KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Lukas Vanpoucke, Bart De Moor Title: Multivariate polynomials and autonomous, scalar, linear, shift-invariant models of finite order 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



Multivariate polynomials and autonomous, scalar, linear, shift-invariant models of finite order

Lukas Vanpoucke, Bart De Moor *

1 Introduction

The current work serves as prelude to the contributions in the remainder of this thesis. To this end, it introduces the theoretical foundations underpinning this thesis: the close connection between multivariate polynomials and autonomous, scalar, linear, shift-invariant models, which is leveraged throughout this work to represent such models and to fit them using polynomial optimization problems. The latter models are a straightforward generalization of linear, time-invariant (LTI) models that can similarly be characterized in three different ways: through the so-called kernel, image and state-space representations, as visualized in Figure 1.

While the results contained in this work have mostly been stated for more general models, these statements often rely on abstract notions from algebraic geometry and module theory, and are thus not always that accessible to an engineering audience. As such, the aim of this work is to introduce the model class under consideration, highlight its close link with multivariate polynomials and provide a brief overview of the different model representations and the linear algebra-based transitions between them. Any reader familiar with this material can safely skip ahead to the later works, in which the most important concepts for each individual work are revised as well.

*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.

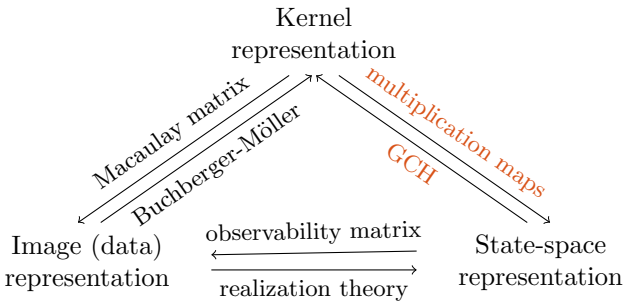


Figure 1: Different representation of autonomous, scalar, linear, shift-invariant models and the linear algebra-based techniques to transition between them. The red color of the 'GCH' (Generalized Cayley-Hamilton) and 'multiplication maps' techniques indicates that this will not be discussed in the current work. Instead, it is a contribution that can be found in works ?? and ?? of this thesis, respectively.

2 Kernel representation

2.1 Difference equations

Discrete, mD autonomous, single-output, multishift-invariant, linear models generate discrete multidimensional (mD) signals which exists on a discrete grid of multi-indices $\boldsymbol{\kappa} \in \mathbb{N}_0^m$. More precisely, any function that associates a grid point $\boldsymbol{\kappa}$ with some numerical value can be viewed as a multidimensional (mD) signal[1] $y_{\boldsymbol{\kappa}} : \mathbb{N}_0^m \to \mathbb{R}^p; \boldsymbol{\kappa} \mapsto y_{\boldsymbol{\kappa}}$. Using so-called shift operators $\sigma_i : \sigma_i y_{\boldsymbol{\kappa}} = y_{\boldsymbol{\kappa}+\boldsymbol{e}_k}, k = 1, \ldots, m$, where $\boldsymbol{e}_k$ is the kth column of the identity matrix, the models under consideration can be characterized by a set of difference equations $\boldsymbol{r}(\boldsymbol{\sigma}) \in \mathbb{R}^{q\times p}[\boldsymbol{\sigma}]$ by requiring that all model-compliant signals $\hat{y}_{\boldsymbol{\kappa}}$ satisfy this set of equations at each grid points:

$$\forall \boldsymbol{\kappa} \in \mathbb{N}_0^m : \ \boldsymbol{r}(\boldsymbol{\sigma})\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{0}. \tag{1}$$

Here, $\boldsymbol{r}(\boldsymbol{\sigma})$ is essentially a set of multivariate polynomials in the shift operators.

