A new framework for $\mathcal{H}_2$-optimal model reduction and its applications

Alessandro Castagnotto

in collaboration with: Heiko Peuscher (né Panzer), Boris Lohmann, Siyang Hu

related publications: “Fast $\mathcal{H}_2$-Optimal Model Order Reduction Exploiting the Local Nature of Krylov-Subspace Methods”, 2016 European Control Conference, Aalborg
Model order reduction (MOR)

\[
\begin{align*}
E \ \frac{dx}{dt} &= A \ x + B \ u \\
\begin{bmatrix} x \\ u \end{bmatrix} &= C \begin{bmatrix} x \\ u \end{bmatrix} + D \ u \\
\end{align*}
\]

\[
\begin{align*}
E_r \ \frac{dx_r}{dt} &= A_r \ x_r + B_r \ u \\
y_r &= C_r \ x_r + D \ u \\
x_r &\in \mathbb{R}^n, \ n \ll N
\end{align*}
\]

- high-fidelity approximation
- preservation of properties
- numerically efficient

Source(s): nasa.gov, wikimedia.org, dailymail.co.uk
Projective MOR

Approximation in the subspace $\mathcal{V} = \text{Im}(V)$

$$x = V \hat{x}_r + e, \quad V \in \mathbb{R}^{N \times n}$$

Petrov-Galerkin Projection:

$$\Pi = EV(W^\top EV)^{-1}W^\top$$

Notation

$$\Sigma = \{E, A, B, C\}$$
$$\Sigma_r = W^\top < \Sigma > V$$

[DeVillemagne/Skelton '87]
Moment matching (rational interpolation)

Moments of a transfer function

\[ G(s) = C(sE - A)^{-1}B \]
\[ = G(s_0 + \Delta s) = -\sum_{i=0}^{\infty} M_i(s_0) \Delta s^i \]

\( s_0 \): Interpolation frequency (shift)
\( M_i(s_0) \): i-th moment about \( s_0 \)

Rational Krylov (RK) subspaces

Choose \( V \) and \( W \) such that:

\[ \mathcal{K}_q \left((A - s_0 E)^{-1} E, (A - s_0 E)^{-1} B\right) \subseteq \text{Im} (V) \]
\[ \mathcal{K}_r \left((A - s_0 E)^{-\top} E^\top, (A - s_0 E)^{-\top} C^\top\right) \subseteq \text{Im} (W) \]

\[ M_i(s_0) = M_{r,i}(s_0) \quad \text{for} \ i = 0, \ldots, q + r - 1 \]

\[ AV - EV S_V = BR \]
\[ A^\top W - E^\top W S_W^\top = C^\top L \]
\[ \lambda_i(S_V) = \lambda_i(S_W) = s_0 \]
\[ R \equiv [I_m, 0, \ldots, 0], \ L \equiv [I_p, 0, \ldots, 0] \]

[Grime '97, Gallivan/ '04a]
\( H_2 \)-optimal model order reduction

\[
\|G - G_r\|_{H_2} = \min_{\text{dim}(\tilde{G}_r)=n} \|G - \tilde{G}_r\|_{H_2}
\]

Iterative Rational Krylov Algorithm

\[
\Sigma := \{E, A, B, C\} \quad \Sigma_r := \{E_r, A_r, B_r, C_r\}
\]

**Algorithm** Iterative Rational Krylov Algorithm (IRKA)

**Input:** \( \Sigma, s_0, \text{tol} \)

**Output:** locally \( H_2 \)-optimal reduced model \( \Sigma_r, s_0^\ast \)

1. while relative change of \( s_0 < \text{tol} \) do
2. \( \Sigma_r \leftarrow \text{RK}(\Sigma, s_0) \) // Hermite reduction
3. \( s_0 \leftarrow -\lambda(\Sigma_r) \) // mirror reduced eigenvalues
4. end while
5. \( s_0^\ast \leftarrow s_0 \) // return optimal shifts

Gradient-based methods

Expressions for gradient and Hessian can be derived and used for trust-region optimization

[Meier/Luenberger ’67, Gugercin/Antoulas/Beattie ’08, Gugercin/Beattie ’12, Panzer/Jaensch/Wolf/Lohmann ’14]
Agenda

