathbb{C}$ and the parameter interpolation point $\boldsymbol{\pi} \in \mathbb{R}^{\nu}$, suppose that $\mathcal{B}(s, \mathbf{p})$, $\mathcal{C}(s, \mathbf{p})$, and $\mathcal{K}(s, \mathbf{p})$ are analytic with respect to s at and $\mu \in \mathbb{C}$, and are continuously differentiable with respect to \mathbf{p} in a neighborhood of $\boldsymbol{\pi}$. Also let $\mathcal{K}(\sigma, \boldsymbol{\pi})$ and $\mathcal{K}(\mu, \boldsymbol{\pi})$ have full rank. Also, let $\mathbf{r} \in \mathbb{C}^m$ and $\mathbf{l} \in \mathbb{C}^{\ell}$ be the nontrivial tangential directions vectors. Then,*

(a) *If*

$$\mathcal{K}(\sigma, \boldsymbol{\pi})^{-1}\mathcal{B}(\sigma, \boldsymbol{\pi})\mathbf{r} \in \text{Ran}(\mathcal{V}) \quad (7.8)$$

then

$$\mathcal{H}(\sigma, \boldsymbol{\pi})\mathbf{r} = \mathcal{H}_r(\sigma, \boldsymbol{\pi})\mathbf{r}. \quad (7.9)$$

(b) *If*

$$\left(\mathbf{l}^T \mathcal{C}(\mu, \boldsymbol{\pi}) \mathcal{K}(\mu, \boldsymbol{\pi})^{-1}\right)^T \in \text{Ran}(\mathcal{W}), \quad (7.10)$$

then

$$\mathbf{l}^T \mathcal{H}(\mu, \boldsymbol{\pi}) = \mathbf{l}^T \mathcal{H}_r(\mu, \boldsymbol{\pi}). \quad (7.11)$$

(c) *If both (7.8) and (7.10) hold and if $\sigma = \mu$, then*

$$\mathbf{l}^T \mathcal{H}'(\sigma, \boldsymbol{\pi})\mathbf{r} = \mathbf{l}^T \mathcal{H}'_r(\sigma, \boldsymbol{\pi})\mathbf{r} \quad (7.12)$$

and

$$\nabla_{\mathbf{p}} \mathbf{l}^T \mathcal{H}(\sigma, \boldsymbol{\pi})\mathbf{r} = \nabla_{\mathbf{p}} \mathbf{l}^T \mathcal{H}_r(\sigma, \boldsymbol{\pi})\mathbf{r} \quad (7.13)$$

assuming $\mathcal{K}_r(\sigma, \boldsymbol{\pi}) = \mathcal{W}^T \mathcal{K}(\sigma, \boldsymbol{\pi}) \mathcal{V}$, and $\mathcal{K}_r(\mu, \boldsymbol{\pi}) = \mathcal{W}^T \mathcal{K}(\mu, \boldsymbol{\pi}) \mathcal{V}$ have full rank.

Once again, the basic interpolatory projection theorem extends directly to a more general setting, in this case to the reduction of parametric systems. Possibly the most important property here is that, as (7.13) shows, interpolatory projection provides matching the parameter sensitivity without ever computing them, i.e. the subspaces \mathcal{V} and \mathcal{W} do not contain any information about the parameter sensitivity. Nonetheless, the two-sided projection forces a match with this quantity. Indeed, the Hessian with respect to the parameter vector can be matched similarly by adding more vectors to the subspace; see [14] for details.

A simple example. Consider a mass-spring-damper system where two masses m_1 and m_2 are connected with a spring-dashpot pair with spring constant k_2 and the damping constant p_2 . Further assume that the mass m_1 is connected to ground by another spring-dashpot pair with spring constant k_1 and the damping constant p_1 . Also, suppose that a point external force $u(t)$ is applied to m_1 and we are interested in the displacement of the mass m_2 . Let the state-vector $\mathbf{x}(t) = [\mathbf{x}_1(t) \ \mathbf{x}_2(t)]$ consists of the displacements of both masses. Then, the corresponding differential equation is given by

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

where $\mathbf{b} = [1 \ 0]^T$, $\mathbf{c} = [0 \ 1]$,

$$\mathbf{M} = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} p_1 + p_2 & -p_2 \\ -p_2 & p_2 \end{bmatrix}, \quad \text{and} \quad \mathbf{K} = \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix}.$$

