KU LEUVEN

STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Lukas Vanpoucke, Sibren Lagauw Bart De Moor Title: Globally optimal least squares misfit identification of multidimensional, autonomous, linear, shift-invariant systems Journal: Vol: Pages:
Archived version	■ To be submitted ☐ 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	
Copyright statement	
Author contact	lukas.vanpoucke@esat.kuleuven.be



Globally optimal least squares misfit identification of multidimensional, autonomous, linear, shift-invariant systems

Lukas Vanpoucke [1], Sibren Lagauw, Bart De Moor [a]

[a] *Center for Dynamical Systems, Signal Processing, and Data Analytics (STADIUS), Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium.*

Abstract

The problem of least squares misfit identification of autonomous, scalar, linear, shift-invariant, multidimensional systems from data is considered. The local minima of this optimization problem are characterized as common roots of two different sets of multivariate polynomial equations, one for the image- and one for the kernel-based representation of the model-compliant sequence. Moreover, it is shown that lower-order models cannot be locally optimal, allowing their polynomial characterization to be simplified. It is demonstrated that these optimal models constitute isolated solutions, such that all minima can be found through polynomial root finding techniques, resulting in a deterministic methodology to obtain the global minimum. The presence of an infinite number of lower-order solutions to these polynomial characterizations is shown as well. Finally, the effectiveness of the proposed methodology is illustrated with numerical examples.

Key words: Multidimensional linear systems, Least-squares realization, Global optimization

1 Introduction

The relative simplicity of linear dynamical models, combined with the insight that such models, in many practical cases, provide a sufficiently accurate approximation of the true system dynamics, caused them to become popular modeling tools. The current work focuses on multidimensional (mD) generalizations of the well-known linear time-invariant (LTI) models. More precisely, we focus on autonomous models. Multiple such *linear, shift-invariant model* classes have been proposed in the literature, for example, the state-space models in [2,52,19,7,3,32,34], some more general than others. On the other hand, the behavioral framework for systems theory focuses on the study of models governed by mD difference equations [20,43,51,65,68]. The latter point of view is particularly relevant to this paper.

As it is not always feasible to obtain such a model through first-principles, a common approach is to fit them to given input-output data, especially for LTI models. Moreover, in the one-dimensional case, it has been shown for a variety of modeling approaches that such a fitting problem can be solved exactly when using a least-squares distance measure, that is, there exists a deterministic algorithm guaranteed to find the globally optimal model w.r.t. a given data sequence. Such approaches rely on characterizing stationary points of the underlying optimization problem as common roots of a set of polynomial equations, which can all be found through well-established computational tools [63,4,12]. The current work extends this methodology to the least squares misfit identification of autonomous, scalar, linear, shift-invariant models, extending insights from the 1D autonomous case [36,12] to the mD setting. Note that the contribution is of a theoretical nature, providing insight on the properties of the underlying optimization problem, and, more precisely, its stationary points.

This approach stands in stark contrast to the identification methods for mD systems present in engineering practice, which all involve some form of heuristics. They either rely on local optimization algorithms—which can at best guarantee that a local optimum is found and the results of which depends strongly on the chosen initialization—or on statistical assumptions that are of-

Email addresses: lukas.vanpoucke@esat.kuleuven.be (Lukas Vanpoucke), sibren.lagauw@esat.kuleuven.be (Sibren Lagauw), bart.demoor@esat.kuleuven.be (Bart De Moor).

[1] Corresponding author.

ten unverifiable in practice. Such approaches focus on providing a sufficiently good approximation for engineering applications, rather than completely solving the underlying optimization problem.

Among these 'heuristic' multidimensional identification methods there exist different branches. Most notably, subspace identification methods, which, based on statistical assumptions, construct a state-space model, through efficient linear algebra operations, [48,47,21,30,49]. Optimization-based methods to estimate the model parameters have also been considered in [18,17,9,25,58], extending the well-established prediction-error-methods for LTI models. Another common approach for modeling mD systems 'circumvents' the multidimensionality through distributed parameter models, that is, the models are described as a set of linked 1D models or a single, larger 1D model, to subsequently identified using 1D techniques, see e.g. [21,9,25].

These techniques have been successful in generating useful models in many signal processing applications [7], for example, image processing [38,48] and texture analysis [14] among others, as well as for multidimensional control techniques [53,13,66], and for the parameter estimation of PDE's [17,45]. For the autonomous models considered here, the problem of multidimensional harmonic retrieval is particularly relevant [59,60,54,16,55,31], the techniques of which have been applied in nuclear magnetic resonance spectroscopy [40,22,67], the analysis of wireless communications channels [24,42], and direction of arrival problems [23].

Because a different problem is tackled altogether, it should be emphasized that the current work does not aim to be computationally competitive with such 'heuristic' mD system identification methods. Rather, we strive for theoretical insight into such problems, which may serve to improve these methods and provide a means to benchmark them.

1.1 Problem formulation

We consider the problem of the least-squares misfit identification of autonomous, scalar, linear, shift-invariant models from a given, finite multi-indexed data sequence. An nth order model in this class allows state-space representation of the form [20,26,15]:

$$\begin{aligned} \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_1} &= \boldsymbol{A}_1 \boldsymbol{x}_{\boldsymbol{\kappa}} \\ \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_2} &= \boldsymbol{A}_2 \boldsymbol{x}_{\boldsymbol{\kappa}} \\ &\vdots \\ \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_m} &= \boldsymbol{A}_m \boldsymbol{x}_{\boldsymbol{\kappa}} \\ \hat{y}_{\boldsymbol{\kappa}} &= \boldsymbol{c}\boldsymbol{x}_{\boldsymbol{\kappa}} \\ \text{where} \quad \boldsymbol{A}_k \boldsymbol{A}_l &= \boldsymbol{A}_l \boldsymbol{A}_k \;\; \forall k,l \in \{1,\ldots,m\}, \end{aligned} \tag{1}$$

and $\boldsymbol{A}_k \in \mathbb{C}^{n\times n}$, $k \in \{1,\ldots,m\}$, $\boldsymbol{c} \in \mathbb{C}^{1\times n}$, $\boldsymbol{x_0} \in \mathbb{C}^n$ and $\boldsymbol{e}_k$ denotes the kth column of the identity matrix. For an initial state $\boldsymbol{x} = \boldsymbol{x_0}$, the output of such a system can thus be written as $\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}\boldsymbol{x_0}$, with [2] $\boldsymbol{A}^{\boldsymbol{\kappa}} = \prod_{k=1}^{m} \boldsymbol{A}_k^{\kappa_k}$. Furthermore, the output trajectory remains unaltered under a common similarity transform of the form $\boldsymbol{A}_k^{\text{new}} = \boldsymbol{P}^{-1}\boldsymbol{A}_k\boldsymbol{P}, \boldsymbol{c}^{\text{new}} = \boldsymbol{c}\boldsymbol{P}, \boldsymbol{x}_{\boldsymbol{\kappa}}^{\text{new}} = \boldsymbol{P}^{-1}\boldsymbol{x}_{\boldsymbol{\kappa}}$, where $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ is an invertible matrix. Also note that $\boldsymbol{c}$ and $\boldsymbol{x}$ can be scaled interchangeably, so that one can choose $\boldsymbol{c} = \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix}$ without loss of generality.

The commuting set of state-transition matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ can be brought into an upper triangular form through a common unitary similarity transformation [28]. This triangular format allows the eigenvalues of $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ on the diagonal entries to be grouped into n m-tuples, which we refer to as eigentuples $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n \in \mathbb{C}^m$. For notational convenience, the eigentuples are sometimes referred to in matrix format, to this end, let $\begin{bmatrix} \boldsymbol{\lambda}_1 & \ldots & \boldsymbol{\lambda}_n \end{bmatrix}^{\mathsf{T}} = \boldsymbol{\Lambda} \in \mathbb{C}^{n\times m}$.

By lack of a more general parameterization, we confine ourselves to fitting models with distinct eigentuples of multiplicity one. This assumption implies that the state transition matrices can be chosen as the diagonal matrices $(\boldsymbol{A}_k)_{i,i} = (\boldsymbol{\lambda}_i)_k$ [28]. Models with eigentuples at infinity are also excluded by the model class (1) [15].

As will be detailed in Section 2, these models also adhere to a set of difference equations with constant coefficients. These equations are essentially a set of multivariate polynomials, whose common roots correspond to the eigentuples [15]. In this work, we are interested in real models, that is, those governed by difference equations with real coefficients. For such difference equations, or rather, multivariate polynomials $\boldsymbol{r}(\boldsymbol{\lambda})$, it holds that $\overline{\boldsymbol{r}(\boldsymbol{\lambda})} = \boldsymbol{r}(\overline{\boldsymbol{\lambda}})$, such that its eigentuples and initial states arise in complex-conjugate pairs. Definition 1.1 formalizes this notion for the model parameters $(\boldsymbol{\Lambda}, \boldsymbol{x})$.

Definition 1.1 (Real models)**.** Let $\mathcal{R}$ be the set of model parameters $(\boldsymbol{\Lambda}, \boldsymbol{x})$ for which each subset of indices $\mathcal{I} \subseteq \{1,\ldots,n\}$ that satisfies $\forall i,s \in \mathcal{I} : (\boldsymbol{\lambda}_i, x_i) = (\boldsymbol{\lambda}_s, x_s)$, there exists a subset of indices $\mathcal{J} \subseteq \{1,\ldots,n\}$ such that $\forall i \in \mathcal{I}, \exists j \in \mathcal{J} : (\boldsymbol{\lambda}_j, x_j) = \overline{(\boldsymbol{\lambda}_i, x_i)}$, and $\#\mathcal{I} = \#\mathcal{J}$. $(\boldsymbol{\Lambda}, \boldsymbol{x})$ is said to define *a real model* if $(\boldsymbol{\Lambda}, \boldsymbol{x}) \in \mathcal{R}$.

Given a finite multidimensional data sequence [3] $y_{\boldsymbol{\kappa}} : \mathcal{K} \to \mathbb{R} : \boldsymbol{\kappa} \mapsto y_{\boldsymbol{\kappa}}$ over some finite set of grid points $\mathcal{K} \subseteq \mathbb{N}_0^m$, where $\#\mathcal{K} = N$, and a predefined model order n, the goal of misfit identification is to find model parameters $(\boldsymbol{\Lambda}, \boldsymbol{x})$ associated with the model-compliant signal $\hat{y}_{\boldsymbol{\kappa}}$ that lies nearest to $y_{\boldsymbol{\kappa}}$. Using a least-squares

[2] *Notation:* A similar multi-exponent notation will be used throughout this work: $\boldsymbol{z}^{\boldsymbol{\kappa}} = \prod_{k=1}^{m} z_k^{\kappa_k}$, where $\boldsymbol{x} \in \mathbb{C}^m$

[3] $\mathbb{N}_0$ denotes the set of natural numbers including zero.

measure for distance, the misfit identification problem can be formulated as [39]:

$$\begin{aligned} \min_{\boldsymbol{\Lambda},\boldsymbol{x}} \ & \|\tilde{\boldsymbol{y}}\|_2^2 \qquad (2) \\ \text{subject to: } \forall \boldsymbol{\kappa} \in \mathcal{K} : \hat{y}_{\boldsymbol{\kappa}} &= \prod_{k=1}^{m} \boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}\boldsymbol{x} \\ \hat{\boldsymbol{y}} &= \boldsymbol{y} - \tilde{\boldsymbol{y}} \\ (\boldsymbol{\Lambda}, \boldsymbol{x}) &\in \mathcal{R}, \end{aligned}$$

where $\boldsymbol{y} = \begin{bmatrix} y_{\boldsymbol{\kappa}_1} \dots y_{\boldsymbol{\kappa}_N} \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^N$ is the sequence of given data points and $\tilde{\boldsymbol{y}}, \hat{\boldsymbol{y}} \in \mathbb{R}^N$ are defined similarly. Here, $\tilde{\boldsymbol{y}}$ is the sequence of residuals at each grid point, also referred to as the *misfit*.

The particular ordering [4] of these sequences is of no importance for this problem. Rather, to refer to specific elements of these vectors we employ the multi-index $\boldsymbol{\kappa}$ directly instead of a sequential index over $\mathcal{K}$.

1.2 Contributions and paper outline

For the considered model class, (2) amounts to a non-convex optimization problem, such that there can be many local minima. This complicates the search for the strict solution of this problem, that is, its global minimum. The approach presented in this work characterizes the local minima of this optimization problem as common roots of a set of multivariate polynomial equations. All solutions to these polynomial problems can then be computed using numerical polynomial solvers. By extension, all local minima can thus be obtained, including the global minimum. This methodology provides a deterministic means of solving the misfit problem (2).

A brief overview of the main sections of the paper and the contributions therein is provided in Figure 1, which we now briefly elaborate on. First, Section 2 introduces the necessary preliminaries regarding the considered model class and its representations.

Next, a relaxed version of the problem is considered in Section 3. More precisely, the constraint that the model parameters should define a real model is disregarded, as it is hard to impose analytically. Through manipulation of the first-order necessary conditions for optimality (FONC) that follow from this, the aforementioned novel polynomial characterization for the local minima of (2) is obtained. A formal proof that these indeed include all local minima of the non-relaxed problem (2) is deferred to Section 4. Finally, by exploiting duality between the kernel and image representation of a model-compliant

[4] In this text we prefer the use of graded inverse lexicographic ordering when needed. For example, in 2D, this amounts to $(0,0), (1,0), (0,1), (2,0), \dots$ [10].

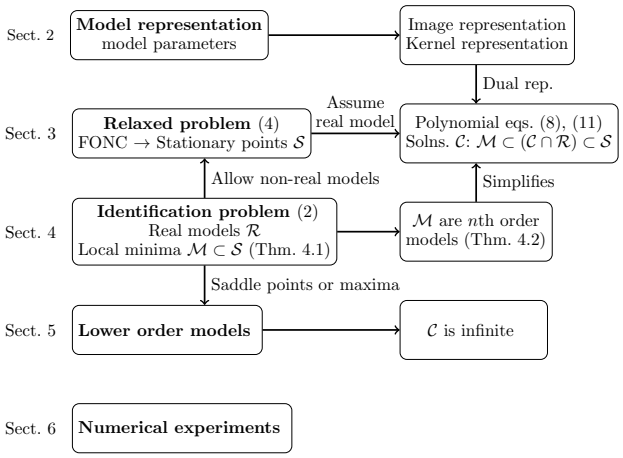


Fig. 1. Overview of the main sections of this paper and the contributions therein.

sequence, an alternative polynomial characterization is derived.

Section 4 then formally relates the local minima of the original problem (2) to the stationary points of the relaxed problem. Moreover, it is shown that these local minima cannot be described using a lower model order, providing theoretical insight into the system theoretic properties of the minima akin to the LTI case [50,12]. This property allows us to eliminate part of the non-optimal stationary points from the solution sets of the polynomial characterizations, similar to the approaches in [36,35]. Lastly, this section shows that these local minima are almost always a finite set of isolated points, allowing them to be found using off-the-shelf numerical polynomial root finding techniques.

In Section 5, the non-optimal, *lower-order* stationary points are investigated. It is illustrated that these arise as infinite sets of equivalent models leading to a positive dimensional solution component in the polynomial characterizations. Finally, additional numerical experiments are presented in Section 6, after which a conclusion is drawn in Section 7.

2 Autonomous, shift-invariant, linear models

In this section, we briefly explore the properties of the considered model class (1). More specifically, an alternative difference equations-based representation of the model is considered, and it is explored how the model compliant-signals can be represented over a finite set of grid points using observability and Macaulay matrices.

2.1 Observability matrix

By the linearity of the model class (1), the set of output signals generated by such a model forms a finite-dimensional vector space over $\mathbb{C}$ of dimension at most n,

referred to as the *behavior* of the model (1) [65,43,20]. When considering signals over a finite subset of grid points $\mathcal{K} \subset \mathbb{N}_0^m$, this vector space of model-compliant signals is the image of an *observability matrix* $\boldsymbol{\Gamma} \in \mathbb{R}^{N \times n}$, defined as [5] $\boldsymbol{\Gamma}_{\boldsymbol{\kappa},:} = \boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}, \boldsymbol{\kappa} \in \mathcal{K}$. For our choice of $\boldsymbol{c}$ and diagonal $\boldsymbol{A}_i$, this can be written out in terms of the model parameters, i.e. the eigentuples $\boldsymbol{\lambda}_i$, as follows:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}_1} & \dots & \boldsymbol{\lambda}_n^{\boldsymbol{\kappa}_1} \\ \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}_2} & \dots & \boldsymbol{\lambda}_n^{\boldsymbol{\kappa}_2} \\ \vdots & & \vdots \\ \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}_N} & \dots & \boldsymbol{\lambda}_n^{\boldsymbol{\kappa}_N} \end{bmatrix}.$$

The image of this matrix then describes all model-compliant sequences over $\mathcal{K}$, that is, $\hat{\boldsymbol{y}} = \boldsymbol{\Gamma}\boldsymbol{x}$. As such this will be referred to as the *image representation*.

