Interpolatory Weighted-$H_2$ Model Reduction

Branimir Anić¹, Christopher Beattie², Serkan Gugercin², and Athanasios C. Antoulas³,

²Department of Mathematics, Karlruhe Institute of Technology, Germany
³Department of Mathematics, Virginia Tech, Blacksburg, VA, 24061-0123, USA

1 Introduction

Consider a single input/single output (SISO) linear dynamical system with a realization

\[ E \dot{x}(t) = Ax(t) + b u(t), \quad y(t) = c^T x(t) \quad (1) \]

for \( E, A \in \mathbb{R}^{n \times n} \) and \( b, c \in \mathbb{R}^n \). \( x(t) \in \mathbb{R}^n \), \( u(t) \in \mathbb{R} \), \( y(t) \in \mathbb{R} \), are respectively the state, input, and output of the system. The transfer function of this system is \( G(s) = c^T(sE - A)^{-1}b \). Following common usage, the underlying system will also be denoted by \( G \). For many examples, the state-space dimension \( n \) is quite large, leading to untenable demands on computational resources. Model reduction attempts to address this by finding a reduced-order system of the form,

\[ E_r \dot{x}_r(t) = A_r x_r(t) + b_r u(t), \quad y_r(t) = c_r^T x_r(t) \quad (2) \]

with \( G_r(s) = c_r^T(sE_r - A_r)^{-1}b_r \) for \( E_r, A_r \in \mathbb{R}^{r \times r} \) and \( b_r, c_r \in \mathbb{R}^r \) with \( r \ll n \) such that \( y_r(t) \approx y(t) \) over a large class of inputs \( u(t) \). \( G_r \) is a low order, yet high fidelity, approximation to \( G \). We construct \( G_r \) via state-space projection: Two matrices (“reduction bases”) \( V_r, W_r \in \mathbb{R}^{n \times r} \) are chosen. Then, system dynamics are approximated by \( x(t) \approx V_r x_r(t) \) and forcing a Petrov-Galerkin condition (“orthogonal residuals”) \( W_r^T(E(V_r x_r(t) - AV_r x_r(t) - b_r u(t)) = 0 \), together with the output equation \( y_r(t) = c_r^T V_r x_r(t) \) to produce

\[ E_r = W_r^T E V_r, \quad b_r = W_r^T b, \quad A_r = W_r^T A V_r, \quad \text{and} \quad c_r = c_r^T V_r. \quad (3) \]

See [2,3] for more information.

1.1 Model Reduction by Interpolation

The reduction bases, \( V_r \) and \( W_r \), used in (3) will be chosen to force interpolation: \( G_r(s) \) will interpolate \( G(s) \) (possibly together with higher order derivatives) at selected interpolation points. This approach to rational interpolation has been considered in [20,21,5,8,7,3] and depends on the following result.

Theorem 1 Given two sets of interpolation points \( \{\sigma_k\}_{k=1}^r \) and \( \{\zeta_k\}_{k=1}^r \), that are each closed under conjugation, and a dynamical system \( G \) as in (1), consider matrices \( V_r \) and \( W_r \) such that

\[ \text{Range}(V_r) = \text{span} \{ (\sigma_i E - A)^{-1} b \} \quad \text{for } i = 1, \ldots, r. \quad (4) \]

\[ \text{Range}(W_r) = \text{span} \{ (\zeta_i E - A)^{-T} c \} \]

Preprint submitted to Automatica 21 December 2013
Then, $V_r$ and $W_r$ can be chosen to be real; $G_r(s) = c_r^T(sE_r - A_r)^{-1}b_r$ defined by (2)-(3) is a real dynamical system that satisfies $G(\sigma_k) = G_r(\sigma_k)$ and $G(\zeta_k) = G_r(\zeta_k)$ for $k = 1, \ldots, r$; and, if $\sigma_j = \zeta_j$ for some $j$, then $G'(\sigma_j) = G'_r(\sigma_j)$, as well where $G'$ denotes the derivative of $G(s)$ with respect to $s$.

Theorem 1 can be generalized to higher-order derivative interpolation as well, see [20,21,5,8,7,3]. The subspaces of Theorem 1 are rational Krylov subspaces and so, interpolatory model reduction methods for SISO systems are sometimes referred to as rational Krylov methods.