Let $m_1 = m_2 = 1$, $k_1 = 2$, $k_2 = 2$. Also let the damping constants be parametric and vary as $p_1 \in [0.15, 0.25]$ and $p_2 \in [0.25, 0.35]$. Define the parameter vector $\mathbf{p} = [p_1 \ p_2]^T$. Then, the damping matrix can be written as

$$\mathbf{G}(\mathbf{p}) = p_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + p_2 \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = p_1 \mathcal{K}_1 + p_2 \mathcal{K}_2.$$

Then, the underlying system becomes a parametric dynamical system with a transfer function of the form (7.1), i.e. $\mathcal{H}(s, \mathbf{p}) = \mathcal{C}(s, \mathbf{p})\mathcal{K}(s, \mathbf{p})^{-1}\mathcal{B}(s, \mathbf{p})$ with

$$\mathcal{K}(s, \mathbf{p}) = \underbrace{s^2 \mathbf{M} + \mathbf{K}}_{\mathcal{K}^{[0]}(s)} + \underbrace{p_1}_{k_1(\mathbf{p})} \underbrace{s \mathbf{K}_1}_{\mathcal{K}^{[1]}(s)} + \underbrace{p_2}_{k_2(\mathbf{p})} \underbrace{s \mathbf{K}_2}_{\mathcal{K}^{[2]}(s)} \quad (7.14)$$

$$\mathcal{B}(s, \mathbf{p}) = \mathbf{b} = \mathcal{B}^{[0]}(s), \quad \text{and} \quad \mathcal{C}(s, \mathbf{p}) = \mathbf{c} = \mathcal{C}^{[0]}(s). \quad (7.15)$$

We would like to construct a degree-1 parametric reduced model using the frequency interpolation point $\sigma = 1$ and the parameter interpolation vector $\boldsymbol{\pi} = [0.2 \ 0.3]^T$. Note that since the system is SISO, no direction vectors are needed. Then,

$$\begin{aligned} \mathcal{V} &= \mathcal{K}(1, \boldsymbol{\pi})^{-1} \mathcal{B}(1, \boldsymbol{\pi}) = \begin{bmatrix} 2.5661 \times 10^{-1} \\ 1.7885 \times 10^{-1} \end{bmatrix}, \\ \mathcal{W} &= \mathcal{K}(1, \boldsymbol{\pi})^{-T} \mathcal{C}(1, \boldsymbol{\pi})^T = \begin{bmatrix} 1.7885 \times 10^{-1} \\ 4.2768 \times 10^{-1} \end{bmatrix}. \end{aligned}$$

This leads to a reduced parametric model $\mathcal{H}_r(s, \mathbf{p}) = \mathcal{C}_r(s, \mathbf{p})\mathcal{K}_r(s, \mathbf{p})^{-1}\mathcal{B}_r(s, \mathbf{p})$ with

$$\begin{aligned} \mathcal{K}_r(s, \mathbf{p}) &= \left(s^2 \mathcal{W}^T \mathbf{M} \mathcal{V} + \mathcal{W}^T \mathbf{K} \mathcal{V} \right) + p_1 \left(s \mathcal{W}^T \mathbf{K}_1 \mathcal{V} \right) + p_2 \left(s \mathcal{W}^T \mathbf{K}_2 \mathcal{V} \right) \\ \mathcal{B}_r(s, \mathbf{p}) &= \mathcal{W}^T \mathbf{b}, \quad \text{and} \quad \mathcal{C}_r(s, \mathbf{p}) = \mathbf{c} \mathcal{V}. \end{aligned}$$

One can directly check that at $\sigma = 1$ and $\boldsymbol{\pi} = [0.2 \ 0.3]^T$,

$$\mathcal{H}(\sigma, \boldsymbol{\pi}) = \mathcal{H}_r(\sigma, \boldsymbol{\pi}) = 1.7885 \times 10^{-1},$$

thus (7.9) holds. Since the system is SISO (7.9) and (7.11) are equivalent. Note that

$$\mathcal{H}'(s, \mathbf{p}) = -\mathbf{c} \mathcal{K}(s, \mathbf{p})^{-1} (2s \mathbf{M} + p_1 \mathbf{K}_1 + p_2 \mathbf{K}_2) \mathcal{K}(s, \mathbf{p})^{-1} \mathbf{b}$$