By this characterization, any model-compliant signal must lie in the kernel of $\boldsymbol{r}(\boldsymbol{\sigma})$, hence it is referred to as *kernel representation*. In the behavioral framework for systems theory, this kernel is referred to as *the behavior* $\mathfrak{B} = \ker \boldsymbol{r}(\boldsymbol{\sigma})$. The behavior forms a linear subspace for the considered model class, which forms the basis for the image representation [16, 43, 60, 64]. In this thesis, we will confine ourselves to such models with a finite-dimensional behavior. To clarify this abstract characterization, consider the example below.

[1] *Notation:* $\mathbb{N}_0$ designates the set of natural numbers, including zero.

Example 2.1. Consider the following set of second-degree difference equations:

$$\boldsymbol{r}(\boldsymbol{\sigma}) = \begin{bmatrix} \sigma_1 - \sigma_1\sigma_2 \\ \sigma_2 - \sigma_1\sigma_2 \\ \sigma_1\sigma_2 \end{bmatrix}.$$

Notice that $\boldsymbol{r}(\boldsymbol{\sigma})$ is a polynomial in the 'variables' $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$. To obtain a model-compliant output signal, we can propose a solution signal of the form $\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{\lambda}^{\boldsymbol{\kappa}} = \prod_{i=1}^{m} \lambda_i^{\kappa_i}$, where $\boldsymbol{\lambda} \in \mathbb{C}^m$, which, by (1), results in the equations:

$$\begin{aligned} \forall \boldsymbol{\kappa} \in \mathbb{N}_0^2 : \boldsymbol{r}(\boldsymbol{\sigma})\hat{y}_{\boldsymbol{\kappa}} &= \boldsymbol{r}(\boldsymbol{\sigma})\boldsymbol{\lambda}^{\boldsymbol{\kappa}} \\ &= \begin{bmatrix} \boldsymbol{\lambda}^{\boldsymbol{\kappa}+(1,0)} - \boldsymbol{\lambda}^{\boldsymbol{\kappa}+(1,1)} \\ \boldsymbol{\lambda}^{\boldsymbol{\kappa}+(0,1)} - \boldsymbol{\lambda}^{\boldsymbol{\kappa}+(1,1)} \\ \boldsymbol{\lambda}^{\boldsymbol{\kappa}+(1,1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda}^{(1,0)} - \boldsymbol{\lambda}^{(1,1)} \\ \boldsymbol{\lambda}^{(0,1)} - \boldsymbol{\lambda}^{(1,1)} \\ \boldsymbol{\lambda}^{(1,1)} \end{bmatrix} \boldsymbol{\lambda}^{\boldsymbol{\kappa}} = \boldsymbol{r}(\boldsymbol{\lambda}) = \mathbf{0}. \end{aligned}$$

For complex-valued $\boldsymbol{\lambda}$, the above equations is satisfied if and only if $\boldsymbol{\lambda}$ satisfies the system of multivariate polynomial equations $\boldsymbol{r}(\boldsymbol{\lambda})$. These solutions $\boldsymbol{\lambda}$ will also be interchangeably referred to as *eigentuples* or (common) roots throughout this text. In this simple case, the $\boldsymbol{\lambda}^{(1,1)}$ term can be eliminated in the first two equations leading to the single solution $(\lambda_1, \lambda_2) = (0, 0)$ and the associated signal $\hat{y}_{\boldsymbol{\kappa}} = 0^{\kappa_1}0^{\kappa_2}$.

In general, it can be shown that $\operatorname{null} \boldsymbol{r}(\boldsymbol{\sigma})$ is spanned by exponential signals associated with the roots of $\boldsymbol{r}(\boldsymbol{\sigma})$ [9, 44, 45, 53], as illustrated in Example 2.1. While a full blown proof for this statement would lead us to far, to build some intuition on the matter we consider the converse problem: any set of model-compliant signals spanned by multivariate exponentials associated with the eigentuples $V = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n\}$ can be equivalently represented as the kernel of a set of difference equations[2].

