Projection methods for model reduction of large-scale dynamical systems

Serkan Gugercin
Dept. of Math., Virginia Tech., USA

jointly with **Athanasios C. Antoulas**
ECE Dept., Rice University, USA

and **Danny Sorensen**
CAAM Dept., Rice University, USA

Oberwolfach, October 2003

Introduction

- Linear time invariant (LTI) systems:

$$\mathbf{\Sigma} : \begin{cases} \dot{x}(t) & = & Ax(t) + Bu(t) \\ y(t) & = & Cx(t) + Du(t) \end{cases} \Leftrightarrow \mathbf{\Sigma} = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \in \mathbb{R}^{(n+p)\times(n+m)}$$

- In many applications, n is quite large and $m, p \ll n$. $\Longrightarrow$

- $\mathbf{\Sigma}_r := \left[\begin{array}{c|c} A_r & B_r \\ \hline C_r & D_r \end{array}\right] \in \mathbb{R}^{(r+p)\times(r+m)}$ with $r \ll n$ such that

  1. $\|y - y_r\|$ is *small*, and there exists a *global* error bound.
  2. System properties, like *stability*, *passivity* are preserved.
  3. The procedure is *computationally efficient*.

- Model reduction through projection: a unifying framework.
- Construct $\Pi = VZ^T$, where $V, Z^T \in \mathbb{R}^{n\times r}$ with $Z^TV = I_r$:

$$\dot{x}_r = \underbrace{Z^TAV}_{:=A_r} x_r(t) + \underbrace{Z^TB}_{:=B_r} u(t), \quad y_r(t) = \underbrace{CV}_{:=C_r} x_r(t) + \underbrace{D}_{:=D_r} u(t)$$

Model Reduction Methods:

- SVD Based Methods: Lyapunov Balancing
  - A modified low-rank Smith method
    (Gugercin/Sorensen/Antoulas [2002])
- Krylov Based Methods: Lanczos, Arnoldi, Rational Krylov
  - How to choose the interpolation points (shifts)?
    (Gugercin/Antoulas [2003])
- SVD-Krylov based methods
  - Least-Squares Reduction (Gugercin/Antoulas [2003])

Balanced Truncation

- $\mathcal{P}$ and $\mathcal{Q}$: the unique Hermitian positive definite solutions to

$$A\mathcal{P} + \mathcal{P}A^T + BB^T = 0, \quad A^T\mathcal{Q} + \mathcal{Q}A + C^TC = 0$$

- The Hankel singular values of $\mathbf{\Sigma}$: $\sigma_i(\mathbf{\Sigma}) := \sqrt{\lambda_i(\mathcal{P}\mathcal{Q})}$.
- $\mathcal{P} = UU^T$, $\mathcal{Q} = LL^T \Rightarrow U^TL = ZSY^T$ $S = diag(\sigma_1, \ldots, \sigma_n)$.
- $W_r := LY_rS_r^{-1/2}$, $V_r := UZ_rS_r^{-1/2}$. $V_rW_r^T$: <u>oblique projector</u>
- $\mathbf{\Sigma}_r = \left[\begin{array}{c|c} W_r^TAV_r & W_r^TB_r \\ \hline CV_r & D \end{array}\right]$, of order r satisfies:
- $\mathbf{\Sigma}_r$ is stable, balanced and $\|\mathbf{\Sigma} - \mathbf{\Sigma}_r\|_\infty \leq 2(\sigma_{r+1} + \cdots + \sigma_n)$.
- Requires U and L $\Longrightarrow$ $\mathcal{O}(n^3)$.

Iterative implementations of balancing

- Sparsity

(a) The ADI Iteration

- $\omega(\lambda) = \frac{\mu^* - \lambda}{\mu + \lambda}$, $\mu \in \mathbb{C}_-$: Continuous time $\Rightarrow$ Discrete time

- $\mathcal{P}_i^A = (A - \mu_i^* I)(A + \mu_i I)^{-1} \mathcal{P}_{i-1}^A [(A - \mu_i^* I)(A + \mu_i I)^{-1}]^* - 2\rho_i (A + \mu_i I)^{-1} B B^T (A + \mu_i I)^{-*}$

- $\rho_{ADI} = \rho \left(\prod_{i=1}^l (A - \mu_i^* I)(A + \mu_i I)^{-1} \right) < 1$

- $\|\mathcal{P} - \mathcal{P}_l^A\|_F \leq K \rho_{ADI}^2 \|\mathcal{P}\|_F$, where $A = V \Lambda V^{-1}$, $K = (\kappa(V))^2$.

- Smith's Method: uses one shift, i.e. $\mu_1 = \cdots = \mu_k$.

- Smith(ℓ): uses ℓ cyclic shift: $\mu_{i+j\ell} = \mu_i$. Faster convergence

- All the above methods have storage requirement $\mathcal{O}(n^2)$.
- Low-Rank versions $\Longrightarrow$ Storage: $\mathcal{O}(nr)$.

(b) LR-Smith(l) Iteration (Li/White [1999], Penzl [2000])

- $A_d = \prod_{i=1}^{\ell}(A - \mu_i I)(A + \mu_i I)^{-1}$ and $B_d B_d^T = P_\ell^A$
- Initialized by $Z_1^{Sl} = B_d$, and followed by
- $Z_k^{Sl} = [\ B_d \ \ A_d B_d \ \ A_d^2 B_d \ \ \cdots \ \ A_d^{k-1} B_d\ ]$. $\mathcal{P}_k^{Sl} = Z_k^{Sl}(Z_k^{Sl})^T$.

Remarks:

- Z_k^{Sl}: $m \times \ell \times k$ columns.
- For $\rho(A_d) \approx 1$ and/or $m > 1$, the method can fail.
- $\mathbf{\Sigma} = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]^{(n+p)\times(n+m)} \Rightarrow \mathbf{\Sigma}_d = \left[\begin{array}{c|c} A_d & B_d \\ \hline C_d & D_d \end{array}\right]^{(n+\ell p)\times(n+\ell m)}$

(c) The Modified LR-Smith(ℓ) Iteration

- Given tolerance τ, replace Z_k^{Sl} with best low-rank approximant.
- $Z_k^{Sl} = V\Sigma W^T \Longrightarrow \tilde{Z}_k = V\Sigma$ is also a square root factor for $\mathcal{P}_k^{Sl}$.
- At the $(k+1)^{st}$ step, $Z_{k+1}^{Sl} = [Z_k^{Sl} \quad A_d^k B_d]$.
- Find $A_d^k B_d = V\Gamma + \hat{V}\Theta$ where $V^T\hat{V} = 0$ and $\hat{V}^T\hat{V} = I$.
- $\hat{Z}_{k+1} = [\ V \quad \hat{V}\ ] \begin{bmatrix} \Sigma & \Gamma \\ 0 & \Theta \end{bmatrix} = [\ V \quad \hat{V}\ ] T\hat{\Sigma}Y^T \Longrightarrow$
- $\tilde{Z}_{k+1} = \tilde{V}\hat{\Sigma}$ where $\tilde{V} = [\ V \quad \hat{V}\ ]T$
- $\tilde{Z}_{k+1} = \tilde{V}\hat{\Sigma} = [\ \tilde{V}_1 \quad \tilde{V}_2\ ]\mathrm{diag}(\hat{\Sigma}_1, \hat{\Sigma}_2)$ so that $\frac{\hat{\Sigma}_2(1,1)}{\hat{\Sigma}_1(1,1)} < \tau$.
- $\tilde{Z}_{k+1} \approx \tilde{V}_1\hat{\Sigma}_1$ and $\tilde{\mathcal{P}}_{k+1} = \tilde{Z}_{k+1}(\tilde{Z}_{k+1})^T$.