2.2 Difference equations

Linear, shift-invariant models can also be represented using difference equations with constant coefficients. These multidimensional difference equations can be succinctly written out using *shift operators* $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_m)$, which act upon a signal as follows: $\sigma_k y_{\boldsymbol{\kappa}} = y_{\boldsymbol{\kappa}+\boldsymbol{e}_k}$. A model can then be characterized by a set of difference equations, represented by a set of multivariate polynomials in these shift operators $\boldsymbol{r}(\boldsymbol{\sigma}) \in \mathbb{C}^p[\boldsymbol{\sigma}]$, in the sense that a signal $\hat{y}_{\boldsymbol{\kappa}}$ is model-complaint if and only if it satisfies these equations everywhere, i.e. $\boldsymbol{r}(\boldsymbol{\sigma})\hat{y} = 0, \forall \boldsymbol{\kappa} \in \mathbb{N}_0^m$ [20,43,51,68]. The vector space of model-compliant signals then corresponds to the kernel of the difference equations $\boldsymbol{r}(\boldsymbol{\sigma})$. Because of this, it will be referred to as the *kernel representation*.

Example 2.1. Consider the following set of difference equations: $\forall \boldsymbol{\kappa} \in \mathbb{N}_0^2, \boldsymbol{r}(\boldsymbol{\sigma})\hat{y}_{\boldsymbol{\kappa}} =$

$$\begin{bmatrix} \sigma_1^2 - 3\sigma_1 + 2 \\ \sigma_1 - \sigma_2 + 2 \end{bmatrix} \hat{y}_{\boldsymbol{\kappa}} = \begin{bmatrix} \hat{y}_{\boldsymbol{\kappa}+(2,0)} - 3\hat{y}_{\boldsymbol{\kappa}+(1,0)} + 2\hat{y}_{\boldsymbol{\kappa}} \\ \hat{y}_{\boldsymbol{\kappa}+(1,0)} - \hat{y}_{\boldsymbol{\kappa}+(0,1)} + 2\hat{y}_{\boldsymbol{\kappa}} \end{bmatrix} = \boldsymbol{0}.$$

Any signal $y_{\boldsymbol{\kappa}}$ satisfying this set of equations is then a valid output of the model characterized by the difference equations $\boldsymbol{r}(\boldsymbol{\sigma}) \in \mathbb{C}^2[\boldsymbol{\sigma}]$, which are essentially polynomials in $\boldsymbol{\sigma}$. The polynomial nature of these equations can be leveraged to construct solutions in the form of 2D exponentials $\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{\lambda}^{\boldsymbol{\kappa}}, \boldsymbol{\lambda} \in \mathbb{C}^2$ [29]:

$$\begin{bmatrix} \boldsymbol{\lambda}^{(2,0)} - 3\boldsymbol{\lambda}^{(1,0)} + 2 \\ \boldsymbol{\lambda}^{(1,0)} - \boldsymbol{\lambda}^{(0,1)} + 2 \end{bmatrix} \boldsymbol{\lambda}^{\boldsymbol{\kappa}} = \boldsymbol{r}(\boldsymbol{\lambda})\boldsymbol{\lambda}^{\boldsymbol{\kappa}} = \boldsymbol{0}, \forall \boldsymbol{\kappa} \in \mathbb{N}_0^2.$$

[5] The subscript ':" denotes that all rows or columns are selected.

Clearly, these equations can only be satisfied when $\boldsymbol{\lambda}$ is a solution of $\boldsymbol{r}(\boldsymbol{\lambda}) = \boldsymbol{0}$. The common roots of the difference equations $\boldsymbol{r}(\boldsymbol{\sigma})$ when viewed as a set of multivariate polynomials, $(1, 3)$ and $(2, 4)$ in this case, thus generate a set of exponential basis signals that comply with the model. It can be shown that these basis signals span the full kernel of $\boldsymbol{r}(\boldsymbol{\sigma})$, so that all model compliant signals are of the form $\hat{y}_{\boldsymbol{\kappa}} = x_1(1, 3)^{\boldsymbol{\kappa}} + x_2(2, 4)^{\boldsymbol{\kappa}}$ [15,57].

Notice that these model-compliant signals correspond precisely to those generated by a state space model (1) with $\boldsymbol{A}_1 = \text{diag}(1, 2)$, $\boldsymbol{A}_2 = \text{diag}(3, 4)$ and $\boldsymbol{c} = \begin{bmatrix} 1 & 1 \end{bmatrix}$, which thus provides an alternative description for this model. Moreover, notice that the eigentuples $\boldsymbol{\lambda}_1 = (1, 3)$ and $\boldsymbol{\lambda}_2 = (2, 4)$ of this state-space model are exactly the common roots of $\boldsymbol{r}(\boldsymbol{\lambda})$.

Example 2.1 illustrates how a basis for the signal space can be obtained through solving the difference equations $\boldsymbol{r}(\boldsymbol{\sigma})$ as if it were a set of multivariate polynomial equations. Due to this close connection with multivariate polynomials, difference equations and multivariate polynomials will often be treated interchangeably. The illustrated relation between the common roots of the difference equations and the eigentuples of an equivalent minimal state-space model also holds in general [15,26]. Because of this, the state-transition matrices always satisfy the polynomial equations [6]: $\boldsymbol{r}(\boldsymbol{A}_1, \dots, \boldsymbol{A}_m) = \boldsymbol{0}$, such that any output of (1) satisfies the difference equations: $\boldsymbol{r}(\boldsymbol{\sigma})\boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}\boldsymbol{x}_{\boldsymbol{0}} = \boldsymbol{c}\boldsymbol{r}(\boldsymbol{A}_1, \dots, \boldsymbol{A}_m)\boldsymbol{A}^{\boldsymbol{\kappa}}\boldsymbol{x}_{\boldsymbol{0}} = \boldsymbol{0}$ [15].

The number of basis signals in the kernel of $\boldsymbol{r}(\boldsymbol{\sigma})$, i.e., the model order n, then equals its total number of common roots, counting multiplicities [57,5,11]. Indeed, multiple roots result in additional basis signals of a different format: For example, it can be worked out that [7] $\boldsymbol{r}(\boldsymbol{\sigma})\kappa_k\boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k} = \boldsymbol{r}(\boldsymbol{\lambda})\kappa_i\boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k} + \frac{\partial \boldsymbol{r}(\boldsymbol{\lambda})}{\partial \lambda_k}\boldsymbol{\lambda}^{\boldsymbol{\kappa}}$, which vanishes if and only if $\forall \boldsymbol{\kappa} \in \mathbb{N}_0^m : \boldsymbol{r}(\boldsymbol{\lambda}) = \frac{\partial \boldsymbol{r}(\boldsymbol{\lambda})}{\partial \lambda_i} = 0$, so that $\boldsymbol{\lambda}$ is a common root of multiplicity at least two. Moreover, linear combinations of differential operators can also lead to multiplicities: if $\boldsymbol{r}(\boldsymbol{\lambda}) = \boldsymbol{0}$ and there exists a $\boldsymbol{v} \in \mathbb{C}^m \backslash \{\boldsymbol{0}\} : \boldsymbol{J}_{\boldsymbol{r}}(\boldsymbol{\lambda})\boldsymbol{v} = \boldsymbol{0}$, where $\boldsymbol{J}_{\boldsymbol{r}}$ is the Jacobian of $\boldsymbol{r}(\boldsymbol{\lambda})$, then $\sum_{k=1}^m v_k\kappa_k\boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k}$ satisfies $\boldsymbol{r}(\boldsymbol{\sigma})$. The vanishing of linear combinations of higher-order partial derivative operators in a common root result in even higher multiplicities [56]. When such multiple roots occur, the state-transition matrices of an equivalent state-space model are not jointly diagonalizable, violating the assumptions made in Section 1.1.

[6] For models with simple eigentuples, the $\boldsymbol{A}_k$ are jointly diagonalizable [44]. Applying this diagonalizing similarity transform then yields the desired result: $\boldsymbol{P}^{-1}\boldsymbol{r}(\boldsymbol{A}_1, \dots, \boldsymbol{A}_m)\boldsymbol{P} = \boldsymbol{r}(\boldsymbol{P}^{-1}\boldsymbol{A}_1\boldsymbol{P}, \dots, \boldsymbol{P}^{-1}\boldsymbol{A}_m\boldsymbol{P}) = \boldsymbol{0}$, since the eigentuples are common roots of $\boldsymbol{r}(\boldsymbol{\sigma})$.

[7] It suffices to look at the action of σ_k: $\sigma_k\kappa_k\boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k} = (\kappa_k + 1)\boldsymbol{\lambda}^{\boldsymbol{\kappa}} = \kappa_k\lambda_k\boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k} + \boldsymbol{\lambda}^{\boldsymbol{\kappa}}$.

2.3 Generalized Cayley-Hamilton equations

Given a set of distinct eigentuples $\mathbf{\Lambda}$, it is not so straightforward to construct a set of multidimensional difference equations that precisely describe these points. To this end, the *generalized Cayley-Hamilton equations* are introduced in Definition 2.1 and their common roots are characterized in Theorem 2.1 below.

Definition 2.1 (Generalized Cayley-Hamilton (GCH) equations, [33])**.** Given m pairwise commuting matrices $\boldsymbol{A}_k \in \mathbb{C}^{n\times n}, k = 1,\dots,m$, introduce the variables $\boldsymbol{q} \in \mathbb{C}^m$ and consider the characteristic polynomial $\chi(\alpha, \boldsymbol{q}) = \det(\sum_{k=1}^m q_i \boldsymbol{A}_i - \alpha \boldsymbol{I})$. Then, $\chi(\boldsymbol{q}^\mathsf{T}\boldsymbol{\sigma}, \boldsymbol{q})$ can be written as a polynomial in $\boldsymbol{q}$ with polynomial coefficients in $\boldsymbol{\sigma}$:

$$\chi(\boldsymbol{q}^\mathsf{T}\boldsymbol{\sigma}, \boldsymbol{q}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}_n} r_{\boldsymbol{\alpha}}(\boldsymbol{\sigma})\boldsymbol{q}^{\boldsymbol{\alpha}}, \tag{3}$$

where $\mathcal{A}_n = \{\boldsymbol{\alpha} \in \mathbb{N}_0^m \mid \|\boldsymbol{\alpha}\|_1 = n\}$. The (polynomial) coefficients $r_{\boldsymbol{\alpha}}(\boldsymbol{\sigma})$ then form a set of difference equations called the (GCH) equations of order n: $\boldsymbol{r}(\boldsymbol{\sigma}) = \begin{bmatrix} r_{\boldsymbol{\alpha}_1} & r_{\boldsymbol{\alpha}_2} & \dots \end{bmatrix}^\mathsf{T}$.

Lemma 2.1 ([61], Thm. 2)**.** *Given diagonal matrices $\boldsymbol{A}_k \in \mathbb{C}^{n\times n}, k = 1,\dots,m$, with eigentuples $\boldsymbol{\lambda}_i$. The common roots of the GCH equations $\boldsymbol{r}(\boldsymbol{\sigma})$, as constructed in* (3)*, are then given precisely by the set $\{\boldsymbol{\lambda}_1,\dots,\boldsymbol{\lambda}_n\}$. Moreover, each solution has a multiplicity of one whenever this set consists of n distinct tuples.*

Using the model parameters $\mathbf{\Lambda}$, one can construct a set of diagonal, parametric state-transition matrices $\boldsymbol{A}_i$ as discussed in Section 1.1. Applying Lemma 2.1 to these matrices then results in a set of parametric difference equations in terms of $\mathbf{\Lambda}$. The resulting polynomials then precisely describe the given eigentuples as their solution set as long as $\mathbf{\Lambda}$ consists of distinct eigentuples. This construction is illustated in Example 2.2 below.

Example 2.2. Consider a second order parametric state-space model with eigentuples $\boldsymbol{\lambda}_1 = (\lambda_{1,1}, \lambda_{1,2})$ and $\boldsymbol{\lambda}_2 = (\lambda_{2,1}, \lambda_{2,2})$, resulting in the state-transition matrices:

$$\boldsymbol{A}_1 = \begin{bmatrix} \lambda_{1,1} & \\ & \lambda_{2,1} \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} \lambda_{1,2} & \\ & \lambda_{2,2} \end{bmatrix}, \boldsymbol{c} = \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

Now, consider the kernel representation. The GCH equations for $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ then lead to the following set of difference equations:

$$\boldsymbol{r}(\boldsymbol{\sigma}) = \begin{bmatrix} \boldsymbol{\sigma}^{(2,0)} - \boldsymbol{\sigma}^{(1,0)}a + b \\ 2\boldsymbol{\sigma}^{(1,1)} - \boldsymbol{\sigma}^{(0,1)}a - \boldsymbol{\sigma}^{(1,0)}c + d \\ \boldsymbol{\sigma}^{(0,2)} - \boldsymbol{\sigma}^{(0,1)}c + e \end{bmatrix},$$

where we introduced the auxiliary variables $a = \lambda_{1,1} + \lambda_{1,2}, b = \lambda_{1,1}\lambda_{1,2}, c = \lambda_{2,1} + \lambda_{2,2}, d = \lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}\lambda_{2,1}, e = \lambda_{2,1}\lambda_{2,2}$.

2.4 Macaulay matrix

Over a finite set of grid points $\mathcal{K}$, the difference equations define a system of linear equations and can thus be written in matrix format as $\boldsymbol{M}_{\boldsymbol{r}}\hat{\boldsymbol{y}} = \boldsymbol{0}$. The coefficient matrix $\boldsymbol{M}_{\boldsymbol{r}}$ then is a *Macaulay matrix*, originating from linear algebra techniques for manipulation of and root finding for multivariate polynomials. Definition 2.2 compactly desbribes this type of matrix using multi-indices [5,15,41]. The construction is subsequently illustrated in Example 2.3.

Definition 2.2 (Macaulay matrix)**.** Given a set of difference equations $\boldsymbol{r}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} \boldsymbol{v}_{\boldsymbol{\alpha}}\boldsymbol{\sigma}^{\boldsymbol{\alpha}} \in \mathbb{C}^p[\boldsymbol{\sigma}], \boldsymbol{v}_{\boldsymbol{\alpha}} \in \mathbb{C}^p$, where $\mathcal{A} \subseteq \mathbb{N}_0^m$ is a finite set of a total degree $d = \max_{\boldsymbol{\alpha}\in\mathcal{A}} \|\boldsymbol{\alpha}\|_1$. Using multi-indices, the Macaulay matrix $\boldsymbol{M}_{\boldsymbol{r}}^{pq_1\times q_2}$ of degree $d_t \geq d$ associated with $r(\boldsymbol{\sigma})$ is then defined block-elementwise as:

$$(\boldsymbol{M}_r)_{\boldsymbol{\rho},\boldsymbol{\kappa}} = \boldsymbol{v}_{\boldsymbol{\kappa}-\boldsymbol{\rho}}, \quad \text{where } \boldsymbol{\kappa}, \boldsymbol{\rho} \in \mathbb{N}_0^m; \|\boldsymbol{\kappa}\|_1 \leq d_t, \|\boldsymbol{\rho}\|_1 \leq d_t - d.$$

Here, whenever $\boldsymbol{\kappa} - \boldsymbol{\rho} \notin \mathcal{A}$, the coefficient $\boldsymbol{v}_{\boldsymbol{\kappa}-\boldsymbol{\rho}}$ is assumed to be zero.

