2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Exact Characterization of the Global Optima of Least Squares Realization of Autonomous LTI Models as a Multiparameter Eigenvalue Problem

Sibren Lagauw*, Lukas Vanpoucke*, Bart De Moor*, *Fellow, IEEE & IFAC & SIAM*
**Center for Dynamical Systems, Signal Processing, and Data Analytics (STADIUS)*
Department of Electrical Engineering (ESAT), KU Leuven
Leuven, Belgium
Email addresses: {sibren.lagauw, lukas.vanpoucke, bart.demoor}@esat.kuleuven.be

***Abstract*—We consider the problem of finding the best least squares realization of an autonomous single-output linear time-invariant dynamical system, given a sequence of non-model-compliant output data. We characterize the solution set of the identification problem and derive novel properties of the optimal models. We show how the global minima of the problem follow from the eigentuples of a multiparameter eigenvalue problem and illustrate this result using several numerical 'toy examples' in which we compute the globally optimal solution(s) explicitly.**

***Index Terms*—Discrete-time systems; Modeling; Linear systems; Parameter estimation; Model/Controller reduction;**

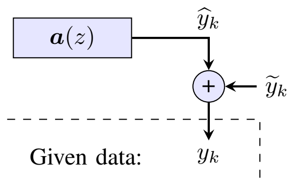

Fig. 1. Schematic overview of the autonomous LTI single-output model. The given output data $\boldsymbol{y} \in \mathbb{R}^N$ are modified using the so-called misfits $\widetilde{\boldsymbol{y}} \in \mathbb{R}^N$, so that the modified data $\widehat{\boldsymbol{y}} \in \mathbb{R}^N$ are model-compliant. The coefficients $\boldsymbol{a} \in \mathbb{R}^{n+1}$ of the nth degree polynomial $a(z)$ in the forward-shift z are the unknown model parameters.

I. Introduction

A data sequence that can be generated by a specific mathematical model will be called *model-compliant*. Said in other words, the data belong to the *behavior* of that model, which is the set of all model-compliant data sequences [1]. For a user-specified model, however, given data are almost never model-compliant: they do not belong to the behavior of the model. There could be many reasons for this: e.g., observational errors, measurement inaccuracies, missing data, outliers, unobserved disturbances, or model mismatch.

One could try to expand the model class, but there is an almost infinite set of mathematical models to choose from. So, unless one has a priori information about the relevant models, this is not a very practical option. Indeed, mathematical models typically only allow a 'thin' set of data trajectories, indicating that the model forbids more than it allows [2], [3]. That is why in engineering applications, models are a matter of inspiration rather than deduction [1]. This naturally leads to the alternative consideration of trying to modify the given data as little as possible, so that the modified data are model-compliant with the pre-specified model, where the modification of the given data will be called the *misfit* between the model and the given data[1]. In order to quantify its size, the choice for a least squares criterion seems to be a natural one.

We will confine our attention to *single-output, linear time-invariant (LTI), causal, lumped parameter dynamical models in discrete time, with a pre-specified model order* n *(corresponding to the number of states)*. For this model class, model-compliant data $\widehat{\boldsymbol{y}} = [\widehat{y}_0, \dots, \widehat{y}_{N-1}]^\mathsf{T} \in \mathbb{R}^N$, assuming $N > n$, must satisfy a difference equation of the form:

$$a_0\widehat{y}_{k+n} + a_1\widehat{y}_{k+n-1} + \cdots + a_n\widehat{y}_k = a(z)y_k = 0,$$

for all $k = 0, \dots, N-n-1$, where $a(z) = a_0z^n + a_1z^{n-1} + \dots + a_n$ is a degree n polynomial in the forward-shift operator z (i.e., $zy_k = y_{k+1}$). This implies that,

$$\begin{bmatrix} a_n & \dots & \dots & a_1 & a_0 & 0 & \dots & 0 \\ 0 & a_n & \dots & \dots & a_1 & a_0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & a_n & \dots & \dots & a_1 & a_0 \end{bmatrix} \widehat{\boldsymbol{y}} = \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\widehat{\boldsymbol{y}} = \boldsymbol{0}, \quad (1)$$

indicating that the behavior of the autonomous model to which the data $\widehat{\boldsymbol{y}}$ belong can be characterized as the n-dimensional kernel of the banded-Toeplitz matrix $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}} \in$

This work was supported in part by the KU Leuven Research Fund under Project iBOF/23/064 and C16/15/059, and several LRD bilateral industrial projects; in part by the Flemish Government agencies: FWO through the EOS Project under Grant G0F6718N (SeLMA) and EWI through the Flanders AI Research Program; and in part by the European Commission (ERC Adv. Grant) under Grant 885682. The work of S. Lagauw was supported by the FWO Fundamental Research Fellowship under Grant 11K5623N.

[1]In the statistical literature, often a priori probabilistic assumptions on the inaccuracies that perturb the model-compliant data are made, for instance that they follow a multivariate Gaussian distribution [4]. Via maximum-likelihood, this then leads to so-called *errors-in-variables* models [5], but as these assumptions are unverifiable in practice, we prefer the purely deterministic approach of this paper.

$\mathbb{R}^{(N-n)\times N}$. The Toeplitz-vector product in the left-hand side of (1) can be rewritten as $\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}\boldsymbol{a}$, with $\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}} \in \mathbb{R}^{(N-n)\times(n+1)}$ the Hankel matrix constructed from the elements of $\widehat{y}$ and $\boldsymbol{a} = [a_n, a_{n-1}, \dots, a_0]^\mathsf{T} \in \mathbb{R}^{n+1}$. Hence, one can easily retrieve the model parameters $\boldsymbol{a}$ associated with the model-compliant data via the kernel of $\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}$.