and similarly for $\mathcal{H}'_r(s, \mathbf{p})$. Then, by substituting $s = \sigma = 1$ and $\mathbf{p} = \boldsymbol{\pi} = [0.2 \ 0.3]^T$, we obtain

$$\mathcal{H}'(\sigma, \boldsymbol{\pi}) = \mathcal{H}'_r(\sigma, \boldsymbol{\pi}) = -2.4814 \times 10^{-1},$$

thus (7.12) holds. We are left with the parametric sensitivity matching condition (7.13). One can directly compute the parameter gradients as

$$\nabla_{\mathbf{p}}\mathcal{H}(s, \mathbf{p}) = \begin{bmatrix} -\mathbf{c}\mathcal{K}(s, \mathbf{p})^{-1}(s\mathbf{K}_1)\mathcal{K}(s, \mathbf{p})^{-1}\mathbf{b} \\ -\mathbf{c}\mathcal{K}(s, \mathbf{p})^{-1}(s\mathbf{K}_2)\mathcal{K}(s, \mathbf{p})^{-1}\mathbf{b} \end{bmatrix}$$

A direct computation yields that at $s = \sigma = 1$ and $\mathbf{p} = \boldsymbol{\pi} = [0.2 \ 0.3]^T$,

$$\nabla_{\mathbf{p}}\mathcal{H}(\sigma, \boldsymbol{\pi}) = \nabla_{\mathbf{p}}\mathcal{H}_r(\sigma, \boldsymbol{\pi}) = \begin{bmatrix} -4.5894 \times 10^{-2} \\ 1.9349 \times 10^{-2} \end{bmatrix}$$

As this simple example illustrates, by adding one vector to each subspace, in addition to matching the transfer function and its s -derivate, we were able to match the parameter gradients for free; once again we emphasize that no parameter gradient information was added to the subspaces. However, we still match them by employing a two-sided Petrov-Galerkin projection.

Theorem 7.1 reveals how to proceed in the case of multiple frequency and parameter interpolation points. If one is given two sets of frequency points, $\{\sigma_i\}_{i=1}^K \in \mathbb{C}$ and $\{\mu_i\}_{i=1}^K \in \mathbb{C}$, the parameter points $\{\boldsymbol{\pi}^{(j)}\}_{j=1}^L \in \mathbb{C}^q$ together with the right directions $\{\mathbf{r}_{ij}\}_{i=1, j=1}^{K, L} \in \mathbb{C}^m$ and the left directions $\{\mathbf{l}_{ij}\}_{i=1, j=1}^{K, L} \in \mathbb{C}^p$, compute

$$\mathbf{v}_{ij} = \mathcal{K}(\sigma_i, \boldsymbol{\pi}^{(j)})^{-1}\mathcal{B}(\sigma_i, \boldsymbol{\pi}^{(j)})\mathbf{r}_{i,j} \text{ and } \mathbf{w}_{ij} = \mathcal{K}(\mu_i, \boldsymbol{\pi}^{(j)})^{-T}\mathcal{C}(\mu_i, \boldsymbol{\pi}^{(j)})^T\mathbf{l}_{i,j}$$

for $i = 1, \dots, K$ and $j = 1, \dots, L$ and construct

$$\mathcal{V} = [\mathbf{v}_{11}, \dots, \mathbf{v}_{1L}, \mathbf{v}_{21}, \dots, \mathbf{v}_{2L}, \dots, \mathbf{v}_{K1}, \dots, \mathbf{v}_{KL}] \in \mathbb{C}^{n \times (KL)}$$

and

$$\mathcal{W} = [\mathbf{w}_{11}, \dots, \mathbf{w}_{1L}, \mathbf{w}_{21}, \dots, \mathbf{w}_{2L}, \dots, \mathbf{w}_{K1}, \dots, \mathbf{w}_{KL}] \in \mathbb{C}^{n \times (KL)}.$$

and apply projection as in (7.5). In practice, \mathcal{V} and \mathcal{W} might have linearly dependent columns. In these cases applying a rank-revealing QR or an SVD to remove these linearly independent columns will be necessary and will also help decrease the reduced model dimension.