Example 2.3. Reconsider the GCH equations constructed in Example 2.2. The associated Macaulay matrix of degree three can then be used to define model-compliant sequences $\hat{\boldsymbol{y}}$:

$$\begin{array}{r} \\ \boldsymbol{r}(\boldsymbol{\sigma}) \\ \\ \\ \boldsymbol{0} = \sigma_1\boldsymbol{r}(\boldsymbol{\sigma}) \\ \\ \\ \sigma_2\boldsymbol{r}(\boldsymbol{\sigma}) \\ \\ \end{array} \begin{array}{c} \begin{matrix} (0,0) & (1,0) & (0,1) & (2,0) & (1,1) & (0,2) & (3,0) & (2,1) & (1,2) & (3,0) \end{matrix} \\ \begin{bmatrix} b & a & & 1 & & & & & & \\ d & c & a & & 2 & & & & & \\ e & & c & & & 1 & & & & \\ & b & & a & & & 1 & & & \\ & d & & c & a & & & 2 & & \\ & e & & & c & & & & 1 & \\ & & b & & a & & & 1 & & \\ & & d & & c & a & & & 2 & \\ & & e & & & c & & & & 1 \end{bmatrix} \end{array} \begin{bmatrix} y_{0,0} \\ y_{(1,0)} \\ y_{(0,1)} \\ y_{(2,0)} \\ y_{(1,1)} \\ y_{(0,2)} \\ y_{(3,0)} \\ y_{(2,1)} \\ y_{(1,2)} \\ y_{(0,3)} \end{bmatrix} = \boldsymbol{M}_{\boldsymbol{r}}\hat{\boldsymbol{y}}$$

Note that for these GCH equations, $\boldsymbol{M}_{\boldsymbol{r}}$ depends on the model parameters $\mathbf{\Gamma}$.

While different sets of grid points can be considered, we confine ourselves to those up to some fixed degree: $\mathcal{K} = \{\boldsymbol{\kappa} \mid \boldsymbol{\kappa} \in \mathbb{N}_0^m, \|\boldsymbol{\kappa}\|_1 \leq d\}$, as in Definition 2.2. Given such a set of grid points of a large enough degree, the null space of a Macaulay matrix corresponds precisely to an observability matrix, formalized in Theorem 2.2 below. Indeed, the kernel of $\boldsymbol{M}_{\boldsymbol{r}}$ describes the model compliant sequences $\hat{\boldsymbol{y}}$, just like range $\mathbf{\Gamma}$ in the image representation. For too small a degree, range $\mathbf{\Gamma}$ may be

a strict subspace of null $\boldsymbol{M_r}$. For example, if signals were considered over $\|\boldsymbol{\kappa}\|_1 \leq 2$ in Example 2.3, the Macaulay matrix would be of size 3×6 so that its nullity would be at least three, which is larger than the model order $n = 2$ in this case.

Lemma 2.2 ([5], Thm. 4.1)**.** *Let $\boldsymbol{r}(\boldsymbol{\sigma}) \in \mathbb{C}^p[\boldsymbol{\sigma}]$ be a set of mD difference equations, with a finite number n of common roots, counting multiplicities, and without roots at infinity and let $\boldsymbol{M_r}$ be the corresponding Macaulay matrix of degree d, as per Definition 2.2. Let $\boldsymbol{\Gamma}_{\boldsymbol{\kappa},:} = \boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}, \|\boldsymbol{\kappa}\|_1 \leq d$ be the observability matrix associated with any equivalent nth order state-space model (1). Then for some finite degree of regularity d^*, it holds that: when $d \geq d^*$, then*

$$\text{null } \boldsymbol{M_r} = \text{range } \boldsymbol{\Gamma}, \quad \textit{and} \quad \dim(\text{range } \boldsymbol{\Gamma}) = n.$$

Moreover, $d^* \leq \min_{\mathcal{I} \subseteq \{1,\ldots,p\}, \#\mathcal{I}=m} \left(\sum_{i \in \mathcal{I}} d_i - m + 1\right)$, where d_i is the total degree of the ith element of $\boldsymbol{r}(\boldsymbol{\sigma})$.

3 Polynomial characterization of the stationary points

We now derive the polynomial description for the local minima of the least-squares misfit problem (2) through a relaxation of the problem. First, a characterization based on the image representation is considerd in Section 3.1. This characterization implies a decomposition of the ambient space for each optimal model, which is detailed in Section 3.2. From this decomposition, another characterization based on the kernel representation is derived in Section 3.3.

3.1 Image representation

Consider the following relaxed version of (2), where the model parameters $(\boldsymbol{\Lambda}, \boldsymbol{x}) \notin \mathcal{R}$ are allowed as well. Using the insights from Section 2, the model-compliancy constraint is expressed simultaneously for all grid points through an observability matrix:

$$\begin{aligned} \min_{\boldsymbol{\Lambda}, \boldsymbol{x}} \ & \|\tilde{\boldsymbol{y}}\|_2^2 \\ \text{subject to } & \hat{\boldsymbol{y}} = \boldsymbol{y} - \tilde{\boldsymbol{y}} \\ & \hat{\boldsymbol{y}} = \boldsymbol{\Gamma}\boldsymbol{x}. \end{aligned} \tag{4}$$

Substituting in the constraints leads to the the unconstrained objective function (5):

$$\|\tilde{\boldsymbol{y}}\|_2^2 = \|\boldsymbol{y} - \boldsymbol{\Gamma}\boldsymbol{x}\|^2 = \boldsymbol{y}^{\mathsf{H}}\boldsymbol{y} + \boldsymbol{x}^{\mathsf{H}}\boldsymbol{\Gamma}^{\mathsf{H}}\boldsymbol{\Gamma}\boldsymbol{x} - \boldsymbol{x}^{\mathsf{H}}\boldsymbol{\Gamma}^{\mathsf{H}}\boldsymbol{y} - \boldsymbol{y}^{\mathsf{H}}\boldsymbol{\Gamma}\boldsymbol{x}. \tag{5}$$

As this objective is generally a multivariate polynomial of degree larger than two, it is non-convex and can thus have multiple local minima.

Since we are interested in real models after all, we focus on the stationary points of this relaxed problem that do in fact define real models. Indeed, it is obvious that any stationary point of (4) that lies in $\mathcal{R}$ is also a stationary point of (2), since, if the derivatives vanish in all directions, they also vanish in the directions that respect the realness contraint. However, the converse is not immediately obvious: if the derivatives vanish in a restricted set of 'real' directions, they do not necessarily vanish in all directions. As such, it has yet to be established that all local minima of (2) are also stationary points of (4). Theorem 4.1 later asserts that this is indeed the case.

The FONC for this relaxed problem can now be derived by equating all partial derivatives of the objective function (5) w.r.t. the decision variables to zero. Wirtinger calculus provides a straightforward methodology to deal with the intricacies of optimization problems over complex-valued decision variables [1]: Treat the complex conjugates of the decision variables, i.e. $\overline{\boldsymbol{\Lambda}}, \overline{\boldsymbol{x}}$, as independent w.r.t. $\boldsymbol{\Lambda}, \boldsymbol{x}$. This results in:

$$\frac{\partial \|\tilde{\boldsymbol{y}}\|_2^2}{\partial \boldsymbol{x}} = (\boldsymbol{x}^{\mathsf{H}}\boldsymbol{\Gamma}^{\mathsf{H}} - \boldsymbol{y}^{\mathsf{H}})\boldsymbol{\Gamma} = \boldsymbol{0}, \tag{6a}$$

$$\begin{aligned} \frac{\partial \|\tilde{\boldsymbol{y}}\|_2^2}{\partial \lambda_{i,k}} &= (\boldsymbol{x}^{\mathsf{H}}\boldsymbol{\Gamma}^{\mathsf{H}} - \boldsymbol{y}^{\mathsf{H}})\frac{\partial \boldsymbol{\Gamma}}{\partial \lambda_{i,k}}\boldsymbol{x} = 0 \\ & \forall (i,k) \in \{1,\ldots,n\} \times \{1,\ldots,m\}. \end{aligned} \tag{6b}$$

We denote the set of stationary points, that is, the common solutions of (6), as $\mathcal{S}$ from here on.

Note that $\frac{\partial \boldsymbol{\Gamma}}{\partial (\lambda_i)_k}\boldsymbol{x} = \begin{bmatrix} \frac{\partial \boldsymbol{\lambda}_i^{\boldsymbol{\kappa}_1}}{\partial \lambda_{i,k}} & \ldots & \frac{\partial \boldsymbol{\lambda}_i^{\boldsymbol{\kappa}_N}}{\partial \lambda_{i,k}} \end{bmatrix}^{\mathsf{T}} x_i$. To more compactly express the following equations, we introduce the following $N \times nm$ polynomial matrix, which stacks these derivatives along one another: $\boldsymbol{\Gamma}^{(\boldsymbol{\lambda})} =$

$$\begin{bmatrix} \kappa_1 \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}_1 - \mathrm{e}_1} & \ldots & \kappa_m \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}_m - \mathrm{e}_m} & \ldots & \kappa_1 \boldsymbol{\lambda}_n^{\boldsymbol{\kappa}_1 - \mathrm{e}_1} & \ldots & \kappa_m \boldsymbol{\lambda}_n^{\boldsymbol{\kappa}_m - \mathrm{e}_m} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \kappa_1 \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}_N - \mathrm{e}_1} & \ldots & \kappa_m \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}_N - \mathrm{e}_m} & \ldots & \kappa_1 \boldsymbol{\lambda}_n^{\boldsymbol{\kappa}_N - \mathrm{e}_1} & \ldots & \kappa_m \boldsymbol{\lambda}_n^{\boldsymbol{\kappa}_N - \mathrm{e}_m} \end{bmatrix}.$$

Now, for any model $(\boldsymbol{\Lambda}, \boldsymbol{x}) \in \mathcal{R}$, it holds that $\boldsymbol{\Lambda}\boldsymbol{x} = \hat{\boldsymbol{y}} \in \mathbb{R}^N$, and consequently, $\boldsymbol{\Lambda}\boldsymbol{x}^{\mathsf{H}} = \boldsymbol{\Lambda}\boldsymbol{x}^{\mathsf{T}}$. As such, the subset of stationary points that define real models, i.e., $\mathcal{S} \cap \mathcal{R}$, satisfies the following polynomial equations, obtained by exchanging $^{\mathsf{H}}$ for $^{\mathsf{T}}$ in (6), where $\otimes$ denotes the Kronecker product:

$$\boldsymbol{0} = (\boldsymbol{x}^{\mathsf{T}}\boldsymbol{\Gamma}^{\mathsf{T}} - \boldsymbol{y}^{\mathsf{T}})\boldsymbol{\Gamma}, \tag{7a}$$

$$\boldsymbol{0} = (\boldsymbol{x}^{\mathsf{T}}\boldsymbol{\Gamma}^{\mathsf{T}} - \boldsymbol{y}^{\mathsf{T}})\boldsymbol{\Gamma}^{(\boldsymbol{\lambda})}(\text{diag}(\boldsymbol{x}) \otimes \boldsymbol{I}_m). \tag{7b}$$

For general 'complex' models, it is clear that this substitution of $^{\mathsf{H}}$ by $^{\mathsf{T}}$ does not retain the equality. Because of this, the solutions of (7) outside of $\mathcal{R}$ do not necessarily belong to $\mathcal{S}$.

Finally, if one is only interested in the local minima of the misfit problem (2), one can apply Theorem 4.2, which states that for these minima, $x_i \neq 0$ for all i. This ensures $(\operatorname{diag}(\boldsymbol{x}) \otimes \boldsymbol{I}_m)$ is invertible, allowing us to simplify (7b) by right multiplying with the inverse of this matrix. The preceding observations then result in the simplified characterization of the local minima (8) and Theorem 3.1 below.

Theorem 3.1. *If $(\boldsymbol{\Lambda}, \boldsymbol{x}) \in \mathcal{S}$ is such that $\forall i : x_i \neq 0$ and $(\boldsymbol{\Lambda}, \boldsymbol{x}) \in \mathcal{R}$, then it is a solution to:*

$$\begin{bmatrix} \boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\Gamma} & \boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{y} \\ \left(\boldsymbol{\Gamma}^{(\boldsymbol{\lambda})}\right)^{\mathsf{T}}\boldsymbol{\Gamma} & \left(\boldsymbol{\Gamma}^{(\boldsymbol{\lambda})}\right)^{\mathsf{T}}\boldsymbol{y} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ -1 \end{bmatrix} = \boldsymbol{0}. \tag{8}$$

We denote the set of common roots of (8) *as $\mathcal{C}$. Conversely, $(\mathcal{C} \cap \mathcal{R}) \subseteq \mathcal{S}$.*

Proof. The first statement was derived in the text leading up to this theorem. To prove the second part of the statement, the steps taken in the derivation leading up to (8) can be reverted. □

By Theorem 3.1, some stationary points with $x_i = 0$ might not arise as solutions to (8). In such models, the ith eigentuple does not contribute to $\hat{\boldsymbol{y}}$, so that it may just as well be expressed with a model of order $n-1$. However, such *lower-order* solutions may persist in $\mathcal{C}$. Indeed, take $\boldsymbol{y} \perp \text{range} \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Gamma}^{(\boldsymbol{\lambda})} \end{bmatrix}$ for some fixed $\boldsymbol{\Lambda}$. The model $(\boldsymbol{\Lambda}, \boldsymbol{0})$ is then a solution to (8).

Let us now consider a first-order, 2D example below to illustrate the presented methodology.

Example 3.1. We consider the misfit problem (2) for fitting a first-order model, i.e. $n = 1$, to the following 2D data sequence:

$$\boldsymbol{y}^{\mathsf{T}} = \begin{bmatrix} \overset{(0,0)}{0.7060} & \overset{(1,0)}{0.0318} & \overset{(0,1)}{0.2769} & \overset{(2,0)}{0.0462} & \overset{(1,1)}{0.0971} & \overset{(0,2)}{0.8235} \end{bmatrix}.$$

We now set up the equations (8) for this problem. There is one eigentuple, the components of which we denote as $\boldsymbol{\lambda}_1 = (\alpha, \beta)$ for this example, to avoid too many super- and subscripts, resulting in:

index	$\boldsymbol{\lambda}_1 = (\alpha, \beta)$	x	$\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2$
1	$(0.1268, 1.1478)$	0.5201	0.3990
2	$(0.0027, -0.3106)$	0.6325	0.9074
3	$(-87.8665, 5.1478)$	$5.5823 \cdot 10^{-6}$	1.1243

Table 1
Solutions of (9) that correspond to real models, ordered by misfit norm. Since they are real models, these are stationary points of (4), that is, they satisfy (6). By extension, they are stationary points of (2).

index	$\boldsymbol{\lambda}_1 = (\alpha, \beta)$	$\|\boldsymbol{f}\|_2$
4-5	$(-0.1579 \mp 1.8369i, -2.4644 \pm 0.4170i)$	7.6877
6-7	$(-0.0540 \mp 1.0321i, 1.1074 \mp 0.0214i)$	0.5863
8-9	$(-0.1372 \mp 0.1135i, -1.4475 \mp 0.8220i)$	3.1840
10-11	$(0.2252 \mp 1.1657i, -1.4219 \mp 0.3115i)$	5.0424
12-13	$(-0.0168 \mp 0.6927i, -0.2462 \pm 0.0686i)$	1.3986

Table 2
Complex-valued solutions of (9), where $\boldsymbol{f}(\boldsymbol{\lambda}_1, \boldsymbol{x})$ is the left-hand-side of (6). This function evaluation shows that these solutions are not part of $\mathcal{S}$. x is omitted as it follows directly from $\boldsymbol{\lambda}_i$, as will be discussed in Section 3.2.

$$\begin{aligned} 0 &= (1 + \beta^2 + \beta^4 + \alpha^2 + \alpha^2\,\beta^2 + \alpha^4)x \\ &\quad - (0.70 + 0.28\,\beta + 0.82\,\beta^2 + 0.03\,\alpha + 0.097\,\alpha\beta + 0.046\,\alpha^2), \end{aligned} \tag{9}$$

$$\begin{aligned} 0 &= (\alpha + \alpha\,\beta^2 + 2\,\alpha^3)x - (0.0318 + 0.0971\,\beta + 0.0924\,\alpha), \\ 0 &= (\beta + 2\,\beta^3 + \alpha^2\,\beta)x - (0.2769 + 1.6470\,\beta + 0.0971\,\alpha). \end{aligned}$$

Solving this using MacaulayLab [62] results in 13 affine complex solutions for $\boldsymbol{\lambda}_1$, as shown in Tables 1 and 2. By Theorem 3.1, the three real-valued solutions in Table 1 are then stationary points of the relaxed problem (4), and by extension of (2). Note that existence of multiple stationary points indicates the nonconvexity of the problem. From this, we conclude that the globally optimal model is corresponds to the eigentuple $\boldsymbol{\lambda}_1 = (0.1268, 1.1478)$.