1.2 Weighted Model Reduction

The $H_\infty$ norm of a stable linear system associated with a transfer function, $G(s)$, is defined as $\|G\|_{H_\infty} = \max_{\omega \in \mathbb{R}} |G(i\omega)|$. The $H_2$ norm of $G$ is defined as $\|G\|_{H_2} := \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \|G(i\omega)\|^2 \, d\omega\right)^{1/2}$. The vector spaces of meromorphic functions that are analytic in the right halfplane, having either bounded $H_\infty$ norm or bounded $H_2$ norm will be denoted simply as $H_\infty$ or $H_2$, respectively. Let $W \in H_\infty$ be given. The ($W$-)weighted $H_2$ norm is defined as $\|G\|_{H_2(W)} = \|G \cdot W\|_{H_2}$. We are interested in finding a reduced-order model $G_r$ that minimizes a $W$-weighted $H_2$ norm, i.e., that solves

$$\|G - G_r\|_{H_2(W)} = \min_{\text{dim}(G_r)=r} \|G - G_r\|_{H_2(W)} \quad (5)$$

The introduction of $W(s)$ allows one to penalize the error in certain frequency ranges more heavily than in others.

An illustrative example: controller reduction

Consider a linear dynamical system, $P$ (the plant) together with an associated stabilizing controller, $G$, that is connected to $P$ in a feedback loop. Many control design methodologies, such as LQG and $H_\infty$ methods, lead ultimately to controllers whose order is generally as high as the order of the plant, see [17,22] and references therein. Thus, high-order plants generally lead to high-order controllers. However, high-order controllers are usually undesirable in real-time applications due to complex hardware, degraded accuracy, and degraded computational speed. Thus, one prefers to use a reduced controller $G_r$ to replace $G$. Requiring $G_r$ to be a good approximation to $G$ is often not enough in terms of closed-loop performance; plant dynamics need to be taken into account during the reduction process. This may be achieved through frequency weighting: Given a stabilizing controller $G$, if $G$ has the same number of unstable poles as $G_r$ and if $\|\|G - G_r\|P[I + PG]^{-1}\|_{H_\infty} < 1$, then $G_r$ will also be a stabilizing $H_2$ controller [1,22]. Hence the controller reduction problem may be formulated as finding a reduced-order controller $G_r$ that minimizes or reduces the weighted error $\|\|G - G_r\|W\|_{H_\infty}$ with $W(s) := P(s)(I + P(s)G(s))^{-1}$; i.e., controller reduction becomes an application of weighted model reduction. This approach has been considered in [17,1,14,10,6,19,12,18,16] and references therein, leading to variants of frequency-weighted balanced truncation. Conversely, the methods in [11] and [15] are tailored instead towards minimizing a weighted-$H_2$ error as in (5).

2 Weighted-$H_2$ model reduction

The methods proposed in [11] and [15] for approximating (5) require solving a sequence of large-scale Lyapunov or Riccati equations; they rapidly become computationally intractable as the system order, $n$, increases. We will approach (5) within an interpolatory model reduction framework requiring only the solution of (generally sparse) linear systems and no need for dense matrix computations or solution of large-scale Lyapunov or Riccati equations. Interpolatory approaches can be effectively applied even when $n$ reaches the tens of thousands.

2.1 A representation of the weighted-$H_2$ norm

Given transfer functions $G$, $H \in H_2$, and $W \in H_\infty$, define the weighted-$H_2$ inner product as

$$\langle G, H \rangle_{H_2(W)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(i\omega)W(i\omega)H(i\omega) \, d\omega$$

so that $\|G\|_{H_2} = \sqrt{\langle G, G \rangle_{H_2(W)}}$. The following lemma gives a compact expression for the weighted-$H_2$ inner product based on the poles and residues of $G(s)$, $H(s)$ and $W(s)$. By $\text{res}[M(s), \pi]$, we denote the residue of $M(s)$ at $\pi \in \mathbb{C}$.

