Rational Krylov Methods for Model Reduction of Large-scale Dynamical Systems

Serkan G"ugercin
Department of Mathematics, Virginia Tech, USA

Aerospace Computational Design Laboratory,
Dept. of Aeronautics & Astronautics, MIT,
October 21, 2005, Cambridge, MA
Outline

1. Introduction and Problem Statement
2. Motivating Examples
3. Rational Krylov-Interpolation Framework
5. Conclusions
Consider an $n^{th}$ order single-input/single-output system $G(s)$:

\[
G(s) : \begin{cases} 
\dot{x}(t) &= Ax(t) + bu(t) \\
y(t) &= cx(t) 
\end{cases} \iff G(s) = c(sI_n - A)^{-1}b = \frac{n(s)}{d(s)}
\]

- $u(t) \in \mathbb{R}$: input, $x(t) \in \mathbb{R}^n$: state, $y(t) \in \mathbb{R}$: output

- $A \in \mathbb{R}^{n \times n}$, $b, c^T \in \mathbb{R}^n$. Will assume $\Re(\lambda_i(A)) < 0$

- Need for improved accuracy $\implies$ Include more details in the modeling stage

- In many applications, $n$ is quite large, $n \approx \mathcal{O}(10^6, 10^7)$,

- Untenable demands on computational resources $\implies$
Model Reduction Problem: Find

\[
\dot{x}_r(t) = A_r x_r(t) + b_r u(t) \quad \Rightarrow \quad G_r(s) = c_r (sI_r - A_r)^{-1} b_r
\]

\[
y_r(t) = c_r x_r(t)
\]

- where \( A_r \in \mathbb{R}^{r \times r}, \quad b_r, c_r^T \in \mathbb{R}^r \), with \( r \ll n \) such that
  1. \( \|y - y_r\| \) is small.
  2. The procedure is computationally efficient.

\[\begin{array}{c}
  A_r \\
  B_r \\
  C_r \\
  D_r \\
\end{array}\]

\( G_r(s) \): used for simulation or designing a reduced-order controller
• Model reduction through projection: a unifying framework.

• Construct $\Pi = VZ^T$, where $V, Z \in \mathbb{R}^{n \times r}$ with $Z^TV = I_r$:

$$
\begin{align*}
\dot{x}_r &= Z^TAVx_r(t) + Z^Tbu(t), \\
&=: A_r \\
y_r(t) &= cVx_r(t) \\
&=: c_r
\end{align*}
$$

What is the approximation error $e(t) := y(t) - y_r(t)$?

• $G(s)$: Associate a convolution operator $S$:

$$
S : u(t) \mapsto y(t) = (Su)(t) = (g \ast u)(t) = \int_{-\infty}^{t} g(t - \tau)u(\tau)d\tau.
$$

• $g(t) = ce^{At}b$ for $t \geq 0$: Impulse response.

• Transfer function: $G(s) = (Lg)(s) = c(sI - A)^{-1}b$. 

5
The $\mathcal{H}_\infty$ Norm: 2-2 induced norm of $S$:

$$\|G(s)\|_{\mathcal{H}_\infty} = \sup_{u \neq 0} \frac{\|y\|_2}{\|u\|_2} = \sup_{u \neq 0} \frac{\|Su\|_2}{\|u\|_2} = \sup_{w \in \mathbb{R}} \|G(jw)\|_2$$

$$\|G - G_r\|_\infty = \text{Worst output error } \|y(t) - y_r(t)\|_2 \quad \forall \quad \|u(t)\|_2 = 1.$$ 

The $\mathcal{H}_2$ Norm: $\mathcal{L}_2$ norm of $g(t)$ in time domain:

$$\|G(s)\|_{\mathcal{H}_2}^2 = \int_0^\infty \text{trace}[g^T(t)g(t)]dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \text{trace}[G^*(jw)G(jw)]dw$$
Motivating Example: Simulation

A Tunable Optical Filter: (Data: D. Hohlfeld, T. Bechtold, and H. Zappe)

- An optical filter, tunable by thermal means.
  - Silicon-based fabrication.
  - The thin-film filter: membrane to improve thermal isolation
  - Wavelength tuning by thermal modulation of resonator optical thickness

- The device features low power consumption, high tuning speed and excellent optical performance.
Modeling:

- A simplified thermal model to analyze/simulate important thermal issues: 2D model and 3D model
- Meshed and discretized in ANSYS 6.1 by the finite element methods
- The Dirichlet boundary conditions at the bottom of the chip.
- A constant load vector corresponding to the constant input power of 1 mW for 2D model and 10 mW for 3D model
- The output nodes located in the membrane

\[
E \dot{x}(t) = A \, x(t) + b \, u(t), \quad y(t) = c \, x(t)
\]