Given data $\boldsymbol{y} = [y_0, \dots, y_{N-1}]^\mathsf{T} \in \mathbb{R}^N$, however, are generally not model-compliant, such that $\boldsymbol{Y}_{\boldsymbol{N-n}}$ is of full column rank. In the *least squares realization problem*, the given output data $\boldsymbol{y}$ are modified using the so-called misfits $\widetilde{\boldsymbol{y}} = [\widetilde{y}_0, \dots, \widetilde{y}_{N-1}]^\mathsf{T} \in \mathbb{R}^N$, the 2-norm of which is to be minimized, such that the modified data $\widehat{\boldsymbol{y}} = \boldsymbol{y} - \widetilde{\boldsymbol{y}}$ are model-compliant,

$$\begin{aligned} \min_{\boldsymbol{a},\widehat{\boldsymbol{y}}} \quad & \frac{1}{2}\|\widetilde{\boldsymbol{y}}\|_2^2 = \frac{1}{2}\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|_2^2, \\ \text{s.t.} \quad & \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\widehat{\boldsymbol{y}} = \boldsymbol{0},\ \boldsymbol{e}^\mathsf{T}\boldsymbol{a} = 1, \end{aligned} \tag{2}$$

where $\boldsymbol{e} \in \mathbb{R}^{n+1}$ is some given, fixed, non-zero vector. The second constraint is necessary[2] to avoid the trivial solution ($\boldsymbol{a} = \boldsymbol{0}$). This modeling setup[3] is depicted in Fig. 1.

Even though the model class is linear, (2) is a nonlinear, nonconvex optimization problem, implying that (many) local optima can exist. Consequently, applying iterative optimization algorithms (see, e.g., [4], [6], [8] and references therein) to the realization problem brings along several complications: e.g., the performance depends on the chosen initial point, reproducibility of the obtained results is not always guaranteed and certification of global optimality if a 'sufficiently good' solution is found is generally impossible. By contrast, we deem the realization problem 'solved' if and only if the globally optimal model(s) have been identified by means of a deterministic procedure. In accordance with this reasoning, it was shown in [3] that (2) is essentially a rectangular multiparameter eigenvalue problem (MEP) [9]–[11], the eigentuples of which lead to the globally optimal model parameters.

Contributions: In Theorem 2, we formalize a finite-dimensional version of what is called 'Walsh's Theorem' in [12, Theorem 3.14]. This characterization of the optimal misfits, as the result of filtering an unknown signal twice by the same finite impulse response (FIR) filter, was initially observed in [3]. Then, inspired by [3] and encouraged by our previous work [13], where we exploited [12, Theorem 3.14] to derive a novel methodology for globally optimal SISO H_2-norm model reduction, we use the obtained characterization of the misfits to derive a novel, alternative MEP that is smaller than the one obtained in [3]. Although we provide numerical 'toy examples' to validate our findings, our contribution is of theoretical nature.

[2] We will see that the Lagrange multiplier associated with the non-triviality constraint is zero, such that the results in this paper remain the same for other choices of normalization, e.g., the quadratic constraint $\boldsymbol{a}^\mathsf{T}\boldsymbol{a} = 1$.

[3] Besides its use in system identification, the formulation in (2) also arises in the context of *shape-from-moment* problems and/or the estimation of the *direction of arrival (DOA)* in array processing when inexact data is considered, see, e.g., [3], [6], [7] and references therein.

II. THE LEAST SQUARES REALIZATION PROBLEM

In this section, we use the first-order necessary conditions for optimality (FONC) of the realization problem (2) to characterize the optimal misfits. The obtained results constitute the foundations of the methodology proposed in Section III.

Consider the Lagrangian of (2),

$$\mathcal{L}(\boldsymbol{a}, \widehat{\boldsymbol{y}}, \boldsymbol{l}, \lambda) = \frac{1}{2}\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|_2^2 + \boldsymbol{l}^\mathsf{T}\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\widehat{\boldsymbol{y}} + \lambda(\boldsymbol{e}^\mathsf{T}\boldsymbol{a} - 1),$$

where the variables $\boldsymbol{l} \in \mathbb{R}^{N-n}$ and $\lambda \in \mathbb{R}$ are Lagrange multipliers. The FONC of (2) can now be obtained as,

$$\begin{aligned} \partial\mathcal{L}/\partial\widehat{\boldsymbol{y}} &= \widehat{\boldsymbol{y}} - \boldsymbol{y} + (\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^\mathsf{T}\boldsymbol{l} = \boldsymbol{0}, \\ \partial\mathcal{L}/\partial\boldsymbol{a} &= (\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}})^\mathsf{T}\boldsymbol{l} - \boldsymbol{e}\lambda = \boldsymbol{0}, \\ \partial\mathcal{L}/\partial\boldsymbol{l} &= \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}\boldsymbol{a} = \boldsymbol{0}, \\ \partial\mathcal{L}/\partial\lambda &= 1 - \boldsymbol{e}^\mathsf{T}\boldsymbol{a} = 0. \end{aligned} \tag{3}$$

Pre-multiplying the second equation with $\boldsymbol{a}^\mathsf{T}$, and using the third and fourth equation indicates that $\lambda = 0$, such that the FONC in (3) are equivalent to,

$$\widetilde{\boldsymbol{y}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = (\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^\mathsf{T}\boldsymbol{l}, \tag{4}$$

$$\boldsymbol{l}^\mathsf{T}\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}} = \boldsymbol{0}, \tag{5}$$

$$\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\widehat{\boldsymbol{y}} = \boldsymbol{0}, \tag{6}$$

$$\boldsymbol{e}^\mathsf{T}\boldsymbol{a} = 1.$$

These relations, the real-valued solutions of which are the *stationary points* of (2), lead to the following results.