- $0 \leq \mathbf{Tr}\,(\mathcal{P}_k^{Sl} - \tilde{\mathcal{P}}_k) \leq K\tau^2 \sum n_i(\sigma_{max}(\tilde{Z}_i))^2$

- $0 \leq \mathbf{Tr}\,(\mathcal{P} - \tilde{\mathcal{P}}_k) \leq K\,m\,l\,(\rho(A_d))^{2k}\mathbf{Tr}\,(\mathcal{P}) + K\tau^2 \sum n_i(\sigma_{max}(\tilde{Z}_i))^2$

- $0 \leq \sum \sigma_i^2 - \sum \tilde{\sigma}_i^2 \leq K\,l\,(\rho(A_d))^{2k}\,(\eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5)$

$\eta_1 = K\ \min(m,p)(\rho(A_d))^{2k}\mathbf{Tr}\,(\mathcal{P})\mathbf{Tr}\,(\mathcal{Q})$

$\eta_2 = m\,\mathbf{Tr}\,(\mathcal{P}) \sum_{i=0}^{k-1} \|C_d A_d^j\|_2^2, \qquad \eta_3 = p\,\mathbf{Tr}\,(\mathcal{Q}) \sum_{i=0}^{k-1} \|A_d^j B_d\|_2^2$

$\eta_4 = \tau_{\mathcal{P}}^2 \|\mathcal{Q}_k^{Sl}\| \sum n_i^{\mathcal{P}}\ (\sigma_{max}(\tilde{Z}_i))^2$

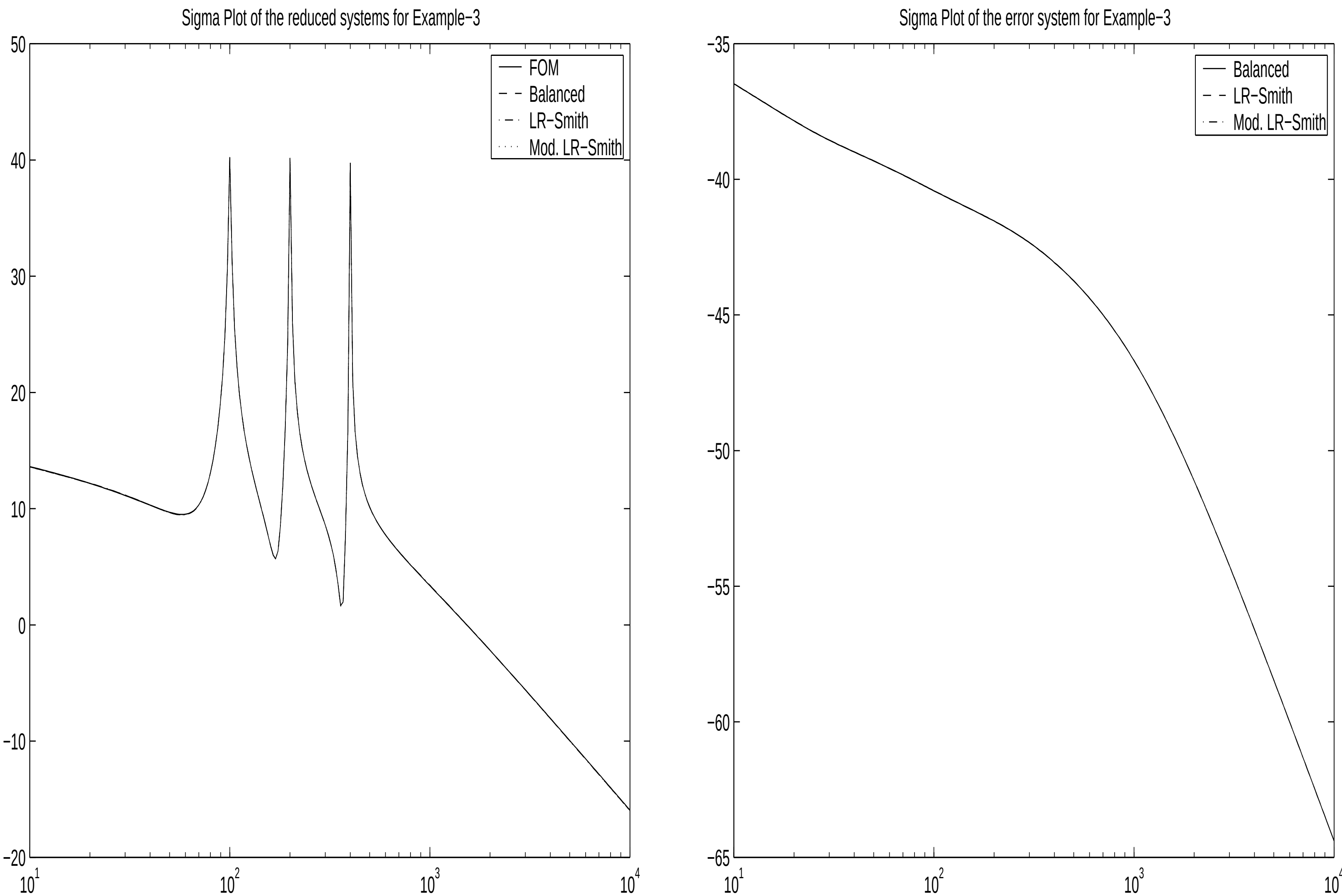
$\eta_5 = \tau_{\mathcal{Q}}^2 \|\mathcal{P}_k^{Sl}\| \sum n_i^{\mathcal{Q}}\ (\sigma_{max}(\tilde{Y}_i))^2$

An Arbitrary Model from Penzl [1999]

- FOM has 1006 states, $m = p = 1$. 1000 real and 6 complex poles.
- 8 shifts chosen. 6 of them are the complex eigenvalues.
- LR-Smith(l) yields Z_k^{Sl} and Y_k^{Sl} with 300 columns.
- Modified Smith yields Z_k^{Sl} and Y_k^{Sl} with 19 columns.