- 2D: \( n = 1668, \, \text{nnz}(A) = 6209, \, \text{nnz}(E) = 1668 \)
- 3D: \( n = 108373, \, \text{nnz}(A) = 1406808, \, \text{nnz}(E) = 1406791 \)
Motivating Example: Control

International Space Station (ISS) Modules:

- ISS: a complex structure of many modules.
- More than 40 Shuttle flights to complete the assembly.
- Each module modeled by $\approx 1000$ state variables
- For ISS, controllers will be needed for many reasons:
  1. to reduce the oscillatory motions;
  2. to fix the orientation of the space station with respect to some desired direction
- Controllers to be implemented on-board $\Rightarrow$
- Controllers of low complexity needed due to:
  - hardware, radiation, throughput and testing issues
Flex Structure Variation During Assembly

Figure 1: ISS: Evolution of the frequency response as more modules are added (Data: Draper Labs)
SVD based approximation methods

- Singular Value Decomposition (SVD) provides low-rank approximation of matrices.
- What are the singular values of $G(s)$?
- Two Lyapunov Equations:
  \[
  AP + PA^T + bb^T = 0, \quad 0 < P : \text{ Reachability Gramian}
  \]
  \[
  A^TQ + QA + c^Tc = 0, \quad 0 < Q : \text{ Observability Gramian}
  \]
- Hankel Singular Values:
  \[
  \sigma_i(\Sigma) := \sqrt{\lambda_i(PQ)} \text{ for } i = 1, \ldots, n.
  \]
- States difficult to reach correspond to small eigenvalues of $P$.
- States difficult to observe correspond to small eigenvalues of $Q$. 
Balanced Truncation Method

- States difficult to reach ⇒ States difficult to observe?
- Make $P = Q$ by a basis change: $\bar{x} = Tx$, $\det T \neq 0$
  \[ \bar{A} = TAT^{-1}, \quad \bar{b} = Tb, \quad \bar{c} = cT^{-1}. \]
- $P = UU^T$, $U^TQU = R\Sigma^2R^T \implies T_b := \Sigma^{1/2}R^TU^{-1}$
- In the balanced basis: $\bar{P} = \bar{Q} = \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$

**Projection Framework:**

- $P = UU^T$, $Q = LL^T$, $U^T L = W\Sigma Y^T$, $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$.
- $Z := LY_1\Sigma_1^{-1/2}$, $V := UW_1\Sigma_1^{-1/2}$. $VZ^T$: oblique projector
- $G_r(s) = \begin{bmatrix} A_r & b_r \\ c_r & 0 \end{bmatrix} = \begin{bmatrix} Z^T AV & Z^T b \\ cV & 0 \end{bmatrix}$. 
Properties of the reduced model:

1. \( G_r(s) \) is asymptotically stable.

2. \( \| G(s) - G_r(s) \|_\infty \leq 2 (\sigma_{r+1} + \cdots + \sigma_n) \).

Drawbacks in large-scale settings: Requires \( U \) and \( L \)

1. \( \mathcal{O}(n^3) \) arithmetic operations

2. \( \mathcal{O}(n^2) \) memory requirement

Remedy: \( \sigma_i(\Sigma) \), \( \lambda_i(P) \) and \( \lambda_i(Q) \) decay rapidly.

\[ \Rightarrow \]

- Find \( \tilde{U}, \tilde{L} \in \mathbb{R}^{n \times k} \) s.t. \( \tilde{P} := \tilde{U}\tilde{U} \approx P \) and \( \tilde{Q} := \tilde{L}\tilde{L}^T \approx Q \).

- (Approximate) balancing using \( \tilde{U} \) and \( \tilde{L} \).

- \( \mathcal{O}(n^2k) \) arithmetic operations, \( \mathcal{O}(nk) \) memory requirement.