A new framework for $H_2$-optimal MOR and its applications

- Model order reduction by projection
- Rational interpolation and $H_2$-optimal reduction
- What is the cost of $H_2$-optimal reduction?
- A new framework to exploit the local nature of rational interpolation
- Numerical results and further applications
The cost of $\mathcal{H}_2$-optimal reduction

Fundamental difficulty: Complexity of sparse direct/iterative methods

Complexity of reduction...

\[ C_{N,n} (\text{RK}) \approx n \cdot C_N (\text{LSE}) \]
\[ C_{N,n} (\text{IRKA}) \approx n \cdot C_N (\text{LSE}) \cdot k_{IRKA} \]

Is there a way of making the optimization cost independent of $\alpha$? [Golub '12]

...for a specific LSE

\[ C_{N,n} (\text{RK}) \approx \alpha \]
\[ C_{N,n} (\text{IRKA}) \approx \alpha \cdot k_{IRKA} \]

Cost of finding optimal solution
Model function framework – the main idea

What does this have to do with model reduction?

- Rational interpolation is intrinsically **local** in nature (cp. moment matching)
- All $H_2$-optimal methods can guarantee only **local** optimality at convergence

These properties are exploited in the **model function** framework!
Model functions – exploit local nature of method

\[ \Sigma \]

\[ \Sigma k_m \]

\[ \Sigma m \]

\[ \Sigma n_m \]

\[ \Sigma r \]

\[ \Sigma m, r \]

[Panzer '14, C./Panzer/Lohmann '16]
Model functions – exploit local nature of method

The model function framework as...

- **...a surrogate optimization approach**
  \[ \Sigma_m \] is used to build a surrogate of the cost function
  \[ \|G - G_r\|_{\mathcal{H}_2} \approx \|G_m - G_r\|_{\mathcal{H}_2} \]

- **...a shift optimization within a subspace**
  \[ \Sigma_m = \{E_m, A_m, B_m, C_m\} \] results from the projection
  \[ W_m^T < \Sigma > V_m \], where the subspaces \( \mathcal{R}(W_m) \) and \( \mathcal{R}(V_m) \) are constantly updated

Can we claim anything about the optimality of \( \Sigma_{m,r} \)?
Result 1: Let $\Sigma_m$ be a Hermite interpolant of $\sum$ at $s_0^{m,j}$, define $S^m := \{s_0^{m,j}\}_j$. Let $\Sigma_{m,r}$ be an $\mathcal{H}_2$-optimal approximation of $\Sigma_m$ with optimal shifts $s_{0,i}^*$, let $S^* := \{s_{0,i}^*\}_i$. Then $\Sigma_{m,r}$ is an $\mathcal{H}_2$-optimal approximation of $\sum$ if $S^* \subseteq S^m$.

Update of the model function $\Sigma_m$

Central to the optimality proof is the update of $\Sigma_m$.

At convergence it must hold: $S^* \subseteq S^m$

Recall:

\[
G(-\lambda_{r,i}) = G_r(-\lambda_{r,i}) \\
G'(-\lambda_{r,i}) = G'_r(-\lambda_{r,i})
\]

Proof:

Assumptions:

1. $\mathcal{H}_2$-optimal
\[
G_m(-\lambda_{m,r,i}) = G_{m,r}(-\lambda_{m,r,i}) \\
G'_m(-\lambda_{m,r,i}) = G'_{m,r}(-\lambda_{m,r,i})
\]

2. Update of the model function
\[
G(s_{0,i}^*) = G_m(s_{0,i}^*) = G_{m,r}(s_{0,i}^*) \\
G'(s_{0,i}^*) = G'_m(s_{0,i}^*) = G'_{m,r}(s_{0,i}^*)
\]

Note: analogously for MIMO systems

[C./Panzer/Lohmann '16]
Result 2: Let all assumptions of result 1 hold, in particular $S^* \subseteq S^m$. Let $\Sigma_r$ be a Hermite interpolant of $\Sigma$ at $s_{0,i}^*$. Then $\Sigma_{m,r} = \Sigma_r$.