Remark 7.2. We have focussed here on a global basis approach to interpolatory parametric model reduction in the sense that we assume that the reduction bases \mathcal{V} and \mathcal{W} are constant with respect to parameter variation and rich enough to carry global information for the entire parameter space. As for other parametric model reduction approaches, interpolatory model reduction can also be formulated with \mathbf{p} -dependent model reduction bases, $\mathcal{V}(\mathbf{p})$ and $\mathcal{W}(\mathbf{p})$. These parameter dependent bases can be constructed in several ways, say by interpolating local bases that correspond to parameter samples $\boldsymbol{\pi}^{(i)}$. Such considerations are not specific to interpolatory approaches and occur in other parametric model reduction approaches where the bases might be computed via POD, Balanced Truncation, etc. Similar questions arise in how best to choose the parameter samples $\boldsymbol{\pi}^{(i)}$. This also is a general consideration for all parametric model reduction methods. The common approaches such as greedy sampling can be applied here as well. For a detailed discussion of these general issues related to parametric model reduction, we refer the reader to [25]. We mention in passing that [14] introduced an *optimal* joint parameter and frequency interpolation point selection strategy for a special case of parametric systems.

8. Conclusions

We have provided here a brief survey of interpolatory methods for model reduction of large-scale dynamical systems. In addition to a detailed discussion of basic principles for generic first-order realizations, we have presented an interpolation framework for more general system classes that include generalized coprime realizations and parameterized systems. Reduction of systems of differential algebraic equations are also discussed. An overview of optimal interpolation methods in the \mathcal{H}_2 norm including the weighted case, has also been provided.

References

- [1] M.I. Ahmad and P. Benner. Interpolatory model reduction for second order descriptor systems. In *Model Reduction of Complex Dynamical Systems 2013*, Magdeburg, Germany, December 2013.
- [2] M.I. Ahmad, D.B. Szyld, and M.B. van Gijzen. Preconditioned multishift bigc for \mathcal{H}_2 -optimal model reduction. Technical report, Research Report 12-06-15, Department of Mathematics, Temple University, 2012.
- [3] K. Ahuja, E. de Sturler, K. Feng, and P. Benner. Recycling BiCGSTAB with an application to parametric model order reduction. Technical Report MPIMD/13-21, Max Planck Institute, Magdeburg, November 2013.
- [4] K. Ahuja, E. de Sturler, S. Gugercin, and E.R. Chang. Recycling BiCG with an application to model reduction. *SIAM Journal on Scientific Computing*, 34(4):A1925–A1949, 2012.
- [5] D. Amsallem and C. Farhat. Interpolation method for the adaptation of reduced-order models to parameter changes and its application to aeroelasticity. *AIAA Journal*, 46:1803–1813, July 2008.
- [6] B. Anić, C. Beattie, S. Gugercin, and A.C. Antoulas. Interpolatory weighted- \mathcal{H}_2 model reduction. *Automatica*, 2013.
- [7] H. Antil, M. Heinkenschloss, and R. H. W. Hoppe. Domain decomposition and balanced truncation model reduction for shape optimization of the Stokes system. *Optimization Methods and Software*, 26(4–5):643–669, 2011.
- [8] H. Antil, M. Heinkenschloss, R. H. W. Hoppe, C. Linsenmann, and A. Wixforth. Reduced order modeling based shape optimization of surface acoustic wave driven microfluidic biochips. *Mathematics and Computers in Simulation*, 82(10):1986–2003, 2012.
- [9] A.C. Antoulas. *Approximation of Large-Scale Dynamical Systems (Advances in Design and Control)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2005.
- [10] A.C. Antoulas, C.A. Beattie, and S. Gugercin. Interpolatory model reduction of large-scale dynamical systems. In J. Mohammadpour and K. Grigoriadis, editors, *Efficient Modeling and Control of Large-Scale Systems*, pages 2–58. Springer-Verlag, 2010.
- [11] E. Arian, M. Fahl, and E. Sachs. Trust-region proper orthogonal decomposition models by optimization methods. In *Proceedings of the 41st IEEE Conference on Decision and Control*, pages 3300–3305, Las Vegas, NV, 2002.
- [12] Z. Bai. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics*, 43(1–2):9–44, 2002.