- Penzl [2000], Li/White [2000]: LR-Smith, G/Sorensen/Antoulas [2002]: Modified LR-Smith, Willcox/Peraire [2002]: Balanced POD
Model Reduction via Interpolation

**Rational Interpolation:** Given $G(s)$, find $G_r(s)$ so that

$G_r(s)$ interpolates $G(s)$ and certain number of its derivatives at selected frequencies $\sigma_k$ in the complex plane

\[
\left. \frac{(-1)^j}{j!} \frac{d^j G(s)}{d s^j} \right|_{s=\sigma_k} = \left. \frac{(-1)^j}{j!} \frac{d^j G_r(s)}{d s^j} \right|_{s=\sigma_k}, \quad \text{for } k = 1, \ldots, K,
\]

\[
\left. \frac{(-1)^j}{j!} \frac{d^j G(s)}{d s^j} \right|_{s=\sigma_k} = c(\sigma_k I - A)^{-(j+1)} b:
\]

- $j^{th}$ moment of $G(s)$ at $\sigma_k$
- $\eta_{\sigma_k}(j)$.
• For $\sigma_1 = \sigma_2 = \ldots = \sigma_K = \sigma = \infty$
  
  $\eta_\sigma^{(j)} = cA^{j-1}b : j^{th}$ Markov Parameter

  - Moment matching problem: **Partial realization**

  - Solution: *Lanczos* and *Arnoldi* procedures

• For arbitrary $\sigma \neq \infty$

  $\eta_\sigma^{(j)} = c(\sigma I - A)^{-(j+1)}b$

  - Moment matching problem: **Rational Interpolation**

  - Solution: *Rational Lanczos/Arnoldi* procedures

• For multiple points $\implies$ **rational Krylov method**
• Moments are very ill-conditioned to compute

Ball et al. [1990], Antoulas et al. [1990], Pillage et al. [1993]:
Moment matching problem with explicit moment computation.
\[ \implies \text{Total loss of accuracy in large-scale settings} \]

• Feldman and Freund [1995]: Padé via Lanczos (Krylov setting)
  – Moment matching without moment computation
  – Iterative implementation
The Arnoldi Procedure for $\sigma = \infty$}

1. Given $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, and $c^T \in \mathbb{R}^n$,

2. $v_1 := \frac{b}{\|b\|}$, $w := Av_1$; $\alpha_1 := v_1^T w$

   $f_1 := w - v_1 \alpha_1$; $V_1 := (v_1)$; $H_1 := (\alpha_1)$

3. For $j = 1, 2, \cdots, r - 1$

   $\beta_j := \|f_j\|$, $v_{j+1} := \frac{f_j}{\beta_j}$

   $V_{j+1} := (V_j \ v_{j+1})$, $\hat{H}_j = \begin{pmatrix} H_j \\ \beta_j e_j^T \end{pmatrix}$

   $w := Av_{j+1}$, $h := V_{j+1}^T w$, $f_{j+1} = w - V_{j+1} h$

   $H_{j+1} := \begin{pmatrix} \hat{H}_j \\ h \end{pmatrix}$

- $AV = VH + fe_r^T$, $V^T V = I_r$, $V^T f = 0$, $V \in \mathbb{R}^{n \times r}$
- $A_r = H$, $b_r = \|b\|e_1$, $c_r = cV$
- \( G_r(s) \) matches the first \( r \) Markov parameters of \( G(s) \) without computing them

- Only matrix-vector multiplications. No SVD, or Schur decomposition required

- \( \text{Range}(V) = \text{Span} ([b, Ab, \ldots, A^{r-1}b]) \)

- \textbf{Lanczos Procedure}: Two-sided Arnoldi.
  - \( \text{Range}(V) = \text{Span} ([b, Ab, \ldots, A^{r-1}b]) \)
  - \( \text{Range}(Z) = \text{Span} ([c^T, A^Tc^T, \ldots, (A^T)^{r-1}c^T]) \)
  - \( G_r(s) \) matches the first \( 2r \) Markov parameters of \( G(s) \) without computing them

- For arbitrary \( \sigma \neq \infty \),

  \[
  A \rightarrow (\sigma I - A)^{-1}, \quad b \rightarrow (\sigma I - A)^{-1}b, \quad c \rightarrow c(\sigma I - A^T)^{-1}
  \]
Approximation of a Building Model

- Los Angeles University Hospital: $n = 48$ to $r = 6$. 

\[ \| H(jw) \| \text{(dB)} \]
Rational Krylov Method

- Lanczos/Arnoldi: Moment matching at a single point.
- Large error around other frequencies
- Match the moments around various frequencies ⇒
- Better approximation over a broad range
- Solution by projection: First by Skelton et. al. [1985]
- Grimme [1997]: Efficient computation by Krylov projection
• Given $2r$ interpolation points: $\{\sigma_i\}_{i=1}^{2r}$

• Set $V = \text{Span} \left[ (\sigma_1 I - A)^{-1} b, \cdots, (\sigma_r I - A)^{-1} b \right]$, and

• $Z = \text{Span} \left[ (\sigma_{r+1} I - A^T)^{-1} c^T, \cdots, (\sigma_{2r} I - A^T)^{-1} c^T \right]$, and

• Make $Z^T V = I_r$.