The remaining ten solutions in Table 2 do not define real models. As such, these models are not necessarily stationary points. This is evidenced by the last column of the table, where the norm of the the FONC eavluated in these points is given.

Notice how all the solutions satisfy $x \neq 0$. However, if we turn our attention to (7) instead and search for solutions with $x = 0$, these equations reduce to:

$$\begin{aligned} 0 = y^{\mathsf{T}}\boldsymbol{\Gamma} =& 0.7060 + 0.0318\alpha + 0.2769\beta + 0.0462\alpha^2 \\ &+ 0.0971\alpha\beta + 0.8235\beta^2. \end{aligned}$$

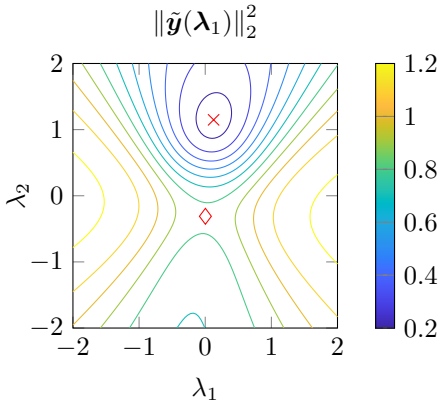


Fig. 2. Objective for the misfit problem in Example 3.1. The optimal value of x is determined based on $\boldsymbol{\lambda}$ through a linear projection of $\boldsymbol{y}$ onto range $\boldsymbol{\Gamma}$. The global minimum is marked by $\times$ and the saddle point by $\diamond$.

Clearly, there is an infinite set of solutions to this bivariate polynomial equation, which gets eliminated in the simplified characaterization (9). In Section 5, we briefly discuss how such infinite solution sets may hinder the convergence of some polynomial solvers. Note that this observation can be easily generalized to higher values of n.

3.2 Misfit structure

As previously discussed, range $\boldsymbol{\Gamma} \subseteq \mathbb{R}^N$ forms the linear subspace of model-compliant sequences over $\mathcal{K}$. For a fixed value of $\boldsymbol{\Lambda}$, the closest sequence $\hat{\boldsymbol{y}}$ in a least-squares sense is thus readily obtained by orthogonally projecting $\boldsymbol{y}$ onto this subspace, defining $\boldsymbol{x}$ in terms of $\boldsymbol{\Lambda}$ in the process. This is reflected in (6a), which constitute the normal equations for this projection. Using this projection to define x in terms of $\boldsymbol{\lambda}$, a contour plot of the objective function underlying Example 3.1 is plotted in Figure 2, clearly indicating the non-convexity of the optimization problem.

In this sense, $\boldsymbol{\Lambda}$ are the only decision variables of importance for optimality, while $\boldsymbol{x}$, on which the objective only depends linearly, can be considered auxiliary variables. Each value of $\boldsymbol{\Lambda}$ thus decomposes the ambient space $\mathbb{R}^N$ into two orthogonal parts: A model-compliant subspace, range $\boldsymbol{\Gamma}$, and its orthogonal complement, containing the residuals. In case of distinct eigentuples and an appropriate set of grid points, this complementary space is row $\boldsymbol{M}_{\boldsymbol{r}}$ by Lemmas 2.1 and 2.2, such that $\tilde{\boldsymbol{y}} \in$ row $\boldsymbol{M}_{\boldsymbol{r}}$.

To also characterize the local optimality of $\boldsymbol{\Lambda}$, the whole of (8) should be considered, which can be rewritten as:

$$\tilde{\boldsymbol{y}}^\mathsf{T} \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Gamma}^{(\boldsymbol{\lambda})} \end{bmatrix} = \tilde{\boldsymbol{y}}^\mathsf{T} \boldsymbol{G} = \boldsymbol{0}. \tag{10}$$

For real models, this condition further restricts the misfit $\tilde{\boldsymbol{y}}$ to lie in the orthogonal complement of range $\boldsymbol{G}$, a strict subspace of (range $\boldsymbol{\Gamma})^\perp$ by its construction. Similarly to the 1D case [12], each optimal model $\boldsymbol{\Lambda}$ then gives rise to a further decomposition of $\mathbb{R}^N$ into three parts: the model-compliant space range $\boldsymbol{\Gamma}$, the misfit space (range $\boldsymbol{G})^\perp$, and the complementary subspace range $\boldsymbol{\Gamma}^{(\boldsymbol{\lambda})}$.

Finally, define the polynomial $\tilde{y}(\boldsymbol{\lambda}) = \sum_{\boldsymbol{\kappa} \in \mathcal{K}} \tilde{y}_{\boldsymbol{\kappa}} \boldsymbol{\lambda}^{\boldsymbol{\kappa}}$. Equation (10) can then be re-expressed as the interpolation condition $\forall i, k : \tilde{y}(\boldsymbol{\lambda}_i) = 0 = \frac{\partial \tilde{y}(\boldsymbol{\lambda})}{\partial \lambda_k}|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_i}$ [50,12], or, in a system-theoretic sense, the signals $\boldsymbol{\lambda}^{\boldsymbol{\kappa}}, \kappa_k \boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k}$ satisfy $\tilde{y}(\boldsymbol{\sigma})$. Indeed, $\tilde{\boldsymbol{y}}$ lies in the orthogonal complement of the observability matrix $\boldsymbol{G}$ associated with these signals. By the discussion in Section 2, this complement is spanned by the rows of a Macaulay matrix generated by a set of polynomials with this specific set of common roots and multiplicity structure, given a suitable $\mathcal{K}$.

3.3 Kernel representation

The Macaulay matrix whose row space forms the orthogonal complement of the observability matrix $\boldsymbol{G}$ can be used to characterize $\hat{\boldsymbol{y}}$ through the Kernel representation, leading to an alternative set of polynomial equations which are satisfied by the minima of (2). Let us now look into how one can obtain difference equations with the required multiplicity structure, that is, for which all first-order partial derivatives vanish in their common roots. To this end, consider Example 3.2.

Example 3.2. To illustrate the procedure, let us constrain ourselves to a simple 2D, first-order example. First, consider a set of difference equations describing precisely one exponential mode $\boldsymbol{\lambda}_1^{\boldsymbol{\kappa}} = (\alpha, \beta)^{\boldsymbol{\kappa}}$, which is readily constructed as:

$$\boldsymbol{r}(\boldsymbol{\sigma}) y_{\boldsymbol{\kappa}} = \begin{bmatrix} r_1(\boldsymbol{\sigma}) \\ r_2(\boldsymbol{\sigma}) \end{bmatrix} y_{\boldsymbol{\kappa}} = \begin{bmatrix} \sigma_1 - \alpha \\ \sigma_2 - \beta \end{bmatrix} y_{\boldsymbol{\kappa}} = \boldsymbol{0}.$$

Now consider the difference equations resulting from all pairwise products of the elements of this $\boldsymbol{r}(\boldsymbol{\sigma})$: $\boldsymbol{r}^2(\boldsymbol{\sigma}) y_{\boldsymbol{\kappa}} =$

$$\begin{bmatrix} r_1(\boldsymbol{\sigma})^2 \\ r_1(\boldsymbol{\sigma}) r_2(\boldsymbol{\sigma}) \\ r_2(\boldsymbol{\sigma})^2 \end{bmatrix} y_{\boldsymbol{\kappa}} = \begin{bmatrix} \alpha^2 - 2\alpha\sigma_1 + \sigma_1^2 \\ \alpha\beta - \alpha\sigma_2 - \beta\sigma_1 + \sigma_1\sigma_2 \\ \beta^2 - 2\beta\sigma_2 + \sigma_2^2 \end{bmatrix} y_{\boldsymbol{\kappa}} = \boldsymbol{0}.$$

Clearly, $\boldsymbol{r}^2$ then has the same solution $\boldsymbol{\lambda}_1$, though with an altered multiplicity. Indeed, the difference equations $\boldsymbol{r}^2(\boldsymbol{\sigma})$ precisely describes the signal space spanned by $\boldsymbol{\lambda}_1^{\boldsymbol{\kappa}}, \kappa_1 \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}-\boldsymbol{e}_1}$ and $\kappa_2 \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}-\boldsymbol{e}_2}$.

Let us verify this using the Macaulay framework: For $\|\boldsymbol{\kappa}\|_2 \leq 2$, the null space of the associated Macaulay matrix constitutes signals in the range of the observability

matrix $\boldsymbol{G}$: $\boldsymbol{M}_{\boldsymbol{r}^2}\boldsymbol{G} =$

$$\begin{bmatrix} \alpha^2 & -2\,\alpha & 0 & 1 & 0 & 0 \\ \alpha\,\beta & -\beta & -\alpha & 0 & 1 & 0 \\ \beta^2 & 0 & -2\,\beta & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ \beta & 0 & 1 \\ \alpha^2 & 2\,\alpha & 0 \\ \alpha\,\beta & \beta & \alpha \\ \beta^2 & 0 & 2\,\beta \end{bmatrix} = \boldsymbol{0},$$

Clearly, $\boldsymbol{M}_{\boldsymbol{r}^2}$ is of rank 3, such that this $\boldsymbol{G}$ spans its full null space, indicating no further multiplicities. It can be verified that a similar observation holds for higher degree Macaulay matrices.

Now, since range $\boldsymbol{\Gamma} \subset$ range $\boldsymbol{G}$, the converse holds for their orthogonal complements: row $\boldsymbol{M}_{\boldsymbol{r}} \supset$ row $\boldsymbol{M}_{\boldsymbol{r}^2}$. This indicates that $\boldsymbol{M}_{\boldsymbol{r}^2}$ can be expressed as linear combinations of the rows of $\boldsymbol{M}_{\boldsymbol{r}}$. Indeed, notice that the polynomial multiplications can be expressed using Macaulay matrices $\boldsymbol{P}_{r_1}, \boldsymbol{P}_{r_2}$ matching the degree of the row shifts in $\boldsymbol{M}_{\boldsymbol{r}}$, i.e. one in this case:

$$\boldsymbol{P}_{r_1} = \begin{bmatrix} -\alpha & 1 & 0 \end{bmatrix}, \boldsymbol{P}_{r_2} = \begin{bmatrix} -\beta & 0 & 1 \end{bmatrix},$$

$$\boldsymbol{M}_{\boldsymbol{r}^2} = \begin{bmatrix} \boldsymbol{P}_{r_1} & & \\ \boldsymbol{P}_{r_2} & & \\ & \boldsymbol{P}_{r_2} & \end{bmatrix} \boldsymbol{M}_{\boldsymbol{r}} = \boldsymbol{P}\boldsymbol{M}_{\boldsymbol{r}}$$

$$= \left[\begin{array}{ccc|ccc} -\alpha & 1 & 0 & 0 & 0 & 0 \\ -\beta & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta & 0 & 1 \end{array}\right] \begin{bmatrix} -\alpha & 1 & 0 & 0 & 0 & 0 \\ 0 & -\alpha & 0 & 1 & 0 & 0 \\ 0 & 0 & -\alpha & 0 & 1 & 0 \\ -\beta & 0 & 1 & 0 & 0 & 0 \\ 0 & -\beta & 0 & 0 & 1 & 0 \\ 0 & 0 & -\beta & 0 & 0 & 1 \end{bmatrix}.$$

We now generalize the observations of Example 3.2 to general multidimensional difference equations.

Lemma 3.1. *Let $\boldsymbol{r}(\boldsymbol{\sigma}) \in \mathbb{R}^p[\boldsymbol{\sigma}], p \geq m$, be a set of difference equations with a kernel that is spanned by n basis signals $\mathcal{B} = \{\boldsymbol{\lambda}_1^{\boldsymbol{\kappa}}, \ldots, \boldsymbol{\lambda}_n^{\boldsymbol{\kappa}}\}$. Define the set of polynomials $\boldsymbol{r}^2(\boldsymbol{\sigma}) = \left[r_s(\boldsymbol{x})r_t(\boldsymbol{x}) \mid s = 1, \ldots, n; t \leq s\right]$. The kernel of $\boldsymbol{r}^2(\boldsymbol{\sigma})$ is then spanned by the basis $\mathcal{B} \cup \{\kappa_k \boldsymbol{\lambda}_i^{\boldsymbol{\kappa}-\boldsymbol{e}_k} \mid i = 1, \ldots, n; k = 1, \ldots, m\}$.*

Proof. Since all signals in the kernel are of the form $\boldsymbol{\lambda}_i^{\boldsymbol{\kappa}}$, the common roots $\boldsymbol{\lambda}_i$ are multiplicity-free. This means that $\boldsymbol{r}(\boldsymbol{\lambda}_i) = \boldsymbol{0}$ and $\forall i, \boldsymbol{v} \in \mathbb{C}^m \backslash \{\boldsymbol{0}\} : \boldsymbol{J}_{\boldsymbol{r}}(\boldsymbol{\lambda}_i)\boldsymbol{v} \neq \boldsymbol{0}$, where $\boldsymbol{J}_{\boldsymbol{r}}$ is the Jacobian of $\boldsymbol{r}(\boldsymbol{\lambda})$ [57,11].

Since $\boldsymbol{r}(\boldsymbol{\lambda}_i) = \boldsymbol{0}$, $\frac{\partial r_s r_t}{\partial \lambda_k}|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_i} = \frac{\partial r_s}{\partial \lambda_k} r_t(\boldsymbol{\lambda}_i) + \frac{\partial r_t}{\partial \lambda_k} r_s(\boldsymbol{\lambda}_i) = 0$ for any pair s, t, so that $\kappa_k \boldsymbol{\lambda}_i^{\boldsymbol{\kappa}-\boldsymbol{e}_k}$ satisfies $\boldsymbol{r}^2(\boldsymbol{\sigma})$.

Finally, due to the closedness property of the dual space, to show the absence of all higher-order multiplicities it suffices to show that nontrivial linear combinations of second-order partial derivatives vanish in $\boldsymbol{\lambda}_i$ [56]. It can be worked out that: $\sum_k^m \sum_{l \geq k}^m h_{(k,l)} \frac{\partial^2 r_s r_t}{\partial \lambda_k \partial \lambda_l}(\boldsymbol{\lambda}_i) = (\nabla r_s(\boldsymbol{\lambda}_i))^{\mathsf{T}} \boldsymbol{H} (\nabla r_t(\boldsymbol{\lambda}_i))$, where $\boldsymbol{H}$ is a symmetric matrix containing the $h_{(k,l)}$. To vanish for all pairs (s, t), it should thus hold that $\boldsymbol{J}_{\boldsymbol{r}}(\boldsymbol{\lambda}_i)\boldsymbol{H}\boldsymbol{J}_{\boldsymbol{r}}(\boldsymbol{\lambda}_i)^{\mathsf{T}} = \boldsymbol{0}$. However, since $\boldsymbol{H}$ is nonzero and $\boldsymbol{J}_{\boldsymbol{r}}(\boldsymbol{\lambda}_i)$ is of full column rank, this cannot be the case. □

Corollary 3.1. *Let $\boldsymbol{r}(\boldsymbol{\sigma})$ be the parametric m GCH equations of order n constructed from the eigentuples $\boldsymbol{\Lambda}$ and let $\boldsymbol{r}^2(\boldsymbol{\sigma})$ be as defined in Lemma 3.1. Whenever $\boldsymbol{\Lambda}$ comprises n distinct eigentuples, the right null space of the degree $d > (2n-1)m$ Macaulay matrix $\boldsymbol{M}_{\boldsymbol{r}^2}$ is then* range $\left[\boldsymbol{\Gamma}\ \boldsymbol{\Gamma}^{(\boldsymbol{\lambda})}\right] =$ range $\boldsymbol{G}$.

