KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Sibren Lagauw, Oscar Mauricio Agudelo, Bart De Moor Title: Globally Optimal SISO H_2-Norm Model Reduction Using Walsh's Theorem Journal/Conference/Book: IEEE Control Systems Letters (L-CSS) Vol/Book Chapter: 7 Pages: 1670–1675
Archived version	☐ Submitted version ■ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ☐ Published ☐ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ☐ EU H22020 ☐ EU Horizon Europe ■ FWO ■ KULeuven Internal Funds ☐ VLAIO ☐ AI ■ ERC ☐ Other
Published version	https://ieeexplore.ieee.org/document/10124192
Copyright statement	978-1-5386-5541-2/18/$31.00 2018 IEEE
Author contact	sibren.lagauw@esat.kuleuven.be mauricio.agudelo@esat.kuleuven.be bart.demoor@esat.kuleuven.be



Globally Optimal SISO H_2-Norm Model Reduction Using Walsh's Theorem

Sibren Lagauw, Oscar Mauricio Agudelo, and Bart De Moor *Fellow, IEEE & IFAC & SIAM*

***Abstract*— We present a novel methodology for single-input single-output (SISO) $\mathbf{H_2}$-norm model reduction that guarantees global optimality of the obtained solution(s). By exploiting Walsh's theorem, which is an elegant formulation of the first-order necessary conditions for optimality, we reformulate the model reduction problem as a multiparameter eigenvalue problem (MEVP), the real-valued eigentuples of which characterize the globally optimal solution(s) of the model reduction problem. While aiming for global optimality comes at the cost of a combinatorial growth of the problem complexity for increasing model orders, the novel methodology allows us to tackle larger problems compared to the few other globally optimal approaches in the literature. In particular, the degree of the obtained MEVP is independent of the order of the original higher order and obtained reduced-order model, a property that is favorable from a computational point of view. We perform three numerical experiments to illustrate the effectiveness of the methodology.**

Model/Controller reduction; Linear systems; Optimization

I. Introduction

We consider the model reduction problem for systems in $\mathcal{M}$, the class of minimal, stable, causal single-input single-output (SISO) linear time-invariant (LTI) continuous-time systems with real-valued impulse response, a problem that already has been studied extensively, see e.g., [1], [2] and references therein. The model reduction problem can be summarized as follows: given a strictly proper transfer function $H(s) \in \mathcal{M}$ of order n,

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

we search for a model $\hat{H}(s) \in \mathcal{M}$ of order $m < n$,

$$\hat{H}(s) = \frac{\hat{b}(s)}{\hat{a}(s)} = \frac{\hat{b}_{m-1}s^{m-1} + \cdots + \hat{b}_1 s + \hat{b}_0}{s^m + \hat{a}_{m-1}s^{m-1} + \cdots + \hat{a}_1 s + \hat{a}_0}, \quad (2)$$

so that $\hat{H}(s)$ is a 'good approximation' of $H(s)$. Define the coefficient vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ as $(a_{n-1}, \ldots, a_0)^\mathsf{T}$ and $(b_{n-1}, \ldots, b_0)^\mathsf{T}$ respectively, and define $\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}} \in \mathbb{R}^m$ similarly using the coefficients of respectively $\hat{a}(s)$, $\hat{b}(s)$.

Sibren Lagauw (corresponding author), Oscar Mauricio Agudelo and Bart De Moor are with the Center for Dynamical Systems, Signal Processing, and Data Analytics (STADIUS), Dept. of Electrical Engineering (ESAT), KU Leuven. Kasteelpark Arenberg 10, B-3001 Leuven, Belgium. E-mail addresses: {sibren.lagauw, mauricio.agudelo, bart.demoor}@esat.kuleuven.be

This work was supported in part by the KU Leuven: Research Fund (projects iBOF/23/064, C16/15/059), and several LRD bilateral industrial projects, in part by Flemish Government agencies: FWO (EOS project G0F6718N (SeLMA)), EWI (Flanders AI Research Program), and in part by the European Commission (ERC Adv. Grant grant No 885682). The work of S. Lagauw was supported by the FWO Fundamental Research Fellowship under grant 11K5623N.

In this letter, we seek for an approximant $\hat{H}(s)$ that minimizes the H_2-norm of the approximation error $E(s) = H(s) - \hat{H}(s)$. The problem under consideration becomes:

$$\hat{H}(s) \in \underset{\hat{H}(s)\in\mathcal{M}}{\operatorname{argmin}}\ J^2, \quad (3)$$

where

$$\begin{aligned} J^2 = \|E(s)\|_{H_2}^2 &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(j\omega) - \hat{H}(j\omega)|^2 \,\mathrm{d}\omega \\ &= \int_0^{\infty} \Big(h(t) - \hat{h}(t)\Big)^2 \,\mathrm{d}t, \end{aligned} \quad (4)$$

with $h(t)$ and $\hat{h}(t)$ the impulse responses of H, $\hat{H}$, respectively. Assuming that H and $\hat{H}$ are stable implies that the roots of $a(s)$ and $\hat{a}(s)$ lie in left half-plane of the complex s-plane, guaranteeing that the H_2-norm is well-defined [2].

The optimization problem in (3) over the decision variables $\{\hat{a}_i, \hat{b}_i\}_{i=0,\ldots,m-1}$ is generally non-convex, implying that the solution space can contain (many) local optima. This non-convexity makes it very challenging to aim for global optimality. Instead, state-of-the-art techniques 'solve' the problem by iteratively computing a solution that satisfies the first-order necessary conditions for optimality (FONC), i.e., a stationary point of the optimization problem in (3). These FONC have been formulated both in terms of Lyapunov equations [3], [4] and interpolatory conditions [5], [6], dividing the landscape of high-performance algorithms into interpolation-based methods (e.g., [6], [7]) and Lyapunov-based methods (e.g., [8]). Note that both frameworks for the FONC have been shown to be equivalent [6, Lemma 3.7].