• $A_r = Z^T A V$, $b_r = Z^T b$, $c_r = c V$

$\implies G(\sigma_i) = G_r(\sigma_i)$ for $i = 1, \ldots, 2r$

• Moment matching without explicit moment computation
• Given \( r \) interpolation points: \( \{\sigma_i\}_{i=1}^r \)

• Set \( V = \text{Span} \left[ (\sigma_1 I - A)^{-1} b, \ldots, (\sigma_r I - A)^{-1} b \right] \), and

• \( Z = \text{Span} \left[ (\sigma_1 I - A^T)^{-1} c^T, \ldots, (\sigma_r I - A^T)^{-1} c^T \right] \),

• Make \( Z^T V = I_r \).

• \( A_r = Z^T A V \), \( b_r = Z^T b \), \( c_r = c V \)

\[ \implies G(\sigma_i) = G_r(\sigma_i), \text{ and } \frac{d}{ds} G(s) \bigg|_{s=\sigma_i} = \frac{d}{ds} G_r(s) \bigg|_{s=\sigma_i} \]

• Moment matching without explicit moment computation

• Efficient computations of \( V \) and \( Z \): Dual Rational Arnoldi, Rational Lanczos, Rational Power Krylov. (Grimme [1997])
Why to choose model reduction via rational interpolation?

- van Dooren et al. [2004], Halevi [2005]: Generically, any reduced model $G_r(s)$ can be obtained via interpolation.

- Interpolation points = Zeroes of $G(s) - G_r(s)$.

- Example: $G(s) = \frac{1}{s^2 + 3s + 2}$ and $G_r(s) = \frac{0.2}{s + 0.4}$

- $G(s) - G_r(s) = \frac{-0.2s(s - 2)}{s^3 + 3.4s^2 + 3.2s + 0.8} \Rightarrow \sigma_1 = 0, \sigma_2 = 2$

- $V = (\sigma_1 - A)^{-1}b = A^{-1}b, \quad Z = (\sigma_2 - A)^{-T}c^T = (2I - A)^{-T}c^T$

- BUT:

What is a good selection of interpolation points?

- Similar to polynomial approximation of complex functions.
Advantages of Krylov based methods:

- Iterative implementation
- Only matrix-vector multiplications and sparse LU decompositions
  \[\mathcal{O}(r^2n)\] arithmetic operation and \(\mathcal{O}(nr)\) storage requirements

Drawbacks:

- Stability is not guaranteed
- No global error bound
- Selection of \(\sigma_i\) is an ad-hoc process.
Stability: Implicit Restart (Sorensen et al. [1995])

- Let $A_r, b_r, c_r$ be obtained via $r$ steps of Arnoldi algorithm.
- Let $A_r$ have one unstable eigenvalue $\mu$, i.e., $\Re(\mu) > 0$.
- Define $\hat{b} := (\mu I_n - A)b$
- Run $r - 1$ steps of Arnoldi process on the pair $(A, \hat{b})$
- The new reduced system will be asymptotically stable.
- Matching the moments of $\hat{G}(s) = c(sI - A)^{-1}\hat{b}$
- Trade-off between guaranteed stability and exact moment matching.
• **Least-squares model reduction (G./Antoulas [2003])**
  – Bilinear transformation and reduction in the discrete-time
  – Combines the SVD and Krylov reduction
  – \( V \) : Krylov-side, \( Z \) : SVD-side
  – Stability and moment matching

• **Fourier model reduction (Willcox/Megretski [2005])**
  – Bilinear transformation and reduction in the discrete-time
  – Stability and moment matching
An $\mathcal{H}_2$ Error Expression (G./Antoulas [2002])

- $\|G(s)\|_{\mathcal{H}_2}^2 = cPc^T = b^TQb$.
- $\phi_i := G(s)(s - \lambda_i) \mid_{s = \lambda_i}$ and $\hat{\phi}_j := G_r(s)(s - \hat{\lambda}_j) \mid_{s = \hat{\lambda}_j}$.

$$\|G(s) - G_r(s)\|_{\mathcal{H}_2}^2 = \sum_{i=1}^{n} \phi_i \left(G(-\lambda_i) - G_r(-\lambda_i)\right) + \sum_{j=1}^{r} \hat{\phi}_j \left(G_r(-\hat{\lambda}_j) - G(-\hat{\lambda}_j)\right).$$

- Error due to mismatch at $-\lambda_i$ and $-\hat{\lambda}_j$.
- Try to kill the first term $\Rightarrow \sigma_i = -\lambda_i(A)$.
- Select $\lambda_i(A)$ with large $\phi_i$. 