Sketch of proof:

$$\Sigma_m = W_m^T < \Sigma > V_m$$

$$\Sigma_{m,r} = (W_m W_{m,r})^T < \Sigma > V_m V_{m,r}$$

$$\Sigma_r = W_r^T < \Sigma > V_r$$

Equivalence Krylov – Sylvester

$$\text{span} (V_r) = \bigcup_i K_{n_i} ((A - s_{0,i}^* E)^{-1} E, (A - s_{0,i}^* E)^{-1} B)$$

Computation of $V_{m,r}$

$$A_m V_{m,r} - E_m V_{m,r} S_{V}^* - B_m R^* = 0$$

$$W_m^T (A V_{m,r} - E V_{m,r} S_{V}^* - B R^*) = 0$$

$$W_m^T \left( A V_m - E V_m S_{V}^* - B R^* \right) = 0$$

$$C \left( A - s_{0,i}^* E \right)^{-1} \left( A - s_{0,i}^* E \right) \tilde{V}_{r,i} - B \right) = 0$$

$$C \left( \tilde{V}_{r,i} - \left( A - s_{0,i}^* E \right)^{-1} B \right) = 0 \quad \forall C$$

$$\tilde{V}_{r,i} = \left( A - s_{0,i}^* E \right)^{-1} B = V_{r,i}$$

Note: analogously for MIMO and non-primitive Krylov bases

[C./Panzer/Lohmann '16]
Model functions – a general framework

- Can be applied to different $\mathcal{H}_2$-optimal reduction methods
- Applicable to further classes of models (e.g. DAE, irrational, data driven, …)
- Particularly advantageous, the more expensive the evaluation of the FOM gets

Application to IRKA:

**Algorithm** Confined Iterative Rational Krylov Algorithm (CIRKA)

**Input:** $\Sigma$, $s_0$, tol

**Output:** locally $\mathcal{H}_2$-optimal reduced model $\Sigma_r$

1. Initialize $\Sigma_m$ to empty
2. while relative Change of $s_0 < $ tol do
3. $\Sigma_m \leftarrow$ updateModelFct($\Sigma, \Sigma_m, s_0$)
4. $[\Sigma_r, s_0^*] \leftarrow$ IRKA($\Sigma_m, s_0$)
5. $s_0 \leftarrow s_0^*$
6. end while

Possible implementations
- Update for all $s_0$
- Update only new $s_0$
- New model function about $s_0$

[C./Panzer/Lohmann '16]
Agenda

A new framework for $\mathcal{H}_2$-optimal MOR and its applications

- Model order reduction by projection
- Rational interpolation and $\mathcal{H}_2$-optimal reduction
- What is the cost of $\mathcal{H}_2$-optimal reduction?
- A new framework to exploit the local nature of rational interpolation

- Numerical results and further applications
Complexity of CIRKA

When is CIRKA expected to be faster than IRKA?

\[ C_{N,n}(\text{IRKA}) \approx k_{IRKA} \cdot n \cdot C_N(\text{LSE}) \]

\[ C_{N,n}(\text{CIRKA}) \approx \sum_{i=1}^{k_m} n_{m,+}^i \cdot C_N(\text{LSE}) + \sum_{i=1}^{k_m} k_{IRKA,i} \cdot n \cdot C_{n_m^i}(\text{LSE}) \]

assuming \( C_{n_m}(\text{LSE}) \ll C_N(\text{LSE}) \)

\[ \sum_{i=1}^{k_m} n_{m,+}^i \ll n \cdot k_{IRKA} \]
The cost of one LU decomposition

\[ C_{n,m}^{i} \text{ (LSE)} \ll C_{N} \text{ (LSE)} \]
Numerical results

![Graph showing numerical results with speedup and magnitude plots. The graphs compare IRKA and CIRKA methods for various system components.]
Numerical results

- **Graph 1:**
  - Title: Speedup vs. t/s.
  - X-axis: Time (t/s) in logarithmic scale.
  - Y-axis: Speedup.
  - Legend: IRKA and CIRKA.

- **Graph 2:**
  - Title: Magnitude (dB) vs. Frequency (rad/s).
  - X-axis: Frequency in logarithmic scale.
  - Y-axis: Magnitude in linear scale.
  - Legend: $G(s)$ and $G_r(s)$.