Although the above-mentioned iterative techniques are very useful in practice, we consider them suboptimal from a mathematical point of view: their performance depends heavily on the heuristic choice of a 'good' initial point and at most local optimality with respect to (3) can be guaranteed. By contrast, we deem the model reduction problem mathematically 'solved' if and only if the globally optimal reduced model(s) have been identified up to machine precision by means of a deterministic procedure, e.g., by solving a system of linear equations or an eigenvalue problem.

The approach of the few existing globally optimal methods [9]–[12] is, roughly speaking, to compute all reduced models that satisfy the FONC of (3), so that this set of stationary points is guaranteed to contain the globally optimal reduced model(s). Inspired by iterative rational Krylov algorithms (e.g., [6]), the author of [11] shows that all optimal sets of interpolation points, which implicitly define the stationary points $\hat{H}(s)$, can be retrieved via a multiparameter eigenvalue problem (MEVP) [13]. In particular, each eigentuple of the MEVP is used to construct a univariate polynomial, the roots of which correspond to a set of

interpolation points. The coefficient matrices of the MEVP are composed using a state-space model of the higher order system. The methodology is a generalization of the results in [9], where only first- and second-order reduced models are considered. Alternatively, the authors of [12] exploit the Lyapunov equation to rewrite the objective function (4) in terms of a controllability Grammian, where they use the control canonical form to describe the reduced model in terms of the model parameters $\{\hat{a}_i, \hat{b}_i\}_{i=0,\dots,m-1}$. In turn, the authors reformulate the FONC of this obtained objective function as an MEVP, the eigentuples/-vectors of which contain the optimal model parameters. For the case of order-one reductions, the authors of [10] retrieve the globally optimal solution(s) via the common roots of a system of quadratic polynomial equations.

Aiming for global optimality comes at the cost of an (enormous) increase in computational burden. The efficiency of the globally optimal methods [11], [12] is completely determined by the difficulty involved with the computation of all the eigentuples of the obtained MEVP. So, it is clear that the viability of these approaches depends on the degree and the size of the to-be-solved MEVPs.

Contributions: In this letter, we propose a novel methodology for globally optimal SISO H_2-norm model reduction. The methodology combines properties of the two globally optimal approaches in the literature: we use the interpolatory framework of the FONC (Walsh's theorem) to reformulate the problem as an MEVP, similarly as in [11], and the optimal model parameters can be retrieved directly from the eigentuples and eigenvectors of the obtained MEVP, similarly as in [12]. However, contrary to the existing approaches [11], [12], the proposed methodology solely uses the frequency-domain representation (transfer function) of the models, eliminating the use of state-space descriptions. Furthermore, we argue that the MEVPs obtained with the proposed methodology have favorable properties from a computational point of view, especially noticeable for growing model orders. Lastly, this letter provides a concise summary of the literature related to globally optimal H_2 model reduction.

Remark: Note that we do not insinuate that the above-mentioned 'approximate' methods [6]–[8] are not useful in practice. The globally optimal approaches should not be considered as meant to be competitive with respect to these established techniques in the context of large-scale models. Rather, we address a theoretical problem, the solution of which could potentially be used for benchmarking purposes.

II. Characterizing first-order optimality

In this Section, we briefly describe the interpolatory formulation of the FONC of the model reduction problem in (3). In the context of systems theory and control, these optimality conditions date back to the 1940s, where they arose from a network-synthesis problem considering real-valued and simple poles [14]. Later, the framework has been generalized to multiple poles that are allowed to be complex-valued [5]. Model reduction for linear systems interacts in many ways with the mathematical discipline of approximation on the complex domain, a connection investigated in [15]. In particular, the model reduction problem in (3) and its equivalent for discrete-time systems can be considered as rational l_2-approximation problems, see e.g., [1]. So, it comes as no surprise that the interpolatory conditions derived in [5], [14] were already known in the field of classical analysis in the 1920s, e.g., by Walsh [16].

A. Interpolatory conditions for optimality

Theorem 1 (Meier and Luenberger [5]): Given a stable SISO model $H(s) \in \mathcal{M}$ of order n, let $\hat{H}(s)$ of order m ($m<n$) be a stationary point of the H_2-norm model reduction problem in (3). Then for all poles p_i of $\hat{H}(s)$,

$$H(-p_i)^{(j)} = \hat{H}(-p_i)^{(j)}, \quad j = 0, \dots, d_i, \tag{5}$$

where d_i is the multiplicity of the pole p_i and the superscript j denotes the jth derivative with respect to s, i.e.,

$$F(a)^{(j)} = \left.\frac{\mathrm{d}^j F(s)}{\mathrm{d}s^j}\right|_{s=a},$$

for any function $F(s)$ and $a \in \mathbb{C}$.