$$\frac{\|\mathcal{P}_k^{Sl} - \tilde{\mathcal{P}}_k\|}{\|\mathcal{P}_k^{Sl}\|} = 1.90 \times 10^{-8}, \text{ and } \qquad \frac{\|\mathcal{Q}_k^{Sl} - \tilde{\mathcal{Q}}_k\|}{\|\mathcal{Q}_k^{Sl}\|} = 3.22 \times 10^{-8}.$$

- Reduce to $r = 11$ using $\underbrace{(\mathcal{P}, \mathcal{Q})}_{\Sigma_r}$, $\underbrace{(\mathcal{P}_k^{Sl}, \mathcal{Q}_k^{Sl})}_{\Sigma_r^{Sl}}$, and $\underbrace{(\tilde{\mathcal{P}}_k, \tilde{\mathcal{Q}}_k)}_{\tilde{\Sigma}_r}$


Sigma Plot of the reduced systems for Example−3
FOM
Balanced
LR−Smith
Mod. LR−Smith
50
40
30
20
10
0
−10
−20
10^1
10^2
10^3
10^4
Sigma Plot of the error system for Example−3
Balanced
LR−Smith
Mod. LR−Smith
−35
−40
−45
−50
−55
−60
−65
10^1
10^2
10^3
10^4

(a) (b)

Heat Distribution on a Plate

- 2-dimensional heat equation on a square plate consisting of 9 sub-plates.
- FOM of order 20736 with $m = 9$ and $p = 9$.
- $k = 40$ steps using $l = 2$ shifts.
- Modified Smith yields Z_k^{Sl} and Y_k^{Sl} with 85 columns.

$$\begin{aligned} \|\mathcal{P}_{red} - \operatorname{diag}(\mathcal{P}_{red})\| &= 7.28 \times 10^{-9} \\ \|\mathcal{Q}_{red} - \operatorname{diag}(\mathcal{Q}_{red})\| &= 1.05 \times 10^{-12} \\ \|\mathcal{P}_{red} - \mathcal{Q}_{red}\| &= 1.04 \times 10^{-9}. \end{aligned}$$

- Reduce the order to $r = 11$.

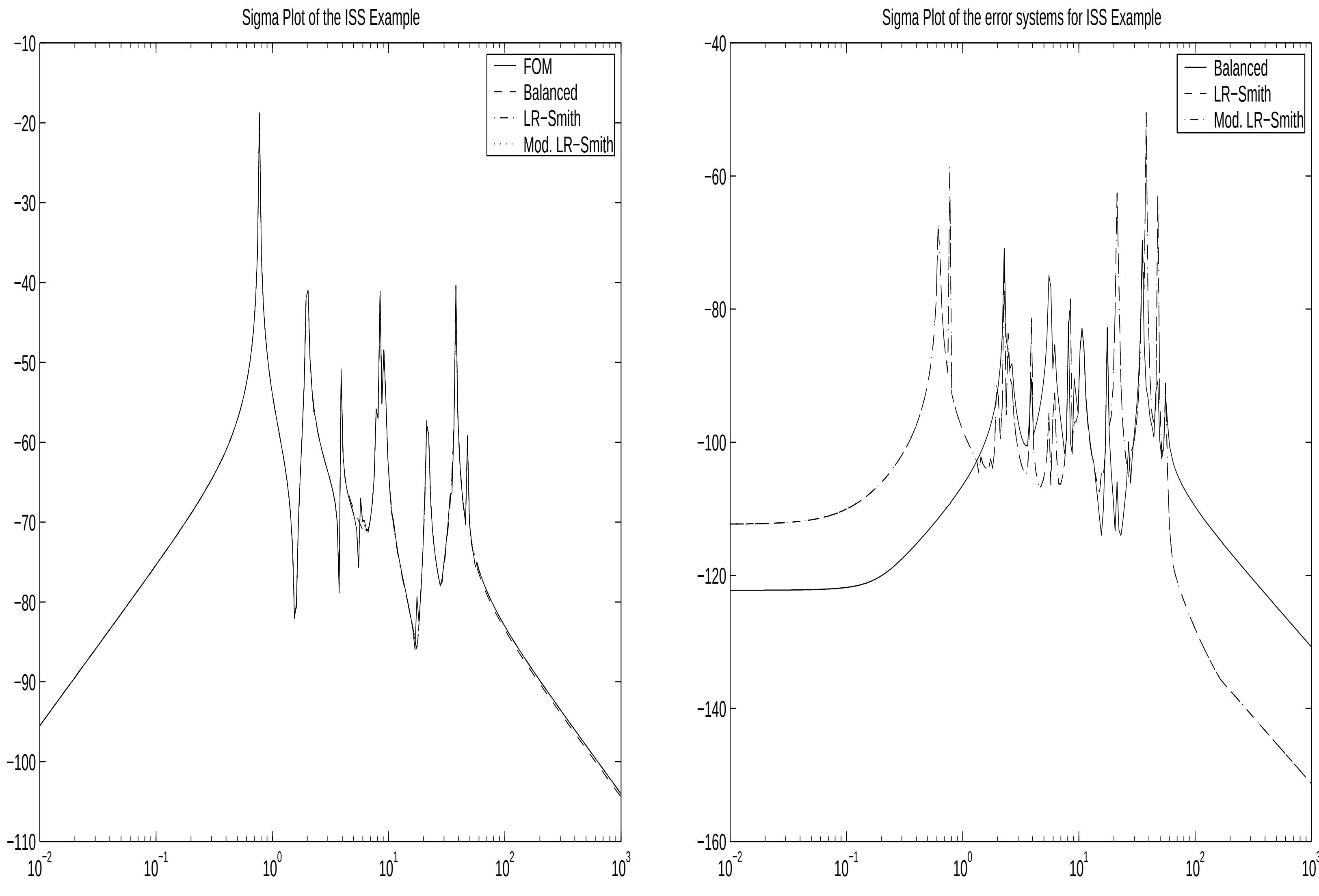
ISS 1R Model

- $n = 270$, $m = 3$ and $p = 3$.
- Even $l = 20$, yields $\rho = 0.9973$.
- Single shift with $k = 70$ iterations.
- Z_k^{Sl} and Y_k^{Sl} has 210 columns. $\tilde{Z}_k$ and $\tilde{Y}_k$ have 106 columns.

$$\frac{\|\mathcal{P}_k^{Sl} - \tilde{\mathcal{P}}_k\|}{\|\mathcal{P}_k^{Sl}\|} = 4.59 \times 10^{-8}$$

$$\frac{\|\mathcal{Q}_k^{Sl} - \tilde{\mathcal{Q}}_k\|}{\|\mathcal{Q}_k^{Sl}\|} = 2.18 \times 10^{-8}.$$

- Reduce the order to 26.