**Lemma 2** Suppose $G$, $H \in H_2$ have poles denoted respectively as $\{\lambda_1, \ldots, \lambda_n\}$ and $\{\mu_1, \ldots, \mu_m\}$, and suppose $W \in H_\infty$ has poles denoted as $\{\gamma_1, \ldots, \gamma_p\}$. Assume that $H(s)$ and $W(s)$ have no common poles, and the poles of $W(s)$ are simple. Then

$$\langle G, H \rangle_{H_2(W)} = \sum_{k=1}^{m} \text{res}[G(-s)W(-s)W(s)H(s), \mu_k] + \sum_{i=1}^{p} G(-\gamma_i)W(-\gamma_i)H(\gamma_i) \cdot \text{res}[W(s), \gamma_i].$$

Define $\chi_k = \text{res}[G(-s)W(-s)W(s)H(s), \mu_k]$. If $\mu_k$ is a simple pole of $H(s)$, then

$$\chi_k = G(-\mu_k)W(-\mu_k)W(\mu_k) \cdot \text{res}[H(s), \mu_k].$$
• If $\mu_k$ is a double pole of $H(s)$, then
\[
\chi_k = G(-\mu_k)W(-\mu_k)W(\mu_k) \cdot \text{res}[H(s), \mu_k]
- h_{-2}(\mu_k) \cdot \frac{d}{ds} [G(s)W(s)W(-s)]|_{s = -\mu_k},
\]
where $h_{-2}(\mu_k) = \lim_{s \to -\mu_k} (s - \mu_k)^2 H(s)$.

**Proof:** $T(s) = G(-\lambda)W(-\lambda)W(s)H(s)$ has poles at
\[
\{-\lambda_1, \ldots, -\lambda_n\} \cup \{\pm \gamma_1, \ldots, \pm \gamma_p\} \cup \{\mu_1, \ldots, \mu_m\}.
\]
For any $R > 0$, define a semicircular contour in the left half-plane: $\Gamma_R = \{z \mid \omega \in [-R, R] \} \cup \{z \mid z = Re^{\theta} \text{ with } \theta \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]\}$. For $R$ large enough, the region bounded by $\Gamma_R$ contains $\{-\gamma_1, \ldots, -\gamma_p\} \cup \{\mu_1, \ldots, \mu_m\}$, constituting all the poles of $W(s)H(s)$, and hence all the stable poles of $G(-\lambda)W(-\lambda)W(s)H(s)$. Then, the Residue Theorem yields
\[
\langle G, H \rangle_{H_2(W)} = \frac{1}{2\pi i} \int_{-\infty}^{\infty} G(-\omega)W(-\omega)W(\omega)H(\omega) d\omega
= \lim_{R \to \infty} \left[ \frac{1}{2\pi i} \int_{\Gamma_R} G(-\omega)W(-\omega)W(s)H(s) ds \right]
= \sum_{k=1}^{p} \text{res}[G(-\omega)W(-\omega)W(s)H(s), \mu_k]
+ \sum_{i=1}^{m} \text{res}[G(\omega)W(-\omega)W(s)H(s), \gamma_i].
\]
This leads to the first assertion. If $\mu_k$ is a simple pole for $H(s)$, then
\[
\text{res}[G(-\omega)W(-\omega)W(s)H(s), \mu_k]
= \lim_{s \to -\mu_k} ([s - \mu_k]G(-\omega)W(-\omega)W(s)H(s)]
= G(-\mu_k)W(-\mu_k)W(\mu_k) \lim_{s \to -\mu_k} (s - \mu_k)H(s).
\]
Similarly, if $\mu_k$ is a double pole for $H(s)$, then it is also a double pole for $G(-\lambda)W(-\lambda)W(s)H(s)$ and
\[
\text{res}[G(-\omega)W(-\omega)W(s)H(s), \mu_k]
= \lim_{s \to -\mu_k} \frac{d}{ds} [[s - \mu_k]^2G(-\omega)W(-\omega)W(s)H(s)]
= G(-\mu_k)W(-\mu_k)W(\mu_k) \cdot \text{res}[H(s), \mu_k]
- h_{-2}(\mu_k) \cdot \frac{d}{ds} [G(s)W(s)W(-s)]|_{s = -\mu_k}.
\]

**Corollary 3** If $G(s)$ and $W(s)$ in Lemma 2 each have simple poles, then
\[
\|G\|_{H_2(W)}^2 = \sum_{k=1}^{n} G(-\lambda_k)W(-\lambda_k)W(\lambda_k) \cdot \text{res}[G(s), \lambda_k]
+ \sum_{k=1}^{p} G(-\gamma_k)W(-\gamma_k)G(\gamma_k) \cdot \text{res}[W(s), \gamma_k].
\]
This new formula (6) for the weighted-$H_2$ norm contains as a special case (with $W(s) = 1$), a similar expression for the (unweighted) $H_2$ norm introduced in [9].