If the model parameters $\boldsymbol{a}$ were to be known (cf. *the projection onto the behavior problem*), the equality $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y} = \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\widetilde{\boldsymbol{y}}$ would constitute an underdetermined linear system in the variables $\widetilde{\boldsymbol{y}}$. Since the matrix $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}$ is of full row rank, the unique minimal 2-norm solution of this system could be computed via the pseudo-inverse of $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}$. In this regard, the 2-norm of the optimal misfit $\widetilde{\boldsymbol{y}}$ can be interpreted as a measure of the 'distance' between the model determined by the parameters $\boldsymbol{a}$, and the given data $\boldsymbol{y}$.

Theorem 1 (Projection onto the behavior)**.** *For given output data $\boldsymbol{y} \in \mathbb{R}^N$ and a model order n, with $N > n$, then the minimal norm misfit $\widetilde{\boldsymbol{y}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} \in \mathbb{R}^N$ in the least squares realization problem* (2) *can be expressed as the orthogonal projection of $\boldsymbol{y}$ onto* $\mathrm{row}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})$,

$$\widetilde{\boldsymbol{y}} = (\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^\mathsf{T}(\boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{-1}\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y}, \tag{7}$$

where $\boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}} = \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^\mathsf{T} \in \mathbb{R}^{(N-n)\times(N-n)}$. Consequently, the optimal model-compliant data corresponds to,

$$\widehat{\boldsymbol{y}} = \left(\boldsymbol{I} - (\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^\mathsf{T}(\boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{-1}\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\right)\boldsymbol{y}. \tag{8}$$

Proof. Use (6) and (4), respectively, to show that,

$$\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y} = \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\widetilde{\boldsymbol{y}} = \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^\mathsf{T}\boldsymbol{l} = \boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{l}, \tag{9}$$

where $\boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}}$ is a positive definite matrix, such that,

$$\boldsymbol{l} = (\boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{-1}\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y}. \tag{10}$$

Combine this with (4) and $\widehat{\boldsymbol{y}} = \boldsymbol{y} - \widetilde{\boldsymbol{y}}$ to obtain (7) and (8). $\square$

Observe that Theorem 1 decomposes the ambient data space $\mathbb{R}^N$ into two orthogonal subspaces[4]: the optimal misfit $\widetilde{\boldsymbol{y}}$ resides in the $(N-n)$-dimensional row space of $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}$, whereas the model-compliant data sequence $\widehat{\boldsymbol{y}}$ lies in the n-dimensional kernel of $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}$. Obviously, this implies that the optimal model-compliant data and the optimal misfits are orthogonal with respect to each other, i.e.,

$$\widehat{\boldsymbol{y}}^{\mathsf{T}}\widetilde{\boldsymbol{y}} = 0.$$

Additionally, we can deduce from Theorem 1 that the model parameters $\boldsymbol{a}$ suffice to describe a particular stationary point of the realization problem (2). Indeed, for given data $\boldsymbol{y}$, the model parameters $\boldsymbol{a}$ implicitly define a unique $\widehat{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{y}}$ via the projections in (7) and (8), respectively.

Theorem 2. *Given a sequence of output data $\boldsymbol{y} \in \mathbb{R}^N$ and a model order n, with[5] $N > 2n$, and a stationary point $\boldsymbol{a}$ of* (2) *for which the model-compliant data $\widehat{\boldsymbol{y}}$ has $\operatorname{rank}(\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}) = n$ (i.e., $\widehat{\boldsymbol{y}}$ has nth order LTI dynamics), then the misfit $\widetilde{\boldsymbol{y}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} \in \mathbb{R}^N$ in the least squares realization problem* (2) *can be expressed as,*

$$\widetilde{\boldsymbol{y}} = (\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})^{\mathsf{T}}\boldsymbol{g},$$

for some $\boldsymbol{g} \in \mathbb{R}^{N-2n}$, where $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}} \in \mathbb{R}^{(N-2n)\times(N-n)}$ is a banded Toeplitz matrix defined similarly to the matrix $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}} \in \mathbb{R}^{(N-n)\times N}$ from (1).

Proof. Consider a stationary point $\boldsymbol{a}$ for which the matrix $\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}$ has rank n. Then, we know from the properties of a model-compliant data Hankel matrix that the $(N-2n)$-dimensional left null space of $\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}$ is spanned by the rows of the banded Toeplitz matrix $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}} \in \mathbb{R}^{(N-2n)\times(N-n)}$,

$$\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}}\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}} = \boldsymbol{0}.$$

Combined with (5), from which we know that $\boldsymbol{l}^{\mathsf{T}}$ lies in this left null space of $\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}$, it is clear that there must exist a vector $\boldsymbol{g} \in \mathbb{R}^{N-2n}$ such that,

$$\boldsymbol{l} = (\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})^{\mathsf{T}}\boldsymbol{g}. \tag{11}$$

Substituting the above into (4) gives the required result. □

The result in Theorem 2, which was initially encountered in [3, Section 9.3], can be seen to be a finite-dimensional version of [12, Theorem 3.14], since it expresses the misfit, i.e., the approximation error, in the stationary points of the least squares realization problem as the result of filtering an unknown signal $\boldsymbol{g}$ twice by the anti-causal FIR filter $a(z^{-1})$, where $a(z)$ determines the LTI dynamics of the optimal model. The original result [12, Theorem 3.14], named after Walsh in [12] to recognize its origins in rational approximation theory [17], states that for the first-order optimal solutions of the SISO H_2-norm model reduction problem[6], the approximation error function can always be expressed in the form $[z^n\widehat{a}(z^{-1})]^2 g(z)$ for some $g(z) \in \mathcal{H}_2$, where the nth order monic polynomial $\widehat{a}(z)$ is the denominator of the transfer function of the optimal, lower-order approximant. Note that contrary to the result in the context of model reduction, Theorem 2, where N is assumed to be finite, does not assume any form of stability of the estimated model.