27
CD Player and Selection of $\sigma_i$

- Dynamics of a CD Player with 120 states, $m = p = 1$.

- Reduce the order to $r = 14$ by Rational Krylov

- $G_r^*$: Choosing $\sigma_i = -\lambda_i(A)$.

- $G_1, G_2, G_3$: Arbitrary $\sigma_i$. Best 3 among 2000 models.

<table>
<thead>
<tr>
<th></th>
<th>$G_r^*$</th>
<th>$G_1$</th>
<th>$G_2$</th>
<th>$G_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{H}_2$</td>
<td>$4.86 \times 10^{-1}$</td>
<td>$1.16 \times 10^0$</td>
<td>$1.35 \times 10^0$</td>
<td>$1.56 \times 10^0$</td>
</tr>
<tr>
<td>$\mathcal{H}_\infty$</td>
<td>$7.33 \times 10^{-2}$</td>
<td>$3.21 \times 10^{-1}$</td>
<td>$3.95 \times 10^{-1}$</td>
<td>$5.53 \times 10^{-1}$</td>
</tr>
</tbody>
</table>

Table 1: $\mathcal{H}_2$ and $\mathcal{H}_\infty$ errors for the CD Player Example

- Does there exist an optimal shift selection?
Problem: Given a stable dynamical system $G(s)$, find a reduced model $G_r(s)$ that satisfies

$$G_r(s) = \arg \min_{\deg(\tilde{G}) = r, \tilde{G} : \text{stable}} \| G(s) - \tilde{G}(s) \|_{\mathcal{H}_2}.$$  

- First-order conditions: (Meier and Luenberger [1967])

$$G(-\hat{\lambda}_i) = G_r(-\hat{\lambda}_i), \quad \text{and} \quad \frac{d}{ds} G(s) \bigg|_{s=-\hat{\lambda}_i} = \frac{d}{ds} G_r(s) \bigg|_{s=-\hat{\lambda}_i}$$

- Match the first two moments at the mirror images of the Ritz values.

- First-order conditions as interpolation. ⇒

- Rational Krylov Framework
• For the $\mathcal{H}_2$ problem, **simply** set $\sigma_i = -\hat{\lambda}_i$

• $\hat{\lambda}_i$ NOT known a priori $\implies$ Needs iterative rational steps

**An Iterative Rational Krylov Algorithm (IRKA):**
(G..Beattie/Antoulas [2004])

1. Choose $\sigma_i$ for $i = 1, \ldots, r$.

2. $V = \text{Span } [(\sigma_1I - A)^{-1}b, \ldots, (\sigma_rI - A)^{-1}b]$,

3. $Z = \text{Span } [(\sigma_1I - A^T)^{-1}c^T, \ldots, (\sigma_rI - A^T)^{-1}c^T]$, $Z^TV = I_r$.

4. while [relative change in $\sigma_j$] > $\epsilon$
   
   (a) $A_r = Z^TAV$,
   
   (b) $\sigma_i \leftarrow -\lambda_i(A_r)$ for $i = 1, \ldots, r$
   
   (c) $V = \text{Span } [(\sigma_1I - A)^{-1}b, \ldots, (\sigma_rI - A)^{-1}b]$.
   
   (d) $Z = \text{Span } [(\sigma_1I - A^T)^{-1}c^T, \ldots, (\sigma_rI - A^T)^{-1}c^T]$, $Z^TV = I_r$.

5. $A_r = Z^TAV$, $b_r = Z^Tb$, $c_r = cV$

• **Upon convergence**, first-order conditions satisfied via Krylov projection framework, no Lyapunov solvers
\( n = 1412 \). Reduce to \( r = 2 : 2 : 60 \)

- Compare with balanced truncation
Example: Optimal Cooling of Steel Profiles (P. Benner)

- $G(s) = c(sE - A)^{-1}b$, $n = 79,841$
- $r = 6$ via IRKA

$\|G(s) - G_1(s)\|_{\infty} = 6.01 \times 10^{-5}$
• van Dooren et al. [2005]: Model reduction of interconnected systems via rational Krylov
  – Reduce each subsystem via interpolation
  – Reduced-interconnection interpolates the full-model