Suppose $W \in H_{\infty}$ has simple poles at $\{-\gamma_1, \ldots, -\gamma_p\}$ and define a linear mapping $\tilde{\mathfrak{F}} : H_2 \to H_2$ by
\[
\tilde{\mathfrak{F}}[G](s) = G(s)W(s)W(-s) + \sum_{k=1}^{p} G(-\gamma_k)W(-\gamma_k) \cdot \text{res}[W(s), \gamma_k],
\]
Then $\tilde{\mathfrak{F}}[G](s)$ has poles only in the left half plane and indeed $\tilde{\mathfrak{F}} : H_2 \to H_2$.

**Corollary 4** Suppose $G$ and $W$ are stable with poles $\{\lambda_1, \ldots, \lambda_n\}$ and $\{\gamma_1, \ldots, \gamma_p\}$, respectively. Choose $\mu$ arbitrarily in the left half plane distinct from these points. Then for $F(s) = \tilde{\mathfrak{F}}[G](s)$, $\langle G, \frac{1}{s-\mu} \rangle_{H_2(W)} = F(-\mu)$ and $\langle G, \frac{1}{(s-\mu)^2} \rangle_{H_2(W)} = -F'(-\mu)$.

**Proof:** By Lemma 2,
\[
\langle G, \frac{1}{s-\mu} \rangle_{H_2(W)} = G(-\mu)W(-\mu)W(\mu)
+ \sum_{k=1}^{p} G(-\gamma_k)W(-\gamma_k) \frac{\text{res}[W(s), \gamma_k]}{\gamma_k - \mu} = F(-\mu),
\]
and
\[
\langle G, \frac{1}{(s-\mu)^2} \rangle_{H_2(W)} = \frac{d}{ds} [G(s)W(s)W(-s)]|_{s = -\mu}
+ \sum_{k=1}^{p} G(-\gamma_k)W(-\gamma_k) \frac{\text{res}[W(s), \gamma_k]}{(\gamma_k - \mu)^2} = -F'(-\mu).
\]

**2.2** Weighted-$H_2$ optimality conditions

Consider the problem of finding a reduced order system, $G_r$, that solves (5). The feasible set for (5) is nonconvex, so finding a true (global) minimizer is generally intractable. Nonetheless, we are able to obtain descriptive necessary conditions for $G_r$ to satisfy (5).
Theorem 5 If $G_r$ has simple poles, $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_r\}$, and solves (5), then $G_r$ must satisfy: for $k = 1, \ldots, r$,

$$F_r(-\hat{\lambda}_k) = F(-\hat{\lambda}_k) \quad \text{and} \quad F'_r(-\hat{\lambda}_k) = F'(-\hat{\lambda}_k) \quad (8)$$

where $F = \mathcal{F}[G]$ and $F_r = \mathcal{F}[G_r]$ are defined from (7).

Proof: Suppose by way of contradiction that, for some $\mu \in \{\hat{\lambda}_1, \ldots, \hat{\lambda}_r\}$, $\left\langle G - G_r, \frac{1}{(s - \mu)^2} \right\rangle_{\mathcal{H}_2(W)} = \alpha_1 \neq 0$. By hypothesis, $G_r$ can be represented as $G_r(s) = \sum_{i=1}^{r} \frac{\varphi_i}{s - \lambda_i}$ and for some index $k$, $\mu = \lambda_k$. Define $\vartheta_0 = \arg(\alpha_0)$ and with $\varepsilon > 0$, define

$$\tilde{G}_r^{(\varepsilon)}(s) = \frac{\varphi_k + \varepsilon e^{-i\theta_0}}{s - \mu} + \sum_{i \neq k} \frac{\varphi_i}{s - \lambda_i}.$$

Then

$$\|G_r - \tilde{G}_r^{(\varepsilon)}\|_{\mathcal{H}_2(W)} = \left\| \frac{-e^{-i\theta_0}}{s - \mu} \right\|_{\mathcal{H}_2(W)} \leq \|W\|_{\mathcal{H}_\infty} \frac{\varepsilon}{\sqrt{2} |\text{Re}(\mu)|}$$

so that $\|G_r(s) - \tilde{G}_r^{(\varepsilon)}(s)\|_{\mathcal{H}_2(W)} = O(\varepsilon)$ as $\varepsilon \to 0$. Since $G_r$ solves (5),

$$\|G - \tilde{G}_r^{(\varepsilon)}\|_{\mathcal{H}_2(W)}^2 \leq \|G - \tilde{G}_r^{(\varepsilon)}\|_{\mathcal{H}_2(W)}^2 + 2 \text{Re} \left\langle G - G_r, \tilde{G}_r^{(\varepsilon)} \right\rangle_{\mathcal{H}_2(W)}$$

Thus,

$$0 \leq 2 \text{Re} \left\langle G - G_r, \tilde{G}_r^{(\varepsilon)} \right\rangle_{\mathcal{H}_2(W)} + \|G_r - \tilde{G}_r^{(\varepsilon)}\|_{\mathcal{H}_2(W)}^2.$$

This implies first that $0 \leq -\varepsilon |\alpha_0| + O(\varepsilon^2)$, which then leads to a contradiction, $\alpha_0 = 0$.