The set of all polynomials that vanish on the eigentuples can be straightforwardly defined as $I = \{p(\boldsymbol{\lambda}) \in \mathbb{C}[\boldsymbol{\lambda}] \mid \forall i \in \{1, \ldots, n\} : p(\boldsymbol{\lambda}_i) = 0\}$, a so-called *ideal*. Now, for any two polynomials $g_1(\boldsymbol{\lambda})$ and $g_2(\boldsymbol{\lambda})$ with $\forall i : g_1(\boldsymbol{\lambda}_i) - g_2(\boldsymbol{\lambda}_i) = 0$ it then holds that $g_1(\boldsymbol{\lambda}) - g_2(\boldsymbol{\lambda}) = r(\boldsymbol{\lambda}) \in I$. As such, interpreting these polynomials as difference equations, it follows that $\forall \hat{y}_{\boldsymbol{\kappa}} \in \ker I : g_1(\boldsymbol{\sigma})\hat{y}_{\boldsymbol{\kappa}} = (g_2(\boldsymbol{\sigma}) + r(\boldsymbol{\sigma}))\hat{y}_{\boldsymbol{\kappa}} = g_2(\boldsymbol{\sigma})\hat{y}_{\boldsymbol{\kappa}}$. Now, consider $b_i(\boldsymbol{\lambda})$ so that the vectors $\boldsymbol{b}_i = \begin{bmatrix} b_i(\boldsymbol{\lambda}_1) & \ldots & b_i(\boldsymbol{\lambda}_n) \end{bmatrix}, i \in \{1, \ldots, n\}$ form a basis for $\mathbb{C}^n$, then, $\forall \boldsymbol{\kappa}$ there always exists some $\boldsymbol{c}_{\boldsymbol{\kappa}}$ such that $\begin{bmatrix} \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}} & \ldots & \boldsymbol{\lambda}_n^{\boldsymbol{\kappa}} \end{bmatrix} = \sum_{i=1}^{n} (c_{\boldsymbol{\kappa}})_i \boldsymbol{b}_i$. Consequently, $\forall \boldsymbol{\kappa} : y_{\boldsymbol{\kappa}} = \boldsymbol{\sigma}^{\boldsymbol{\kappa}} y_0 = (\sum_{i=1}^{n} (c_{\boldsymbol{\kappa}})_i b_i(\boldsymbol{\sigma}))\, y_0$, i.e. any model-compliant signal can be written as linear combinations of the n basis signals in $\boldsymbol{c}_{\boldsymbol{\kappa}}$. Conversely, by a similar construction as in Example 2.1, $\boldsymbol{\lambda}_i^{\boldsymbol{\kappa}}$ clearly satisfies all difference equations in I and since they from n linearly independent signals, they form a basis. Finally, Hilbert's basis theorem then states that any polynomial ideal can be described by a finite set of polynomials as follows $I = \langle p_1(\boldsymbol{\sigma}), \ldots, p_q(\boldsymbol{\sigma}) \rangle = \{\sum_{i=1}^{q} g_i p_i \mid g_i \in \mathbb{C}[\boldsymbol{\sigma}]\}$, such that it suffices to satisfy these generating equations for a signal to be model-compliant.

[2] Adapted from the derivations leading up to [12], Proposition 5.2.3.

The constructions above only covers the case where the eigentuples are simple, i.e. all common roots are of multiplicity one. In case of a multiple common root, said root will lead to the same number of linearly independent basis signals, each of a different format. We will get back to the case of multiplicities later in Section 3. Let us first look into the more practical case where only a finite number of grid points are considered.

2.2 Homogenization, roots at infinity and the Macaulay matrix

Example 2.1 illustrates how every common root of the multivariate polynomials $\boldsymbol{r}(\boldsymbol{\lambda})$ generates a model-compliant signal component $\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{\lambda}^{\boldsymbol{\kappa}}$. However, we often consider signals over a finite set of grid points. In this work, sets of the form $\mathcal{K} = \{\boldsymbol{\kappa} \in \mathbb{N}_0^m \mid \|\boldsymbol{\kappa}\|_1 \leq d\}$ are most often considered, that is all grid points up to a certain 'degree' d. In this case, not only the affine roots, but the projective roots determine the number of solution 'modes' of the form $\boldsymbol{\lambda}^{\boldsymbol{\kappa}}$. To obtain the projective roots of a set polynomials, which may contain so-called *roots at infinity*, we must consider their homogenization. Polynomials can be homogenized by adding a new variable λ_0 as follows: $\lambda_0^d \boldsymbol{r}(\frac{\boldsymbol{\lambda}}{\lambda_0})$, where d is the degree of $\boldsymbol{r}(\boldsymbol{\lambda})$. This causes all terms of the resulting $m+1$-variate polynomial to be precisely of degree d. Conversely, a homogeneous polynomial is a polynomial where all nonzero terms are of the same total degree. To see why these solutions at infinity must be accounted for, we reconsider the running example.

