Globally optimal SISO H_2-norm model reduction

Sibren Lagauw* and Bart De Moor*

KU Leuven, Department of Electrical Engineering (ESAT-STADIUS)

June 14, 2022

Abstract H_2-model reduction is an established discipline in the systems and control community, and today's state-of-the-art methods achieve great performance on large-scale problems. However, the methods are generally suboptimal, i.e., only local optimality can be guaranteed. By contrast, the approach that we present identifies all the stationary points of the model reduction problem, allowing one to select the globally optimal reduced model. It is well-known that the first-order necessary conditions for optimality of the H_2 model reduction problem impose interpolatory conditions on the optimal lower-order approximant. We use these conditions to rephrase the model reduction problem to a root-finding problem of a system of multivariate polynomial equations, the real-valued common roots of which characterize all the stationary points of the model reduction problem. The methodology derives ultimately from (rational) approximation theory and outperforms existing techniques for globally optimal H_2-norm model reduction in terms of computational costs. We repeat several numerical examples of H_2 optimal model reduction used in the literature to illustrate the methodology.

Problem formulation

Given a 'higher-order' LTI system $H(z) \in H_2$, with H_2 the class of stable and causal LTI models with real-valued impulse response:

$$H(z) = \frac{b(z)}{a(z)} = \frac{b_{n-1}z^{n-1} + \cdots + b_1 z + b_0}{z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0},$$

model reduction techniques search for model $\hat{H}(z) \in H_2$ of order $m < n$, while minimizing the approximation error in some measure. We consider the H_2-norm:

$$\min_{\hat{H}(z)} J = ||H(z) - \hat{H}(z)||^2_{H_2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} |H(e^{j\omega}) - \hat{H}(e^{j\omega})|^2 d\omega = \sum_{k=0}^{\infty}(h_k - \hat{h}_k)^2,$$

with $\{h_k\}_{k=0,\dots,\infty}$, $\{\hat{h}_k\}_{0,\dots,\infty}$ the impulse response of $H(z)$ and $\hat{H}(z)$ respectively.

Walsh's theorem

Walsh's theorem has first been derived in the field of rational approximation theory [8]. Later, the same result has been encountered in the context of systems and control [1, 4] (continuous-time), [2, 5] (discrete-time).

Theorem 1 (Walsh's theorem) *The stationary points $\hat{H}(z)$ of the model reduction problem satisfy:*

$$H(z) - \hat{H}(z) = \frac{b(z)}{a(z)} - \frac{\hat{b}(z)}{\hat{a}(z)} = \left[\frac{\hat{a}_r(z)}{\hat{a}(z)}\right]^2 G(z),$$

with $G(z) \in H_2$ the z-transform of some real-valued, stable and causal signal and $\hat{a}_r(z)$ defined as:

$$\hat{a}_r(z) = \hat{a}_0 z^m + \hat{a}_1 z^{m-1} + \cdots + \hat{a}_{m-1}z + 1,$$

the polynomial which has the reciprocals of the roots of $\hat{a}(z)$ as its roots.

Proof outline: Use a partial fractions representation for $\hat{H}(z)$:

$$\hat{H}(z) = \frac{c_1}{z-\pi_1} + \frac{c_2}{z-\pi_2} + \cdots + \frac{c_m}{z-\pi_m},$$

and show that the **first-order necessary conditions for optimality** formulate $2m$ orthogonality constraints:

$$\begin{cases} \dfrac{\partial J}{\partial \overline{c_i}} = -\dfrac{1}{2\pi}\left\langle H(z) - \hat{H}(z), \dfrac{1}{z-\pi_i}\right\rangle = 0 & \text{for } i = 1,\dots,m, \\ \dfrac{\partial J}{\partial \overline{\pi_i}} = -\dfrac{1}{2\pi}\left\langle H(z) - \hat{H}(z), \dfrac{c_i}{(z-\pi_i)^2}\right\rangle = 0 & \text{for } i = 1,\dots,m. \end{cases}$$