Proof. By Lemma 2.1, these equations meet the conditions of Lemma 3.1 when the eigentuples are distinct. As discussed in Section 2.2, all signals of the form $\kappa_k \boldsymbol{\lambda}_i^{\boldsymbol{\kappa}-\boldsymbol{e}_k}$ are then model-compliant, so that the statement follows from Lemma 2.2. Note that the observability matrix $\boldsymbol{G}$ is to be constructed as in Theorem 2.2 with $\mathcal{K} = \{\boldsymbol{\kappa} \in \mathbb{N}_0 \mid \|\boldsymbol{\kappa}\|_1 \leq d\}$. The degree bound follows from the upper bound on the degree of regularity of $\boldsymbol{r}^2(\boldsymbol{\sigma})$, defined in Theorem 2.2. □

Note that, as was already illustrated in Example 3.1, the row space of $\boldsymbol{M}_{\boldsymbol{r}^2}$ is a subspace of row $\boldsymbol{M}_{\boldsymbol{r}}$. Indeed, it is readily verified that the multiplication property holds in general: Given a Macaulay matrix $\boldsymbol{M}_{r_t}$ of degree d, with r_t of degree d_t and let $\boldsymbol{M}_{r_s}$ be another Macaulay matrix of degree $d - d_t$, then it holds that $\boldsymbol{M}_{r_s r_t} = \boldsymbol{M}_{r_s}\boldsymbol{M}_{r_t}$ is the Macaulay matrix of degree d associated with $r_s(\boldsymbol{\sigma})r_t(\boldsymbol{\sigma})$. From this relation it can be worked out that $\boldsymbol{M}_{\boldsymbol{r}^2} = \boldsymbol{P}\boldsymbol{M}_{\boldsymbol{r}} = \text{blockdiag}(\boldsymbol{M}_{\boldsymbol{r}_{1:p}}, \boldsymbol{M}_{\boldsymbol{r}_{2:p}}, \ldots, \boldsymbol{M}_{r_p})\boldsymbol{M}_{\boldsymbol{r}}$ for an appropriate choice of degrees of the Macaulay matrices in $\boldsymbol{P} \in \mathbb{C}^{q \times t}$.

Lemma 3.1 provides us with an explicit expression for the left kernel of $\boldsymbol{G} = \left[\boldsymbol{\Gamma}\ \boldsymbol{\Gamma}^{(\boldsymbol{\lambda})}\right]$ in terms of the eigentuples given a suitable $\mathcal{K}$. This allows (10) to be expressed as $\tilde{\boldsymbol{y}} \in$ row $\boldsymbol{M}_{\boldsymbol{r}^2}$. Using this property, Theorem 3.1 is rephrased by leveraging the kernel representation $\boldsymbol{M}_{\boldsymbol{r}}\hat{\boldsymbol{y}} = \boldsymbol{0}$, similarly to the 1D misfit and model-order-reduction problems [36,35].

Theorem 3.2. *Let $\mathcal{K}$ be as defined in Corollary 3.1 and let $(\boldsymbol{\Lambda}^*, \boldsymbol{x}^*) \in \mathcal{C}$ be any model with distinct eigentuples.*

Then, there exists a vector $\boldsymbol{g}^* \in \mathbb{C}^q$ such that $(\boldsymbol{\Lambda}^*, \boldsymbol{g}^*)$

satisfies:

$$\begin{bmatrix} \boldsymbol{M_r} \boldsymbol{M_r}^{\mathsf{T}} \boldsymbol{P}^{\mathsf{T}} & \boldsymbol{M_r} \boldsymbol{y} \end{bmatrix} \begin{bmatrix} \boldsymbol{g} \\ -1 \end{bmatrix} = \boldsymbol{0}. \tag{11}$$

Conversely, for any point $(\boldsymbol{\Lambda}^*, \boldsymbol{g}^*)$ *with distinct eigentuples satisfying* (11)*, there exists a vector* $\boldsymbol{x}^*$ *such that* $(\boldsymbol{\Lambda}^*, \boldsymbol{x}^*) \in \mathcal{C}$.

Proof. Both sides of the proof rely on $\operatorname{null} \boldsymbol{M_r} = \operatorname{range} \boldsymbol{\Gamma}$, which holds by Theorems 2.1 and 2.2 .

We first prove that any solution of (8) with distinct eigentuples satisfies (11). Note that (10) follows from (8). By Corollary 3.1 the former implies that $\exists \boldsymbol{g} : \tilde{\boldsymbol{y}} = \boldsymbol{M_r}^{\mathsf{T}} \boldsymbol{P}^{\mathsf{T}} \boldsymbol{g}$. This results in:

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \tilde{\boldsymbol{y}} = \boldsymbol{\Gamma} \boldsymbol{x} + \boldsymbol{M_r}^{\mathsf{T}} \boldsymbol{P}^{\mathsf{T}} \boldsymbol{g} \Rightarrow \boldsymbol{M_r} \boldsymbol{M_r}^{\mathsf{T}} \boldsymbol{P}^{\mathsf{T}} \boldsymbol{g} = \boldsymbol{M_r} \boldsymbol{y}.$$

Conversely, choosing $\hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{M_r}^{\mathsf{T}} \boldsymbol{P}^{\mathsf{T}} \boldsymbol{g}$ in (11) results in $\boldsymbol{M_r} \hat{\boldsymbol{y}} = \boldsymbol{0}$. As such $\hat{\boldsymbol{y}}$ lies in the kernel of $\boldsymbol{M_r}$, i.e. $\exists \boldsymbol{x}^* : \hat{\boldsymbol{y}} = \boldsymbol{\Gamma} \boldsymbol{x}^*$. Consequently, by Corollary 3.1, it holds that:

$$\hat{\boldsymbol{y}} + \tilde{\boldsymbol{y}} = \boldsymbol{\Gamma} \boldsymbol{x}^* + \boldsymbol{M_r}^{\mathsf{T}} \boldsymbol{P}^{\mathsf{T}} \boldsymbol{g} = \boldsymbol{y} \Rightarrow \boldsymbol{G}^{\mathsf{T}} \boldsymbol{\Gamma} \boldsymbol{x}^* = \boldsymbol{G}^{\mathsf{T}} \boldsymbol{y}.$$

□

Example 3.3. Reconsider the same data and parameters as in Example 3.1. Working out (11) results in the system of polynomial equations:

$$\begin{bmatrix} 0.0318 - 0.7060\,\alpha & 2\,\alpha + \alpha^3 & \beta + \alpha^2\,\beta & \alpha\,\beta^2 \\ 0.0462 - 0.0318\,\alpha & -1 - 2\,\alpha^2 & -\alpha\,\beta & 0 \\ 0.0971 - 0.2769\,\alpha & 0 & -1 - \alpha^2 & -2\,\alpha\,\beta \\ 0.2769 - 0.7060\,\beta & \alpha^2\,\beta & \alpha + \alpha\,\beta^2 & 2\,\beta + \beta^3 \\ 0.0971 - 0.0318\,\beta & -2\,\alpha\,\beta & -1 - \beta^2 & 0 \\ 0.8235 - 0.2769\,\beta & 0 & -\alpha\,\beta & -1 - 2\,\beta^2 \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{g} \end{bmatrix} = \boldsymbol{0}. \tag{12}$$

4 Real models

Now that the proposed methodology has been outlined, it remains to be shown that the local minima of (2) are in fact part of the solutions $\mathcal{C}$. Indeed, the polynomial characterizations were obtained under the assumptions that local minima of (2) are also stationary points $\mathcal{S}$ of the relaxation (4), that $x_i \neq 0$ for $i = 1, \ldots, n$, and for (11), that the eigentuples are distinct. In the following subsections, we formally prove these statements. Moreover, it is also shown that the solutions of interest are almost always isolated solutions of the polynomial characterization, such that numerical solvers can indeed find these points. Finally, we briefly highlight a limitation of the proposed methodology: the absence of a global minimum and how it can be detected.

index	g_1	g_2	g_3
1	0.0378	0.0214	0.1383
2	0.0462	0.0976	0.7625
3	0.0031	0.0996	0.8234
4-5	$0.5502 \mp 0.1130\mathrm{i}$	$0.1939 \pm 0.6935\mathrm{i}$	$-0.1242 \pm 0.0539\mathrm{i}$
6-7	$0.8834 \pm 0.1322\mathrm{i}$	$0.2092 \mp 0.9014\mathrm{i}$	$-0.1459 \mp 0.0883\mathrm{i}$
8-9	$0.0395 \pm 0.0040\mathrm{i}$	$0.0289 \pm 0.0263\mathrm{i}$	$0.1530 \pm 0.1317\mathrm{i}$
10-11	$0.5930 \mp 0.0687\mathrm{i}$	$0.1979 \pm 0.6682\mathrm{i}$	$0.0076 \pm 0.1439\mathrm{i}$
12-13	$0.4549 \pm 0.0180\mathrm{i}$	$0.0538 \pm 0.1446\mathrm{i}$	$0.7757 \mp 0.0286\mathrm{i}$

Table 3
$\boldsymbol{g}$-values of the solutions of (12). The index refers to the entries row in Tables 1 and 2, which contain the corresponding values of $\boldsymbol{\lambda}$ and $\|\tilde{\boldsymbol{y}}\|$.

4.1 Locally minimal real models

It was already highlighted in Section 3 that it is not immediately obvious that all local minima of (2) are stationary points of (4), while all real models that are stationary points of the latter relaxed problem are clearly stationary points of the former. This is due to the additional constraint that restricts the 'movement' of the parameters to directions that respect the complex-conjugate structure of real models. With respect to this restricted set of directions, the gradient of the objective function may vanish, while not necessarily vanishing in general directions. To show that this is not an issue for the misfit problem (2), we first formalize what it means to be a local minimum of this problem.

Definition 4.1. A real model $(\boldsymbol{\Lambda}^*, \boldsymbol{x}^*) \in \mathcal{R}$ is a *locally minimal real model* of (2) when there exists a $\delta \in \mathbb{R}^+$ such that for any $(\boldsymbol{\Lambda}, \boldsymbol{x})$, where $\|\boldsymbol{\Lambda}^* - \boldsymbol{\Lambda}\|_{\mathrm{F}} < \delta$ and $(\boldsymbol{\Lambda}, \boldsymbol{x}) \in \mathcal{R}$, it holds that $\|\tilde{\boldsymbol{y}}^*\|_2 \leq \|\tilde{\boldsymbol{y}}\|_2$. The set of locally minimal real models will be denoted as $\mathcal{M}$.

A globally minimal real model is, of course, also a locally minimal real model as it satisfies this definition for $\delta = +\infty$.

Observe that the real local minimality of a model only imposes that no better value of the eigentuples $\boldsymbol{\Lambda}$ can be found arbitrarily close, disregarding $\boldsymbol{x}$. Indeed, recall that given $\boldsymbol{\Lambda}$, the corresponding optimal value for $\boldsymbol{x}$ of is readily obtained by solving a linear projection, essentially making $\boldsymbol{\Lambda}$ the only decision variable. However, in case $\boldsymbol{\Gamma}$ is not of full column rank, i.e. when $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_j$, this projection does not define $\boldsymbol{x}$ uniquely, though it does uniquely define the objective value. As such, to correctly handle these 'singular' cases, the local minimality of a model should only depend on the optimality of $\boldsymbol{\Lambda}$.

Example 4.1, briefly highlights why we need to discern

λ_1, λ_2	x_1, x_2	$\|\tilde{\boldsymbol{y}}\|_2^2$
$-0.5774, 1.0948$	$3.098, 4.793$	18.155
$1.2088, 15.4705$	$4.8414, -9.3275 \times 10^{-5}$	23.326
$1.0306, 1.0306$	$4.1903, 1.6417$	30.425
$-0.2463 \pm 0.9047i$	$1.4926 \mp 0.9746i$	197.55
$-2.1340, -2.1340$	$0.2510, -0.1245$	214.19

Table 4
Real models that satisfy the characterization (8) for the data $\boldsymbol{y}$ presented in Example 4.1

between locally minimal real models and the minima of the relaxed optimization problem (4). Here the globally optimal real model is a saddle point of (4), rather than a minimum. This shows that, in general, $\mathcal{M}$ are not minima of the relaxed problem.

Example 4.1. Consider the problem of fitting an order $n = 2$, 1D model to the data sequence $\boldsymbol{y}^\mathsf{T} = \begin{bmatrix} 7 & 2 & 9 & 8 & 5 \end{bmatrix}$. Up to permutation of the poles, the real models that solve the corresponding set of equations (8) are listed in Table 4.

Using the projection discussed in 3.2 to determine $\boldsymbol{x}$, Figure 3 then plots the objective $\|\tilde{\boldsymbol{y}}\|_2^2$ around the global minimum $(\lambda_1, \lambda_2) = (-0.5774, 1.0948)$. The left figure, where only real eigentuples (λ_1, λ_2) are considered, demonstrates that this point is indeed a local minimum for this parameterization. Since no models with complex-conjugate eigentuples are directly adjacent to this model, this suffices to conclude that it is a local minimum of the misfit problem (2).

However, relaxing the problem as in (4) creates additional dimensions to achieve a better fit: the complex part of (λ_1, λ_2). The righ-hand-side figure plots the objective function over eigentuples of the form $(-0.5774 + \mathrm{i}a, 1.0948 + \mathrm{i}b)$ for varying $a, b \in \mathbb{R}$. This figure clearly shows that the $(-0.5774, 1.0948)$ is not a local minimum of the relaxed problem. Indeed, it corresponds to a saddle point in this plot, and can thus be locally improved upon using such complex-valued eigentuples.

From Example 4.1 it can be concluded that, while the best-fitting real model can be found through solving (8), the global minimum of the relaxation (5) is not always found in this way. Moreover, the global minimum of (2), while a stationary point, may not even be a local minimum of this relaxation. Intuitively, this makes sense, as the absence of the realness constraint enlarges the search space and provides more search directions to improve the fit.

Notice that, using the proposed methodology, the globally minimal real model was indeed found in Example 4.1. The following theorem ensures that this is always the case, such that it makes sense to search for them amongst the stationary points $\mathcal{S}$ by solving (8) or (11).

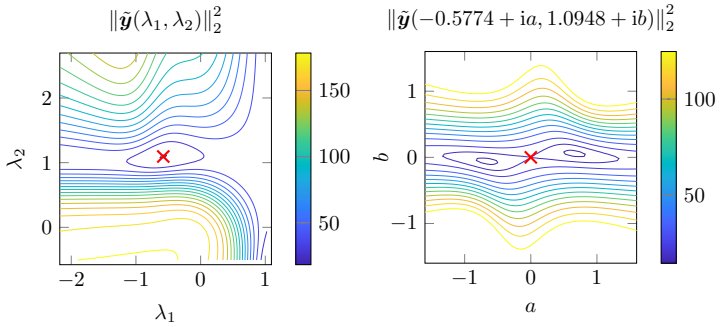


Fig. 3. Plots of the misfit $\|\tilde{\boldsymbol{y}}\|_2^2$ for varying real (left) and imaginary (right) values of (λ_1, λ_2) associated with the data $\boldsymbol{y}$ given in Example 4.1. The globally minimal real model $(-0.5774, 1.0948)$ is indicated by $\times$.

Theorem 4.1. *Any locally minimal real model satisfies* (6)*, i.e.,* $\mathcal{M} \subseteq \mathcal{S}$.

Proof. The proof is by contradiction. Suppose that $(\boldsymbol{\Lambda}^*, \boldsymbol{x}^*) \in \mathcal{M}$ does not satisfy the FONC (6), then, by abuse of notation, there exists some subset of variables $\boldsymbol{z} \subseteq (\boldsymbol{\Lambda}^*, \boldsymbol{x}^*)$ such that $\nabla_{\boldsymbol{z}} \|\tilde{\boldsymbol{y}}^*\|_2^2 \neq \boldsymbol{0}$. Due to the nonzero gradient, there then exists a direction $\Delta\boldsymbol{z}$ such that the directional derivative w.r.t $\Delta\boldsymbol{z}$ is: $\Delta\boldsymbol{z}^\mathsf{T} \nabla_{\boldsymbol{z}} \|\tilde{\boldsymbol{y}}^*\|_2^2 + \overline{\Delta\boldsymbol{z}}^\mathsf{T} \nabla_{\overline{\boldsymbol{z}}} \|\tilde{\boldsymbol{y}}^*\|_2^2 = \gamma < 0$. Also note that $\nabla_{\overline{\boldsymbol{z}}} \|\tilde{\boldsymbol{y}}^*\|_2^2 = \overline{\nabla_{\boldsymbol{z}} \|\tilde{\boldsymbol{y}}^*\|_2^2}$ [1].