- **Graph 3:**
  - Title: Frequency vs. Amplitude.
  - X-axis: Frequency.
  - Y-axis: Amplitude.
  - Legend: $s_0$ IRKA and $s_0$ CIRKA.
Further applications of the framework

Once you have a hammer…

- Error estimation
- Global $H_2$-optimal reduction
- Adaptive choice of reduced order

Source(s): tiger-supplies, homedepot.com, for.unipi.it
Error estimation using model function $\Sigma_m$

Rigorous error bounds are an open problem

$\varepsilon \leq \|G - G_r\|_{\mathcal{H}_p} \leq \bar{\varepsilon}$

Exploit the model function

$\|G - G_r\|_{\mathcal{H}_p} \approx \|G_m - G_r\|_{\mathcal{H}_p}$
Global $\mathcal{H}_2$-optimal reduction

\[ N \]

\[ \Sigma \]

\[ \Sigma k_m \]

\[ \Sigma i_m \]

\[ \Sigma r \]

\[ \Sigma i_{m,r} \]

\[ \Sigma i_{m,r} \]

Graph showing magnitude in dB against frequency and real part against imaginary part.
Toolboxes for sparse, large-scale models in

\[ \text{sys} = \text{sss}(A, B, C, D, E); \]

\[ \text{sysr} = \text{tbr}(\text{sys}, n) \]
\[ \text{sysr} = \text{rk}(\text{sys}, s0) \]
\[ \text{sysr} = \text{irka}(\text{sys}, s0) \]
\[ \text{sysr} = \text{cure}(\text{sys}) \]
\[ \text{sysr} = \text{cirka}(\text{sys}, s0) \]

bode(sys), sigma(sys)
step(sys), impulse(sys)
norm(sys, 2), norm(sys, inf)
c2d, lsim, eigs, connect,…

Powered by: M-M.E.S.S. toolbox [Saak, Köhler, Benner] for Lyapunov equations
Available at www.rt.mw.tum.de/?sssmor
[C./Cruz Varona/Jeschek/Lohmann: „sss & sssMOR: Analysis and Reduction of Large-Scale Dynamic Systems in MATLAB“, 2017 at-Automatisierungstechnik]
Comprehensive documentation with examples and references

**sssMOR App**
graphical user interface

completely **free**
and **open source**
(contributions welcome)
Chair of Automatic Control
Department of Mechanical Engineering
Technical University of Munich

Comprehensive documentation with examples and references

**sssMOR App**
Graphical user interface

**completely free and open source**
(contributions welcome)
Comprehensive documentation with examples and references

**sssMOR App**
graphical user interface

**completely free and open source**
(contributions welcome)
A new framework for $\mathcal{H}_2$-optimal MOR and its applications

- Model reduction as projection
- Cost of $\mathcal{H}_2$-optimal methods
  - Reduction vs optimization cost
- A new framework: model functions
  - Exploit local nature of Krylov methods
  - Cost: from optimization to model function update
  - Different algorithms and classes of models
- Optimality and speedup
  - Optimality guaranteed through update
  - Speedup especially for very high order models
- Further applications
  - Error estimation
  - Global $\mathcal{H}_2$-optimal reduction
  - Adaptive choice of reduced order
A new framework for $\mathcal{H}_2$-optimal model reduction and its applications

Alessandro Castagnotto

in collaboration with: Heiko Peuscher (né Panzer), Boris Lohmann, Siyang Hu

related publications: “Fast $\mathcal{H}_2$-Optimal Model Order Reduction Exploiting the Local Nature of Krylov-Subspace Methods”, 2016 European Control Conference, Aalborg
References

[C. et al. ‘17] sss & sssMOR: Analysis and reduction of large-scale dynamic systems in MATLAB
[C./Panzer/Lohmann ‘16] Fast H2-optimal model order reduction exploiting the local nature of Krylov subspace methods
[De Villemagne/Skelton ‘87] Model reductions using a projection formulation
[Gallivan et al. ‘04] Sylvester equations and projection-based model reduction
[Golub ‘12] Matrix computations
[Gugercin/Antoulas/Beattie ‘08] H2-optimal model reduction for large-scale linear dynamical systems
[Meier/Luenberger ‘67] Approximation of linear constant systems