Technical note: For stationary points $\boldsymbol{a}$ for which $\operatorname{rank}(\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}) < n$ (indicating that the model-compliant data obtained with (8) has LTI dynamics of an order strictly lower than n), the vector $\boldsymbol{l}$ from (10) is not guaranteed to lie exclusively in $\operatorname{row}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})$, since the latter is only a subspace of the left null space of $\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}$. As such, the relation in (11), and by consequence Theorem 2, is not guaranteed[7] to hold for these points. Extreme examples of this phenomenon are the *global maximizers* $\boldsymbol{a}$ of the realization problem: $\widehat{\boldsymbol{y}} = \boldsymbol{0}$, for which $\boldsymbol{y} \in \operatorname{row}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})$. In these cases, $\operatorname{rank}(\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}) = 0$, such that the optimality constraint (5) does not put any restriction on the vector $\boldsymbol{l}$. We will consider these *lower-order solutions* in more detail in future work.

III. A Multiparameter Eigenvalue Problem

In this section, we leverage Theorems 1-2 to compose a multiparameter eigenvalue problem, the eigentuples of which contain the global minimizer(s) of the realization problem.

Remark: Throughout the rest of this paper, we will assume that the vector $\boldsymbol{e} \in \mathbb{R}^{n+1}$ in the non-triviality constraint in (2) is equal to $[0, \dots, 0, 1]^{\mathsf{T}}$, implying that $a_0 = 1$. This choice simplifies the derivations, as substituting a_0 in the matrices $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}$ and $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}}$ suffices to ensure that the constraint is met, and is favorable from a computational point of view, as it eliminates one decision variable from the optimization problem (2). Nevertheless, the proposed methodology remains similar when other non-triviality constraints are used, e.g., $\boldsymbol{a}^{\mathsf{T}}\boldsymbol{a} = 1$.

[4]The orthogonal subspaces defined by the Toeplitz matrix $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}$ are reminiscent of the operator-theoretic result of Beurling–Lax–Halmos [14]–[16], which describes how each function in the Hardy space $\mathcal{H}_2$ induces an orthogonal decomposition of that space [12, Chapter 3]. Indeed, for a square summable sequence $\boldsymbol{y} \in \ell_2$ (take $N \to \infty$), the row space of the doubly-infinite matrix $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}} \in \mathbb{R}^{\infty\times\infty}$ constitutes an infinite-dimensional forward shift-invariant subspace. Each row is a forward-shift of the vector $\boldsymbol{a}_{\boldsymbol{\infty}} = [a_n, \dots, a_0, 0, 0, \dots]^{\mathsf{T}} \in \mathbb{R}^{\infty}$, corresponding to the Taylor coefficients of the functions $\{z^{n+k}a(z^{-1})\}_{k=0,1,2,\dots}$. The n-dimensional orthogonal complementary subspace, the so-called *model space*, is backward shift-invariant and corresponds to the infinite-length observability matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{\infty\times n}$ of the autonomous model defined by $a(z)$.

[5]In the case $n < N \leq 2n$, (5) can only be satisfied if $\boldsymbol{l} = \boldsymbol{0}$, which implies by (4) that $\widetilde{\boldsymbol{y}} = \boldsymbol{0}$. Indeed, for these values of N, one can always find model parameters $\boldsymbol{a}$ such that $\boldsymbol{y}$ lies in the behavior of the model: $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y} = \boldsymbol{0}$. For $\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}$ to have a non-trivial left null space, we need $N > 2n$. Remark that $N > n$ suffices for Theorem 1 since it does not rely on the optimality of the parameters $\boldsymbol{a}$, i.e., it does not use the relation in (5).

[6]For $\boldsymbol{y} \in \ell_2$ $(N \to \infty)$, the realization problem (2) becomes equivalent to the SISO H_2-norm model reduction problem [18]: take $\boldsymbol{y}$ the impulse response of a stable mth order SISO model, then, for given $n < m$, the realization problem finds the least squares optimal $\widehat{\boldsymbol{y}}$ for which the Hankel matrix $\widehat{\boldsymbol{Y}} \in \mathbb{R}^{\infty\times(n+1)}$ is rank-deficient, which, by Kronecker's Theorem [19], implies that $\widehat{\boldsymbol{y}}$ is the impulse response of an nth order SISO model.

[7]Remark that even though these *lower-order stationary points* will generally not satisfy Theorem 2, nothing forbids them to 'coincidentally' do so. One can easily construct such an example: given a set of model parameters $\boldsymbol{a}$, construct the data $\boldsymbol{y}$ as $\boldsymbol{y} = (\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})^{\mathsf{T}}\boldsymbol{f}$ for some $\boldsymbol{f} \in \mathbb{R}^{2N-n}$. Then, for this sequence $\boldsymbol{y}$, $\boldsymbol{a}$ is a global maximizer of (2), which nevertheless satisfies Theorem 2.