Example 2.2. Reconsider Example 2.1. The difference equations $\boldsymbol{r}(\boldsymbol{\sigma})$ then describe linear relations along the grid points. As such, when we consider the finite set of points $\mathcal{K} = \{\boldsymbol{\kappa} \in \mathbb{N}_0^m \mid \|\boldsymbol{\kappa}\|_1 \leq 2\}$, these equations can be written out in a matrix format:

$$
\begin{array}{rc}
 & \begin{array}{ccccccccccc} (0,0) & (1,0) & (0,1) & (2,0) & (1,1) & (0,2) & (3,0) & (2,1) & (1,2) & (0,3) \end{array} \\
\boldsymbol{r}(\boldsymbol{\sigma})y_{(0,0)} & \left[\begin{array}{ccccccccccc} & 1 & & & -1 & & & & & \end{array}\right. \\
\boldsymbol{r}(\boldsymbol{\sigma})y_{(0,0)} & \left[\begin{array}{ccccccccccc} & & 1 & & -1 & & & & & \end{array}\right. \\
 & \left[\begin{array}{ccccccccccc} & & & & 1 & & & & & \end{array}\right. \\
 & \left[\begin{array}{ccccccccccc} & & & 1 & & & & -1 & & \end{array}\right. \\
\mathbf{0} = \boldsymbol{M}_{\boldsymbol{r}}\hat{\boldsymbol{y}} = \sigma_1\boldsymbol{r}(\boldsymbol{\sigma})y_{(1,0)} & \left[\begin{array}{ccccccccccc} & & & & 1 & & & -1 & & \end{array}\right. \\
 & \left[\begin{array}{ccccccccccc} & & & & & & & 1 & & \end{array}\right. \\
 & \left[\begin{array}{ccccccccccc} & & & & 1 & & & & -1 & \end{array}\right. \\
\sigma_2\boldsymbol{r}(\boldsymbol{\sigma})y_{(0,1)} & \left[\begin{array}{ccccccccccc} & & & & & 1 & & & -1 & \end{array}\right. \\
 & \left[\begin{array}{ccccccccccc} & & & & & & & & 1 & \end{array}\right.
\end{array}
\begin{bmatrix} \hat{y}_{(0,0)} \\ \hat{y}_{(1,0)} \\ \hat{y}_{(0,1)} \\ \hat{y}_{(2,0)} \\ \hat{y}_{(1,1)} \\ \hat{y}_{(0,2)} \\ \hat{y}_{(3,0)} \\ \hat{y}_{(2,1)} \\ \hat{y}_{(1,2)} \\ \hat{y}_{(0,3)} \end{bmatrix}.
$$

Solving this set of linear equation then results in the model-compliant sequences $\hat{\boldsymbol{y}}$ over $\mathcal{K}$. Notice, however, that by taking row combinations of each of the row blocks, it can be deduced that $0 = \hat{y}_{(1,0)} = \hat{y}_{(0,1)} = \hat{y}_{(2,0)} = \hat{y}_{(1,1)} = \hat{y}_{(0,2)}$, while $\hat{y}_{(3,0)}$ and $\hat{y}_{(0,3)}$ remain unrestricted. Now consider the case where we all grid points are considered: it can then be deduced that

$0 = r_1(\boldsymbol{\sigma}) - r_3(\boldsymbol{\sigma})\hat{y}_{(2,0)} = \sigma_1^2\hat{y}_{(3,0)} = \hat{y}_{(3,0)}$ and, through a similar analysis, $0 = \hat{y}_{(0,3)}$. Clearly $\boldsymbol{M_r}$ does not take these second-degree shifts of $\boldsymbol{r}(\boldsymbol{\sigma})$ into account, as these consider points outside of $\mathcal{K}$, causing two additional degrees of freedom for the solutions $\hat{\boldsymbol{y}}$.