Now, since $(\boldsymbol{\Lambda}^*, \boldsymbol{x}^*)$ defines a real model, there exists a set of complex-conjugate variables $\boldsymbol{w} \subseteq (\boldsymbol{\Lambda}^*, \boldsymbol{x}^*)$ such that $\bar{\boldsymbol{z}} = \boldsymbol{w}$. From (6), it can be worked out that $\nabla_{\boldsymbol{w}} \|\tilde{\boldsymbol{y}}^*\|_2^2 = \overline{\nabla_{\boldsymbol{z}} \|\tilde{\boldsymbol{y}}^*\|_2^2}$. As such, choosing $\Delta\boldsymbol{w} = \overline{\Delta\boldsymbol{z}}$ results in a directional derivative of $\Delta\boldsymbol{w}^\mathsf{T} \nabla_{\boldsymbol{w}} \|\tilde{\boldsymbol{y}}^*\|_2^2 + \overline{\Delta\boldsymbol{w}}^\mathsf{T} \nabla_{\overline{\boldsymbol{w}}} \|\tilde{\boldsymbol{y}}^*\|_2^2 + \gamma = 2\gamma < 0$ w.r.t. the direction $\Delta\boldsymbol{z} + \Delta\boldsymbol{w}$.

Consequently, this is a descent direction which retains the realness of the model. There thus exist arbitrarily small steps along this direction that improve the fit $\|\tilde{\boldsymbol{y}}\|_2^2$, contradicting its local minimality. □

4.2 Model order of real minima

While Theorem 4.1 shows that $\mathcal{M}$ can be found as solutions of the FONC (6), it remains to be shown that the x_i are nonzero, as was assumed in the derivation of (8). Moreover, it also remains unclear whether the condition of distinct eigentuples posed in Theorem 3.2 is satisfied by such a model. The following theorem states that both these conditions are indeed met for locally minimal real models $\mathcal{M}$. This finally proves that all locally minimal real models, including the globally minimal one, can be found by solving (8) or, by Theorem 3.2, solving (11).

Theorem 4.2. *Any* $(\boldsymbol{\Lambda}^*, \boldsymbol{x}^*) \in \mathcal{M}$ *for which* $\|\tilde{\boldsymbol{y}}^*\| \neq 0$ *satisfies:* $\forall i \in \{1, \ldots, n\} : x_i \neq 0$ *and* $\forall i, j \in \{1, \ldots, n\}$ *with* $i \neq j : \boldsymbol{\lambda}_i \neq \boldsymbol{\lambda}_j$.

Proof. The proof is by contradiction. Note that, by Definition 4.1, to contradict the local real minimality of the model, it suffices to show that there always exists a better choice of $\boldsymbol{\Lambda}$ arbitrarily close to $\boldsymbol{\Lambda}^*$, while $\boldsymbol{x}$ can be arbitrarily far from $\boldsymbol{x}^*$.

Suppose that for some i it holds that $x_i^* = 0$ and define the Vandermonde vector $\boldsymbol{v}(\boldsymbol{z}) \in \mathbb{C}^N[\boldsymbol{z}]$ as $\boldsymbol{v}_{\boldsymbol{\kappa}}(\boldsymbol{z}) = \boldsymbol{z}^{\boldsymbol{\kappa}}$. Consider the polynomial $\tilde{y}^*(\boldsymbol{z}) = (\tilde{\boldsymbol{y}}^*)^{\mathsf{T}}\boldsymbol{v}(\boldsymbol{z}) = \sum_{\boldsymbol{\kappa}\in\mathcal{K}} \tilde{\boldsymbol{y}}_{\boldsymbol{\kappa}}\boldsymbol{z}^{\boldsymbol{\kappa}}$, which is nonzero since since $\tilde{\boldsymbol{y}}^* \neq \boldsymbol{0}$. Akin to the interpolation condition (10), it holds that $0 = \tilde{y}^*(\boldsymbol{\lambda}_i^*)$, as the real model satisfies (7a). There exists an arbitrarily small, real perturbation $\boldsymbol{\epsilon} \in \mathbb{R}^m$ so that $\tilde{y}^*(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}) \neq 0$.

For the sake of brevity, we presume that $\mathrm{Re}(\tilde{y}^*(\boldsymbol{\lambda}_i^*+\boldsymbol{\epsilon})) \neq 0$. The argument can be easily adjusted to cover the case where only $\mathrm{Im}(\tilde{y}^*(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon})) \neq 0$.

Since $(\boldsymbol{\Lambda}^*, \boldsymbol{x}^*)$ defines a real model, there is a complex-conjugate eigentuple to $\boldsymbol{\lambda}_i$, that is, $\exists j : x_j = 0$ and $(\boldsymbol{\lambda}_i^*, x_i^*) = \overline{(\boldsymbol{\lambda}_j^*, x_j^*)}$. Then consider $\tilde{y}^*(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}) + \tilde{y}^*(\boldsymbol{\lambda}_j^* + \overline{\boldsymbol{\epsilon}}) = \tilde{y}^*(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}) + \overline{\tilde{y}^*(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon})} = 2\mathrm{Re}(\tilde{y}^*(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}))$. When $\mathrm{Re}(\tilde{y}^*(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon})) \neq 0$, the component along $\mathrm{Re}(\boldsymbol{v}(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}))$ can be projected away so that: $\|\tilde{\boldsymbol{y}}^{\mathrm{new}}\|_2^2 =$

$$\left\| \tilde{\boldsymbol{y}}^* - 2\mathrm{Re}(\tilde{y}^*(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon})) \frac{\mathrm{Re}(\boldsymbol{v}(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}))}{\|\mathrm{Re}(\boldsymbol{v}(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}))\|_2^2} \right\|_2^2 < \|\tilde{\boldsymbol{y}}^*\|_2^2.$$

It can be verified that $\|\tilde{\boldsymbol{y}}^{\mathrm{new}}\|_2^2$ is attained by the real model:

$$\boldsymbol{\lambda}_i^{\mathrm{new}} = \boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}, \quad x_i^{\mathrm{new}} = \frac{\mathrm{Re}(\tilde{y}^*(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}))}{\|\mathrm{Re}(\boldsymbol{v}(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}))\|_2^2}$$
$$\boldsymbol{\lambda}_j^{\mathrm{new}} = \boldsymbol{\lambda}_j^* + \boldsymbol{\epsilon}, \quad x_j^{\mathrm{new}} = \frac{\mathrm{Re}(\tilde{y}^*(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}))}{\|\mathrm{Re}(\boldsymbol{v}(\boldsymbol{\lambda}_i^* + \boldsymbol{\epsilon}))\|_2^2},$$

while keeping other model parameters the same as in $(\boldsymbol{\Lambda}^*, \boldsymbol{x}^*)$. Since $\|\boldsymbol{\Lambda}^{\mathrm{new}} - \boldsymbol{\Lambda}^*\|_{\mathrm{F}}^2 = 2\|\boldsymbol{\epsilon}\|^2$ can be arbitrarily small, this contradicts that $(\boldsymbol{\Lambda}^*, \boldsymbol{x}^*) \in \mathcal{M}$.

Lastly, suppose $\exists i, j$ with $i \neq j$ for which $\boldsymbol{\lambda}_i^* = \boldsymbol{\lambda}_j^*$. This case can be reduced to an equivalent model with $x_i = 0$: Let (s, t) be such that $\overline{(\boldsymbol{\lambda}_i^*, x_i)} = (\boldsymbol{\lambda}_s^*, x_s)$ and $\overline{(\boldsymbol{\lambda}_j^*, x_j)} = (\boldsymbol{\lambda}_t^*, x_t)$, and choose $x_j^{\mathrm{new}} = x_j^* + x_i^*$ and $x_i^{\mathrm{new}} = 0$ and $x_t^{\mathrm{new}} = x_t^* + x_s^*$ and $x_s^{\mathrm{new}} = 0$, while keeping the rest of the parameters the same. By the first part of the proof, this then contradicts the local minimality of the model.

□

Notice that a locally minimal real model thus has n nonzero contributions of n distinct eigentuples, ensuring that a model of order at least n is needed to describe the corresponding model-compliant data sequence $\hat{\boldsymbol{y}}$. Informally, one could state that the model has *nth order dynamics*. Indeed, any model that does not meet these restrictions can be equivalently described using a model of order at most $n-1$. Either remove the eigentuple with no contribution, i.e. for which $x_i = 0$, or merge the contributions of a pair of equal eigentuples, as was done in the proof of Theorem 4.2. This proof also indicates that such lower-order models are saddle points, in the sense that there is always a better choice of $(\boldsymbol{\Lambda}, \boldsymbol{x})$ arbitrarily closeby.

Consequently, such models can just as well be found by solving (8) for a lower model order. Indeed, substituting $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_j$ into these equations results in a set of equations equivalent to the $(n-1)$th order version of (8), so that optimal solutions to a lower-order problem 'carry over' to saddle points of the nth order problem.

Remark: Since a perfect fit, i.e. $\tilde{\boldsymbol{y}} = \boldsymbol{0}$, cannot be improved upon, posing a problem for a proof by contradiction, this case is excluded in theorem 4.2. In case a real model with $\tilde{\boldsymbol{y}} = \boldsymbol{0}$ exists and has *nth order dynamics*, it still satisfies both (8) and (11). If not, it can be equivalently described using a lower model order, so that at least one equivalent model appears as a solution to (8).

For (11), adding any eigentuple $\boldsymbol{\lambda}_n$ to a $(n-1)$th order model ensures that, by Theorems 2.1 and 2.2, $\hat{\boldsymbol{y}}$ still lies in the kernel of the corresponding Macaulay matrix. Since $\tilde{\boldsymbol{y}} = \boldsymbol{0}$ lies in the row space of any Macaulay matrix, choosing $\boldsymbol{g} = \boldsymbol{0}$ then provides a solution to the higher order problem as well. As such, any globally minimal real model is always a solution to both characterizations.

4.3 Real, local minima are isolated

In case of a continuum of solutions, numerical polynomial solvers, of course, only compute a finite number of 'representatives', called a *witness set*, for each irreducible part of the solution set [8]. While we now know that $\mathcal{M}$ models are indeed solutions to our polynomial characterizations, this does not guarantee that they do not lie on a continuum of complex solutions, for which such a witness set can be provided at best. This set may not comprise real models, while the continuum may contain the real model of interest. Thus, to ensure that this is not the case so that the models of interest can actually be found using polynomial root finding techniques, it remains to be shown that they are isolated solutions. Theorem 4.3 shows that, for *almost all* $\boldsymbol{y}$, this is indeed the case.

Theorem 4.3. *There exists a dense, open subset $\mathcal{O} \subseteq \mathbb{C}^N$, such that for $\boldsymbol{y} \in \mathcal{O}$, any $(\boldsymbol{\Lambda}, \boldsymbol{x}) \in \mathcal{M}$ is an isolated solution of* (8)*. Secondly, when $\mathcal{K}$ meets the condition of Theorem 3.2, the corresponding $\boldsymbol{\Lambda}$ are isolated points of the set $\{\boldsymbol{\Lambda} \in \mathbb{C}^{n\times m} \mid \exists \boldsymbol{g} \in \mathbb{C}^q : (\boldsymbol{\Lambda}, \boldsymbol{g})$ satisfies* (11)*$\}$.*

Proof. To show the first statement, it suffices to show that for a dense, open subset of $\boldsymbol{y}$-values, the Jacobian of (8) has a nullity of zero in the solutions with distinct eigentuples and nonzero x_i, as this upper bounds the local dimension of the solution set [27]. By Theorem 4.2, these solutions include $\mathcal{M}$.

First, we eliminate the solutions where $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_j$ or $x_i = 0$ by applying Rabinowitsch's trick. Let $\mathbf{0} \neq \boldsymbol{v} \in \mathbb{C}^m$ and introduce the new variables u_{ij} and w_i. Then, by abuse of notation, define the set of equations $\boldsymbol{c} = \{\forall i, j; i \neq j : 1 - u_{ij}\boldsymbol{v}^\mathsf{T}(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_j) = 0\} \cup \{\forall i : 1 - w_i x_i = 0\}$ and construct the extended the set of equations: $\boldsymbol{p} = \{(8)\} \cup \boldsymbol{c}$. For every solution $(\boldsymbol{\Lambda}, \boldsymbol{x}, \boldsymbol{y})$ of (8) with $x_i \neq 0$ and $\boldsymbol{v}^\mathsf{T}(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_j) \neq 0$, there then exist unique values for u_{ij} and w_i that solve $\boldsymbol{p}$. Conversely, any solution of $\boldsymbol{p}$ implies a corresponding solution of (8) by construction. Now considering $\boldsymbol{y}$ as variables, this allows us to define the variety $V = \{\boldsymbol{z} = (\boldsymbol{\Lambda}, \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}, \boldsymbol{y}) \in \mathbb{C}^t \mid \boldsymbol{p}(\boldsymbol{z}) = \mathbf{0}\}$.

Next, consider a model $(\boldsymbol{\Lambda}^*, \boldsymbol{x}^*)$ with $x_i^* \neq 0$ and distinct eigentuples. Selecting $\boldsymbol{y}^* = \boldsymbol{\Gamma}^* \boldsymbol{x}^*$, there then exist values $\boldsymbol{u}^*, \boldsymbol{w}^*$ so that $\boldsymbol{z}^* = (\boldsymbol{\Lambda}^*, \boldsymbol{x}^*, \boldsymbol{u}^*, \boldsymbol{w}^*, \boldsymbol{y}^*) \in V$. Then, the Jacobian $\boldsymbol{J}_{\boldsymbol{p}}^*$ in $\boldsymbol{z}^*$ of $\boldsymbol{p}$ is of the form:

$$\boldsymbol{J}_{\boldsymbol{p}}^* = \begin{array}{c} \begin{matrix} \nabla_{(\boldsymbol{\Lambda},\boldsymbol{x})} & \nabla_{(\boldsymbol{u},\boldsymbol{w})} & \nabla_{\boldsymbol{y}} \end{matrix} \\ \begin{bmatrix} \boldsymbol{J}_1^* & \mathbf{0} & (\boldsymbol{G}^*)^\mathsf{T} \\ \boldsymbol{B}_1 & \boldsymbol{D} & \boldsymbol{B}_2 \end{bmatrix} \end{array} \begin{matrix} \\ \boldsymbol{J}_{(8)} \\ \boldsymbol{J}_{\boldsymbol{c}} \end{matrix},$$

where $\boldsymbol{J}_1^* = (\boldsymbol{G}^*)^\mathsf{T} \left[\boldsymbol{\Gamma}^* \left(\boldsymbol{\Gamma}^{(\boldsymbol{\lambda})}\right)^* (\boldsymbol{I}_m \otimes \operatorname{diag}(\boldsymbol{x}^*))\right] = (\boldsymbol{G}^*)^\mathsf{T} \boldsymbol{H}$ and $\boldsymbol{D}$ is diagonal and full rank.

Remark that for any $\boldsymbol{z} \in V$, the corresponding $\boldsymbol{G}$ is of full column rank, since its columns span the null space of[8] $\boldsymbol{M}_{\boldsymbol{r}^2}$, which has a nullity of $(m+1)n$. Since the columns of $\boldsymbol{H}^*$ are scaled versions of the columns of $\boldsymbol{G}^*$, $\boldsymbol{J}_1^*$ is of full rank. Together with $\boldsymbol{D}$ being invertible, this implies that $\boldsymbol{J}_{\boldsymbol{p}}^*$ is of full row-rank as well.

From the implicit function theorem it now follows that there exists an open subset $U \subseteq \mathbb{C}^N$ around $\boldsymbol{y}^*$ and a function $\boldsymbol{f}$ such that $\forall \boldsymbol{y} \in U : (\boldsymbol{\Lambda}, \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}) = \boldsymbol{f}(\boldsymbol{y})$ and $\boldsymbol{p}(\boldsymbol{f}(\boldsymbol{y}), \boldsymbol{y}) = \mathbf{0}$. This implies that $\{(\boldsymbol{f}(\boldsymbol{y}), \boldsymbol{y}) \mid \boldsymbol{y} \in U\} \subseteq V$ which is thus N-dimensional at $\boldsymbol{z}^*$.

Now define the morphism $\pi : V \to \pi(V) : \boldsymbol{z} \mapsto \boldsymbol{y}$. As $U \subseteq \pi(V)$, it is N-dimensional. By the closure theorem [10], its closure $\overline{\pi(V)}$ then is an N-dimensional variety in $\mathbb{C}^N$, implying $\overline{\pi(V)} = \mathbb{C}^N$ and that $\pi(V)$ is a dense, open subset of[9] $\mathbb{C}^N$.