Show, based on Cauchy's integral formula, that these orthogonality constraints are satisfied if $\hat{H}(z)$, assumed to have real coefficients, satisfies the following interpolatory conditions [8]:

$$\begin{cases} \hat{H}(1/\overline{\pi_i}) = H(1/\overline{\pi_i}) & \text{for } i = 1,\dots,m, \\ \dfrac{\partial \hat{H}(1/\overline{\pi_i})}{\partial \xi} = \dfrac{\partial H(1/\overline{\pi_i})}{\partial \xi} & \text{for } i = 1,\dots,m. \end{cases}$$

I.e., $\hat{H}(z)$ interpolates $H(z)$ in $2m$ interpolation points, the reciprocals of the poles of $\hat{H}(z)$.

Methodology

Multiplying out the denominators in Walsh's equality gives:

$$b(z)\hat{a}(z) - a(z)\hat{b}(z) = [\hat{a}_r(z)]^2 G'(z).$$

with $G'(z) = [G(z)a(z)]/\hat{a}(z)$. This shows that $G'(z)$ is an $(n+m-1)$'th order polynomial. Bring everything to the left-hand side to get:

$$l(z) = b(z)\hat{a}(z) - a(z)\hat{b}(z) - [\hat{a}_r(z)]^2\left(g_{n-m-1}z^{n-m-1} + \cdots + g_1 z + g_0\right) = 0,$$

where we parametrized $G'(z)$ in the parameters $\{g_i\}_{i=0,\dots,n-m-1}$.

Algorithm

- Construct $l(z)$, with coefficients parametrized in the $n+m$ parameters $(\{\hat{a}_i, \hat{b}_i\}_{i=0,\dots,m-1}$ and $\{g_i\}_{i=0,\dots,n-m-1})$.
- Equate each coefficient of $l(z)$ to zero to obtain a system of $n+m$ multivariate polynomial equations in $n+m$ variables. The system contains one linear, one quadratic and $n+m-2$ cubic equations.
- Identify the real-valued common roots for which the resulting model $\hat{H}(z)$ is stable and evaluate J for each solution. Select the globally optimal minimizer of the model reduction problem.

Continuous-time equivalent

A similar derivation can be applied in the continuous-time case. The interpolation points are the reflections over the imaginary axis of the poles of $\hat{H}(s)$.

Theorem 2 (Walsh's theorem, continuous-time) *With $H(s)$ the transfer function of a stable and causal LTI model, the stationary points $\hat{H}(s)$ of the model reduction problem satisfy:*

$$H(s) - \hat{H}(s) = \frac{b(s)}{a(s)} - \frac{\hat{b}(s)}{\hat{a}(s)} = \left[\frac{\hat{a}(-s)}{\hat{a}(s)}\right]^2 G(s),$$

with $G(s)$ the Laplace-transform of some real-valued, stable and causal signal.