Sigma Plot of the ISS Example
FOM
Balanced
LR−Smith
Mod. LR−Smith
Sigma Plot of the error systems for ISS Example
Balanced
LR−Smith
Mod. LR−Smith

(a) (b)

Model reduction by Krylov Projection

- Expand $H(s) = C(sI - A)^{-1}B$ around some $\sigma \in \mathbb{C}$:

$$H(s) = \eta_\sigma^{(0)} + \eta_\sigma^{(1)}(\sigma - s) + \eta_\sigma^{(2)}(\sigma - s)^2 + \eta_\sigma^{(3)}(\sigma - s)^3 + \cdots.$$

- $\eta_\sigma^{(j)}$: the j^{th} *moment* of $\mathbf{\Sigma}$ at σ, for $j \geq 0$.

- **The moment matching problem:** Find $\mathbf{\Sigma}_r = \left[\begin{array}{c|c} A_r & B_r \\ \hline C_r & 0 \end{array}\right]$ with

$$H_r(s) = \hat{\eta}_\sigma^{(0)} + \hat{\eta}_\sigma^{(1)}(\sigma - s) + \hat{\eta}_\sigma^{(2)}(\sigma - s)^2 + \hat{\eta}_\sigma^{(3)}(\sigma - s)^3 + \cdots.$$

such that for an appropriate r,

$$\boxed{\eta_\sigma^{(j)} = \hat{\eta}_\sigma^{(j)} \quad j = 0, 1, 2, \cdots, r.}$$

- Moments are ill-conditioned to compute
- Krylov projection: Solves the moment matching problem recursively in a numerical reliable way.
- Construct V and Z such that they span (union of) certain Krylov subspaces.
- Then the reduced model

$$\mathbf{\Sigma}_r = \left[\begin{array}{c|c} A_r & B_r \\ \hline C_r & 0 \end{array}\right] = \left[\begin{array}{c|c} Z^T A V & Z^T B \\ \hline C V & 0 \end{array}\right]$$

solves the moment matching problem.

- For $\sigma = \infty$
  - $\eta_\sigma^{(j)} = CA^{j-1}B$ and is called the 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 $\Longrightarrow$ *rational Krylov method*
- Krylov based methods
  - Moment matching without moment computation
  - Iterative implementation
- Numerically reliable and efficient.

Rational Krylov Method

- Lanczos/Arnoldi: Moment matching at a single point.
- Large error around other frequencies
- Match the moments around various frequencies $\Rightarrow$
- Better approximation over a broad range
- **Problem:** Given $\mathbf{\Sigma}$ with $\eta_{\sigma_k}^{(j)}$, find $\mathbf{\Sigma}_r$ with $\hat{\eta}_{\sigma_k}^{(j)}$ such that

$$\boxed{\eta_{\sigma_k}^{(j)} = \hat{\eta}_{\sigma_k}^{(j)}, \quad j = 1, \cdots, J_k \text{ and } k = 1, \cdots, K.}$$

- Solution by projection: First by Skelton and De Villemagne [1987]
- Grimme [1997], Van Dooren *et. al.* [1998, 1999]:
  Efficient computation by Krylov projection.

$$\bigcup_{k=1}^{K} \mathcal{K}_{b_k}\left((\sigma_k I - A)^{-1}, (\sigma_k I - A)^{-1} B\right) \subseteq \mathcal{V} = \mathrm{Im}(V)$$

$$\bigcup_{k=1}^{K} \mathcal{K}_{c_k}\left((\sigma_k I - A)^{-T}, (\sigma_k I - A)^{-T} C^T\right) \subseteq \mathcal{Z} = \mathrm{Im}(Z)$$

- Efficient computations of V and Z: Dual Rational Arnoldi, Rational Lanczos, Rational Power Krylov. (Grimme [1997])
- MIMO case: Deflation must be employed in forming *Block* Krylov subspaces: Boley [1994], Freund *et al.* [2000], Bai *et al.* [1999].

Advantages of Krylov based methods:

- Iterative implementation
- Only matrix-vector multiplications and sparse LU decompositions

$$\Downarrow$$

- $\mathcal{O}(\alpha r^2 n)$ arithmetic operation and $\mathcal{O}(nr)$ storage requirements

Drawbacks:

- Stability is not guaranteed: Remedy: *Implicit Restart*, but !!!
- No global error bound
- Selection of σ_i is an ad-hoc process.

Least Squares Reduction: An SVD-Krylov Method

- Current research in model reduction: Combine SVD and Krylov based methods
- Jaimoukha and Kasenally [1997], Penzl [1999,2000], Antoulas and Sorensen [2001], ...
- Proposed method: A two sided projection method, $\Pi = V Z^T$
- V: Krylov subspace, Z^T: Based on $\mathcal{Q}$ (SVD side)
- Main motivation: Combine local behavior (moment matching) with global properties, such as stability and error criterion.
- Approximately match $K > 2r$ moments.
- Related to Prony's method for IIR filter design.

Prony's Method for IIR filter design

- Design an IIR filter with response $h(n)$ $n = 0, \cdots, K$.
- $H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_M z^{-M}}{1 + a_1 z^{-1} + \cdots + a_N z^{-N}}$, $N \geq M$
- Define $\mathbf{b} := \begin{bmatrix} b_0 & \cdots & b_M \end{bmatrix}^T$, $\mathbf{a} := \begin{bmatrix} a_1 & \cdots & a_N \end{bmatrix}^T$.
- Using the first $K+1$ terms of $h(n)$, one can write

$$\left[\begin{array}{c} \mathbf{b} \\ \hline 0 \end{array}\right] + \left[\begin{array}{c} \mathbf{e}_1 \\ \hline \mathbf{e}_2 \end{array}\right] = \underbrace{\left[\begin{array}{c|c} \multicolumn{2}{c}{H_1} \\ \hline h_{21} & H_2 \end{array}\right]}_{:=\mathbf{H}} \left[\begin{array}{c} 1 \\ \hline \mathbf{a} \end{array}\right] = \mathbf{H}\,\bar{\mathbf{a}}.$$

- $h_{21} \in \mathbb{R}^{K-M}$, $H_1 \in \mathbb{R}^{(M+1)\times(N+1)}$, and $H_2 \in \mathbb{R}^{(K-M)\times(N)}$.

- $\boxed{1) \quad \mathbf{b} + \mathbf{e}_1 = H_1 \bar{\mathbf{a}}}$ and $\boxed{2) \quad \mathbf{e}_2 = h_{21} + H_2 \mathbf{a}}$

- If $K = N + M \quad \Rightarrow \quad H_2$ is square $\Rightarrow$

  $\mathbf{e}_1 = \mathbf{e}_2 = 0 \quad \Rightarrow$ Partial Realization

- If $K > N + M$,
  - To find $\mathbf{a}$, LS on 2) to minimize $\|\mathbf{e}_2\|^2$
  - Plug $\mathbf{a}$ into 1) to make $\mathbf{e}_1 = 0$.