Start from (9) and use (11) to derive the following cubic n-parameter eigenvalue problem in the parameters $a_1, \ldots, a_n$,

$$\underbrace{\begin{bmatrix} \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y} & \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})^{\mathsf{T}} \end{bmatrix}}_{\mathcal{M}(\boldsymbol{a})} \begin{bmatrix} -1 \\ \boldsymbol{g} \end{bmatrix} = \boldsymbol{0}. \quad (12)$$

The matrix to the left, $\mathcal{M}(\boldsymbol{a}) = \sum_{\{\boldsymbol{\alpha}\}} \boldsymbol{M}_{\boldsymbol{\alpha}}\boldsymbol{a}^{\boldsymbol{\alpha}}$, is a matrix polynomial in the monomials $\boldsymbol{a}^{\boldsymbol{\alpha}} = a_1^{\alpha_1} \ldots a_n^{\alpha_n}$, with coefficient matrices $\boldsymbol{M}_{\boldsymbol{\alpha}} \in \mathbb{R}^{(N-n)\times(N-2n+1)}$, the size of which indicates that (12) is overdetermined when $n > 1$. The values $\boldsymbol{a} \in \mathbb{C}^n$ for which $\mathcal{M}(\boldsymbol{a})$ becomes rank-deficient, such that there exists a vector $\boldsymbol{g} \in \mathbb{C}^{N-2n}$ for which these equations are satisfied, are the *affine eigentuples* of this MEP [9].

Theorem 3. *For a given model order n and non-model-compliant data $\boldsymbol{y} \in \mathbb{R}^N$, such that $\operatorname{rank}(\boldsymbol{Y}_{\boldsymbol{N-n}}) = n + 1$, with $N > 2n$, it holds that,*

1) *each stationary point $\boldsymbol{a} \in \mathbb{R}^n$ of (2), for which $\operatorname{rank}(\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}) = n$ (i.e., $\widehat{\boldsymbol{y}}$ has nth order LTI dynamics), is an affine eigentuple of the MEP (12), and,*
2) *each real-valued affine eigentuple $\boldsymbol{a}$ of the MEP constitutes a stationary point of the realization problem (2),*

such that the set of real-valued affine eigentuples $\boldsymbol{a}$ of the cubic n-parameter eigenvalue problem (12) is guaranteed to contain the global minimizer(s) of the realization problem (2).

Proof. By the combination of Theorems 1-2, we know that for a stationary point $\boldsymbol{a}$ of (2) for which $\operatorname{rank}(\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}) = n$, there must exist a vector $\boldsymbol{g} \in \mathbb{R}^{N-2n}$ such that,

$$(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}(\boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{-1}\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y} = (\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})^{\mathsf{T}}\boldsymbol{g}. \quad (13)$$

Since $(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}$ has full column rank, (13) is equivalent to,

$$\begin{aligned} &\Longleftrightarrow (\boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{-1}\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y} - (\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})^{\mathsf{T}}\boldsymbol{g} = \boldsymbol{0}, \\ &\Longleftrightarrow \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y} - \boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})^{\mathsf{T}}\boldsymbol{g} = \boldsymbol{0}, \quad (14) \end{aligned}$$

for which 'separating' out the variables in $\boldsymbol{g}$ into the eigenvector gives (12). This proves the first claim. Secondly, because there is a one-to-one[8] correspondence between the affine eigentuples $\boldsymbol{a}$ of the MEP (12) and the affine common roots $(\boldsymbol{a}, \boldsymbol{g})$ of (14), it suffices to show that each tuple $(\boldsymbol{a}, \boldsymbol{g})$ which satisfies (14), also satisfies the FONC (4)-(6). Substitution of $\boldsymbol{l}{=}(\boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{-1}\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y}$, and $\widetilde{\boldsymbol{y}}{=}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})^{\mathsf{T}}\boldsymbol{g}$ in (13)-(14) gives the required result. Lastly, one can show that the global minimizer(s) must have $\operatorname{rank}(\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}) = n$. Combined with the above-mentioned claims, this concludes the proof. ☐

In the case n=1, the MEP in (12) becomes a polynomial eigenvalue problem (PEP) in the variable a_1 with square coefficient matrices. As such, all its eigenvalues could be obtained from its *secular equation*, $\det(\mathcal{M}(\boldsymbol{a})) = 0$, which leads to a univariate polynomial of degree $3N - 5$ in the variable a_1. For $n = 2$, the matrix polynomial $\mathcal{M}(a_1, a_2)$ has dimensions $(N-2)\times(N-3)$, such that its eigentuples can be computed as the common roots of the system of polynomial equations obtained by equating all $(N-3)\times(N-3)$ minors[9] of that matrix to zero. This reformulation, which eliminates the $N-2n$ 'linear' variables $\boldsymbol{g}$ at the cost of higher polynomial degrees, is possible for arbitrary problem sizes (N, n). However, the number of minors grows quickly with (N, n).

Numerical algorithms to find all the (real-valued) affine[10] eigentuples of the MEP (12) are available, e.g., the (block) Macaulay framework described in [10] or the methods from [20]. Alternatively, the system of multivariate polynomial equations obtained in (14) can be solved via off-the-shelf polynomial root-finding techniques, e.g., [21]. Then, the globally optimal solution(s) of the realization problem (2) can be selected by evaluating the objective function for each obtained stationary point. Note that because (multiparameter) eigenvalue-solvers and polynomial root-finding techniques generally work over the field of complex numbers, the complex-valued eigentuples or common roots have to be pruned away: they have no meaningful interpretation in the context of the realization problem (2).