Numerical examples

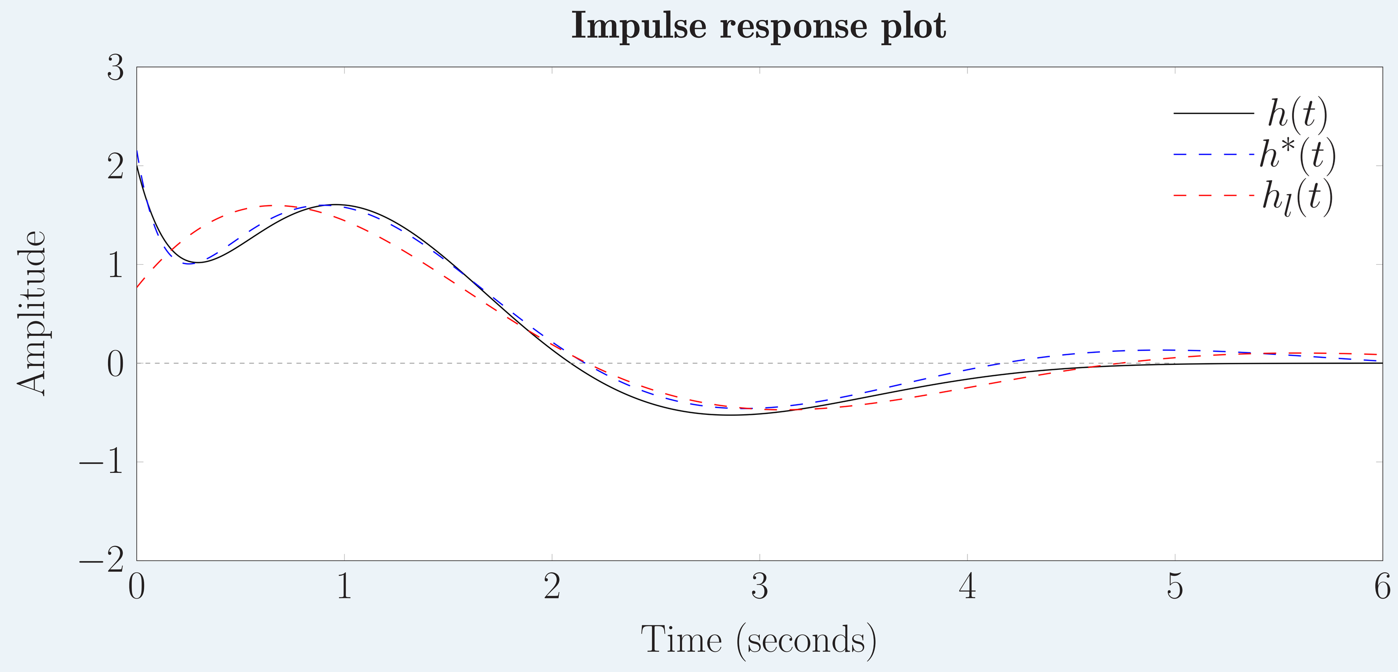
Example ($n = 7,\ m = 3$): We consider the higher order model used in [3]:

$$H(s) = \frac{2s^6 + 11.5s^5 + 57.75s^4 + 178.625s^3 + 345.5s^2 + 323.625s + 94.5}{s^7 + 10s^6 + 46s^5 + 130s^4 + 239s^3 + 280s^2 + 194s + 60}.$$

Searching for the optimal third order reduced model gives two stationary points corresponding to stable models, for which the impulse responses are visualized in the figure below.

Reduced model	$J/\|\|H(s)\|\|_{H_2}$
$\hat{H}^*(s) = \frac{2.155s^2+3.343s+33.8}{s^3+7.457s^2+10.51s+17.57}$	0.1171
$H_l(s) = \frac{0.7669s^2+3.562s+0.4614}{s^3+1.217s^2+2.083s+0.3007}$	0.2338

Impulse response plot



Toy example ($n = 3,\ m = 1$): Consider the higher-order model:

$$H(s) = \frac{s^2 + 9s - 10}{s^3 + 12s^2 + 49s + 78},$$

for which we search the optimal first order approximation. Using the described methodology gives:

$$l(s) = (1 - g_1 - \hat{b}_0)s^3 + (\hat{a}_0 - 12\hat{b}_0 - g_0 + 2\hat{a}_0 g_1 + 9)s^2 + (9\hat{a}_0 - 49\hat{b}_0 + 2\hat{a}_0 g_0 - \hat{a}_0^2 g_1 - 10)s + (-g_0\hat{a}_0^2 - 10\hat{a}_0 - 78\hat{b}_0)$$

from which we compose the following system of 4 equations in 4 unknowns:

$$\begin{cases} 0 = 1 - g_1 - \hat{b}_0, \\ 0 = \hat{a}_0 - 12\hat{b}_0 - g_0 + 2\hat{a}_0 g_1 + 9, \\ 0 = 9\hat{a}_0 - 49\hat{b}_0 + 2\hat{a}_0 g_0 - \hat{a}_0^2 g_1 - 10, \\ 0 = -g_0\hat{a}_0^2 - 10\hat{a}_0 - 78\hat{b}_0. \end{cases}$$