- [13] U. Baur and P. Benner. Model reduction for parametric systems using balanced truncation and interpolation. *at-Automatisierungstechnik*, 57(8):411–420, 2009.
- [14] U. Baur, P. Benner, C.A. Beattie, and S. Gugercin. Interpolatory projection methods for parameterized model reduction. *SIAM Journal on Scientific Computing*, 33:2489–2518, 2011.
- [15] U. Baur, P. Benner, A. Greiner, J.G. Korvink, J. Lienemann, and C. Moosmann. Parameter preserving model order reduction for MEMS applications. *Mathematical and Computer Modelling of Dynamical Systems*, 17(4):297–317, 2011.
- [16] C. Beattie and S. Gugercin. Inexact solves in Krylov-based model reduction. In *Decision and Control, 2006 45th IEEE Conference on*, pages 3405–3411. IEEE, 2006.
- [17] C. Beattie and S. Gugercin. Interpolatory projection methods for structure-preserving model reduction. *Systems & Control Letters*, 58(3):225–232, 2009.
- [18] C. Beattie and S. Gugercin. Realization-independent \mathcal{H}_2 approximation. In *Proceedings of the 51st IEEE Conference on Decision & Control*, pages 4953–4958. IEEE, 2012.
- [19] C. Beattie, S. Gugercin, and S. Wyatt. Inexact solves in interpolatory model reduction. *Linear Algebra and its Applications*, 436(8):2916–2943, 2012.
- [20] C.A. Beattie, G. Flagg, and S. Gugercin. An interpolation-based approach to optimal \mathcal{H}_∞ model reduction. In *SIAM Conference on Computational Science and Engineering*, Miami, March 2009.
- [21] C.A. Beattie and S. Gugercin. Krylov-based minimization for optimal \mathcal{H}_2 model reduction. *46th IEEE Conference on Decision and Control*, pages 4385–4390, Dec. 2007.
- [22] C.A. Beattie and S. Gugercin. A trust region method for optimal \mathcal{H}_2 model reduction. *48th IEEE Conference on Decision and Control*, Dec. 2009.
- [23] P. Benner and T. Breiten. Interpolation-based \mathcal{H}_2 -model reduction of bilinear control systems. *SIAM Journal on Matrix Analysis and Applications*, 33(3):859–885, 2012.
- [24] P. Benner and L. Feng. Recycling Krylov subspaces for solving linear systems with successively changing right-hand sides arising in model reduction. In P. Benner, M. Hinze, and E.J.W. ter Maten, editors, *Model Reduction for Circuit Simulation*, pages 125–140. Springer, 2011.
- [25] P. Benner, S. Gugercin, and K. Willcox. A survey of model reduction methods for parametric systems. Technical Report MPIMD/13-14, Max Planck Institute Magdeburg, August 2013.
- [26] P. Benner, M. Hinze, and E.J.W. ter Maten, editors. *Model Reduction for Circuit Simulation*, volume 74 of *Lecture Notes in Electrical Engineering*. Springer-Verlag, Dordrecht, NL, 2011.
- [27] P. Benner, M. Köhler, and J. Saak. Sparse-dense Sylvester equations in \mathcal{H}_2 -model order reduction. Technical Report MPIMD/11-11, Max Planck Institute Magdeburg Preprints, December 2011.
- [28] P. Benner and V.I. Sokolov. Partial realization of descriptor systems. *Systems Control Lett.*, 55(11):929–938, 2006.