- Define $\varepsilon(n) := h(n) - \hat{h}(n). \quad \Rightarrow \quad \boxed{\mathbf{e} = \mathbf{A}_c \varepsilon}$ Weighted error

- A_c is the $(K+1) \times (K+1)$ convolution matrix.

- The weighted error $\|\mathbf{e}\|$ is minimized.

Least-squares reduction in Discrete time

- Apply to approximate dynamical systems, not the time signals
- $N = M = r =$ the reduced order and $K > 2r$.
- Exactly match the first r Markov parameters
- Approximately match $(r+1)^{\text{st}}$ to K^{th} Markov parameters
- But, first need to compute the K parameters explicitly
- Not suitable for large scale settings.
- Reformulate the problem using state-space as $K \to \infty$

- Recall that $\eta_i = CA^{i-1}B$ and $\mathcal{R}_r = [\ B \quad AB \quad \cdots \quad A^{r-1}B]$.

- $$\mathcal{H}_r = \begin{bmatrix} \eta_1 & \eta_2 & \cdots & \eta_r \\ \vdots & \vdots & & \vdots \\ \eta_r & \eta_{r+1} & \cdots & \eta_{2r-1} \\ \vdots & \vdots & & \vdots \end{bmatrix}, \quad \mathbf{h}_{r+1} = \begin{bmatrix} \eta_{r+1} \\ \vdots \\ \eta_{2r} \\ \vdots \end{bmatrix}$$

- Solution by solving the LS problem: $\mathcal{H}_r \mathbf{x}_{LS} = \mathbf{h}_{r+1} + e_{LS}$

- $$\mathbf{\Sigma}_{LS} = \left[\begin{array}{c|c} Z^T A V & Z^T B \\ \hline CV & D \end{array}\right], \quad Z^T = (\mathcal{R}_r^T \mathcal{Q} \mathcal{R}_r)^{-1} \mathcal{R}_r^T \mathcal{Q}, \quad V = \mathcal{R}_r.$$

- $\mathrm{Im}(V) = \mathcal{R}_r$: Krylov $\qquad \mathrm{Im}(Z^T) = Im(\mathcal{R}_r^T \mathcal{Q})$: SVD

Theorem: Given the above set-up,

1. $\mathbf{\Sigma}_{LS}$ matches the first r Markov parameters of $\mathbf{\Sigma}$.
2. $\mathbf{\Sigma}_{LS}$ is unique, i.e. independent of initial realization.
3. $\mathbf{\Sigma}_{LS}$ is asymptotically stable.
4. $H_{LS}(z) := \frac{B_{LS}(z)}{A_{LS}(z)}$ and $\widehat{H}(z) = \frac{\widehat{B}(z)}{\widehat{A}(z)}$

$$\min_{\widehat{A}(z)} \left\| \widehat{A}(z) \left(H(z) - \widehat{H}(z) \right) \right\|_{\mathcal{H}_2} = \| A_{LS}(z) \left(H(z) - H_{LS}(z) \right) \|_{\mathcal{H}_2}$$

- Krylov based methods: Exactly match the first $2r$ moments, but do not state anything for the rest.
- LS Method: Exactly match the first r Markov parameters, and a weighted least-squares minimization for the rest.
- Global information $\Rightarrow$ Stability
- Related to Prony's method, BUT, by letting $K \to \infty$,
  1. No explicit computation of the moments
  2. Guaranteed asymptotic stability
  3. Minimization of the weighted $\mathcal{H}_2$ error

Moment matching at arbitrary interpolation points

- To have a better response over a broader frequency range
- L interpolation points σ_j with multiplicities b_j, $j = 1, \ldots, L$,
- Construct V_{RL} such that

$$\mathrm{Im}(V_{RL}) = \bigcup_{k=1}^{L} \mathcal{K}_{b_k}((\sigma_k I - A)^{-1}, (\sigma_k I - A)^{-1} b)$$

- $\Pi_{RL} = V_{RL} Z_{RL}^T$ where $Z_{RL}^T = (V_{RL}^T Q V_{RL})^{-1} V_{RL} Q$.

Theorem: Let $\mathbf{\Sigma}_{RL}$ be obtained by Π_{RL}.

1. $\mathbf{\Sigma}_{RL}$ is asymptotically stable,
2. $\mathbf{\Sigma}_{RL}$ matches b_j moments of $\mathbf{\Sigma}$ at σ_j for $j = 1, \ldots, L$.

The least-squares reduction in continuous time

- Bilinear ($z = \frac{1+s}{1-s}$) and inverse bilinear ($s = \frac{z-1}{z+1}$) trans.

$$A_c, B_c, C_c, D_c \quad \overset{z=\frac{1+s}{1-s}}{\longrightarrow} \quad \left\{ \begin{array}{l} A_d = (I + A)(I - A)^{-1} \\ B_d = \sqrt{2}\,(I - A)^{-1} B \\ C_d = \sqrt{2}\, C(I - A)^{-1} \\ D_d = D + C(I - A)^{-1} B \end{array} \right.$$

$$\left. \begin{array}{l} A_c = (A_d + I)^{-1}(A_d - I) \\ B_c = \sqrt{2}\,(A_d + I)^{-1} B_d \\ C_c = \sqrt{2}\, C_d(A_d + I)^{-1} \\ D_c = D_d - C_d(A_d + I)^{-1} B_d \end{array} \right\} \quad \overset{s=\frac{z-1}{z+1}}{\longleftarrow} \quad A_d, B_d, C_d, D_d.$$

- Given $\mathbf{\Sigma}$, apply bilinear transformation to obtain $\mathbf{\Sigma}_d$.
- Least-squares at the interpolation points $\varsigma_k = \frac{1+\sigma_k}{1-\sigma_k}$ with the multiplicities b_k for $k = 1, \ldots, L$ to obtain $\mathbf{\Sigma}_{RL}$.
- Apply inverse bilinear transformation to obtain $\mathbf{\Sigma}_r$.

Lemma: Let $\mathbf{\Sigma}_r$ be obtained least-squares reduction in continuous time. Then,

1. $\mathbf{\Sigma}_r$ is asymptotically stable
2. $\mathbf{\Sigma}_r$ matches the first b_k moments of $\mathbf{\Sigma}$ at σ_k.

Comparison with Balancing and Rational Krylov:

- Balancing: Uses $\mathcal{P}$ and $\mathcal{Q}$, no moment matching
- Rational Krylov: $2b_k$ moments matched at σ_k, but no global information.
- Least-squares: Uses $\mathcal{Q}$ and matches b_k moments at σ_k.
- Lies between Balancing and Rational Krylov

Large-scale implementation issues

- $\mathcal{Q}$: Approximate using (modified) Smith method: $\tilde{\mathcal{Q}}_k = YY^T$ where $Y \in \mathbb{R}^{n\times k}$.