• Bai [2003]: Krylov-reduction of second-order systems:
  – \( \ddot{\mathbf{x}}(t) + \mathbf{G}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = \mathbf{b}\mathbf{u}(t), \quad \mathbf{y}(t) = \mathbf{c}\mathbf{x}(t) \)

\[
\mathbf{A} = \begin{bmatrix}
0 & I \\
-M^{-1}\mathbf{K} & -M^{-1}\mathbf{G}
\end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix}
0 \\
M^{-1}\mathbf{b}
\end{bmatrix}, \quad \mathbf{C}^T = \begin{bmatrix}
\mathbf{c}^T \\
0
\end{bmatrix}
\]

\[
\mathbf{G}(s) : \begin{cases}
\dot{\mathbf{q}}(t) = \mathbf{A}\mathbf{q}(t) + \mathbf{B}\mathbf{u}(t) \\
\mathbf{y}(t) = \mathbf{C}\mathbf{q}(t)
\end{cases}
\]

– Apply model reduction in the first-order state-space
– Transfer back to second-order form \( \Rightarrow \) Structure is lost
• find a matrix $W \in \mathbb{R}^{n \times r}$ such that $W^T W = I_r$

• $M_r = W^T M W$, $G_r = W^T G W$, $K_r = W^T K W$, $b_r = W^T b$, and $c_r = c W$

• $M_r \ddot{x}_r(t) + G_r \dot{x}_r(t) + K_r x_r(t) = b_r u(t)$, $y_r(t) = c_r x_r(t)$

• **Second-order Krylov subspaces (Second-order Arnoldi Iteration):** $W$ spans the required Krylov subspace but constructed directly in the second-order framework

• Interpolation conditions hold.

• Antoulas/Sorensen [2003]: Passivity preserving Krylov-based model reduction
  - $G(jw) + (G(jw))^* > 0$.
  - $G(s) + G^~(s) = H(s)H^~(s)$
  - Interpolation points: Zeros of $H(s)$
  - Passivity is preserved.
Conclusions and Future Work:

- Rational Krylov framework for model reduction
  - In the SISO case, any $G_r(s)$ can be obtained via interpolation
  - Match moments without moment computation
  - Cost is reduced from $O(n^3)$ to $O(r^2n)$

- To guarantee stability
  - Implicit restart
  - Least-squares reduction (G./Antoulas)
  - Fourier reduction (Willcox/Megretski)

- Optimal selection of interpolation points
  - Interpolation condition for optimal $H_2$ approximation
  - Iterative Rational Krylov Algorithm

- On-going and future work
  - Interconnected systems
  - Second-order systems
  - Controller reduction
  - Optimal approximation
  - Guaranteed error estimates
  - Balancing via Krylov
Controller reduction for large-scale systems

- Consider an $n^{th}$ order plant $G(s) = c(sI - A)^{-1}b$

- $n_{\kappa}^{th}$ order stabilizing controller: $K(s) = c_K(sI - A_K)^{-1}b_K + d_K$

- LQG, $\mathcal{H}_\infty$ control designs $\Rightarrow n_{\kappa} = n \Rightarrow$
  (i) Complex hardware
  (ii) Degraded accuracy
  (iii) Degraded computational speed

- Obtain $K_r(s)$ of order $r \ll n_{\kappa}$ to replace $K(s)$ in the closed loop.
Controller reduction via frequency weighting

- Small open loop error $\|K(s) - K_r(s)\|_\infty$ not enough. ⇒

- Minimize the weighted error:

$$\|W_o(s)(K(s) - K_r(s))W_i(s)\|_\infty.$$  

- How to obtain the weights $W_o(s)$ and $W_i(s)$?

- If $K(s)$ and $K_r(s)$ have the same number of unstable poles and if
  $$\|[K(s) - K_r(s)]G(s)[I + G(s)K(s)]^{-1}\|_\infty < 1, \text{ or}$$
  $$\|[I + G(s)K(s)]^{-1}G(s)[K(s) - K_r(s)]\|_\infty < 1,$$

  ⇒ $K_r(s)$ stabilizes $G(s)$.  

37
• For stability considerations:

\[ W_i(s) = I \quad \text{and} \quad W_o(s) = [I + G(s)K(s)]^{-1}G(s) \quad \text{or} \]
\[ W_o(s) = I \quad \text{and} \quad W_i(s) = G(s)[I + G(s)K(s)]^{-1}. \]

• To preserve closed-loop performance:

\[ W_i(s) = [I + G(s)K(s)]^{-1} \quad \text{and} \quad W_o(s) = [I + G(s)K(s)]^{-1}G(s). \]

• Solved by frequency-weighted balancing (Anderson and Liu [1989],
Schelfhout and De Moor [1996], Varga and Anderson [2002]).

• Requires solving two Lyapunov equations of order \( n + n_\kappa \).

\[ A_i P + PA_i^T + b_i b_i^T = 0, \quad A_o^T Q + QA_o + c_o^T c_o = 0, \]

• \( A_i, b_i : K(s)W_i(s), \quad A_o, c_o : W_o(s)K(s) \)

• Balance \( P \) and \( Q \).
Controller-reduction via Krylov Projection

• How to modify IRKA for the controller reduction problem?