Homogenizing the polynomials as follows

$$\boldsymbol{r}(\lambda_0, \lambda_1, \lambda_2) = \begin{bmatrix} \lambda_0\boldsymbol{\lambda}^{(1,0)} - \boldsymbol{\lambda}^{(1,1)} \\ \lambda_0\boldsymbol{\lambda}^{(0,1)} - \boldsymbol{\lambda}^{(1,1)} \\ \boldsymbol{\lambda}^{(1,1)} \end{bmatrix} = \mathbf{0}, \tag{2}$$

unveils that on top of the known root $(\lambda_0, \lambda_1, \lambda_2) = (1, 0, 0)$, there are two additional projective roots: $(0, 1, 0)$ and $(0, 0, 1)$. Using this technique forces all terms to be precisely of degree two in the variables $(\lambda_0, \lambda_1, \lambda_2)$, which nicely encodes that the second-degree equation $0 = \lambda_0\boldsymbol{\lambda}^{(1,0)}\hat{y}_{\boldsymbol{\kappa}}$ is formed through a linear combination involving difference equations of up to degree two. Similarly, introducing a homogenization shift operator σ_0, it then can be verified that the rows of $\boldsymbol{M_r}$ exactly describe the linear relations imposed by $\sigma_0\boldsymbol{r}(\sigma_0, \sigma_1, \sigma_2), \sigma_1\boldsymbol{r}(\sigma_0, \sigma_1, \sigma_2)$ and $\sigma_2\boldsymbol{r}(\sigma_0, \sigma_1, \sigma_2)$ over the homogenized grid of exactly degree three $\mathcal{K}_{\text{hom}} = \{\boldsymbol{\kappa} \in \mathbb{N}_0^{m+1} \mid \|\boldsymbol{\kappa}\|_1 = 3\}$. Indeed, the points in $\mathcal{K}$ and $\mathcal{K}_{\text{hom}}$ can be associated one-to-one with one another: $\forall\boldsymbol{\kappa} \in \mathcal{K} : (3 - \|\boldsymbol{\kappa}\|_1, \boldsymbol{\kappa}) \in \mathcal{K}_{\text{hom}}$ and $\forall\boldsymbol{\kappa} \in \mathcal{K}_{\text{hom}} : \boldsymbol{\kappa}_{1:m} \in \mathcal{K}$, i.e. we simply disregard κ_0. Taking this bijection between the grids into account it can be seen that $\boldsymbol{M}_r$ implicitly considers these homogenized difference equations over $\mathcal{K}_{\text{hom}}$.

Now, looking at solutions to $\sigma_0\boldsymbol{r}, \sigma_1\boldsymbol{r}$ and $\sigma_2\boldsymbol{r}$ over $\mathcal{K}_{\text{hom}}$, a solution of the form $(\lambda_0, \lambda_1, \lambda_2)^{(\kappa_0,\kappa_1,\kappa_2)}$, where $(\kappa_0, \kappa_1, \kappa_2) \in \mathcal{K}_{\text{hom}}$ can again be proposed. Working this out, keeping in mind that $\mathbf{0}$ does not represent a projective point, we conclude that the projective point $(\lambda_0, \lambda_1, \lambda_2)$ must satisfy (2), so that each projective solutions lead to one specific exponential solution signal.