We give a high-level view on how the conditions in (5) follow naturally from the model reduction problem (3), assuming, for the ease of notation, that $\hat{H}(s)$ has simple poles only. By expressing $\hat{H}(s)$ in a partial fraction expansion:

$$\hat{H}(s) = \frac{c_1}{s-p_1} + \frac{c_2}{s-p_2} + \cdots + \frac{c_m}{s-p_m}, \tag{6}$$

the model reduction problem (3) becomes an optimization problem over the $2m$ decision variables c_i and p_i. The FONC could now be constructed by equating all partial derivatives of J^2 with respect to the real and imaginary parts of these decision variables to zero. However, Wirtinger derivatives[1] can be used to deal with the complex decision variables, resulting in an equivalent but more elegant formulation of the FONC. The Wirtinger derivative of J^2 with respect to $\overline{c_i}$, where $\hat{H}(s)$ is parametrized as in (6) and $\overline{c_i}$ denotes the complex conjugate of c_i, gives:

$$\begin{aligned}\frac{\partial J^2}{\partial \overline{c_i}} &= \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \left(H(s)-\hat{H}(s)\right) \frac{\partial}{\partial \overline{c_i}} \overline{\left(H(s)-\hat{H}(s)\right)}\,\mathrm{d}s \\ &= \frac{-1}{2\pi j}\int_{-j\infty}^{j\infty} \left(H(s)-\hat{H}(s)\right)\left(\frac{1}{s-\overline{p_i}}\right)\mathrm{d}s, \end{aligned} \tag{7}$$

and similarly:

$$\frac{\partial J^2}{\partial \overline{p_i}} = \frac{-1}{2\pi j}\int_{-j\infty}^{j\infty} \left(H(s)-\hat{H}(s)\right)\left(\frac{\overline{c_i}}{(s-\overline{p_i})^2}\right)\mathrm{d}s. \tag{8}$$

Because the objective function J^2 is real-valued, we know that the partial derivatives with respect to c_i and p_i are equal to the complex conjugates of (7) and (8), respectively:

$$\overline{\left(\frac{\partial J^2}{\partial c_i}\right)} = \frac{\partial J^2}{\partial \overline{c_i}} \quad \text{and} \quad \overline{\left(\frac{\partial J^2}{\partial p_i}\right)} = \frac{\partial J^2}{\partial \overline{p_i}}.$$

[1]The Wirtinger derivative of $f(s,\overline{s})$ wrt. s (resp. $\overline{s}$) is calculated using the standard rules of differentiation, while considering the variable $\overline{s}$ (resp. s) as a constant, i.e., s and $\overline{s}$ are treated as two independent variables for the purpose of differentiation. This is also called $\mathbb{CR}$-*calculus* [17].

It therefore suffices to equate the expressions in (7)–(8) to zero for all $i = 1, \ldots, m$, to compose the FONC. Using properties of the Laplace transform, these $2m$ equations can be shown to be equivalent to the interpolatory conditions described in Theorem 1. See, e.g., [5], for more details.

B. Walsh's theorem

The conditions in Theorem 1 can be expressed compactly using the polynomial $\hat{a}(-s)$, which has the mirror images of the poles p_i of the approximant $\hat{H}(s)$ as its roots.

Theorem 2 (Regalia [1]): Given a stable SISO model $H(s) \in \mathcal{M}$ of order n, let $\hat{H}(s)$ of order m with $m<n$ be a stationary point of the model reduction problem (3). Then for all $s \in \mathbb{C}$:

$$H(s) - \hat{H}(s) = \frac{b(s)}{a(s)} - \frac{\hat{b}(s)}{\hat{a}(s)} = \left[\frac{\hat{a}(-s)}{\hat{a}(s)}\right]^2 G(s), \qquad (9)$$

with $G(s)$ the Laplace-transform of some real-valued, stable and causal signal, and where $a(s), b(s), \hat{a}(s)$ and $\hat{b}(s)$ are defined as in (1)–(2).

The denominator of the right-hand side of (9) contains the factor $[\hat{a}(s)]^2$ to illustrate that the residual contains an 'all-pass' filter defined by the poles of $\hat{H}(s)$, hinting to the connection with the operator-theoretic result of Beurling–Lax[2]. Because the roots of $\hat{a}(s)$ lie in the complex left half-plane, the factor could equally well be 'absorbed' into $G(s)$.

The formulation in (9) has been encountered in [18] and the equivalent relation phrased for discrete-time models, which is similar to the continuous-time case upon changing the setting from the Laplace- to the z-domain, has been derived as a limiting case of an iterative model reduction algorithm [19], based on polynomial algebra [20] and using the result of Beurling–Lax [1]. The author of [1] labeled the result as *Walsh's theorem*, recognizing its origins within approximation theory. The relation has been extended to multiple-input multiple-output (MIMO) models in [21] and [22], in the continuous- and discrete-time setting respectively.

III. Methodology

In this Section, we show how Theorem 2 can be exploited to retrieve all mth order stationary points $\hat{H}(s)$ of the model reduction problem in (3) via the eigentuples of an MEVP.

A. System of multivariate polynomial equations

Theorem 3: For any given stable SISO model $H(s)$ of order n and mth order approximant $\hat{H}(s)$ with $m<n$ as defined in (1)–(2), define the polynomial,

$$l(s) = b(s)\hat{a}(s) - a(s)\hat{b}(s) - [\hat{a}(-s)]^2\tilde{G}(s), \qquad (10)$$

where $\tilde{G}(s)$ is a polynomial parametrized in the coefficients $\boldsymbol{g} = (g_0, \ldots, g_{n-m-1})^\mathsf{T} \in \mathbb{R}^{m-n}$:

$$\tilde{G}(s) = g_{n-m-1}s^{n-m-1} + \cdots + g_1 s + g_0.$$

[2]The transfer function of systems in $\mathcal{M}$ can be considered as an element of the Hardy space $\mathcal{H}_2$ of functions analytic on the complex right half-plane (for discrete-time systems this becomes the Hardy space of functions analytic on the exterior of the unit-disc). Beurling–Lax's theorem describes how an *inner function* (in this case the all-pass filter) induces an orthogonal decomposition of this space into two shift-invariant subspaces [1].

