Inexact Solves in Krylov-based Model Reduction of Large-scale Dynamical Systems

Chris Beattie and Serkan Gugercin
Department of Mathematics, Virginia Tech.

Supported in parts by the NSF DMS-0505971 and DMS-0513542 and AFOSR FA9550-05-1-0449

Ninth Copper Mountain Conference on Iterative Methods, April 2-7, 2006, Copper Mountain, Colorado
• Consider an $n^\text{th}$ order single-input/single-output system $\mathbf{H}(s)$:

\[
\begin{align*}
\dot{x}(t) &= A \, x(t) + b \, u(t) \\
y(t) &= c \, x(t)
\end{align*}
\]

\[
\mathbf{H}(s) = c(sI_n - A)^{-1}b = \frac{p_{n-1}(s)}{q_n(s)}
\]

• $u(t) \in \mathbb{R}$: 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

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

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

\[
\Leftrightarrow \quad H_r(s) = c_r (sI_r - A_r)^{-1}b_r
\]

- 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.

\( \mathbf{G}_r(s) \): used for simulation or designing a reduced-order controller.
• Model reduction through projection: a unifying framework.

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

\[
\dot{x}_r = W^T A V x_r(t) + W^T b u(t), \quad y_r(t) = c V x_r(t)
\]

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

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

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

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

• Transfer function: $H(s) = (\mathcal{L}h)(s) = c(sI - A)^{-1}b$. 
The $\mathcal{H}_\infty$ Norm: 2-2 induced norm of $S$:

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

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

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

$$\|H(s)\|_{\mathcal{H}_2}^2 = \int_0^\infty \text{trace}[h^T(t)h(t)] dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \text{trace}[H^*(jw)H(jw)] dw$$
Rational Interpolation: Given $H(s)$, find $H_r(s)$ so that

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

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

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

$= j^{th}$ moment of $H(s)$ at $\sigma_k$

$=: \eta_{\sigma_k}(j)$. 

---

Model Reduction via Interpolation
Rational Krylov Method

- Moments are very ill-conditioned to compute
- 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
- $W = \text{Span} \left[ (\sigma_1 I - A^T)^{-1} c^T, \ldots, (\sigma_r I - A^T)^{-1} c^T \right]$. Make $W^T V = I_r$.
- $A_r = W^T A V$, $b_r = W^T b$, $c_r = c V$

\[
\Rightarrow H(\sigma_i) = H_r(\sigma_i), \quad \text{and} \quad \frac{d}{ds} H(s) \bigg|_{s=\sigma_i} = \frac{d}{ds} H_r(s) \bigg|_{s=\sigma_i}
\]

- Moment matching without explicit moment computation
- Efficient computations of $V$ and $W$: Grimme [1997]
Inexact solves in Krylov-based model reduction

• Need for more detail and accuracy in the modeling stage ⇒

• System dimension $n$: $\mathcal{O}(10^6)$ or more ⇒

• $(\sigma I - A)v = b$ cannot be solved directly

• Inexact solves need to be employed in constructing $V$ and $W$

• Questions:
  1. What are the perturbation effects on interpolation?
  2. What are the effective preconditioning, restarting strategies?
  3. What is the effect on (the optimality of) the reduced model?
• For simplicity, consider the one-sided projection, i.e. $V = W$.

• Let $\hat{v}_j$ be an inexact solution for $(\sigma_j I - A)v_j = b$

$$
(\sigma_j I - A)\hat{v}_j - b = \delta b_j
$$

• Define $\delta v_j := \hat{v}_j - v_j = (\sigma_j I - A)^{-1}\delta b_j$, and

$$
\hat{K} := \begin{bmatrix} (\sigma_1 I - A)^{-1}b + \delta v_1, & \cdots & (\sigma_r I - A)^{-1}b + \delta v_r \end{bmatrix}.
$$

• **Inexact** Krylov-based reduced model obtained by

$$
\hat{A}_r = \hat{V}^T A \hat{V}, \quad \hat{b}_r = \hat{V}^T b, \quad \hat{c}_r = c \hat{V}, \quad \text{where} \quad \hat{V}^T \hat{V} = I_r.
$$

• $\hat{V}$: an orthogonal basis for Range($\hat{K}$)

• $\hat{H}_r(s) = \hat{c}_r (sI_r - \hat{A}_r)^{-1}\hat{b}_r$
**Theorem:** The response (interpolation) error at $\sigma_j$ is

$$\varepsilon_j := \hat{H}_r(\sigma_j) - H(\sigma_j) = \|\delta b_j\| \cdot c M_j \frac{\delta b_j}{\|\delta b_j\|}$$

- $M_j = \left[ (\sigma_j I_n - A)^{-1} - \hat{V}(\sigma_j I_r - A_r)^{-1}\hat{V}^T \right]$

- How well $\hat{V}(\sigma_j I_r - A_r)^{-1}\hat{V}^T$ approximates $(\sigma_j I_n - A)^{-1}$
  \[ \Rightarrow \text{Quality of the Ritz approximation } \hat{V}A_r\hat{V}^T \text{ to } A. \]

- Depends on the selection of $\sigma_j$: Good selection $\Rightarrow$ Krylov-based model reduction robust with respect to inexact solves.