Example 2.2 illustrates how the projective roots come into play when compiling the difference equations in a specific matrix format. $\boldsymbol{M}_r$ belongs to the class of Macaulay matrices, which in the jargon of algebraic geometry can be viewed as a finite-dimensional representation of the so-called polynomial ideal associated with the equations [12]. Given a degree d polynomial, or equivalently, difference equation $\boldsymbol{r}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} \boldsymbol{c}_{\boldsymbol{\alpha}}\boldsymbol{\sigma}^{\boldsymbol{\alpha}}$, where $\mathcal{A} = \{\boldsymbol{\alpha} \in \mathbb{N}_0^m \mid \|\boldsymbol{\alpha}\|_1 \leq d\}$ and is finite, the associated Macaulay matrix of degree d_t can then be concisely defined as follows[3], using multi-indices for the block rows and columns:

$$(\boldsymbol{M_r})_{\boldsymbol{\rho},\boldsymbol{\kappa}} = \boldsymbol{c}_{\boldsymbol{\rho}-\boldsymbol{\kappa}}, \text{ where } \boldsymbol{\kappa}, \boldsymbol{\rho} \in \mathbb{N}_0^m; \|\boldsymbol{\kappa}\|_1 \leq d_t, \|\boldsymbol{\rho}\|_1 \leq d_t - d,$$

where $\boldsymbol{c}_{\boldsymbol{\alpha}}$ is considered zero when $\boldsymbol{\alpha} \notin \mathcal{A}$.

When $\boldsymbol{r}(\boldsymbol{\lambda})$ has a finite number of projective roots and given that d_t is large enough, the nullity of this matrix is equal to said number of roots, counting multiplicities. Moreover, the right null space (kernel) the matrix is then spanned by the vectors that can be defined elementwise as $(\boldsymbol{v}_i)_{\boldsymbol{\kappa}} = \boldsymbol{\lambda}_i^{\boldsymbol{\kappa}}$, so-called *generalized*

[3]Different row-permutation of this matrix may be used throughout this thesis, though for its scope, the order of the rows is of no importance.

Vandermonde vectors, where $\boldsymbol{\lambda}_i, i \in \{1, \ldots, n\}$ are the projective common roots of $\boldsymbol{r}(\boldsymbol{\lambda})$ [3, 14, 55, 59]. Since each root contributes one exponential solution mode, n is referred to as the *order* of the model. The Macaulay matrix offers a practical check to see if a signal complies with the model, as the common roots are often not known in practice. Moreover, this class of matrices can be employed to find said roots, as will be illustrated later on [3, 14, 59].

Let us briefly clarify the concept of projective roots. Projective space can be considered an extension to Euclidean space by adding 'points at infinity'. A projective 'point' is defined as the equivalence class $[\boldsymbol{p}] = \{a\boldsymbol{p} \mid a \in \mathbb{C}, \boldsymbol{p} \in \mathbb{C}^{m+1} \backslash \{\boldsymbol{0}\}\}$, where any member of the set can be chosen to represent it: $[a\boldsymbol{p}] = [\boldsymbol{p}]$ for any a. Remark that $\boldsymbol{p} = 0$ is excluded and does not represent any projective point. For notational convenience, we will simply denote projective points by one of their representative coordinates in $\mathbb{C}^{m+1}$, dropping the brackets. For example, the coordinates $(1, 2, 3), (2, 4, 6)$ and $(3, 6, 9)$ all represent the same projective point. As such, the associated affine point in $\mathbb{C}^m$ can by retrieved by normalizing by the homogenization variable: $(\frac{p_1}{p_0}, \ldots, \frac{p_m}{p_0})$. When $p_0 = 0$ the point is thus considered to lie at infinity.

While a set of homogeneous polynomials $\boldsymbol{r}(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \mathbb{C}^{m+1}$ of degree d does not define a function over the projective points – it takes different values when different representatives of the same point are considered: $\boldsymbol{r}(a\boldsymbol{\lambda}) = a^d \boldsymbol{r}(\boldsymbol{\lambda})$ – its common roots are well-defined. As such, the common roots of a set of homogeneous polynomials corresponds precisely to a set of projective points. For a more detailed exposition on the subject, we refer the reader to [3, 12, 55]. Regarding notation, $\boldsymbol{\lambda} \in \mathbb{P}^m$ will designate a projective point and $\boldsymbol{\kappa} \in \mathbb{N}_0^{m+1}$ will indicate a homogenized multi-index for the remainder of this work.