• Let $W_i(s) = I$ and $W_o(s) = [I + G(s)K(s)]^{-1}G(s)$ ⇒

$$A_K P + P A_K^T + b_K b_K^T = 0 \quad \text{unweighted Lyapunov eq.}$$

$$A_T Q + Q A_w + c_w^T c_w = 0 \quad \text{weighted Lyapunov eq.}$$

• $Z = K(A^T, C^T, \sigma_i)$, and $V = K(A, B, \mu_j)$

• $Z$ and $\sigma_i$: Reflect $W_o(s)$: the closed-loop information.

$$\sigma_i = jw_i \text{ over the region where } W_o(jw) \text{ is dominant}$$

• $V$ and $\mu_j$: Obtained in an (optimal) open loop sense.

$\mu_j$: From an iterative rational Krylov iteration
An Iterative Rational Krylov Iteration for Controller Reduction:

1. Choose $\sigma_i = jw_i$, for $i = 1, \ldots, r$ where $w_i$ is chosen to reflect $W_o(jw)$.

2. $Z = \text{Span} \left[ (\sigma_1 I - A_K^T)^{-1} c_K^T \cdots (\sigma_r I - A_K^T)^{-1} c_K^T \right]$ with $Z^T Z = I_r$.

3. $V = Z$

4. while [relative change in $\mu_j$] > $\epsilon$
   
   (a) $A_r = Z^T A_K V$,
   
   (b) $\mu_j \leftarrow -\lambda_i(A_r)$ for $j = 1, \ldots, r$

   (c) $V = \text{Span} \left[ (\mu_1 I - A_K)^{-1} b_K \cdots (\mu_r I - A_K)^{-1} b_K \right]$ with $Z^T V = I_r$.

5. $A_r = Z^T A_K V, \quad b_r = Z^T b_K, \quad c_r = c_K V$

$Z \Rightarrow K_r(s)$ includes the closed-loop information

$V \Rightarrow K_r(s)$ is optimal in a restricted $\mathcal{H}_2$ sense $\Pi = Z V^T$

$\Pi = Z V^T$
International Space Station Module 1R:

- $n = 270$. $G(s)$ is lightly damped $\Rightarrow$ Long-lasting oscillations.
- $K(s)$ is designed to remove these oscillations. $n_K = 270$.

- Reduce the order to $r = 19$ using iterative Rational Krylov and
to $r = 23$ using one-sided frequency weighted balancing
**FWBR**: Frequency-weighted balancing with \( W_i(s) = I \) and \( W_o(s) = [I + G(s)K(s)]^{-1} G(s) \).

**IRK-CL**: Iterative Rational Krylov - Closed Loop version: \( \sigma_i \) reflect the weight \( W_o(s) \).

\[
\sigma_i = j \ast \text{logspace}(-1, 2, 10) \text{ rad/sec}
\]
ISS Example: Bode Plots of reduced closed-loop systems

Relative Errors

<table>
<thead>
<tr>
<th>Relative Errors</th>
<th>$\mathcal{H}_\infty$ error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T - T_{FW20}$</td>
<td>$3.88 \times 10^1$</td>
</tr>
<tr>
<td>$T - T_{FW23}$</td>
<td>$5.63 \times 10^{-1}$</td>
</tr>
<tr>
<td>$T - T_{IRK-CL}$</td>
<td>$1.47 \times 10^{-1}$</td>
</tr>
</tbody>
</table>

Relative Errors

<table>
<thead>
<tr>
<th>Relative Errors</th>
<th>$\mathcal{H}_2$ error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T - T_{FW20}$</td>
<td>$3.90 \times 10^0$</td>
</tr>
<tr>
<td>$T - T_{FW23}$</td>
<td>$1.88 \times 10^{-1}$</td>
</tr>
<tr>
<td>$T - T_{IRK-CL}$</td>
<td>$3.57 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

Weighted Errors

<table>
<thead>
<tr>
<th>Weighted Errors</th>
<th>$\mathcal{H}_2$ error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_i(K - K_{FW20})$</td>
<td>$0.984 &lt; 1$</td>
</tr>
<tr>
<td>$W_i(K - K_{FW23})$</td>
<td>$0.416 &lt; 1$</td>
</tr>
<tr>
<td>$W_i(K - K_{IRK-CL})$</td>
<td>$0.365 &lt; 1$</td>
</tr>
</tbody>
</table>
An Unstable Model:

- \( n = 2000 \). \( K(s) \) of order \( n_\kappa = 2000 \) stabilizes the model.