Example 1. *Consider the numerical example ($N = 4$) described in [3, Section 8.2], where the globally optimal first-order ($n = 1$) autonomous LTI realization is computed for the sequence of given output data $\boldsymbol{y} = [4, 3, 2, 1]^{\mathsf{T}}$. The system of quartic polynomial equations in the variables $\{a_1, g_1, g_2\}$ described in (14) corresponds to,*

$$\begin{cases} 0 = 4a_1 - 2a_1g_1 - a_1^2g_2 - a_1^3g_1 + 3, \\ 0 = 3a_1 - g_1 - 2a_1g_2 - 2a_1^2g_1 - a_1^3g_2 + 2, \\ 0 = 2a_1 - g_2 - a_1g_1 - 2a_1^2g_2 + 1. \end{cases} \quad (15)$$

Observe that the variables $\{g_1, g_2\}$ appear only linearly. Reformulating this system of multivariate polynomial equations from (15) gives the following PEP in the parameter a_1,

$$\begin{aligned} \Bigg(&\underbrace{\begin{bmatrix} 3 & 0 & 0 \\ 2 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}}_{\boldsymbol{M}_0} + \underbrace{\begin{bmatrix} 4 & -2 & 0 \\ 3 & 0 & -2 \\ 2 & -1 & 0 \end{bmatrix}}_{\boldsymbol{M}_1} a_1 \\ &+ \underbrace{\begin{bmatrix} 0 & 0 & -1 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}}_{\boldsymbol{M}_2} a_1^2 + \underbrace{\begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}}_{\boldsymbol{M}_3} a_1^3 \Bigg) \begin{bmatrix} 1 \\ g_1 \\ g_2 \end{bmatrix} = \boldsymbol{0}. \end{aligned} \quad (16)$$

The secular equation of this PEP, obtained by equating the determinant of the matrix polynomial to zero, is given as,

$$2a_1^7 - 5a_1^6 + 12a_1^5 - a_1^4 + 6a_1^3 + 3a_1^2 + 3 = 0. \quad (17)$$

The author of [3] exploits Theorem 1 to derive an alternative, yet equivalent formulation of the objective function of (2), which solely relies on the model parameters $\boldsymbol{a}$,

$$\|\widetilde{\boldsymbol{y}}\|_2^2 = \boldsymbol{y}^{\mathsf{T}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}(\boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{-1}\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}\boldsymbol{y}.$$

[8]The relation $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})^{\mathsf{T}}\boldsymbol{g} = \boldsymbol{0}$ can only be satisfied for $\boldsymbol{g} = \boldsymbol{0}$, since $\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-n}})^{\mathsf{T}}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-2n}})^{\mathsf{T}}$ is of full column rank. The first element of the eigenvector corresponding to an eigentuple $\boldsymbol{a}$ of the MEP in (12) is therefore guaranteed to be non-zero, such that the eigenvector can always be appropriately normalized to retrieve $\boldsymbol{g}$.

[9]The minors are the $\binom{N-n}{n-1}$ determinants of the submatrices obtained by omitting $n-1$ rows of the polynomial matrix of the MEP in (12).

[10]Only the affine solutions are of interest in the context of (2). When working with the (block) Macaulay framework, the effects of the solutions at infinity can be eliminated via a *column compression*. See, e.g., [9], [10].

TABLE I

COMPARISON OF THE SIZES OF THE COEFFICIENT MATRICES OF $\mathcal{M}_1$, THE MEP OBTAINED FROM [3], AND $\mathcal{M}_2$, THE MEP DESCRIBED IN (12), FOR SEVERAL COMBINATIONS OF (N, n).

(N, n)	size($\mathcal{M}_1$)	size($\mathcal{M}_2$)
$(4, 1)$	7×7	3×3
$(16, 6)$	76×71	10×5
$(50, 8)$	386×379	42×35
$(200, 15)$	2975×2961	185×171

TABLE II

THE AFFINE COMMON ROOTS OF THE SQUARE SYSTEM OF MULTIVARIATE POLYNOMIAL EQUATIONS IN (15).

$\lVert\widetilde{\boldsymbol{y}}\rVert_2^2$	a_1	g_1	g_2
0.1486	-0.6764	-0.2525	-0.2734
/	$-0.1589 \mp 0.808j$	$1.3577 \pm 3.8194j$	$1.8359 \pm 3.3491j$
/	$0.4209 \pm 0.6233j$	$3.0425 \mp 2.9959j$	$-0.0785 \pm 1.2013j$
/	$1.3261 \pm 2.0058j$	$-0.2739 \mp 0.6279j$	$0.3793 \mp 0.3849j$

This expression leads to an unconstrained optimization problem over the model parameters $\boldsymbol{a}$, the FONC of which can be used to construct an MEP. However, as this approach introduces many auxiliary variables to cope with the inverse of the matrix $\boldsymbol{D}^{\boldsymbol{a}}_{\boldsymbol{N-n}}$, the coefficient matrices tend to grow very large: $((N-n)(n+1)+n) \times ((N-n)(n+1)+1)$, which is approximately $n+1$ times larger than the coefficient matrices of the proposed n-parameter eigenvalue problem from (12). This becomes especially noticeable for increasing problem sizes (N, n), see, e.g., Table I. It is not straightforward to compare the complexity of different MEPs, because the computational complexity involved with solving an MEP depends on the interplay of multiple attributes: e.g., the highest degree of its parameters, the number of parameters, and the size of the coefficient matrices. We will investigate this in more detail in future work. Also notice that since the MEP from [3] does not exploit Theorem 2, its eigentuples comprise the entire set of stationary points, and therefore, contrary to the eigentuples of the MEP in (12), always include the stationary points for which $\operatorname{rank}(\widehat{\boldsymbol{Y}}_{\boldsymbol{N-n}}) < n$.

IV. NUMERICAL EXAMPLES

In this section, we consider[11] several numerical 'toy examples' to illustrate the results obtained in Theorems 2-3.

Example 1 (continued)**.** *The 7 affine common roots of the system of polynomial equations in* (15) *are depicted in Table II. Computing the affine eigenvalues* a_1 *of the PEP from* (16) *or computing the roots of the univariate polynomial in* (17) *gives equivalent results. The set of real-valued eigenvalues corresponds to a singleton:* $a_1 = -0.6764$*, which is the real global minimizer of the realization problem. Alternatively, when the approach from [3] is used, 10 affine eigenvalues* a_1 *are retrieved, two of which real-valued:* $\{-0.6764, 1.6506\}$*. The latter can be shown to be a global maximizer, i.e.,* $\boldsymbol{y} \in \operatorname{row}(\boldsymbol{T}^{\boldsymbol{a}}_{\boldsymbol{N-1}}|_{a_1=1.6506})$.