Then, $\hat{H}(s)$ is a stationary point of (3) if and only if,

$$\exists\, \boldsymbol{g} \quad \text{s.t.} \quad l(s) = 0,\ \forall s \in \mathbb{C}. \qquad (11)$$

Proof: Define $\tilde{G}(s) = G(s)[a(s)/\hat{a}(s)]$ so that we can rewrite (9) from Theorem 2 as,

$$b(s)\hat{a}(s) - a(s)\hat{b}(s) = [\hat{a}(-s)]^2\tilde{G}(s). \qquad (12)$$

The left-hand side of (12) is a polynomial in s of degree $n+m-1$, so that $\tilde{G}(s)$ must be a polynomial of degree $n-m-1$. It follows that, given a stable SISO model $H(s)$ of order n and a stationary point $\hat{H}(s)$ of order m of the model reduction problem (3), there exists a polynomial $\tilde{G}(s)$ such that (11) holds. This proves the $\Rightarrow$ direction. Conversely, if, for a given approximant $\hat{H}(s)$ of order m, there exists a polynomial $\tilde{G}(s)$ such that (11) is true, then (9) holds as well. The latter implies that $\hat{H}(s)$ satisfies the interpolatory conditions described in Theorem 1, indicating that $\hat{H}(s)$ is a stationary point of the model reduction problem in (3).

Proposition 1: Let f_k be the coefficient corresponding to s^k in the polynomial from (10) and consider the parameter vectors $\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}} \in \mathbb{R}^m$ containing the coefficients of respectively $\hat{a}(s)$ and $\hat{b}(s)$ as in (2). Define the algebraic variety,

$$\mathcal{V}_\mathbb{R} = \big\{(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \boldsymbol{g}) \in \mathbb{R}^{m+n} \ :\ f_k(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \boldsymbol{g}) = 0, \quad \forall k = 0, \ldots, m+n-1\big\}, \qquad (13)$$

and consider the subvariety $\mathcal{V}'_\mathbb{R} \subseteq \mathcal{V}_\mathbb{R}$ that contains the tuples $(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \boldsymbol{g}) \in \mathcal{V}_\mathbb{R}$ for which all roots of $\hat{a}(s)$ lie in the complex left half-plane. Then, we know by Theorem 3 that the tuples in $\mathcal{V}'_\mathbb{R}$ describe all mth order stationary points $\hat{H}(s)$ of (3).

Stationary points for which $\hat{H}(s)$ has one or more pole-zero cancellations – the order of $\hat{H}(s)$ is strictly smaller than m – are not described by the relation in (9). As such, they cannot be retrieved via $\mathcal{V}'_\mathbb{R}$. However, these stationary points have been shown to correspond to saddle points of the H_2-norm model reduction problem [1, Theorem 5.3] and, by consequence, they are of no interest to us in our search for the globally optimal minimizer of (3).

Example 1: Consider the third order model used in [12]:

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

for which we search the optimal first-order ($m = 1$) approximation. The polynomial $l(s)$ from (10) is equal to,

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

The variety $\mathcal{V}_\mathbb{R}$ defined in (13) corresponds to the set of real-valued common roots of the following square system of polynomial equations in the model parameters $\{\hat{a}_0, \hat{b}_0\}$ and auxiliary variables $\{g_0, g_1\}$,

$$\begin{cases} 0 = f_3 = 1 - g_1 - \hat{b}_0, \\ 0 = f_2 = \hat{a}_0 - 12\hat{b}_0 - g_0 + 2\hat{a}_0 g_1 + 9, \\ 0 = f_1 = 9\hat{a}_0 - 49\hat{b}_0 + 2\hat{a}_0 g_0 - \hat{a}_0^2 g_1 - 10, \\ 0 = f_0 = -g_0\hat{a}_0^2 - 10\hat{a}_0 - 78\hat{b}_0. \end{cases} \qquad (14)$$

B. Quadratic m-parameter eigenvalue problem

Closer observation of the system of equations in (14) shows that only the parameter $\hat{a}_0$ appears nonlinearly, justifiably indicating that the nonlinearity of the model reduction problem is solely caused by the search for the optimal poles, i.e., the optimal values for the parameters in $\hat{\boldsymbol{a}}$. We can make this (partially) linear structure of the problem explicit by 'extracting' the variables that appear only linearly (the $\{\hat{b}_i\}_{i=0,\dots,m-1}$ and $\{g_i\}_{i=0,\dots,n-m-1}$) in the system of polynomial equations from Proposition 1, and organize them into an eigenvector, giving a quadratic m-parameter eigenvalue problem:

$$\underbrace{\left(\sum_{\{\boldsymbol{\alpha}\}} \boldsymbol{M}_{\boldsymbol{\alpha}} \hat{\boldsymbol{a}}^{\boldsymbol{\alpha}}\right)}_{\mathcal{M}(\hat{\boldsymbol{a}})} \begin{bmatrix} 1 \\ \hat{\boldsymbol{b}} \\ \boldsymbol{g} \end{bmatrix} = \boldsymbol{0}, \tag{15}$$

where the vectors $\hat{\boldsymbol{b}}, \boldsymbol{g}$ are similarly defined as before, and $\mathcal{M}(\hat{\boldsymbol{a}}) \in \mathbb{R}^{(n+m)\times(n+1)}[\hat{\boldsymbol{a}}]$ is a matrix polynomial[3] in the monomials $\hat{\boldsymbol{a}}^{\boldsymbol{\alpha}} = \hat{a}_0^{\alpha_0} \dots \hat{a}_{m-1}^{\alpha_{m-1}}$.