- **Backward Error:**

  $$\hat{H}_r(\sigma_j) = H^{[\hat{j}]}(\sigma_j) = c(sI_n - A)^{-1}(b + \Delta b_j)$$

- Exact interpolation of a near-by system.
The two-sided case

• $\hat{K}_1 := \begin{bmatrix} (\sigma_1 I - A)^{-1}b + \delta v_1, & \cdots & (\sigma_r I - A)^{-1}b + \delta v_r \end{bmatrix}$.

• Let $\hat{w}_j$ be an inexact solution for $(\sigma_j I - A^T)w_j = c^T$

\[
(\sigma_j I - A^T)\hat{w}_j - c^T = \delta c_j
\]

• Define $\delta w_j := \hat{w}_j - w_j = (\sigma_j I - A)^{-1}\delta c_j$, and

$\hat{K}_2 := \begin{bmatrix} (\sigma_1 I - A^T)^{-1}c^T + \delta w_1, \cdots, (\sigma_r I - A^T)^{-1}c^T + \delta w_r \end{bmatrix}$.

• Inexact Krylov-based reduced model obtained by

$\hat{A}_r = \hat{W}^T A \hat{V}$, \hspace{0.5cm} $\hat{b}_r = \hat{W}^T b$, \hspace{0.5cm} $\hat{c}_r = c \hat{V}$

• $\hat{V}, \hat{W}$: biorthogonal basis for $\text{Range}(\hat{K}_1)$ and $\text{Range}(\hat{K}_2)$

• $\hat{H}_r(s) = \hat{c}_r(sI_r - \hat{A}_r)^{-1}\hat{b}_r$
**Theorem:** The response (interpolation) error at $\{\sigma_j\}_{j=1}^r$ is

$$
\varepsilon_j := \hat{H}_r(\sigma_j) - H(\sigma_j) = \|\delta b_j\| \cdot \|\delta c_j\| \cdot \frac{\delta c_j^T}{\|\delta c_j\|} M_j (\sigma_j I_n - A) M_j \frac{\delta b_j}{\|\delta b_j\|}
$$

- $M_j = \left[ (\sigma_j I_n - A)^{-1} - \hat{V} (\sigma_j I_r - A_r)^{-1} \hat{W}^T \right]$
- How well $\hat{V} (\sigma_j I_r - A_r)^{-1} \hat{W}^T$ approximates $(\sigma_j I_n - A)^{-1}$
- Quality of the Galerkin approximation $\hat{V} A_r \hat{W}^T$ to $A$.
- Depends on the selection of $\sigma_j$
- Quadratic feature of the error
- Similar backward error estimate as the one-sided case
Example: Optimal Cooling of Steel Profiles (P. Benner)

- \( H(s) = c(sE - A)^{-1}b, \; n = 20, 209 \)
- Bad shift selection: \( \sigma_i = \logspace(-8, -4, 6) \)
- \( r = 6 \) via Rational Krylov (RK) and Inexact-RK (I-RK).
- I-RK uses GMRES with \( tol = 10^{-5} \)

\[
\begin{align*}
|H(j\omega)| & = 10^{-3} \\
\|H(s) - H_1(s)\|_\infty & = 2.18 \times 10^{-3} \\
\|H(s) - H_2(s)\|_\infty & = 2.42 \times 10^{-3} \\
\|H_1(s) - H_2(s)\|_\infty & = 9.18 \times 10^{-4}
\end{align*}
\]
• **RK** with optimal \( \{\sigma_i\} \)
• Use these optimal \( \{\sigma_i\} \) in I-RK.
• **I-RK** uses GMRES with \( tol = 10^{-4} \)

\[
\| H(s) - H_1(s) \|_\infty = H(s) - H_1(s) \|_\infty = 1.56 \times 10^{-4}
\]

\[
\| H_1(s) - H_2(s) \|_\infty = 1.82 \times 10^{-5}
\]
• **GMRES:**

1. The same Krylov subspace for each \((\sigma_j I - A)v_j = b\)
\[ AU_k = U_{k+1} \tilde{H}_k \Rightarrow \min \left\| \sigma_j \tilde{I} - \tilde{H}_k - \|b\|e_1 \right\| \]