To show the next assertion, suppose that for some $\mu \in \{\hat{\lambda}_1, \ldots, \hat{\lambda}_r\}$, $\left\langle G - G_r, \frac{1}{(s - \mu)^2} \right\rangle_{\mathcal{H}_2(W)} = \alpha_1 \neq 0$. Then for some $k, \mu = \lambda_k$ and we define $\vartheta_1 = \arg(\varphi_k \cdot \alpha_1)$. For $\varepsilon > 0$ sufficiently small, define

$$\tilde{G}_r^{(\varepsilon)}(s) = \frac{\varphi_k}{s - \mu + \varepsilon e^{-i\vartheta_1}} + \sum_{i \neq k} \frac{\varphi_i}{s - \lambda_i}.$$

As $\varepsilon \to 0$, we have

$$\|G_r - \tilde{G}_r^{(\varepsilon)}\|_{\mathcal{H}_2(W)} = \left\| \frac{-\varepsilon \varphi_k e^{-i\theta_1}}{(s - \mu)^2 - \varepsilon e^{-i\theta_1}} \right\|_{\mathcal{H}_2(W)} = O(\varepsilon)$$

Following a similar argument as before, we find that

$$0 \leq -|\varphi_k \cdot \alpha_1| + O(\varepsilon^2)$$

as $\varepsilon \to 0$, which leads to a contradiction, $\alpha_1 = 0$. \qed

The interpolation conditions described in (8) give first order necessary conditions for $G_r$ to solve the optimal weighted-$\mathcal{H}_2$ model reduction problem (5). Note that for $W(s) = 1$, one obtains $F(s) = G(s)$ and $F_r(s) = G_r(s)$; thus (8) contains the interpolatory $\mathcal{H}_2$ optimality conditions of [9] for the unweighted problem as a special case. Unfortunately, there does not appear to be a straightforward generalization of the corresponding computational approach that was described in [9] for the optimal (unweighted) $\mathcal{H}_2$ model reduction problem. Instead, we consider a different systematic approach to this problem motivated by an expression for the weighted-$\mathcal{H}_2$ error.

2.3 A weighted-$\mathcal{H}_2$ error expression

The weighted-$\mathcal{H}_2$ norm expression in Corollary 3 leads immediately to an expression for the weighted-$\mathcal{H}_2$ error that forms the basis for our computational approach.

Corollary 6 Suppose that $G$, $G_r$ and $W$ are stable with simple poles $\{\lambda_i\}_{i=1}^{n}$, $\{\hat{\lambda}_j\}_{j=1}^{r}$, and $\{\gamma_k\}_{k=1}^{p}$, respectively, and that there are no common poles. Define residues: $\tilde{\phi}_i := \text{res}(G(s), \lambda_i)$; $\hat{\phi}_j := \text{res}(G_r(s), \hat{\lambda}_j)$; and $\psi_k := \text{res}(W(s), \gamma_k)$. The weighted-$\mathcal{H}_2$ error is given by

$$\|G - G_r\|_{\mathcal{H}_2(W)}^2 = \sum_{i=1}^{n} \|G(-\lambda_i) - G_r(-\lambda_i)\|_{\mathcal{H}_2(W)}^2 \psi_k \hat{\phi}_j \tilde{\phi}_i \quad (9)$$

One may recover the (unweighted) $\mathcal{H}_2$ error expression of [9] as a special case by taking $W(s) = 1$. Notice that the weighted error depends on the mismatch of $G$ and $G_r$ at the reflected full system poles $\{\lambda_i\}$, reflected reduced poles $\{\hat{\lambda}_j\}$, and reflected weight poles $\{\gamma_k\}$.