Proposition 2: Given the square system of $m+n$ equations $f_0 = \dots = f_{n+m-1} = 0$ and the variety $\mathcal{V}_{\mathbb{R}}$, both defined as in Proposition 1, consider the corresponding quadratic m-parameter eigenvalue problem as defined in (15). Denote the set of real-valued eigentuples of this MEVP by $\mathcal{V}_{\mathbb{R}}(\hat{a})$. Then, $\mathcal{V}_{\mathbb{R}}(\hat{a})$ is equal to the projection of $\mathcal{V}_{\mathbb{R}}$ onto $\hat{\boldsymbol{a}}$.

Example 1 (continued): Extracting the variables that appear only linearly out of the system in (14) leads to the polynomial matrix $\mathcal{M}(\hat{a}_0)$:

$$\underbrace{\begin{bmatrix} 1 & -1 & -1 & 0 \\ 9+\hat{a}_0 & -12 & 2\hat{a}_0 & -1 \\ -10+9\hat{a}_0 & -49 & -\hat{a}_0^2 & 2\hat{a}_0 \\ -10\hat{a}_0 & -78 & 0 & -\hat{a}_0^2 \end{bmatrix}}_{\mathcal{M}(\hat{a}_0)} \begin{bmatrix} 1 \\ \hat{b}_0 \\ g_1 \\ g_0 \end{bmatrix} = \boldsymbol{0},$$

which can be rewritten as a quadratic polynomial eigenvalue problem (PEVP) in the parameter $\hat{a}_0$:

$$\left(\underbrace{\begin{bmatrix} 1 & -1 & -1 & 0 \\ 9 & -12 & 0 & -1 \\ -10 & -49 & 0 & 0 \\ 0 & -78 & 0 & 0 \end{bmatrix}}_{\boldsymbol{M}_0} + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 9 & 0 & 0 & 2 \\ -10 & 0 & 0 & 0 \end{bmatrix}}_{\boldsymbol{M}_1} \hat{a}_0 + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}}_{\boldsymbol{M}_2} \hat{a}_0^2 \right) \begin{bmatrix} 1 \\ \hat{b}_0 \\ g_1 \\ g_0 \end{bmatrix} = \boldsymbol{0}. \tag{16}$$

[3]Each term in the system of equations from Proposition 1 is mapped to a specific position in one of the coefficient matrices of the MEVP. The term has a monomial in $\hat{\boldsymbol{a}}$ as a factor that determines in which coefficient matrix the coefficient of the term has to be included. The position within that matrix depends on the equation to which the term belongs (the row) and the presence of a variable from the eigenvector of the MEVP in the term (the column).

TABLE I

PROPERTIES OF THE SYSTEM OF POLYNOMIAL EQUATIONS FOR THE GLOBALLY OPTIMAL TECHNIQUES DESCRIBED IN THE LITERATURE.

(n,m)	Approach	#eqs	#vars	$d_{\max}$	n_{B}
$(3,1)$	Agudelo's approach [12]	2	2	8	32
	Alsubaie's approach* [11]	7	7	2	128
	Novel methodology	4	4	3	18
$(6,2)$	Agudelo's approach [12]	4	4	15	10290
	Alsubaie's approach* [11]	14	14	2	16384
	Novel methodology	8	8	3	1458
$(17,4)$	Agudelo's approach [12]	8	8	41	$\mathcal{O}(10^{11})$
	Alsubaie's approach* [11]	38	38	2	$\mathcal{O}(10^{11})$
	Novel methodology	21	21	3	$\mathcal{O}(10^{9})$

*for this methodology additional post-processing is required.

C. Computing the stationary points

From Theorem 3, we know that all stationary points $\hat{H}(s)$ of order m can be computed via a polynomial root finding problem, for which advanced solvers exists (e.g., [23]). However, as these root finding techniques generally work over the field of complex numbers, one has to prune away the complex-valued common roots to obtain $\mathcal{V}_{\mathbb{R}}$. Alternatively, we can exploit the partially linear structure of the problem and calculate $\mathcal{V}_{\mathbb{R}}$ via, $\mathcal{V}_{\mathbb{R}}(\hat{a})$, the real-valued eigentuples of the MEVP, as described in the above-mentioned proposition. By definition, the matrix $\mathcal{M}(\hat{\boldsymbol{a}})$ is rank deficient when evaluated in one of its eigentuples, allowing to calculate for each eigentuple the corresponding values for the parameters $\hat{\boldsymbol{b}}$ via the null space of $\mathcal{M}(\hat{\boldsymbol{a}})$. Specialized solvers (e.g., [24], [25]) can be used to compute the eigentuples of the MEVP.

D. Comparing complexity

There is no standard means to quantify the complexity of a system of polynomial equations before it has been solved, but the size of the system, measured as the number of variables (#vars) and the number of equations (#eqs), and the highest polynomial degree present in the system ($d_{\max}$) serve as good indicators. Bézout's number n_{B}, equal to the product of the degree of each of the polynomials in a square system of equations, scales with the above-mentioned indicators and could therefore be used as an elementary[4] proxy for its complexity. We compare the alternative approaches for globally optimal SISO H_2-norm model reduction [11], [12] in Table I, which depicts the properties of the to-be-solved system of polynomial equations[5] for three scenarios of (n,m) that correspond to the examples tackled in Section IV.

For growing (n,m) the novel approach comes with the lowest complexity. Contrary to the approach proposed in [12], each polynomial in our system of equations is at most *cubic*, which makes the corresponding MEVP *quadratic*, independent of the model orders (n,m). This low degree is advantageous, e.g., in the context of homotopy continuation

[4]There exist more sophisticated bounds that, often based on *mixed-volume* computations, additionally consider the sparsity of the polynomials [26].