- [29] B. Bond and L. Daniel. Parameterized model order reduction of nonlinear dynamical systems. In *IEEE/ACM International Conference on Computer-Aided Design, 2005. ICCAD-2005*, pages 487–494, 2005.
- [30] J. Borggaard, E. Cliff, and S. Gugercin. Model reduction for indoor-air behavior in control design for energy-efficient buildings. In *American Control Conference (ACC), 2012*, pages 2283–2288. IEEE, 2012.
- [31] T. Breiten. A descent method for the frequency weighted \mathcal{H}_2 model reduction problem, 2013. Talk given at the Centre International de Rencontres Mathématiques, Luminy, France.
- [32] T. Breiten, C. Beattie, and S. Gugercin. Near-optimal frequency-weighted interpolatory model reduction. *arXiv preprint arXiv:1309.0136*, 2013.
- [33] T. Bui-Thanh, K. Willcox, and O. Ghattas. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM Journal on Scientific Computing*, 30(6):3270–3288, 2008.
- [34] T. Bui-Thanh, K. Willcox, and O. Ghattas. Parametric reduced-order models for probabilistic analysis of unsteady aerodynamic applications. *AIAA Journal*, 46(10):2520–2529, 2008.
- [35] A. Bunse-Gerstner, D. Kubalinska, G. Vossen, and D. Wilczek. \mathcal{H}_2 -optimal model reduction for large scale discrete dynamical MIMO systems. *Journal of Computational and Applied Mathematics*, 2009. doi:10.1016/j.cam.2008.12.029.
- [36] L. Daniel, O.C. Siong, S.C. Low, K.H. Lee, and J. White. A multiparameter moment matching model reduction approach for generating geometrically parameterized interconnect performance models. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 23(5):678–693, 2004.
- [37] E. de Sturler, S. Gugercin, M.E. Kilmer, S. Chaturantabut, C.A. Beattie, and M. O’Connell. Nonlinear parametric inversion using interpolatory model reduction. *arXiv preprint arXiv:1311.0922*, 2013.
- [38] C. De Villemagne and R. Skelton. Model reductions using a projection formulation. *International Journal of Control*, 46(6):2141–2169, 1987.
- [39] V. Druskin and V. Simoncini. Adaptive rational Krylov subspaces for large-scale dynamical systems. *Systems & Control Letters*, 60(8):546–560, 2011.
- [40] V. Druskin, V. Simoncini, and M. Zaslavsky. Adaptive tangential interpolation in rational Krylov subspaces for MIMO model reduction data. *Preprint, November*, 2012.
- [41] V. Druskin, V. Simoncini, and M. Zaslavsky. Solution of the time-domain inverse resistivity problem in the model reduction framework part i. one-dimensional problem with siso data. *SIAM Journal on Scientific Computing*, 35(3):A1621–A1640, 2013.
- [42] O. Farle, V. Hill, P. Ingelström, and R. Dyczij-Edlinger. Multi-parameter polynomial order reduction of linear finite element models. *Math. Comput. Model. Dyn. Syst.*, 14(5):421–434, 2008.
- [43] P. Feldmann and R.W. Freund. Efficient linear circuit analysis by Padé approximation via the Lanczos process. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 14:639–649, 1995.
- [44] L. Feng and P. Benner. A robust algorithm for parametric model order reduction based on implicit moment matching. *Proceedings in Applied Mathematics and Mechanics*, 7:10215.01–10215.02, 2008.

- [45] L. Feng, E. Rudnyi, and J. Korvink. Preserving the film coefficient as a parameter in the compact thermal model for fast electrothermal simulation. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 24(12):1838–1847, 2005.
- [46] G. Flagg, C. Beattie, and S. Gugercin. Convergence of the Iterative Rational Krylov Algorithm. *Systems & Control Letters*, 61(6):688–691, 2012.
- [47] Garret Flagg, Christopher A Beattie, and Serkan Gugercin. Interpolatory h model reduction. *Systems & Control Letters*, 62(7):567–574, 2013.
- [48] R. Freund. Model reduction methods based on Krylov subspaces. *Acta Numerica*, 12:267–319, 2003.
- [49] D. Galbally, K. Fidkowski, K. Willcox, and O. Ghattas. Nonlinear model reduction for uncertainty quantification in large-scale inverse problems. *International Journal for Numerical Methods in Engineering*, 81(12):1581–1608, 2010.
- [50] K. Gallivan, E.J. Grimme, and P. van Dooren. Asymptotic waveform evaluation via a Lanczos method. *Applied Mathematics Letters*, 7(5):75–80, 1994.
- [51] K. Gallivan, A. Vandendorpe, and P. van Dooren. Model reduction of MIMO systems via tangential interpolation. *SIAM Journal on Matrix Analysis and Applications*, 26(2):328–349, 2005.
- [52] E. Grimme. *Krylov Projection Methods for Model Reduction*. PhD thesis, Coordinated-Science Laboratory, University of Illinois at Urbana-Champaign, 1997.
- [53] S. Gugercin. *Projection methods for model reduction of large-scale dynamical systems*. PhD thesis, Ph. D. Dissertation, ECE Dept., Rice University, December 2002, 2002.
- [54] S. Gugercin. An iterative rational Krylov algorithm (IRKA) for optimal \mathcal{H}_2 model reduction. In *Householder Symposium XVI*, Seven Springs Mountain Resort, PA, USA, May 2005.
- [55] S. Gugercin, A. Antoulas, and C. Beattie. A rational Krylov iteration for optimal \mathcal{H}_2 model reduction. In *Proceedings of MTNS*, 2006.
- [56] S. Gugercin and A.C. Antoulas. An \mathcal{H}_2 error expression for the Lanczos procedure. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, 2003.
- [57] S. Gugercin, A.C. Antoulas, and C.A. Beattie. \mathcal{H}_2 model reduction for large-scale linear dynamical systems. *SIAM Journal on Matrix Analysis and Applications*, 30(2):609–638, 2008.
- [58] S. Gugercin, T. Stykel, and S. Wyatt. Model reduction of descriptor systems by interpolatory projection methods. *SIAM Journal on Scientific Computing*, 35(5):B1010–B1033, 2013.
- [59] P.K. Gunupudi, R. Khazaka, and M.S. Nakhla. Analysis of transmission line circuits using multidimensional model reduction techniques. *IEEE Transactions on Advanced Packaging*, 25(2):174–180, 2002.
- [60] B. Haasdonk and M. Ohlberger. Efficient reduced models and a posteriori error estimation for parametrized dynamical systems by offline/online decomposition. *Mathematical and Computer Modelling of Dynamical Systems*, 17(2):145–161, 2011.
- [61] Y. Halevi. Frequency weighted model reduction via optimal projection. *IEEE Transactions on Automatic Control*, 37(10):1537–1542, 1992.