- \( K(s) \) has four unstable poles.

![Impulse response of G(s)](image)

![Impulse Response of T(s)](image)
• Reduce the order to $r = 14$: Stabilizing controller

• $K_r(s)$ has 4 unstable poles as desired.
Part II: Conclusions and Future Work

- \( n \gg 10^6 \): Forces usage of Inexact Solves in Krylov-based reduction

- Perturbation effects:
  - Backward and forward error analysis framework
  - *Good/ Optimal* shift selection robust with respect to inexact solves
  - **I-IRKA**
    * (Locally) optimal reduced models for \( n > 10^6 \) without user intervention
    * Acceleration strategies

- Open issues:
  - Global \( \mathcal{H}_2 \) and/or \( \mathcal{H}_\infty \) perturbation effects
  - Modifications to GMRES, effective preconditioning strategies
  - Scalable parallel versions
    * A large-scale easy-to-use model reduction toolbox
    * Modify the algorithms to fit into the framework of, e.g., Trilinos
    * Implementation on Virginia Tech.-System X
Effect of Initialization: CD Player Model

- $n = 120$ to $r = 8$ via IRKA (Logarithmic grid)
• **Starting point:** $\mathcal{H}_2$ error expression (Gugercin and Antoulas [2003])

• $G(s)$ with poles $\lambda_i$ and $G_r(s)$ with poles $\hat{\lambda}_j$

• $\phi_i := G(s)(s - \lambda_i) \mid_{s=\lambda_i}$ and $\hat{\phi}_j := G_r(s)(s - \hat{\lambda}_j) \mid_{s=\hat{\lambda}_j}$.

$$\|G(s) - G_r(s)\|_{\mathcal{H}_2}^2 = \sum_{i=1}^{n} \phi_i \left(G(-\lambda_i) - G_r(-\lambda_i)\right) + \sum_{j=1}^{r} \hat{\phi}_j \left(G_r(-\hat{\lambda}_j) - G(-\hat{\lambda}_j)\right).$$

• **Open loop error** due to mismatch at $-\lambda_i$ and $-\hat{\lambda}_j$
The Lanczos Procedure: SISO Case

- Given $\Sigma = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ $\implies v_1 := \frac{1}{\sqrt{|CB|}} B$ and $z_1 := \frac{\text{sgn}(CB)}{\sqrt{|CB|}} C^T$

$\Rightarrow AV = VT + fe_r^T$
$A^TZ = ZT^T + ge_r^T$, $V^TZ = I_r$, $V^Tg = Z^Tf = 0$

- $\Sigma_r = \begin{bmatrix} Z^TAV & Z^TB \\ CV & D \end{bmatrix} = \begin{bmatrix} T & \sqrt{|CB|e_1} \\ \text{sgn}(CB)\sqrt{|CB|e_1^T} & D \end{bmatrix}$

- $\Sigma_r$ is unique and matches $2r$ Markov parameters

- $\text{Im}(V) = \text{Im}\left(\begin{bmatrix} B & AB & \cdots & A^{r-1}B \end{bmatrix}\right) =: \text{Im}\left(\mathcal{R}_r(A, B)\right)$

- $\text{Im}(Z) = \text{Im}\left(\begin{bmatrix} C^T & A^TC^T & \cdots & (A^T)^{r-1}C^T \end{bmatrix}\right) =: \text{Im}\left(\mathcal{O}_r(A, C)^T\right)$
• Define $\mathcal{H}_r := \begin{bmatrix} \eta_1 & \eta_2 & \cdots & \eta_r \\ \eta_2 & \eta_3 & \cdots & \eta_{r+1} \\ \vdots & \vdots & \ddots & \vdots \\ \eta_r & \eta_{r+1} & \cdots & \eta_{2r-1} \end{bmatrix}$, where $\eta_i = CA^{i-1}B$.

• Lanczos requires $\det\mathcal{H}_i \neq 0$ for $i = 1, \cdots, r$.

• Violated for many systems $\Rightarrow$ Look-Ahead Lanczos

• $\det\mathcal{H}_i \neq 0$ can be avoided if tridiagonal structure not required.

• Only requirement: $\det\mathcal{H}_r \neq 0$.

• **Arnoldi Procedure**: The $r^{th}$ order reduced model $\Sigma_r$ matches only $r$ Markov parameters and is not unique.

• For arbitrary $\sigma$,

$$A \to (\sigma I - A)^{-1}, \quad B \to (\sigma I - A)^{-1}, \quad C \to C(\sigma I - A^T)^{-1}$$