[5]Comparing the complexity of MEVPs is even less straightforward. So, to be able to include the methods of [11], [12], which lead to MEVPs, we consider the properties of the system of polynomial equations obtained by writing out the matrix-vector product of the to-be-solved MEVP.

methods the step size of the continuation procedure typically has to be reduced when operating on systems of higher degree to avoid *path jumping* [27], thereby increasing the overall computation time. Alsubaie's approach [11] leads to linear MEVPs, but the obtained numbers of equations are relatively large (see Table I). This indicates that the coefficient matrices have more rows compared to the novel approach, resulting in an overall higher complexity.

E. Additional notes

Discrete-time: The results of this Section can also be derived in the discrete-time setting. The factor $[\hat{a}(-s)]^2$ in (10) has to be replaced by $[z^m\hat{a}(z^{-1})]^2$, adjusting for the change in the definition of the 'mirror images' of the poles.

Secular equations: When $m = 1$, the MEVP from (15) becomes a PEVP with square coefficient matrices $\{\boldsymbol{M}_\alpha\}_{\alpha=0,1,2} \in \mathbb{R}^{n+1\times n+1}$. In this case, we can alternatively calculate the eigenvalues of the PEVP by rooting a univariate polynomial: the square matrix $\mathcal{M}(\hat{a}_0)$ is singular if its determinant is equal to zero. To give an example, the determinant of the PEVP in (16) gives:

$$\det(\mathcal{M}(\hat{a}_0)) = \hat{a}_0^5+15\hat{a}_0^4-89\hat{a}_0^3-1191\hat{a}_0^2-2596\hat{a}_0+780. \quad (17)$$

This reformulation of the system of equations from (13), eliminating the n 'linear' variables at the cost of higher polynomial degrees, is possible for arbitrary model orders (n, m). Indeed, all minors[6] of the rectangular matrix $\mathcal{M}(\hat{\boldsymbol{a}})$ from (15) should vanish when evaluated in an eigentuple of the MEVP. So, by equating these minors to zero we obtain an overdetermined system of equations in the m variables from $\hat{\boldsymbol{a}}$, the common roots of which correspond to the eigentuples of the MEVP. The number of all possible minors grows quickly with (n, m). However, it turns out that not all minors are required to be able to retrieve all the eigentuples of the MEVP. We will consider this in more detail in future work.

At infinity: The nullity of block Macaulay matrices for growing degree d and the rank structure of the rows in their null space, can be used to analyse both the affine solution set of the MEVP in (15) as the solution set at infinity[7] [13]. Although for all of our numerical experiments the affine solution set has been zero-dimensional, the solution set at infinity seems to be $(m-1)$-dimensional: for $m = 2$ the number of solutions at infinity found via the block Macaulay matrix of degree d grows linearly with d (constant increases of 5), for $m = 3$ the growth of the number of solutions at infinity grows linearly with d (constant increases of 7), and so on. Furthermore, we observed that the system of equations from (13) has an m-dimensional solution set at infinity, which indicates that the reformulation as MEVP removes one dimension of the solution set at infinity. Lastly, the set of secular equations of the MEVP in (15), i.e., the minors of the matrix $\mathcal{M}(\hat{\boldsymbol{a}})$, has no solutions at infinity.

[6]The minors are the $\binom{n+m}{m-1}$ determinants of the $(n+1)\times(n+1)$ submatrices, obtained by omitting $m-1$ rows, of the polynomial matrix $\mathcal{M}(\hat{\boldsymbol{a}})$.

[7]Only the affine solutions are of interest in the context of (3). When working with the Macaulay matrix framework, the effects of the solutions at infinity can be eliminated via a *column compression*. See e.g., [13], [24].

IV. Numerical examples

To validate the proposed methodology we compute[8] the numerical solutions of Example 1 and tackle two other examples from the literature. We only consider the affine common roots (eigentuples) of the obtained systems of polynomial equations (resp. the MEVPs).

Example 1 (continued): The univariate polynomial in (17) has five roots: $\{9.679, -16.619, 0.267, -4.164 \pm 0.903j\}$, which coincide with the eigenvalues of the PEVP in (16). The nullity of the block Macaulay matrix [13] composed from (16) stabilizes at $d = 4$, allowing to retrieve the eigenvalues via the shift-invariance property of its null space. The system of equations in (14) has 5 common roots, the values for the variable $\hat{a}_0$ of which correspond to the above-mentioned eigenvalues of the PEVP. Only two solutions $(\hat{a}_0, \hat{b}_0)$ remain after pruning for the stationary points: $\mathcal{V}'_{\mathbb{R}} = \{(9.679, 1.279), (0.267, -0.0437)\}$. We can select the globally optimal minimizer using their respective objective function values: $J{=}\{0.278, 0.398\}$.

Example 2: Consider the *four disk* system (discrete-time, $n = 6$) described in [19] for which we want to calculate the globally optimal second-order approximant ($m = 2$):

$$H(z){=}\tfrac{0.0448z^5+0.2368z^4+0.0013z^3+0.0211z^2+0.2250z+0.0219}{z^6-1.2024z^5+2.3675z^4-2.0039z^3+2.2337z^2-1.0420z+0.8513}.$$