- [62] A. Har, J.T. Borggaard, and D. Pelletier. Local improvements to reduced-order models using sensitivity analysis of the proper orthogonal decomposition. *Journal of Fluid Mechanics*, 629:41–72, 2009.
- [63] M. Heinkenschloss, D.C. Sorensen, and K. Sun. Balanced truncation model reduction for a class of descriptor systems with application to the oseen equations. *SIAM J. Sci. Comput.*, 30(2):1038–1063, 2008.
- [64] D. Hyland and D. Bernstein. The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore. *IEEE Transactions on Automatic Control*, 30(12):1201–1211, 1985.
- [65] A. Kellems, D. Roos, N. Xiao, and S. Cox. Low-dimensional, morphologically accurate models of subthreshold membrane potential. *Journal of Computational Neuroscience*, 27(2):161–176, 2009.
- [66] W. Krajewski, A. Lepschy, M. Redivo-Zaglia, and U. Viaro. A program for solving the \mathcal{L}_2 reduced-order model problem with fixed denominator degree. *Numerical Algorithms*, 9(2):355–377, 1995.
- [67] D. Kubalinska, A. Bunse-Gerstner, G. Vossen, and D. Wilczek. \mathcal{H}_2 -optimal interpolation based model reduction for large-scale systems. In *Proceedings of the 16th International Conference on System Science*, Poland, 2007.
- [68] A.T.-M. Leung and R. Khazaka. Parametric model order reduction technique for design optimization. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1290–1293, May 2005.
- [69] C. Lieberman, K. Willcox, and O. Ghattas. Parameter and state model reduction for large-scale statistical inverse problems. *SIAM Journal on Scientific Computing*, 32(5):2523–2542, August 2010.
- [70] T. Lieu and C. Farhat. Adaptation of aeroelastic reduced-order models and application to an F-16 configuration. *AIAA Journal*, 45(6):1244–1257, 2007.
- [71] T. Lieu, C. Farhat, and M. Lesoinne. Reduced-order fluid/structure modeling of a complete aircraft configuration. *Computer Methods in Applied Mechanics and Engineering*, 195:5730–5742, 2006.
- [72] A.J. Mayo and A.C. Antoulas. A framework for the solution of the generalized realization problem. *Linear Algebra and Its Applications*, 425(2-3):634–662, 2007.
- [73] L. Meier III and D. Luenberger. Approximation of linear constant systems. *Automatic Control, IEEE Transactions on*, 12(5):585–588, 1967.
- [74] K. Moosmann and J.G. Korvink. Automatic parametric mor for mems design. In B. Lohmann and A. Kugi, editors, *Tagungsband GMA-FA 1.30 "Modellbildung, Identifikation und Simulation in der Automatisierungstechnik"*, Workshop am Bostalsee, 27.–29.9.2006, pages 89–99, 2006.
- [75] N.C. Nguyen, A.T. Patera, and J. Peraire. A best points interpolation method for efficient approximation of parametrized functions. *International Journal for Numerical Methods in Engineering*, 73(4):521–543, 2008.
- [76] H.K.F. Panzer, S. Jaensch, T. Wolf, and B. Lohmann. A greedy rational Krylov method for \mathcal{H}_2 -pseudo-optimal model order reduction with preservation of stability. In *American Control Conference (ACC), 2013*, pages 5512–5517, 2013.
- [77] D. Petersson. *A Nonlinear Optimization Approach to \mathcal{H}_2 -Optimal Modeling and Control*. PhD thesis, Linköping University, 2013.