2.4 An algorithm for the weighted-$\mathcal{H}_2$ model reduction problem: W-IRKA

In order to reduce the weighted error, one may eliminate some terms in the error expression, by forcing interpolation at selected (mirrored) poles. Since $r$ is required to be much smaller than $n$, there is not enough degrees of freedom to force interpolation at all the terms in the first and third components of the weighted-$\mathcal{H}_2$ error. However, the second term, i.e. the mismatch at $\lambda_j$, can be completely eliminated by enforcing $G(-\hat{\lambda}_j) = G_r(-\hat{\lambda}_j)$ for $j = 1, \ldots, r$. Hence, as in the unweighted $\mathcal{H}_2$ problem, the mirror images of the reduced-order poles play a crucial role. This motivates an algorithm with iterative rational Krylov steps to enforce the desired interpolation property as outlined in Algorithm 1 below. However, a crucial difference from the unweighted $\mathcal{H}_2$ problem is that we will not enforce interpolation of $G'(s)$ at
We provide two examples related to controller reduction. \( \Phi^{(N)} \) and \( \Psi^{(N)} \) denote the set of normalized residues of \( G(s) \) and \( W(s) \), respectively.

### 3 Numerical examples

We provide two examples related to controller reduction. \( \Phi^{(N)} \) and \( \Psi^{(N)} \) denote the set of normalized residues of \( G(s) \) and \( W(s) \), respectively.

#### 3.1 A building model

The plant, \( P \), is linearized a model for the Los Angeles University Hospital, and has order 48; see [4] for details. An LQG-based controller, \( G \), of the same order, \( n = 48 \), is designed to dampen oscillations in the impulse response. The ten highest normalized residues of \( G(s) \) and of \( W(s) \) are:

\[
\begin{align*}
\Phi^{(N)} & = [ 1.0000 \quad 0.0000 \quad 0.0286 \quad 0.0286 \quad 0.0088 , \quad 0.0088 \quad 0.0080 \quad 0.0060 \quad 0.0060 ] \\
\Psi^{(N)} & = [ 1.0000 \quad 0.8416 \quad 0.8416 \quad 0.3935 , \quad 0.3935 \quad 0.2646 \quad 0.2646 \quad 0.0951 \quad 0.0951 ]
\end{align*}
\]

There is a significant drop in \( \Phi^{(N)} \) values after the second entry, and so we take the first two residues of \( G \) as dominant. \( \Psi^{(N)} \) remains at roughly the same order until the 9th entry. Thus, we choose \( \nu = 2 \); and \( \varpi = r - \nu = r - 2 \) for a given reduction order, \( r \). To illustrate the effect of this dominant pole selection, we apply \( \text{W-IRKA} \), varying \( \nu \) from 0 to 16. Tables 1 below lists the resulting weighted-\( H_2 \) errors for three cases: \( r = 12 \), \( r = 14 \), and \( r = 16 \).

| \( \nu/\varpi \) | \( r = 12 \) | \( r = 14 \) | \( r = 16 \) |
|----------------|-------------|-------------|-------------|
|                | 12/0 | 10/2 | 8/4 | 6/6 | 4/8 | 2/10 | 0/12 | 16/0 | 14/2 | 12/4 | 10/6 | 8/8 |
| \( 1.4921 \)    | 1.1433 | 0.6548 | 0.6863 | 0.3576 | 0.2181 | 0.2853 |
| \( 1.4734 \)    | 1.3436 | 0.6477 | 0.3019 | 0.1538 | 0.1425 | 0.1351 | 0.2224 |
| \( 1.4206 \)    | 1.1934 | 0.7258 | 0.2898 | 0.1917 |
| \( 0.1221 \)    | 0.1154 | 0.1309 | 0.1388 |

Table 1

Weighted-\( H_2 \) error as \( \nu \) and \( \varpi \) vary

The weighted-\( H_2 \) error decreases as we take more dominant poles of \( W(s) \) over those of \( G(s) \); suggesting the importance of the residues of \( W(s) \) in the error expression \( (9) \). Choosing \( \nu = 2 \) is the best choice for most cases. Tables 1 illustrate that while the weighted error initially decreases as \( \nu \) decreases, it starts increasing when \( \nu < 2 \), justifying the choice \( \nu = 2 \). For the case of \( r = 16 \), similar observations hold. Although \( \nu = 2 \) is not the optimal choice when \( r = 16 \), the error for \( \nu = 2 \) is nearly smallest, making \( \nu = 2 \) still a very good candidate for \( \text{W-IRKA} \).