Equating the coefficients with respect to the powers of z in the discrete-time equivalent of (10) to zero, gives a square system of 8 polynomial equations. The polynomial matrix $\mathcal{M}(\hat{a}_1, \hat{a}_0)$ from the MEVP formulation in (15) is equal to:

$$\left[\begin{array}{c:c:c:c:c:c} 0.045 & -1 & 0 & -\hat{a}_0 & 0 & 0 & 0 \\ 0.045\hat{a}_1{+}0.24 & 1.20 & -1 & -2\hat{a}_0\hat{a}_1 & -\hat{a}_0^2 & 0 & 0 \\ 0.045\hat{a}_0{+}0.24\hat{a}_1{+}0.001 & -2.37 & 1.20 & -\hat{a}_1^2{-}2\hat{a}_0 & -2\hat{a}_0\hat{a}_1 & -\hat{a}_0^2 & 0 \\ 0.24\hat{a}_0{+}0.001\hat{a}_1{+}0.02 & 2.01 & -2.37 & -2\hat{a}_1 & -\hat{a}_1^2{-}2\hat{a}_0 & -2\hat{a}_0\hat{a}_1 & -\hat{a}_0^2 \\ 0.001\hat{a}_0{+}0.02\hat{a}_1{+}0.22 & -2.23 & 2.01 & -1 & -2\hat{a}_1 & -\hat{a}_1^2{-}2\hat{a}_0 & -2\hat{a}_0\hat{a}_1 \\ 0.02\hat{a}_0{+}0.22\hat{a}_1{+}0.02 & 1.04 & -2.23 & 0 & -1 & -2\hat{a}_1 & -\hat{a}_1^2{-}2\hat{a}_0 \\ 0.22\hat{a}_0{+}0.02\hat{a}_1 & -0.85 & 1.04 & 0 & 0 & -1 & -2\hat{a}_1 \\ 0.02\hat{a}_0 & 0 & -0.85 & 0 & 0 & 0 & -1 \end{array}\right]$$

which leads to a quadratic 2-parameter eigenvalue problem:

$$\Big(\boldsymbol{M}_0 + \boldsymbol{M}_{10}\hat{a}_0 + \boldsymbol{M}_{01}\hat{a}_1 + \boldsymbol{M}_{20}\hat{a}_0^2 + \boldsymbol{M}_{11}\hat{a}_0\hat{a}_1 + \boldsymbol{M}_{02}\hat{a}_1^2\Big)\begin{bmatrix}1 & \hat{b}_1 & \hat{b}_0 & g_3 & g_2 & g_1 & g_0\end{bmatrix}^\mathsf{T} = \boldsymbol{0},$$

with coefficient matrices $\boldsymbol{M}_{\boldsymbol{\alpha}} \in \mathbb{R}^{8\times 7}$. The block Macaulay matrix of degree 10 suffices to retrieve the 49 eigentuples, 11 of which real-valued. The variety $\mathcal{V}'_{\mathbb{R}}$ contains 5 stationary points, listed in Table II.

Example 3: We search for the globally optimal 4th order reduced model ($m = 4$) of the continuous-time state-space model ($n = 17$) considered in *Example 3* of [28]. The model describes the interaction between a torque activator and an approximately collocated torsional rate sensor for the ACES structure [29]. Because of the high complexity (see Table I) we used the package of [23] to calculate the common roots of the system of equations from (13), which took

[8]We used a MacBook Pro with a 6-core Intel i7 CPU (2019) working at 2.6 GHz with access to 32 GB RAM. All *timings* are averaged over 5 consecutive runs and the numerical results are rounded for displaying purposes. A MATLAB implementation of the (block) Macaulay method [24] is available online at `www.macaulaylab.net`.

TABLE II

THE STATIONARY POINTS CORRESPONDING TO EXAMPLE 2.

$J^\dagger$	$\hat{a}_1$	$\hat{a}_0$	$\hat{b}_1$	$\hat{b}_0$
0.868	-0.293	0.941	0.139	0.266
1.076	0.505	0.930	-0.254	-0.120
1.124	0.267	0.820	-0.294	0.167
1.174	-1.423	0.969	0.069	0.028
1.254	-0.992	0.534	0.132	0.086

†the discrete-time equivalent of (4).

$3\,\mathrm{h}\,24\,\mathrm{min}\,2\,\mathrm{s}$. Contrary to the block Macaulay method [24], homotopy continuation methods can exploit *parallelism* in the computations, allowing to tackle larger problems. The poles p_i of the four stationary points with smallest value for J are depicted in Table III, the first of which corresponds to the globally optimal solution. For comparison, we applied the iterative method described in [6] using random sets of initial interpolation points to this example: the iteration generally converged to locally optimal reduced models. Only in (very) rare cases the globally optimal solution was obtained.

V. CONCLUSION AND FUTURE WORK

We derived a novel methodology for globally optimal SISO H_2-norm model reduction. The parameters of the optimal model(s) are retrieved via the eigentuples of a multiparameter eigenvalue problem (MEVP). The MEVP is quadratic, independent of the model orders, which is favorable from a computational point of view. We compared the complexity of the novel technique with the few alternative globally optimal approaches in the literature, and solved three numerical examples to validate the proposed methodology.

In future work, we want to develop specialized solvers that compute only the real-valued eigentuples of the MEVP, since these are the only ones of interest. Ideally, we would solve for the global optimal minimizer(s) only. The observation that also for MIMO models (tangential) interpolatory conditions for first-order optimality exist, motivates the search for a generalization of the presented methodology.

TABLE III

THE POLES p_i OF THE FOUR STATIONARY POINTS WITH THE LOWEST VALUE OF J CORRESPONDING TO EXAMPLE 3.

J	$p_{1,2}$	$p_{3,4}$
9.14×10^{-3}	$-0.032 \pm 78.54j$	$-0.111 \pm 15.43j$
9.22×10^{-3}	$-0.032 \pm 78.54j$	$-5.713 \pm 52.57j$
1.03×10^{-2}	$-0.032 \pm 78.54j$	$-0.023 \pm 3.842j$
1.09×10^{-2}	$-0.032 \pm 78.54j$	$-4.663 \pm 15.88j$

REFERENCES

[1] P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*. Marcel Dekker Inc., New York, NY, 1995.

[2] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*. SIAM, Philadelphia, PA, 2005.