- $\tilde{Z}^T = (\Omega\Omega^T)^{-1}\Omega Y^T \quad$ where $\quad \Omega := VY \in \mathbb{R}^{k\times n}$

  $\Rightarrow$ No storage of an $n \times n$ matrix.

- $\tilde{\mathcal{Q}}_{k+1} = \tilde{\mathcal{Q}} + (A_d^T)^k C^T C A_d^k$: $(k+1)^{\mathrm{st}}$ Smith iterate. $\Rightarrow$

- $V^T \tilde{\mathcal{Q}}_{k+1} V = V^T \tilde{\mathcal{Q}}_k V + V^T (A_d^T)^k C^T C A_d^k V$: Iterative

- Bilinear transformation: Triangular solvers.

An Arbitrary Model

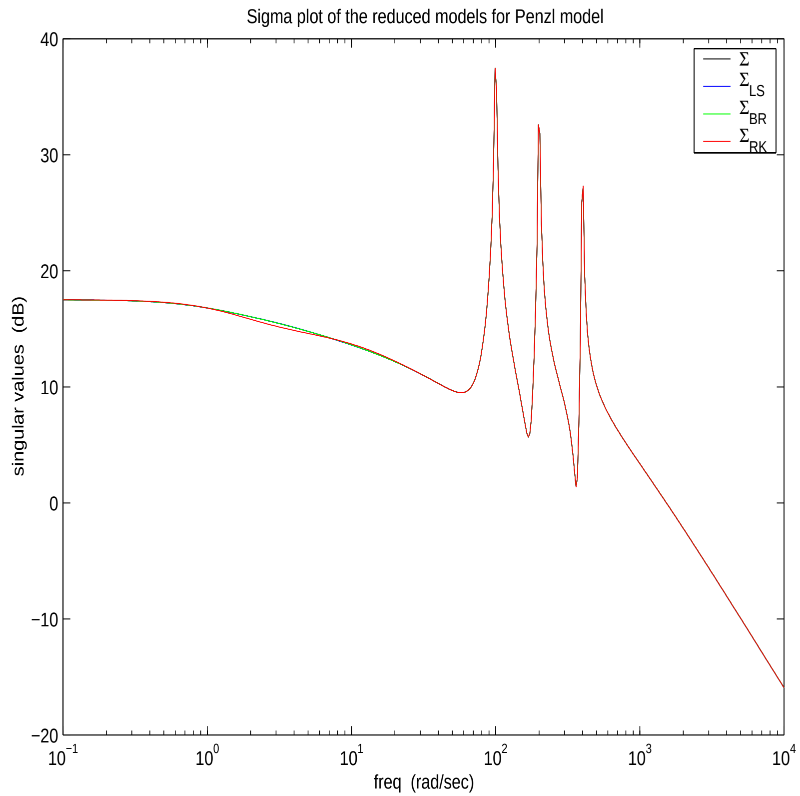
- From Penzl [1999]: $n = 1006$, $m = p = 1$.

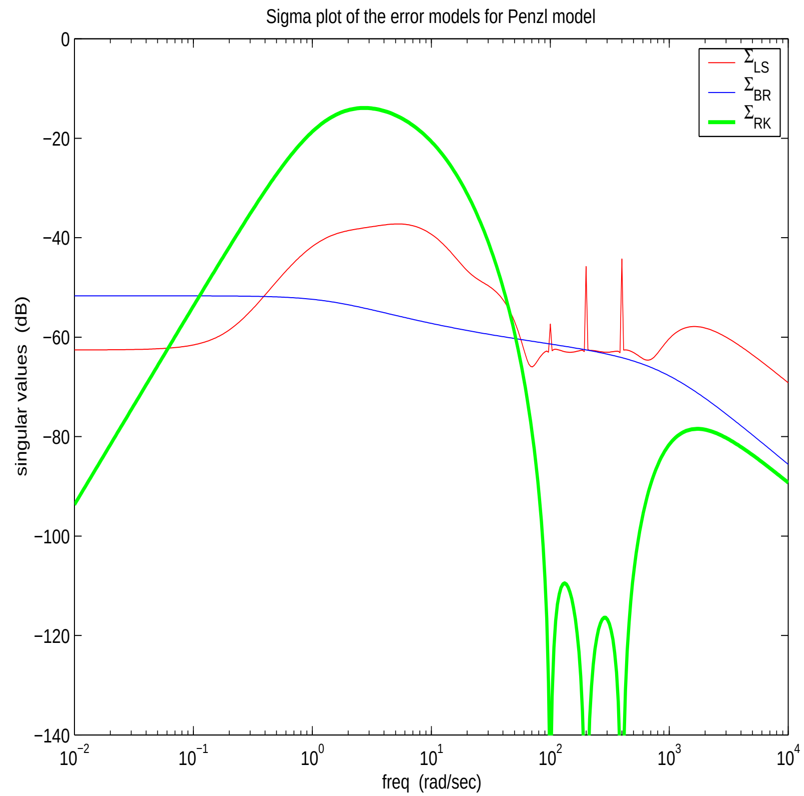
- $B^T = C = [\ \underbrace{10\ \cdots\ 10}_{6}\ \underbrace{1\ \cdots 1}_{1000}]$, $A = \mathrm{diag}(A_1, A_2, A_3, A_4)$ where

$$A_1 = \begin{bmatrix} -1 & 100 \\ -100 & -1 \end{bmatrix},\ A_2 = \begin{bmatrix} -1 & 200 \\ -200 & -1 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} -1 & 400 \\ -400 & -1 \end{bmatrix}, \quad A_4 = diag(-1, \cdots, -1000)..$$

- Reduce to $r = 13$ using **BT** $\to \mathbf{\Sigma}_{BR}$, **LS** $\to \mathbf{\Sigma}_{LS}$, **RK** $\to \mathbf{\Sigma}_{RK}$.