The system of equations can be written as a **multiparameter eigenvalue problem (MEP)** [6] by extracting the variables that appear only linearly $(\hat{b}_i, g_i)$:

$$\left(\begin{bmatrix} 1 & -1 & -1 & 0 \\ 9 & -12 & 0 & -1 \\ -10 & -49 & 0 & 0 \\ 0 & -78 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 9 & 0 & 0 & 2 \\ -10 & 0 & 0 & 0 \end{bmatrix}\hat{a}_0 + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}\hat{a}_0^2\right)\begin{bmatrix} 1 \\ \hat{b}_0 \\ g_1 \\ g_0 \end{bmatrix} = 0.$$

The block-Macaulay method [7] is used to solve the MEP. There are two real-valued solutions that correspond to stable models. The globally optimal model is given as:

$$\hat{H}^*(s) = \frac{1.2799}{s + 9.6796}.$$

Future work

Work towards higher complexity/large-scale:

- Obtain even lower computational costs by searching for real-valued common-roots only.
- Experiment with other rootfinding methods.
- Benchmark the methodology using large-scale numerical examples in a HPC-environment.

References

[1] P.R. Aigrain and E.M. Williams. Synthesis of n-reactance networks for desired transient response. *Journal of Applied Physics*, 20(6):597–600, 1949.

[2] Bart De Moor, Peter Van Overschee, and Geert Schelfhout. H_2 model reduction for SISO systems. In *Proc. of the 12th IFAC World Congress*, pages 227–230, Sydney, Australia, July 1993.

[3] S. Gugercin, A.C. Antoulas, and C. Beattie. $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM Journal on Matrix Analysis and Applications*, 30(2):609–638, 2008.

[4] L. Meier and D.G. Luenberger. Approximation of linear constant systems. *IEEE Transactions on Automatic Control*, 12(5):585 – 588, October 1967.

[5] Phillip A. Regalia. *Adaptive IIR Filtering in Signal Processing and Control.* CRC Press, 1995.

[6] C. Vermeersch and B. De Moor. Globally optimal least-squares ARMA model identification is an eigenvalue problem. *IEEE Control Systems Letters*, 3(4):1062–1067, 2019.

[7] C. Vermeersch and B. De Moor. Two complementary block macaulay algorithms to solve multiparameter eigenvalue problems. Technical report, KU Leuven, ESAT (STADIUS), Leuven, Belgium, 2022.

[8] J.L. Walsh. *Interpolation and Approximation by Rational Functions in the Complex Domain*, volume XX of *American Mathematical Society Colloquium Publications.* American Mathematical Society, 3th edition, 1960.

*KU Leuven, Dept. of Elec. Engineering (ESAT), Center for Dynamical Systems, Signal Processing, and Data Analytics (STADIUS), Kasteelpark Arenberg 10, 3001 Leuven, Belgium (sibren.lagauw@esat.kuleuven.be, bart.demoor@esat.kuleuven.be).

This work was supported in part by the KU Leuven: Research Fund (project C16/15/059), and several Leuven Research and Development bilateral industrial projects, in part by Flemish Government agencies: FWO (EOS project G0F6718N (SeLMA), and PhD grants SB/1SA1319N, SB/1S93918, SB/1S1319N), EWI (Flanders AI Research Program), and in part by the European Commission (EU Research Council under the European Union's Horizon 2020 research and innovation programme (ERC Adv. Grant) under grant 885682).