[3] D. Wilson, "Optimum solution of model-reduction problem," *Proceedings of the Institution of Electrical Engineers*, vol. 117, no. 6, pp. 1161–1165, 1970.

[4] D. Hyland and D. Bernstein, "The optimal projection equations for model reduction and the relationschips among the methods of Wilson, Skelton, and Moore," *IEEE Transactions on Automatic Control*, vol. 30, no. 12, pp. 1201–1211, 1985.

[5] L. Meier and D. Luenberger, "Approximation of linear constant systems," *IEEE Transactions on Automatic Control*, vol. 12, no. 5, pp. 585–588, Oct. 1967.

[6] S. Gugercin, A. Antoulas, and C. Beattie, "$\mathcal{H}_2$ model reduction for large-scale linear dynamical systems," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 2, pp. 609–638, 2008.

[7] P. Van Dooren, K. Gallivan, and P.-A. Absil, "$\mathcal{H}_2$-optimal model reduction with higher-order poles," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 5, pp. 2738–2753, 2010.

[8] J. Spanos, M. Milman, and D. Mingori, "A new algorithm for L_2 optimal model reduction," *Automatica*, vol. 28, no. 5, pp. 897–909, 1992.

[9] M. Ahmad, M. Frangos, and I. Jaimoukha, "Second order $\mathcal{H}_2$ optimal approximation of linear dynamical systems," in *IFAC Proceedings Volumes*, vol. 44, no. 1, 2011, pp. 12 767–12 771.

[10] B. Hanzon, J. M. Maciejowski, and C. T. Chou, "Optimal H2 order-one reduction by solving eigenproblems for polynomial equations," 2007. [Online]. Available: https://arxiv.org/abs/0706.1862

[11] F. F. Alsubaie, "$\mathcal{H}_2$ optimal model reduction for linear dynamical systems and the solution of multiparameter matrix pencil problems," Ph.D. dissertation, Imperial College London, London, UK, 2019.

[12] O. M. Agudelo, C. Vermeersch, and B. De Moor, "Globally optimal H_2-norm model reduction: A numerical linear algebra approach," *IFAC PapersOnLine (Special Issue: Proceedings of MTNS 2020)*, vol. 54, no. 9, pp. 564–571, 2021.

[13] K. De Cock and B. De Moor, "Multiparameter eigenvalue problems and shift-invariance," *IFAC-PapersOnLine (Special Issue: Proceedings of MTNS 2020)*, vol. 54, no. 9, pp. 159–165, 2021.

[14] P. Aigrain and E. Williams, "Synthesis of n-reactance networks for desired transient response," *Journal of Applied Physics*, vol. 20, no. 6, pp. 597–600, 1949.

[15] A. Bultheel and B. De Moor, "Rational approximation in linear systems and control," *Journal of Computational and Applied Mathematics*, vol. 121, no. 1–2, pp. 355–378, 2000.

[16] J. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, 3rd ed. American Mathematical Society, 1960.

[17] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," 2009. [Online]. Available: https://arxiv.org/abs/0906.4835

[18] A. Lepschy, G. A. Mian, G. Pinato, and U. Viaro, "Rational L_2 approximation: A non-gradient algorithm," in *Proceedings of the 30th IEEE Conference on Decision and Control*, Dec. 1991, pp. 2321–2323.

[19] B. De Moor, P. Van Overschee, and G. Schelfhout, "H_2 model reduction for SISO systems," in *Proceedings of the 12th IFAC World Congress*, Sydney, Australia, Jul. 1993, pp. 227–230.

[20] H. Fan and Doroslovački, "On "global convergence" of Steiglitz–McBride adaptive algorithm," *IEEE Transactions on Circuits and Systems–II: Analog and Digital Signal Processing*, vol. 40, no. 2, pp. 73–87, 1993.

[21] W. Krajewski, "Optimality conditions in multivariable L_2 model reduction," *Journal of the Franklin Institute*, vol. 330, no. 3, pp. 431–439, 1993.

[22] A. Bunse-Gerstner, D. Kubalińska, G. Vossen, and D. Wilczek, "h_2-norm optimal model reduction for large scale discrete dynamical MIMO systems," *Journal of Computational and Applied Mathematics*, vol. 223, no. 5, pp. 1202–1216, 2010.

[23] P. Breiding and S. Timme, "HomotopyContinuation.jl: A package for homotopy continuation in Julia," in *International Congress on Mathematical Software*, 2018, pp. 458–465.

[24] C. Vermeersch and B. De Moor, "Two complementary block Macaulay matrix algorithms to solve multiparameter eigenvalue problems," *Linear Algebra and its Applications*, vol. 654, pp. 177–209, 2022.

[25] B. Plestenjak, "MultiParEig," MATLAB Central File Exchange (https://www.mathworks.com/matlabcentral/fileexchange/47844-multipareig), 2023, retrieved March 2, 2023.

[26] T. Li and X. Wang, "The BKK root count in C^n," *Mathematics of Computation*, vol. 65, no. 216, pp. 1477–1484, 1996.

[27] S. Telen, M. Van Barel, and J. Verschelde, "A robust numerical path tracking algorithm for polynomial homotopy continuation," *SIAM Journal on Scientific Computing*, vol. 42, no. 6, pp. 3610–3637, 2020.

[28] D. Žigić, L. T. Watson, E. Collins, and D. S. Bernstein, "Homotopy methods for solving the optimal projection equations for the H_2 reduced order model problem," *International Journal of Control*, vol. 56, no. 1, pp. 173–191, 1992.

[29] E. Collins, D. Phillips, and D. Hyland, "Robust decentralized control laws for the ACES structure," *IEEE Control Systems Magazine*, vol. 11, no. 3, pp. 62–70, 1991.