Finally, the Jacobian $\boldsymbol{J}_1$ of (8) w.r.t the model parameters $(\boldsymbol{\Lambda}, \boldsymbol{x})$ is a square, polynomial matrix in $(\boldsymbol{\Lambda}, \boldsymbol{x}, \boldsymbol{y})$, so that its determinant is a polynomial $\delta(\boldsymbol{z})$ which defines the variety $W = \{\boldsymbol{z} \in \mathbb{C}^t \mid \delta(\boldsymbol{z}) = 0\}$. Since in $\boldsymbol{z}^*$, the Jacobian $\boldsymbol{J}_1^*$ is of full rank, it holds that $V \not\subseteq W$, so that $V \cap W$ is a variety of dimension $d < N$. Thus, dim $\pi(V \cap W) < N$, implying that it is a closed subset of $\overline{\pi(V)}$, so that its complement is a dense, open subset of $\overline{\pi(V)} = \mathbb{C}^N$, where this Jacobian has a nullity of zero.

By Theorem 3.2, the second statement follows directly from the first statement. The solutions of (11) are not isolated in $\boldsymbol{f}$, however, since $\boldsymbol{M}_{\boldsymbol{r}^2}$ generally has a non-trivial null space. □

4.4 Limitations

To conclude this section, we highlight a key limitation of focusing on the stationary points of the optimization problem. It has been shown that all locally minimal real models can be found through the proposed approach, and, clearly, a global minimun is necessarily a local one. However, keep in mind that a globally minimal real model may not exist at all. In such cases, solving (8) or (11) only leads to the locally minimal real models, which may still be improved upon by another real model, as illustrated in Example 4.2 below. To detect such cases, a sum-of-squares decomposition [6,37] can be employed to get a tight lower bound polynomial optimization problems.

Example 4.2. Given $\boldsymbol{y} = \begin{bmatrix} 0.1 & 0 & 0 & 1 \end{bmatrix}^\mathsf{T}$, consider the problem of fitting a first-order 1D model this data. The objective function is then:

$$\|\tilde{\boldsymbol{y}}\|^2 = |0.1 - x|^2 + |\lambda x|^2 + \left|\lambda^2 x\right|^2 + \left|1 - \lambda^3 x\right|^2.$$

Choosing $x = \frac{1}{\lambda^3}$ and taking the limit then results in $\lim_{\lambda \to \pm\infty} \left\|\tilde{\boldsymbol{y}}(\lambda, \frac{1}{\lambda^3})\right\|^2 = 0.01$. This provides an upper bound on the infimum of $\|\tilde{\boldsymbol{y}}\|^2$. If a better fit could be obtained for a finite, real model (λ, x), there would have to be a local real minimum corresponding to such a point. However, solving (8), yields two real models that do not improve upon this bound.

λ	x	$\|\tilde{\boldsymbol{y}}\|^2$
0.0334	0.0999	1.0000
0.0000	0.1000	1.0000

[8] Apply Lemma 2.2: $\boldsymbol{r}^2$ has precisely $(m+1)n$ solutions, counting multiplicities, and the degree of regularity is exceeded.

[9] Note that for any $\boldsymbol{y} \in \pi(V)$ there is at least one corresponding point in V, so that there exists a solution (8) that satisfies the assumptions on $(\boldsymbol{\Lambda}, \boldsymbol{x})$.

Even though, by Theorem 4.1, these models include all locally minimal real models, neither of them define a globally minimal real model, as there exist finite parameter values leading to a lower objective function, e.g. $\lambda = 1, x = 0.1$, leading to $\|\tilde{\boldsymbol{y}}\|^2 = 0.83$.

5 Lower-order models and infinite solutions

A polynomial characterization of the local minima is desirable because there exists a variety of numerical algorithms that can obtain all solutions to such problems. Broadly speaking, there are two classes of polynomial solvers: algebraic methods [26,15,56] and homotopy-based methods [8,64]. We now illustrate the infinite number of *lower-order* solutions that occur in the presented polynomial-based characterizations. This observation limits the choice of viable solvers, as currently, only homotopy-based root finding methods can handle infinite solution sets. Remark that, at the cost of additional variables and the entailing computational load, these infinite solution components may be removed using Rabinowitsch's trick mentioned in the proof of Theorem 4.3.

Example 3.1 already highlighted how an infinite number of solutions with $x_i = 0$ is excluded by leveraging Theorem 4.2 to simplify the equations. Moreover, excluding non-interesting solutions in this way is of interest as it reduces the computational load of solving the equations. However, the solution sets of (8) and (11) do still contain other instances of lower-order models: those for which $\exists i, j, i \neq j : \boldsymbol{\lambda}_i = \boldsymbol{\lambda}_j$.

Let us first take a closer look at how these lower-order models appear as solutions to (8). In case $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_j, i \neq j$, $\boldsymbol{\Gamma}$ is not of full column rank and does not define a unique optimum for $\boldsymbol{x}$ through the projection discussed in Section 3.2. Instead, the optimal value for $\boldsymbol{x}$ is only definined up to null $\boldsymbol{\Gamma}$. Furthermore, such solutions always exist: Each solution of the order $n-1$ problem (8) can be extended to a solution of the corresponding nth order problem by choosing [10] $\boldsymbol{\lambda}_n = \boldsymbol{\lambda}_i$ and $x_i = 0$ for some i, thus inducing an infinite number of equivalent solutions in the higher-order problem. This observation is illustrated in Example 5.1 below.

Example 5.1. Consider the problem of fitting a second order, 1D model to the following data:

$$\boldsymbol{y}^\mathsf{T} = \begin{bmatrix} 1 & \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} & -1 & -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

Now, substitute in a lower-order solution of the form $\lambda_1 = \lambda_2 = 0.2984$ into (8), resulting in:

$$\boldsymbol{G}\boldsymbol{\Gamma}\boldsymbol{x} - \boldsymbol{G}\boldsymbol{y} = \begin{bmatrix} 1.0977 & 1.0977 & 1.1828 \\ 1.0977 & 1.0977 & 1.1828 \\ 0.3596 & 0.3596 & 0.3874 \\ 0.3596 & 0.3596 & 0.3874 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ -1 \end{bmatrix} = \boldsymbol{0}.$$

The kernel of the associated observability matrix $\boldsymbol{\Gamma}$ is spanned by $\begin{bmatrix} 1 & -1 \end{bmatrix}$. As such, all solutions $\boldsymbol{x}$ to these singular, linear equations can thus be described as $\boldsymbol{x}^\mathsf{T} = \begin{bmatrix} 1.0774 & 0 \end{bmatrix} + c \begin{bmatrix} 1 & -1 \end{bmatrix}, c \in \mathbb{C}$, leading to a one-dimensional set of equivalent lower-order solutions for (8).

The characterization (11), on the other hand, relies on describing the misfit as $\tilde{\boldsymbol{y}} = \boldsymbol{M}_{\boldsymbol{r}^2}^\mathsf{T}\boldsymbol{g}$. Here, the coefficient matrix may also have a nontrivial kernel, again causing the auxiliary variables $\boldsymbol{g}$ to be defined non-uniquely by this equation. Moreover, the rows of a Macaulay matrix, and, by extension, $\boldsymbol{M}_{\boldsymbol{r}^2}$, do generally not form a basis [5], regardless of $\boldsymbol{\Lambda}$. Consequently, for most problems, all solutions to (11) entail an infinite number of equivalent values of $\boldsymbol{g}$.

However, assuming distinct eigentuples, Lemma 3.1, implies that the rank of $\boldsymbol{M}_{\boldsymbol{r}^2} \in \mathbb{R}^{p \times N}$ is $N - n(m+1)$. As such, one could heuristically attempt to construct a basis by taking as many linearly independent row combinations, that is, by replacing it with $\boldsymbol{T}\boldsymbol{M}_{\boldsymbol{r}^2}$, where $\boldsymbol{T} \in \mathbb{R}^{N-n(m+1)\times q}$ is some matrix of full row-rank, in (11). For example, one may choose $\boldsymbol{T}$ to be a random matrix. For almost all $\boldsymbol{\Lambda}$, this eliminates the non-uniqueness of $\boldsymbol{g}$ for solutions with distinct eigentuples.

While this approach does eliminate a large number of equivalent models, lower-order solutions with $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_j, i \neq j$ again induce an infinite number of solutions. Indeed, such models cause $\boldsymbol{\lambda}_i$ to become a common root of multiplicity $m+1$ in the GCH equations $\boldsymbol{r}(\boldsymbol{\sigma})$ [61]. Using a similar anlysis as in the proof of Theorem 3.1, its multiplicity as a root of $\boldsymbol{r}^2(\boldsymbol{\sigma})$ then rises to $\binom{m+3}{3}$, increasing the nullity of $(\boldsymbol{T})\boldsymbol{M}_{\boldsymbol{r}^2}$. This rank drop of the coefficient matrix then again leads to an infinite number of equivalent $\boldsymbol{g}$-values.

Remark: In the one-dimensional case, the polynomial characterizations (8) and (11) can be reformulated as a *multiparameter eigenvalue problem* (MEP), which exploits the linearity of the auxiliary variables [12,36]. Such formulations are possible for the mD case as well, which is especially interesting for the kernel-based characterization (11) due to its larger number of auxiliary variables. However, the resulting MEP's also have an infinite number of lower-order solutions, which current numerical MEP solvers cannot handle [46,62].

[10] It can be verified that this choice of eigentuple only generates additional equations that were already present in the lower-order problem.

6 Numerical experiments

In this section, some more elaborate examples are considered to further illustrate the efficacy of the proposed identification approach.

Example 6.1. Consider the problem of fitting a second-order ($n = 2$) model to the randomly generated 2D data sequence $y_{\boldsymbol{\kappa}}$ given below.

$\boldsymbol{\kappa}$	$y_{\boldsymbol{\kappa}}$	$\boldsymbol{\kappa}$	$y_{\boldsymbol{\kappa}}$	$\boldsymbol{\kappa}$	$y_{\boldsymbol{\kappa}}$
$(0,0)$	1.88	$(1,0)$	5.41	$(0,1)$	1.22
$(2,0)$	6.02	$(1,1)$	0.16	$(0,2)$	5.54
$(3,0)$	6.56	$(2,1)$	8.57	$(1,2)$	4.49
$(0,3)$	8.55	$(4,0)$	8.01	$(3,1)$	3.79
$(2,2)$	5.11	$(1,3)$	0.28	$(0,4)$	2.88
$(5,0)$	0.61	$(4,1)$	9.33	$(3,2)$	9.01
$(2,3)$	9.12	$(1,4)$	5.89	$(0,5)$	6.03
$(6,0)$	5.79	$(5,1)$	5.99	$(4,2)$	4.96
$(3,3)$	3.44	$(2,4)$	2.66	$(1,5)$	6.13
$(0,6)$	5.12	$(7,0)$	5.10	$(6,1)$	9.82
$(5,2)$	6.09	$(4,3)$	1.21	$(3,4)$	9.92
$(2,5)$	2.62	$(1,6)$	9.69	$(0,7)$	5.95

Solving (8) for this data sequence using *HomotopyContinuation.jl* [8], results in 69294 complex solutions, of which 75 are real models, of which 29 have real-valued eigentuples and 46 have complex-conjugate eigentuples. For the sake of brevity, only the ten best-fitting real models are presented in Table 5.

Note that the parameterization of $\hat{\boldsymbol{y}}$ and thus the equations are symmetric in $(\boldsymbol{\lambda}_1, x_1)$ and $(\boldsymbol{\lambda}_2, x_2)$, that is, exchanging the contributions as follows: the model $(\boldsymbol{\lambda}_1^{\text{new}}, x_1^{\text{new}}) = (\boldsymbol{\lambda}_2, x_2)$ and $(\boldsymbol{\lambda}_2^{\text{new}}, x_2^{\text{new}}) = (\boldsymbol{\lambda}_1, x_1)$ has the same $\hat{\boldsymbol{y}}$ and thus the same objective function. Consequently, only one of these permutations is provided in Table 5 for each pair of solutions. Under this symmetry, there may still be an uneven number of total solutions, as is the case here, since cases where $(\boldsymbol{\lambda}_1, x_1) = (\boldsymbol{\lambda}_2, x_2)$ cannot be paired with another solution.

While the corresponding kernel formulation (11) cannot be readily tackled by HomotopyContinuation.jl due to the large number of variables, as briefly discussed in Section 5, the claims of Theorem 3.2 can be verified by checking whether the solutions in Table 5 also solve this formulation. Similarly to the discussion in Section 3.2, once $\boldsymbol{\Lambda}$ is known, (11) boils down to solving a linear system for $\boldsymbol{g}$. Obtaining $\boldsymbol{g}$ in such a way for each of the models in Table 5, the corresponding residual norms of (11) can be obtained, the largest of which is 2.388×10^{-10}.

$\boldsymbol{\lambda}_1$	$\boldsymbol{\lambda}_2$	$\|\tilde{\boldsymbol{y}}\|^2$
$(1.062, 1.040)$	$(-0.841, -1.035)$	232.14
$(1.072, 1.076)$	$(-6.557, 8.087)$	241.90
$(1.086, 1.064)$	$(-1.229, 0.565)$	246.17
$(0.504, 0.721)$	$(0.977, 0.987)$	254.23
$(1.838, -0.545)$	$(1.096, 1.052)$	256.34
$(-19.10, 6.251)$	$(1.087, 1.061)$	261.33
$(1.069, 1.055)$	$(-0.436, -0.214)$	262.55
$(1.075, 1.028)$	$(1.063, 1.810)$	265.12
$(1.412, 2.672)$	$(1.073, 1.037)$	265.15
$(1.855, 1.010)$	$(1.102, 1.059)$	265.47

Table 5
Ten best-fitting real models that solve (8) for the problem presented in Example 6.1. The $\boldsymbol{x}$-values can be deduced from the projection discussed in Section 3.2.

Finally, consider the problem of fitting a first-order model to the same data. This results in a best-fitting real model of $\boldsymbol{\lambda}_1 = (1.075, 1.062), x_1 = 3.906$, with a fit of $\|\tilde{\boldsymbol{y}}\|^2 = 266.58$. It can be verified that this model also appears as a lower-order solution to the original second-order equations (8). Indeed, plugging in the second-order model with $\boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_1$ and $x_2 = 0$ into said equations leads to a residual norm of 6.173×10^{-13}. Notice that increasing the model order causes the residual to strictly decrease. Indeed, as argumented in the proof of Theorem 4.2, whenever $\tilde{\boldsymbol{y}} \neq 0$, there always exists an additional eigentuple that results in a better fit, causing the locally minimal real models to be models of precisely order n.

Example 6.2. Consider data of the form $y_{\boldsymbol{\kappa}} = (0.99 + 0.16\text{i}, 0.95 + 0.31\text{i})^{\boldsymbol{\kappa}} + (0.99 - 0.16\text{i}, 0.95 - 0.31\text{i})^{\boldsymbol{\kappa}} + x_3(0.8, 0.9)^{\boldsymbol{\kappa}}$ for $\boldsymbol{\kappa} \in \mathbb{N}_0^2, \|\boldsymbol{\kappa}\|_1 \leq 7$ and for varying values of $x_1 \in \{0.1, 1, 10, 100\}$. We now fit a second order $n = 2$ model to this data. The results of solving (8) and filtering out the best-fitting real models for each x_3 is are presented in Table 6.

Remark that $\boldsymbol{y}$ is model-compliant with a third-order model so that for $x_3 = 0.1$, the best-fitting second order model approximates the 'dominant' eigentuples $(0.99 \pm 0.16\text{i}, 0.95 \pm +0.31\text{i})$ quite well. As x_3 increases, the contributions of these eigentuples become less dominant, as reflected by the fitted 2D eigentuples straying further. For $x_3 = 100$ the third eigentuple $(0.8, 0.9)$ becomes dominant and is decently approximated by the fitted eigentuples.

Example 6.3. Consider the 35-point data 3D data sequence $\forall \boldsymbol{\kappa} \in \mathbb{N}_0^m, \|\boldsymbol{\kappa}\|_1 \leq 4 : \boldsymbol{y} = 0.76(0.95, 0.12, 0.82,)^{\boldsymbol{\kappa}} + 0.92(0.07, 0.57, 0.69)^{\boldsymbol{\kappa}}$. Solving the corresponding equations (8) using $n = 1$ results in four real models, as shown in Table 7. As discussed in Section 4, these models include all real local minima, of which $\boldsymbol{\lambda}_1 = (0.681, 0.307, 0.763)$ achieves the best fit.