These numerical results support the idea of choosing \( \nu \) and \( \varpi \) according to the decay of the normalized residues. Even though this choice seems to yield small weighted errors, there may be variations that are even better. The residues are multiplied by quantities such as \( W(-\lambda_j)W(\lambda_i) \), so one might consider incorporating these multiplied quantities as well.

A satisfactory reduced-order controller should not only approximate the full-order controller, but also provide...
the same closed-loop behavior as the original controller. Let $T$ and $T_r$ denote the full-order and reduced-order closed-loop systems, respectively: $T$ corresponds to the feedback connection of $P$ with $G$; and $T_r$ to the feedback connection of $P$ with $G_r$. Figure 1-(a) depicts the amplitude Bode plots of $G$ and $G_r$ for $r = 14$ obtained with $\nu = 2$. $G_r$ is an accurate match to $G$. Figure 1-(b) shows that the reduced-closed loop behavior $T_r$ almost exactly replicates $T$.

![Bode Plots](image)

Fig. 1. Bode Plots (a) Full and reduced controller (b) Full and reduced closed-loop system

We now compare W-IRKA with Frequency Weighted Balanced Truncation (FWBT) and IRKA of [9] for the (unweighted) $H_2$ problem. Comparison with IRKA is included to illustrate the importance of including weighting in the $H_2$-based model reduction process. We vary the reduction order from $r = 10$ to $r = 20$ in increments of 2, and compute weighted $H_\infty$ and $H_2$ errors for each case. We use $\nu = 2$ for all cases even though it might not be the best choice for W-IRKA. Results are given in Table 2. Note that for every $r$ value, W-IRKA outperforms FWBT with respect to the weighted-$H_2$ norm. This might be anticipated since W-IRKA is designed to reduce the $H_2$ error. But W-IRKA outperforms FWBT with respect to the weighted-$H_\infty$ norm as well in all except the $r = 18$ case. This is significant since balanced truncation approaches generally yield small $H_\infty$ norms. This behavior is similar to the behavior of IRKA where one often observes that IRKA consistently yields satisfactory $H_\infty$ approximants as well [9].

Note that for $r = 10$, the reduced-order controller due to FWBT fails to produce a stable closed-loop system. Table 2 also illustrates that W-IRKA significantly outperforms IRKA in terms of the weighted error norms. This is what we have expected since unlike W-IRKA, IRKA is tailored towards the unweighted $H_2$ model reduction problem. This becomes clearer after inspecting Table 3, which shows that, in terms of the unweighted error $\|G - G_r\|_{H_\infty}$, IRKA outperforms W-IRKA. Thus, while $G_r$ from IRKA is a better approximation to $G$ in an open-loop sense, once the weight is taken into consideration, W-IRKA does what it is designed for, leading to a smaller weighted error.

![Table 2](image)

Table 2 Comparison of W-IRKA, FWBT, and IRKA

![Table 3](image)

Table 3 Comparison of W-IRKA and IRKA: Unweighted error

### 3.2 International Space Station 12A Module

The plant, $P$, is a model for the International Space Station 12A Module with dimension 1412. It is lightly damped and its impulse response exhibits long-lasting oscillations. A state-feedback, full-order, observer-based controller of order $n = 1412$ is designed to dampen these oscillations. The decay rate of the first 50 normalized residues $\Phi^{(N)}$ and $\Psi^{(N)}$ are shown in Figure 2-(a). While there is almost a two order-of-magnitude drop in $\Phi^{(N)}$ between the third and fourth components, $\Psi^{(N)}$ continues to stay significant. Hence, we take $\nu = 3$ and reduce order from $n = 1412$ to $r = 60$ using W-IRKA.

For comparison, we also apply FWBT. We denote the resulting reduced-order closed-loop systems due to W-IRKA and FWBT by $T_r$ and $T_{fwbt}$, respectively. Note that $T_{fwbt}$ was unstable for $r = 60$. Indeed, $r = 88$ is the smallest order FWBT-derived reduced controller that lead to a stable closed-system. All FWBT-derived $G_r$ are stable; however for $r < 88$ when $G_r$ is connected to $P$, the resulting $T_{fwbt}$ is unstable. Hence, we compare below the $r = 60$ case for W-IRKA with the $r = 88$ case for FWBT. In Figure 2-(b), we plot the absolute value of the errors in the impulse responses due to both methods. W-IRKA outperforms FWBT even with a lower-order controller. We also simulate both $T$ and $T_r$ for a sinusoidal input of $u(t) = \cos(2t)$. Results 2-(c) illustrate the superior performance of W-IRKA even more clearly.
We have presented new formulae for the weighted-$H_2$ inner product and norm that explicitly reveal the contribution of poles and residues both of the full-order model and of the weight. One of the major consequences of this new representation are new interpolatory optimality conditions for weighted-$H_2$ approximation. Based on derived weighted-$H_2$ error expressions, we have introduced an approach for producing high-fidelity weighted-$H_2$ reduced models. The effectiveness of this approach has been illustrated with two examples.