2. Span\(\{v_j\}_{j=1}^r\) is important, rather than each \(v_j\)
   \[ \Rightarrow \text{One could afford a less accurate solution for } v_j \text{ if already included in the subspace span } \{v_1, \ldots, v_{j-1}\} \]

• **Preconditioning:**

1. If \(\sigma_j\) is close to \(\sigma_{j+1}\), can re-use preconditioners for different linear systems

2. Cost of recomputing vs cost of using a close-by preconditioner
Problem: Given a stable dynamical system $\mathbf{H}(s)$, find a reduced model $\mathbf{H}_r(s)$ that satisfies

$$\mathbf{H}_r(s) = \arg \min_{\deg(\hat{\mathbf{H}}) = r, \hat{\mathbf{H}}: \text{stable}} \| \mathbf{H}(s) - \hat{\mathbf{H}}(s) \|_{\mathcal{H}_2}.$$ 

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

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

- 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. $V = \text{Span} \left[ (\sigma_1 I - A)^{-1} b, \ldots, (\sigma_r I - A)^{-1} b \right]$
3. $W = \text{Span} \left[ (\overline{\sigma}_1 I - A^T)^{-1} c^T, \ldots, (\overline{\sigma}_r I - A^T)^{-1} c^T \right]$, $W^T V = I_r$
4. while [relative change in $\sigma_j$] $> \epsilon$
   (a) $A_r = W^T A V$
   (b) $\sigma_i \leftarrow -\lambda_i(A_r)$ for $i = 1, \ldots, r$
   (c) $V = \text{Span} \left[ (\sigma_1 I - A)^{-1} b, \ldots, (\sigma_r I - A)^{-1} b \right]$
   (d) $W = \text{Span} \left[ (\overline{\sigma}_1 I - A^T)^{-1} c^T, \ldots, (\overline{\sigma}_r I - A^T)^{-1} c^T \right]$, $W^T V = I_r$
5. $A_r = W^T A V$, $b_r = W^T b$, $c_r = c V$

• Upon convergence, first-order conditions satisfied via Krylov projection framework, no Lyapunov solvers
Inexact IRKA (I-IRKA)

- IRKA requires solving $2r$ linear systems at each step
  $\Rightarrow$ Expensive if $n = \mathcal{O}(10^6)$
- In most cases $\{\sigma_j\}$ converge fast
  \[\downarrow\]
- Use the solution from the previous step as an initial guess for the next step
- Expect faster convergence for a fixed tolerance value
- Optimal reduced model: Expect robustness
An Inexact Iterative Rational Krylov Algorithm (I-IRKA):

1. Make an initial shift selection $\sigma_i$ for $i = 1, \ldots, r$

2. for $i = 1, \ldots, r$
   (a) $v_i = f(A, b, \sigma_i, 0, \epsilon)$ (f(A, b, $\sigma$, 0, $\epsilon$): an iterative solve for $(\sigma I - A)x = b$ )
   (b) $w_i = f(A^T, c^T, \sigma_i, x_0, \epsilon)$

3. $W = \begin{bmatrix} w_1, w_2, \ldots, w_r \end{bmatrix}$, $V = \begin{bmatrix} v_1, v_2, \ldots, v_r \end{bmatrix}$.

4. $W = W(W^T V)^{-T}$ (to make $W^T V = I_r$)

5. while (not converged)
   (a) $A_r = W^T A V$,
   (b) $\sigma_i \leftarrow -\lambda_i(A_r)$ for $i = 1, \ldots, r$
   (c) for $i = 1, \ldots, r$
      i. $v_i = f(A, b, \sigma_i, v_i, \epsilon)$
      ii. $w_i = f(A^T, c^T, \sigma_i, w_i, \epsilon)$
   (d) $W = \begin{bmatrix} w_1, w_2, \ldots, w_r \end{bmatrix}$, $V = \begin{bmatrix} v_1, v_2, \ldots, v_r \end{bmatrix}$.
   (e) $W = W(W^T V)^{-T}$ (to make $W^T V = I_r$)

6. $A_r = W^T A V$, $b_r = W^T b$, $c_r = c V$
• Same model with $n = 79,841$ (Finer discretization)
• $r = 6$ via IRKA and $I - IRKA$ ($tol = 5 \times 10^{-5}$)
• IRKA: Initial guess from the previous step

\[
\|H(s) - H_1(s)\|_\infty = \|H(s) - H_2(s)\|_\infty = 6.01 \times 10^{-5},
\]
\[
\|H_1(s) - H_2(s)\|_\infty = 3.01 \times 10^{-5}.
\]
Conclusions and Future Work

- \( n >> 10^6 \): 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