[11] We used a MacBook Pro with a 6-core Intel i7 CPU (2019) working at 2.6 GHz with access to 32 GB RAM. Numerical values are rounded for displaying purposes and timings are averaged over 5 consecutive runs. We used a MATLAB implementation of the (block) Macaulay method [10, Chapter 6], available online at www.macaulaylab.net, to compute the eigentuples of the MEPs.

TABLE III

THE OBJECTIVE FUNCTION VALUE AND POLES p_i FOR THE STATIONARY POINTS (a_1, a_2) OF EXAMPLE 2 OBTAINED USING THE MEP FROM (12).

$\lVert\widetilde{\boldsymbol{y}}\rVert_2^2$	a_1	a_2	p_1	p_2
0.1327	-1.6255	0.7167	$0.8127 + 0.2369j$	$0.8127 - 0.2369j$
0.1514	-0.0752	-0.5850	0.8033	-0.7282
0.1606	-14.076	10.433	13.291	0.7849
0.5386	-0.7127	1.8381	$0.3564 + 1.3081j$	$0.3564 - 1.3081j$
0.5398	-1.9563	3.3657	$0.9782 + 1.5520j$	$0.9782 - 1.5520j$
0.5405	29.254	43.807	-27.671	-1.5831
0.5425	1.3477	1.4908	$-0.6738 + 1.0182j$	$-0.6738 - 1.0182j$
0.5484	1.9764	1.4075	$-0.9882 + 0.6564j$	$-0.9882 - 0.6564j$
0.5492	-1.3053	1.0564	$0.6527 + 0.7940j$	$0.6527 - 0.7940j$

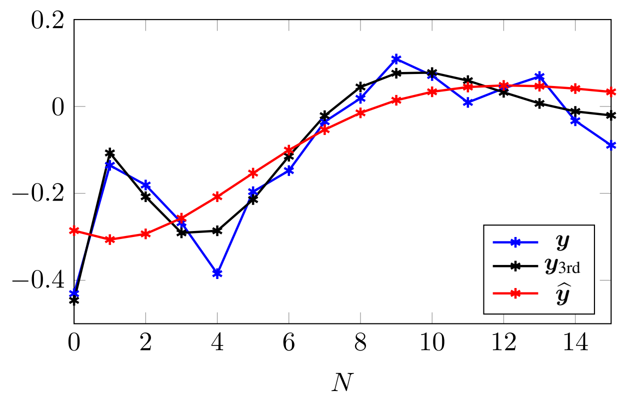

Fig. 2. The original data $\boldsymbol{y}_{3\text{rd}}$, the data $\boldsymbol{y}$ and the model-compliant data $\widehat{\boldsymbol{y}}$ corresponding to the globally optimal solution of Example 2.

Example 2. *In this example we fit a second-order model* ($n = 2$) *to given data* $\boldsymbol{y}$ ($N = 16$)*, where* $\boldsymbol{y}$ *corresponds to the output signal of a third-order autonomous LTI model with poles* $(0.2, 0.7 \pm 0.4j)$*, perturbed using MATLAB's* `randn()` *function (the considered instance of* $\boldsymbol{y}$ *has* $\lVert\boldsymbol{y}\rVert_2^2 = 0.5509$*):*

$$\boldsymbol{y} = \boldsymbol{y}_{3rd} + 0.05 * \texttt{randn(N, 1)}.$$

The MEP (12) *has 739 affine eigentuples, 9 of which are real-valued (see Table III). The globally optimal solution* $\boldsymbol{a} = [0.7167, -1.6255]^{\mathsf{T}}$ *has an objective function value approximately equal to* 0.1327*. The residuals*[12] *of the obtained eigentuples are of the order of magnitude* $\mathcal{O}(10^{-10})$*. The obtained globally optimal data* $\widehat{\boldsymbol{y}}$ *are depicted in Fig. 2.*

Example 3. *We fit models for* $n \in \{1, 2, 3, 4, 5\}$ *to a sequence of given data* $\boldsymbol{y}$ ($N = 10$) *that is generated by a fifth-order autonomous LTI model with poles* $(0.5, 0.25 \pm 0.75j, -0.3 \pm 0.5j)$*. The objective function value of the globally optimal model, the required computation time and the number of affine eigentuples of the MEP are depicted in Table IV. For* $n = 5$*, the MEP has one real-valued eigentuple which corresponds to the model that was used to generate* $\boldsymbol{y}$*.*

[12] We calculate the residual error by substituting the computed eigentuple and eigenvector in the MEP and determining the norm of the residual vector.

TABLE IV
THE OBJECTIVE FUNCTION VALUE, THE REQUIRED COMPUTATION TIME t AND THE NUMBER OF AFFINE EIGENTUPLES n_a CORRESPONDING TO THE DIFFERENT VALUES OF n IN EXAMPLE 3.

n	$\|\widetilde{\boldsymbol{y}}\|_2^2$	t	n_a
1	8.1211	$0.0245\,\mathrm{s}$	5
2	3.7429	$0.9877\,\mathrm{s}$	10
3	1.6775	$9.0798\,\mathrm{s}$	17
4	6.064×10^{-4}	$15.185\,\mathrm{s}$	13
5	$\mathcal{O}(10^{-31})$	$0.0240\,\mathrm{s}$	1

Notice that the computation time increases with n, whereas the number of affine eigentuples n_a shrinks after $n = 3$. The globally optimal model-compliant data $\widehat{\boldsymbol{y}}_{\boldsymbol{n}}$ are depicted in Fig. 3.