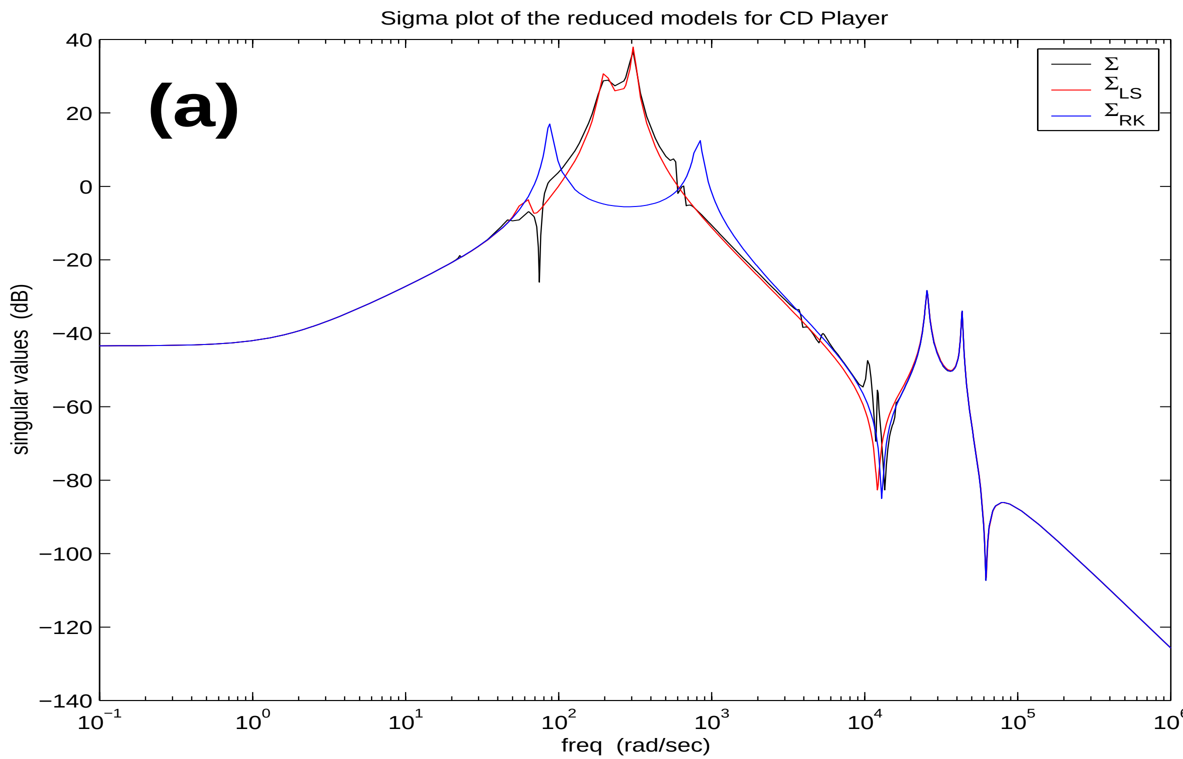




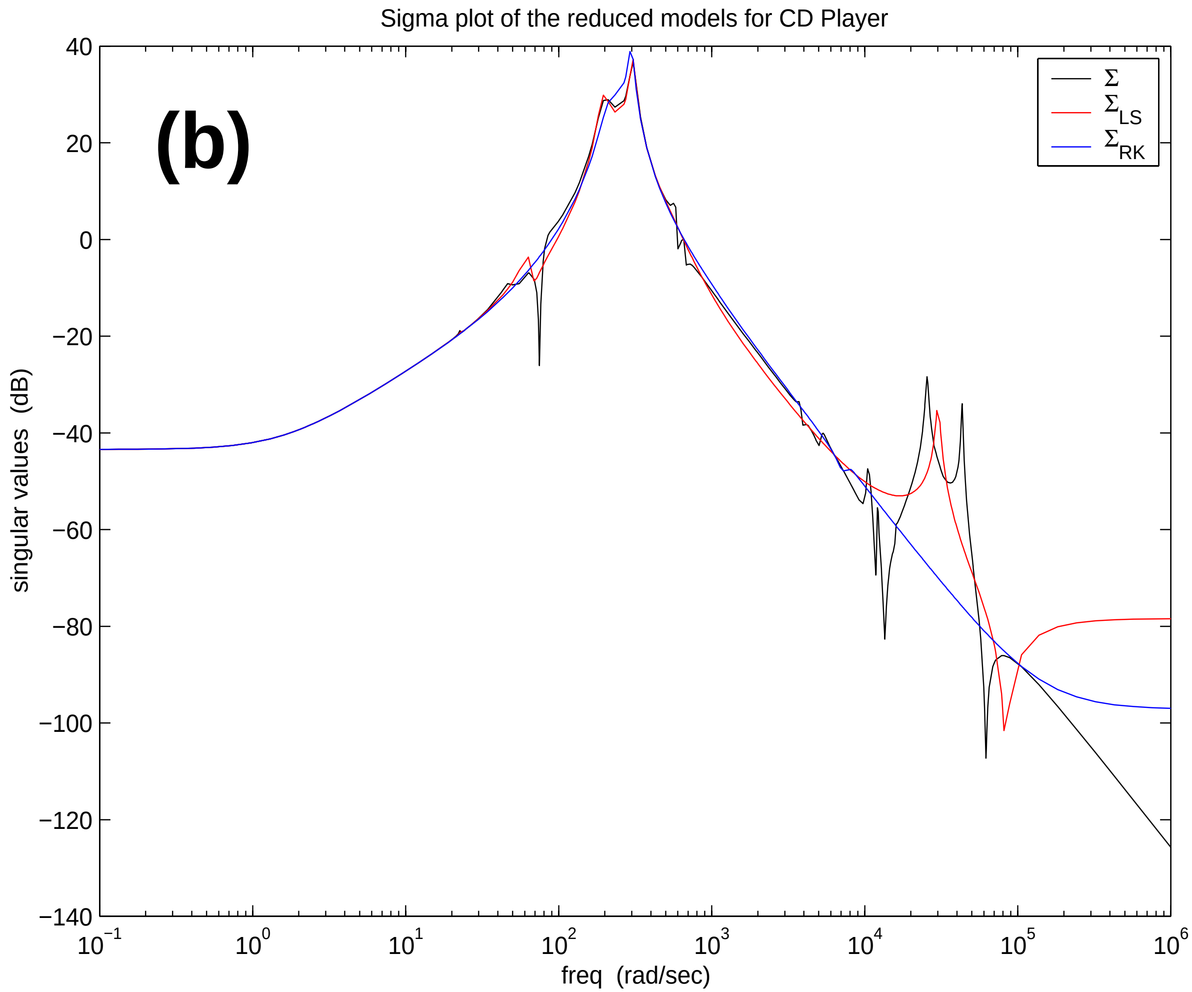
	$\mathbf{\Sigma} - \mathbf{\Sigma}_{LS}$	$\mathbf{\Sigma} - \mathbf{\Sigma}_{BR}$	$\mathbf{\Sigma} - \mathbf{\Sigma}_{RK}$
$\mathcal{H}_\infty$	1.48×10^{-4}	2.82×10^{-5}	2.18×10^{-3}
$\mathcal{H}_2$	5.46×10^{-4}	7.95×10^{-5}	1.58×10^{-3}