4 Acknowledgments

The work S. Gugercin was supported in part by NSF through Grant DMS-0645347. The work of A. C. Antoulas was supported in part by NSF through Grant CCF 1017401, and DFG through Grant AN-693/1-1.

5 Conclusions

We have presented new formulae for the weighted-$H_2$ inner product and norm that explicitly reveal the contribution of poles and residues both of the full-order model and of the weight. One of the major consequences of this new representation are new interpolatory optimality conditions for weighted-$H_2$ approximation. Based on derived weighted-$H_2$ error expressions, we have introduced an approach for producing high-fidelity weighted-$H_2$ reduced models. The effectiveness of this approach has been illustrated with two examples.

References

[1] B.D.O. Anderson and Y. Liu. Controller reduction: concepts and approaches. IEEE Trans. Automat. Control., 34(8):802–812, 2002.
[2] A.C. Antoulas. Approximation of Large-Scale Dynamical Systems. SIAM, 2005.
[3] 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. Springer, 2010.
[4] A.C. Antoulas, D.C. Sorensen, and S. Gugercin. A survey of model reduction methods for large-scale systems. Contemporary Mathematics, 280:193–219, 2001.
[5] C. De Vlilleman and R.E. Skelton. Model reductions using a projection formulation. International Journal of Control, 46(6):2141–2169, 1987.
[6] D.F. Enns. Model reduction with balanced realizations: An error bound and a frequency weighted generalization. In 23rd IEEE Conf. on Decision and Control 1984, volume 23, pages 127–132, 1984.
[7] 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(5):639–649, 1995.
[8] E. Grimme. Krylov Projection Methods for Model Reduction. PhD thesis, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 1997.
[9] S. Gugercin, A. C. Antoulas, and C. A. Beattie. $H_2$ model reduction for large-scale linear dynamical systems. SIAM J. Matrix Anal. Appl., 30(2):609–638, 2008.
[10] S. Gugercin and A.C. Antoulas. A survey of model reduction by balanced truncation and some new results. International Journal of Control, 77(8):748–766, 2004.
[11] Y. Halevi. Frequency weighted model reduction via optimal projection. IEEE Trans. Automat. Control, 37(10):1537–1542, 1992.
[12] C.-A. Lin and T.-Y. Chiu. Model reduction via frequency weighted balanced realization. Control Theory and Advanced Technol., 8:341–351, 1992.
[13] J. Rommes and M. Martins. Efficient computation of multivariable transfer function dominant poles using subspace acceleration. IEEE Trans. on Power Systems, 21:1471–1483, 2006.
[14] G. Schelhout and B. De Moor. A note on closed-loop balanced truncation. IEEE Transactions on Automatic Control, 41(10):1498–1500, 2002.
[15] 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.
[16] V. Sreeram and A. Ghafoor. Frequency weighted model reduction technique with error bounds. In American Control Conference, 2005., volume 4, pages 2584 – 2589, 2005.
[17] A. Varga and B.D.O. Anderson. Accuracy-enhancing methods for balancing-related frequency-weighted model and controller reduction. Automatica, 39(5):919–927, 2003.
[18] G. Wang, V. Sreeram, and WQ Liu. A new frequency-weighted balanced truncation method and an error bound. IEEE Trans. on Automat. Control, 44(9):1734–1737, 1999.
[19] G. Wang, V. Sreeram, and WQ Liu. Balanced performance preserving controller reduction. Systems & Control Letters, 46(2):99–110, 2002.
[20] A. Yousuff and RE Skelton. Covariance equivalent realizations with applications to model reduction of large-scale systems. Control and Dynamic Systems, 22:273–348, 1985.
[21] A. Yousuff, D.A. Wagle, and R.E. Skelton. Linear system approximation via covariance equivalent realizations. J. of mathematical analysis and applications, 106(1):91–115, 1985.
[22] K. Zhou, J. Doyle, and K. Glover. Robust and Optimal Control. Prentice-Hall, 1996.