- [78] C. Prud'homme, D.V. Rovas, K. Veroy, L. Machiels, Y. Maday, A.T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering*, 124:70–80, 2002.
- [79] G. Rozza, D.B.P. Huynh, and A.T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: application to transport and continuum mechanics. *Archives of Computational Methods in Engineering*, 15(3):229–275, 2008.
- [80] A. Ruhe. Rational Krylov algorithms for nonsymmetric eigenvalue problems. II: matrix pair. *Linear algebra and its applications Appl.*, pages 282–295, 1984.
- [81] J.T. Spanos, M.H. Milman, and D.L. Mingori. A new algorithm for L^2 optimal model reduction. *Automatica (Journal of IFAC)*, 28(5):897–909, 1992.
- [82] T. Stykel. Low-rank iterative methods for projected generalized Lyapunov equations. *Electron. Trans. Numer. Anal.*, 30:187–202, 2008.
- [83] P. van Dooren, K. A. Gallivan, and P.-A. Absil. \mathcal{H}_2 -optimal model reduction with higher-order poles. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2738–2753, 2010.
- [84] P. van Dooren, K.A. Gallivan, and P.A. Absil. \mathcal{H}_2 -optimal model reduction of MIMO systems. *Applied Mathematics Letters*, 21(12):1267–1273, 2008.
- [85] K. Veroy, C. Prud'homme, D. Rovas, and A. Patera. A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations. In *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference*, 2003.
- [86] P. Vuillemin, C. Poussot-Vassal, and D. Alazard. A spectral expression for the frequency-limited \mathcal{H}_2 -norm. *arXiv preprint arXiv:1211.1858*, 2012.
- [87] J. Wang and N. Zabararas. Using Bayesian statistics in the estimation of heat source in radiation. *International Journal of Heat and Mass Transfer*, 48:15–29, 2005.
- [88] D. S. Weile, E. Michielssen, E. Grimme, and K. Gallivan. A method for generating rational interpolant reduced order models of two-parameter linear systems. *Appl. Math. Lett.*, 12(5):93–102, 1999.
- [89] D.A. Wilson. Optimum solution of model-reduction problem. *Proc. IEE*, 117(6):1161–1165, 1970.
- [90] S. Wyatt. *Issues in Interpolatory Model Reduction: Inexact Solves, Second-order Systems and DAEs*. PhD thesis, Virginia Polytechnic Institute and State University, 2012.
- [91] W.Y. Yan and J. Lam. An approximate approach to \mathcal{H}_2 optimal model reduction. *Automatic Control, IEEE Transactions on*, 44(7):1341–1358, 1999.
- [92] A. Yousuff and R. Skelton. Covariance equivalent realizations with applications to model reduction of large-scale systems. *Control and Dynamic Systems*, 22:273–348, 1985.
- [93] A. Yousuff, D. Wagie, and R. Skelton. Linear system approximation via covariance equivalent realizations. *Journal of mathematical analysis and applications*, 106(1):91–115, 1985.

- [94] Y. Yue and K. Meerbergen. Using Krylov-Padé model order reduction for accelerating design optimization of structures and vibrations in the frequency domain. *International Journal for Numerical Methods in Engineering*, 90(10):1207–1232, 2012.
- [95] Y. Yue and K. Meerbergen. Accelerating optimization of parametric linear systems by model order reduction. *SIAM Journal on Optimization*, 23(2):1344–1370, 2013.
- [96] D. Zigic, L. Watson, and C. Beattie. Contragredient transformations applied to the optimal projection equations. *Linear Algebra and Its Applications*, 188:665–676, 1993.

Christopher Beattie
Department of Mathematics
Virginia Tech.
Blacksburg, VA, 24061-0123
USA
e-mail: beattie@math.vt.edu

Serkan Gugercin
Department of Mathematics
Virginia Tech.
Blacksburg, VA, 24061-0123
USA
e-mail: gugercin@math.vt.edu