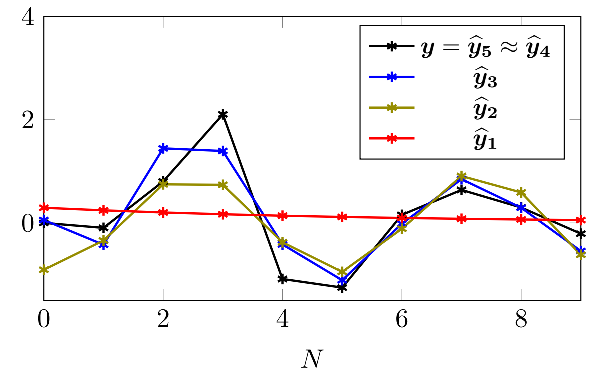

Fig. 3. The given data $\boldsymbol{y}$ and the globally optimal model-compliant data $\widehat{\boldsymbol{y}}$ corresponding to the globally optimal solution(s) of Example 3.

V. Conclusions and Future Work

We showed, based on the first-order necessary conditions for optimality of the least squares realization problem, that the optimal misfits can be characterized via a 'double' FIR filter, which is reminiscent to 'Walsh's Theorem' [12, Theorem 3.14]. We exploited this result to compose a novel multiparameter eigenvalue problem (MEP), the eigentuples of which contain the parameters of the globally optimal model(s). We illustrated our findings using several numerical 'toy examples', and performed a comparison with the alternative globally optimal approach in the literature [3].

Since the computational difficulty of solving the obtained MEP grows exponentially with the problem size (N, n), more research is needed to make our theoretical findings applicable in practice. In future work, we will try to exploit the fact that we are only interested in the real-valued eigentuples of the MEP. The objective function of the realization problem can be shown to admit a purely polynomial form in the variables $(\boldsymbol{a}, \boldsymbol{g})$. In future research we will investigate whether incorporating this objective function into the root-finding/MEP solvers allows to compute the global minimizer(s) only. The Riemannian SVD [22], which can be derived from the FONC (4)-(6) by eliminating $\widehat{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{y}}$, might assist us in this challenge. We also want to perform more numerical experiments to get better insights into the nature of the local/global minimizer(s) of the realization problem, and to investigate the implications of the technical note in Section III. Another challenge involves pushing the problem size (N and/or n) to be as large as possible, e.g., using a supercomputer.

References

[1] J. C. Willems, "From time series to linear system–Part III. Approximate modelling," *Automatica*, vol. 23, no. 1, pp. 87–115, 1987.

[2] C. Heij, *Deterministic Identification of Dynamical Systems*, ser. Lecture Notes in Control and Information Sciences. Berlin, Germany: Springer, 1989, vol. 127.

[3] B. De Moor, "Least squares optimal realisation of autonomous LTI systems is an eigenvalue problem," *Communications in Information and Systems*, vol. 20, no. 2, pp. 163–207, 2020.

[4] L. Ljung, *System Identification: Theory for the User*, 2nd ed., ser. Prentice-Hall Information and System Sciences Series. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[5] T. Söderström, *Errors-in-variables Methods in System Identification*, ser. Communications and Control Engineering (CCE). Springer Cham, 2018.

[6] J. Cadzow, "Singal enhancement: A composite property mapping algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 49–62, 1988.

[7] M. Elad, P. Milanfar, and G. H. Golub, "Shape from moments—an estimation theory perspective," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1814–1829, 2004.

[8] G. H. Golub and V. Pereyra, "Separable nonlinear least squares: The variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. R1–R26, 2003.

[9] 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.

[10] C. Vermeersch, "The (block) Macaulay matrix: Solving systems of multivariate polynomial equations and multiparameter eigenvalue problems," Ph.D. dissertation, Faculty of Engineering, KU Leuven, Leuven, Belgium, 2023.

[11] M. E. Hochstenbach, T. Košir, and B. Plestenjak, "Numerical methods for rectangular multiparameter eigenvalue problems, with applications to finding optimal ARMA and LTI models," *Numerical Linear Algebra with Applications*, vol. 31, no. 2, 2024.

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

[13] S. Lagauw, O. M. Agudelo, and B. De Moor, "Globally optimal SISO H_2-norm model reduction using Walsh's Theorem," *IEEE Control Systems Letters*, vol. 7, pp. 1670–1675, 2023.

[14] A. Beurling, "On two problems concerning linear transformations in Hilbert space," *Acta Mathematica*, vol. 81, pp. 239–255, 1949.

[15] P. D. Lax, "Translation invariant spaces," *Acta Mathematica*, vol. 101, pp. 163–178, 1959.

[16] P. R. Halmos, "Shifts on Hilbert spaces," *Journal für die reine und angewandte Mathematik*, vol. 208, pp. 102–112, 1960.

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

[18] 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.

[19] L. Kronecker, "Zur Theorie der Elimination einer Variablen aus zwei algebraischen Gleichungen," *Monatsberichte der Königlich Preussischen Akademie der Wissenschaften (Berlin)*, pp. 535–600, 1881, reprinted as pp. 113–192 in Mathematische Werke, vol. 2, B.G. Teubner, Leipzig, 1897.

[20] B. Plestenjak, "MultiParEig," MATLAB Central File Exchange (https://www.mathworks.com/matlabcentral/fileexchange/47844-multipareig), 2023, retrieved Jan. 24, 2023.

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

[22] B. De Moor, "Total least squares for affinely structured matrices and the noisy realization problem," *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 3104–3113, 1994.