$\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2$	x_3	$\|\tilde{\boldsymbol{y}}\|^2$
$(0.985 \mp \mathrm{i}0.159, 0.947 \mp \mathrm{i}0.304)$	0.1	5.23×10^{-4}
$(0.949 \pm \mathrm{i}0.150, 0.935 \pm \mathrm{i}0.260)$	1	3.36×10^{-2}
$(0.868 \pm \mathrm{i}0.085, 0.940 \pm \mathrm{i}0.121)$	10	3.17×10^{-1}
$(0.821, 0.922), (0.866, 0.982)$	100	4.74×10^{-1}

Table 6
Best-fitting real models that solve (8) as x_3 increases for the problems presented in Example 6.2 The $\boldsymbol{x}$-values can be defined through the projection discussed in Section 3.2.

$\boldsymbol{\lambda}_1$	x_1	$\|\tilde{\boldsymbol{y}}\|^2$
$(0.681, 0.307, 0.763)$	1.561	0.495
$(-1.132, 0.117, -0.909)$	0.126	10.39
$(1.232, -0.979, -2.175)$	8.412×10^{-3}	10.60
$(-1.127, -4.442, 1.405)$	0.127×10^{-4}	10.68

Table 7
All real models that solve (8) for the problem in Example 6.3.

7 Conclusion and future work

This work presents previously unknown properties of the local minima of the least squares misfit identification problem for multidimensional, autonomous, scalar, linear, shift-invariant models, showing that they are also stationary points of a relaxed misfit problem and that their dynamics cannot exhibit a 'lower model order'. With these observations in mind, a novel polynomial characterization of the local minima is obtained through careful manipulation of the FONC. This leads to a multidimensional extension of the Walsh-like property previously observed for LTI systems [12,50]. By duality between the kernel and image representation, this then leads to an alternative characterization based on the kernel representation, akin to [36,35].

These polynomial characterizations allow the local minima to be found through off-the-shelf polynomial root finding techniques, extending the approach for globally optimal identification pioneered in [12] to the multidimensional case. Moreover, we provide a rigorous theoretical footing for the assumptions made in the latter paper, such as the isolatedness of the solutions. Due to the large computational overhead entailed by this strife for global optimality, the contributions are of a predominantly theoretical nature, offering new insights into the problem's stationary points. This may, for example, be employed to study other, more heuristic algorithms for the problem at hand or to further tailor such methods to the problem.

As this demonstrates that global optimality can be achieved for certain classes of models, we aim to further extend this approach to models including inputs and to models with multiple outputs in the future, as well as to other nonlinear polynomial models. Additionally it may be interesting to investigate alternative parameterizations for autonomous, scalar, linear, shift-invariant models, to eliminate the issues presented by *lower-order* solutions in the current approach.

Computationally, the structure of the polynomial equations may be leveraged to reduce the computational load, most notably the symmetry of the equations w.r.t. the eigentuples can most likely be exploited for this purpose. Applying algorithms that can tackle *multiparameter eigenvalue problems* with an infinite number of solutions may also result in a speedup, as this allows one to exploit the linearity of the auxiliary variables, as was done in the 1D case [36,12].

Acknowledgements

This work was supported in part by the KU Leuven Research Fund (projects iBOF/24/064 and IDN/24/013), the Industrial Research Fund and several LRD bilateral industrial projects; in part by the Flemish Government agencies: FWO (TBM T001725N), EWI through the Flanders AI Research Program and VLAIO (CSBO - HBC.2021.0076); and in part by the European Commission (ERC Adv. Grant) under Grant 885682. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or ERC. Neither the European Union nor the granting authority can be held responsible for them. The work of S. Lagauw was supported by the FWO PhD Fellowship Fundamental Research under Grant 11K5625N.

References

[1] T. Adali and P.J. Schreier. Optimization and estimation of complex-valued signals: Theory and applications in filtering and blind source separation. *IEEE Signal Processing Magazine*, 31(5):112–128, 2014.

[2] S. Attasi. Modelling and recursive estimation for double indexed sequences. In *System Identification Advances and Case Studies*, volume 126, pages 289–348. Elsevier, 1976.

[3] J. Ball and V. Vinnikov. Overdetermined multidimensional systems: State space and frequency domain methods. *Mathematical Systems Theory in Biology, Communications, Computation, and Finance*, pages 63–119, 2003.

[4] K. Batselier, P. Dreesen, and B. De Moor. Prediction error method identification is an eigenvalue problem. *IFAC Proceedings Volumes*, 45(16):221–226, 2012. 16th IFAC Symposium on System Identification.

[5] K. Batselier, P. Dreesen, and B. De Moor. On the null spaces of the Macaulay matrix. *Linear Algebra and its Applications*, 460:259–289, 2014.

[6] G. Blekherman, P. A. Parrilo, and R. R. Thomas. *Semidefinite Optimization and Convex Algebraic Geometry*, volume 13. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013.

[7] N. K. Bose. *Applied Multidimensional Systems Theory*. Springer International Publishing, 2017.

[8] P. Breiding and S. Timme. Homotopycontinuation.jl: A package for homotopy continuation in julia: 6th international conference, south bend, in, usa, july 24-27, 2018, proc., 2018.

[9] D. Coca and S.A. Billings. Identification of finite dimensional models of infinite dimensional dynamical systems. *Automatica*, 38(11):1851–1865, 2002.

[10] D A Cox, J B Little, and D O'Shea. *Ideals, Varieties and Algorithms*. Springer-Verlag, 2007.

[11] B. H. Dayton and Z. Zeng. Computing the multiplicity structure in solving polynomial systems. *Proc. of the International Symposium on Symbolic and Algebraic Computation*, pages 116–123, 2005.

[12] B. De Moor. Least squares optimal realisation of autonomous LTI systems is an eigenvalue problem. *Communications in Information and Systems*, 20(2):163–207, 2020.

[13] M. Dillabough, H. Shang, and P. J. McLellan. A state space approach for boundary control of distributed parameter systems. *IFAC Proceedings Volumes*, 39(2):341–346, 2006. 6th IFAC Symposium on Advanced Control of Chemical Processes.

[14] T. Ding, M. Sznaier, and O. Camps. Robust identification of 2-d periodic systems with applications to texture synthesis and classification. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 3678–3683, 2006.

[15] P. Dreesen, K. Batselier, and B. De Moor. Multidimensional realisation theory and polynomial system solving. *International Journal of Control*, 91(12):2692–2704, 2018.

[16] M. N. El Korso, R. Boyer, A. Renaux, and S. Marcos. Statistical resolution limit for the multidimensional harmonic retrieval model: hypothesis test and cramér-rao bound approaches. *EURASIP Journal on Advances in Signal Processing*, 2011(1):12, 2011.

[17] M. Farah, G. Mercère, R. Ouvrard, and T. Poinot. Combining least-squares and gradient-based algorithms for the identification of a co-current flow heat exchanger. *International Journal of Control*, 92(1):191–203, 2019.

[18] M. Farah, G. Mercère, R. Ouvrard, T. Poinot, and J. Ramos. Identification of 2d roesser models by using linear fractional transformations. In *2014 European Control Conference (ECC)*, pages 382–387, 2014.

[19] E. Fornasini and G. Marchesini. Doubly-indexed dynamical systems: State-space models and structural properties. *Mathematical systems theory*, 12(1):59–72, 1978.

[20] E. Fornasini, P. Rocha, and S. Zampieri. State space realization of 2-D finite-dimensional behaviours. *SIAM Journal on Control and Optimization*, 31(6):1502–1517, 1993.

[21] R. Fraanje and M. Verhaegen. A spatial canonical approach to multidimensional state-space identification for distributed parameter systems. In *Proceedings of the Fourth International Workshop on Multidimensional Systems - NDS 2005*, pages 217–222. IEEE, 2005.

[22] D. Guo, H. Lu, and X. Qu. A fast low rank hankel matrix factorization reconstruction method for non-uniformly sampled magnetic resonance spectroscopy. *IEEE Access*, 5:16033–16039, 2017.

[23] M. Haardt, C. Brunner, and JH. Nossek. Joint estimation of 2-d arrival angles, propagation delays, and doppler frequencies to determine realistic directional simulation models for smart antennas. In *Proc. IEEE Digital Signal Processing Workshop*, 1998.

[24] M. Haardt and J.A. Nossek. 3-d unitary esprit for joint 2-d angle and carrier estimation. In *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 255–258 vol.1, 1997.

[25] Anders Hansson and Michel Verhaegen. Distributed system identification with ADMM. In *Proceedings of the IEEE Conference on Decision and Control*, pages 290–295, 2014.

[26] B. Hanzon. Attasi nD systems and polynomial system solving. *IFAC-PapersOnLine*, 54:145–150, 2021.

[27] R. Hartshorne. *Algebraic Geometry*, volume 52 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1977.

[28] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2013.

[29] P. Hydon. *Difference Equations by Differential Equation Methods*. Cambridge University Press, 2014.

[30] A. Janczak, E. Rogers, and Z. Cai. Subspace identification of process dynamics for iterative learning control. In *Proceedings of the Eight International Workshop on Multidimensional Systems - NDS*, pages 1–6, 2013.

[31] T. Jiang, N. D. Sidiropoulos, and J. Ten Berge. Almost-sure identifiability of multidimensional harmonic retrieval. *IEEE Transactions on Signal Processing*, 49(9):1849–1859, 2001.

[32] T. Kaczorek. The singular general model of 2d systems and its solution. *IEEE Transactions on Automatic Control*, 33(11):1060–1061, 1988.

[33] T. Kaczorek. Generalization of Cayley-Hamilton theorem for n-D polynomial matrices. *IEEE Trans. on Automatic Control*, 50:671–674, 2005.

[34] J. Kurek. Basic properties of q-dimensional linear digital systems. *International Journal of Control*, 42(1):119–128, 1985.

[35] S. Lagauw, O. M. Agudelo, and B. De Moor. Globally optimal siso h2-norm model reduction using walsh's theorem. *IEEE Control Systems Letters*, 7:1670–1675, 2023.

[36] S. Lagauw, L. Vanpoucke, and B. De Moor. Exact solution to the least squares realization problem as a multiparameter eigenvalue problem. *Automatica*, 185:112792, 2026.

[37] B. Legat, T. Weisser, L. Kapelevich, J. Huchette, A. Bhatia, T. Votroubek, M. Forets, A. Moat, C. Coffrin, Chris C., E. Saba, E. Hanson, M. Lubin, Tu. S., and T. Kelman. Sumofsquares.jl: Sum of squares programming for julia, 2021.

[38] S. Lele and J. Mendel. Modeling and recursive state estimation for two-dimensional noncausal filters with applications in image restoration. *IEEE Transactions on Circuits and Systems*, 34(12):1507–1517, 1987.

[39] B. Lemmerling, P. andDe Moor. Misfit versus latency. *Automatica*, 37(12):2057–2067, 2001.

[40] Y. Li, J. Razavilar, and K.J. Ray Liu. A high-resolution technique for multidimensional nmr spectroscopy. *IEEE Transactions on Biomedical Engineering*, 45(1):78–86, 1998.

[41] F S Macaulay. *The algebraic theory of modular systems*. Cambridge University Press, 1916.

[42] K.N. Mokios, N.D. Sidiropoulos, M. Pesavento, and C.E. Mecklenbrauker. On 3d harmonic retrieval for wireless channel sounding. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages ii–89, 2004.

[43] U. Oberst. Multidimensional constant linear systems. *Acta Applicandae Mathematicae*, 20:1–175, 1990.

[44] K. C. O'Meara, J. Clark, and C. I. Vinsonhaler. *Advanced Topics in Linear Algebra*. Oxford University Press, 2011.

[45] D.C. Pham, G. Mercère, R. Ouvrard, T. Poinot, and H. Pálsson. Fouling detection in a parallel flow heat exchanger via a roesser model identification procedure. *IFAC-PapersOnLine*, 50(1):12866–12871, 2017. 20th IFAC World Congress.

[46] B. Plestenjak and M. E. Hochstenbach. Roots of bivariate polynomial systems via determinantal representations. *SIAM Journal on Scientific Computing*, 38(2):A765–A788, 2016.

[47] J.A. Ramos and G. Mercère. Subspace algorithms for identifying separable-in-denominator 2d systems with deterministic–stochastic inputs. *International Journal of Control*, 89(12):2584–2610, 2016.

[48] J.A. Ramos and G. Mercère. A stochastic subspace system identification algorithm for state-space systems in the general 2-d roesser model form. *International Journal of Control*, 91(12):2743–2771, 2018.

[49] P. Rapisarda. Discrete roesser state models from 2d frequency data. *Multidimensional Systems and Signal Processing*, 30, 2019.

[50] P. Regalia. *Adaptive IIR filtering in Signal Processing and Control*, page 129. CRC Press, 1994.

[51] P. Rocha and J.C. Willems. State for 2-D systems. *Linear Algebra and Its Applications*, 122-124:1003–1038, 1989.

[52] R. P. Roesser. A discrete state-space model for linear image processing. *IEEE Trans. on Automatic Control*, 20:1–10, 1975.

[53] E. Rogers, K. Galkowski, W. Paszke, K. L. Moore, P. H. Bauer, L. Hladowski, and P. Dabkowski. Multidimensional control systems: case studies in design and evaluation. *Multidimensional Systems and Signal Processing*, 26(4):895–939, 2015.

[54] S. Sahnoun, K. Usevich, and P. Comon. Multidimensional esprit for damped and undamped signals: Algorithm, computations, and perturbation analysis. *IEEE Transactions on Signal Processing*, 65(22):5897–5910, 2017.

[55] M. Sørensen and L. De Lathauwer. Multidimensional harmonic retrieval via coupled canonical polyadic decomposition—part i: Model and identifiability. *IEEE Transactions on Signal Processing*, 65(2):517–527, 2016.

[56] H. J. Stetter. *Numerical Polynomial Algebra*. SIAM, 2004.

[57] B. Sturmfels. *Solving Systems of Polynomial Equations*. American Mathematical Society, 2002.

[58] D. Umansky and M. Patzold. A two-dimensional autoregressive model for mimo wideband mobile radio channels. In *IEEE GLOBECOM 2008 - Global Telecommunications Conference*, pages 1–6, 2008.

[59] J. Vanderstukken and L. De Lathauwer. Systems of polynomial equations, higher-order tensor decompositions, and multidimensional harmonic retrieval: A unifying framework. part i: The canonical polyadic decomposition. *SIAM Journal on Matrix Analysis and Applications*, 42(2):883–912, 2021.

[60] J. Vanderstukken, P. Kürschner, I. Domanov, and L. De Lathauwer. Systems of polynomial equations, higher-order tensor decompositions, and multidimensional harmonic retrieval: A unifying framework. part II: The block term decomposition. *SIAM Journal on Matrix Analysis and Applications*, 42(2):913–953, 2021.

[61] L. Vanpoucke and B. De Moor. Constructing multidimensional difference equations from a state-space representation using the generalized cayley-hamilton theorem. *IEEE Control Systems Letters*, 8:2259–2264, 2024.

[62] C. Vermeersch. *The (Block) Macaulay Matrix: Solving Systems of Multivariate Polynomial Equations and Multiparameter Eigenvalue Problems*. PhD thesis, KU Leuven, 2023.

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

[64] J. Verschelde. Algorithm 795: PHCpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Transactions on Mathematical Software*, 25(2):251–276, 1999.

[65] J.C. Willems. From time series to linear system - part i. finite dimensional linear time invariant systems. *Automatica*, 22:561–580, 1986.

[66] R. Yang, L. Xie, and C. Zhang. H2 and mixed h2/h∞ control of two-dimensional systems in roesser model. *Automatica*, 42(9):1507–1514, 2006.

[67] J. Ying, H. Lu, Q. Wei, J. Cai, D. Guo, J. Wu, Z. Chen, and X. Qu. Hankel matrix nuclear norm regularized tensor completion for n-dimensional exponential signals. *IEEE Transactions on Signal Processing*, 65(14):3702–3717, 2017.

[68] E. Zerz. *Topics in Multidimensional Linear Systems Theory*. Springer-Verlag London, 2000.