CD Player Model

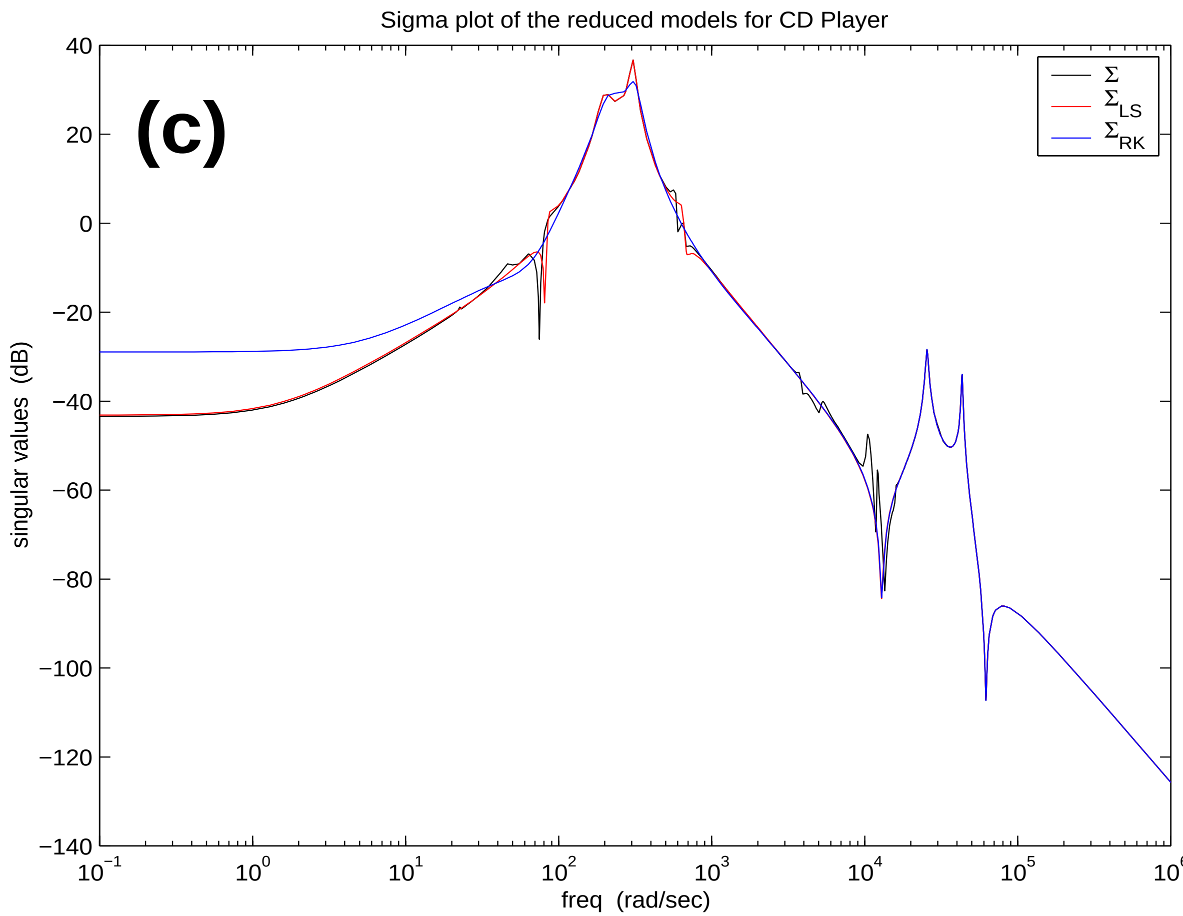
- $n = 120$, $m = p = 1$ and $r = 12$
- Comparison with Rational Krylov
- $\sigma_1 = 0$, $\sigma_2 = \infty$.



$$\sigma_i = \jmath\,\texttt{logspace}(-1, 5, 6)$$



$\sigma_1 = \infty$, $\sigma_2 = 200$



- More robust to changes in σ_i

Choice of interpolation points σ_i

- Choice of σ_i is an ad-hoc process.
- Usually based on the frequency response characteristics
- Intuitively, σ_i relates to $\mathcal{H}_2$ error more than $\mathcal{H}_\infty$ error.

$\mathcal{H}_2$ Norm and $\mathcal{H}_2$ Error Expressions

- $\|\mathbf{\Sigma}\|_{\mathcal{H}_2}^2 = C\mathcal{P}C^T = B^T\mathcal{Q}B$.

Lemma: Let $\lambda_i(A)$ be distinct. Define $\phi_i := H(s)(s-\lambda_i)\mid_{s=\lambda_i}$, residues of $H(s)$. Then

$$\boxed{\|\mathbf{\Sigma}\|_{\mathcal{H}_2}^2 := \sum_{i=1}^{n} \phi_i H(-\lambda_i)}$$

Corollary: Let $\widehat{H}(s)$ be an r^{th} order stable reduced model with distinct reduced poles $\widehat{\lambda}_j(A_r)$. Define $\widehat{\phi}_j := \widehat{H}(s)(s - \widehat{\lambda}_j)\,|_{s=\widehat{\lambda}_j}$.

$$\|\boldsymbol{\Sigma}_e\|_{\mathcal{H}_2}^2 := \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_r\|_{\mathcal{H}_2}^2 = \sum_{i=1}^{n} \phi_i \left(H(-\lambda_i) - \widehat{H}(-\lambda_i) \right) + \sum_{j=1}^{r} \widehat{\phi}_j \left(\widehat{H}(-\widehat{\lambda}_j) - H(-\widehat{\lambda}_j) \right).$$

- Error due to mismatch at $-\lambda_i$ and $-\widehat{\lambda}_j$.
- Try to kill the first term.
- In the rational Krylov method, choose $\boxed{\sigma_i = -\lambda_i(A).}$
- Select $\lambda_i(A)$ with large ϕ_i.

CD Player and Selection of σ_i

- Reduce the order to $r = 14$ by Rational Krylov
- $\mathbf{\Sigma}_r^*$: Choosing $\sigma_i = -\lambda_i(A)$.
- $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3$: Arbitrary σ_i. Best 3 among 2000 models.

	$\mathbf{\Sigma}_r^*$	$\mathbf{\Sigma}_1$	$\mathbf{\Sigma}_2$	$\mathbf{\Sigma}_3$
$\mathcal{H}_2$	4.86×10^{-1}	1.16×10^{0}	1.35×10^{0}	1.56×10^{0}
$\mathcal{H}_\infty$	7.24×10^{-2}	3.21×10^{-1}	3.95×10^{-1}	5.53×10^{-1}

	$\mathbf{\Sigma}_{BR}$	$\mathbf{\Sigma}_r^*$
$\mathcal{H}_2$	1.01×10^{0}	4.86×10^{-1}
$\mathcal{H}_\infty$	3.82×10^{-2}	7.24×10^{-2}

Conclusions and Future Work

SVD-Based Methods

- Modified Cyclic Smith Method
  - Much lower number of columns in the approximant
  - Convergence results. (Almost) no lost of accuracy

Krylov Based Methods

- $\mathcal{H}_2$ error expression for the Lanczos procedure
- Future Direction: How many and what mirror images?

SVD-Krylov Based Methods

- Least-squares model reduction with asymptotic stability, moment matching and error minimization
- Future Direction:
  - Obtaining an error bound
  - Computing $Im(V^T \mathcal{Q})$ without computing $\mathcal{Q}$.