In 1D, LTI models with eigentuples at infinity are called *singular* or *descriptor systems*. Such systems do not arise without explicit homogenization, however. Moreover, using the Weierstrass canonical form, these models can be split into a causal ('forward') and anti-causal ('backward') part [18, 30].

Finally, we note that there exist techniques such as border and Gröbner that eliminate the solutions at infinity by computing a different set of equations with the same non-infinite roots [12, 40, 51]. Alternatively, components due to the solutions at infinity may be eliminated from the null space of a Macaulay matrix using a column compression [14], which is especially useful for polynomial root-finding purposes, where these roots are often not of interest. For the sake of completeness, in this work, these solutions will be considered though.

3 Image representation

3.1 Simple eigentuples

In the previous section, it was already highlighted that the kernel of the difference equations, i.e. the behavior, forms a vector space that is finite-dimensional if the number of projective common roots is finite. Indeed, the shift operators are linear operators, so that for any linear combination of model-compliant

signals is model-compliant itself. As such a basis for the behavior suffices to express all model compliant signals. Moreover, it was demonstrated how such a basis can be constructed by considering exponential signals associated with the roots of the difference equations, i.e. the eigentuples. Focusing on the finite-grid case and difference equations with simple roots, when a large enough degree d is considered, any model-compliant signal can thus be written as an image of these basis signals, leading to the *image representation*:

$$\forall \boldsymbol{\kappa} \in \mathbb{N}_0^{m+1}, \|\boldsymbol{\kappa}\|_1 = d : \hat{y}_{\boldsymbol{\kappa}} = \begin{bmatrix} \boldsymbol{\lambda}_1^{\boldsymbol{\kappa}} & \cdots & \boldsymbol{\lambda}_n^{\boldsymbol{\kappa}} \end{bmatrix} \boldsymbol{a} = \boldsymbol{\Gamma} \boldsymbol{a},$$

where $\boldsymbol{a} \in \mathbb{C}^n$. Clearly, any signal of the above form always lies in the null space of $\boldsymbol{M}_{\boldsymbol{r}}$. However, to guarantee that range $\boldsymbol{\Gamma}$ equals this space generally requires one to consider degrees above the *degree of regularity*, where $\boldsymbol{\Gamma}$ will form a basis [3], e.g. see Example five in [58]. Recall that a κ_0-variable was added to homogenize the m-grid of degree lower than or equal to d, as explained in Example 2.2.

Similarly to autonomous LTI systems, this class of models also describes multivariate sinusoidal signals through complex-conjugate eigentuples. This is illustrated in Figure 3.1, where the output of a fourth-order 2D linear, shift-invariant model is visualized. Of course, when a finite grid is considered, these basis signals can be cast into a matrix format. As previously mentioned, the resulting generalized Vandermonde matrix then forms a basis for the right null space of an appropriately large Macaulay matrix associated with the difference equations. Example 3.1 illustrates this dual representation for the running example.

Example 3.1. Reconsider Example 2.2 and the Macaulay matrix $\boldsymbol{M}_{\boldsymbol{r}}$ within it. Then observe that the columns of the following matrix span the three exponential basis signals over the grid $\mathcal{K}_{\text{hom}}$, where $\boldsymbol{\lambda}_1 = (1,0,0), \boldsymbol{\lambda}_2 = (0,1,0)$ and $\boldsymbol{\lambda}_3 = (0,0,1)$:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\lambda}_1^{(3,0,0)} & \boldsymbol{\lambda}_2^{(3,0,0)} & \boldsymbol{\lambda}_3^{(3,0,0)} \\ \boldsymbol{\lambda}_1^{(2,1,0)} & \boldsymbol{\lambda}_2^{(2,1,0)} & \boldsymbol{\lambda}_3^{(2,1,0)} \\ \boldsymbol{\lambda}_1^{(2,0,1)} & \boldsymbol{\lambda}_2^{(2,0,1)} & \boldsymbol{\lambda}_3^{(2,0,1)} \\ \boldsymbol{\lambda}_1^{(1,2,0)} & \boldsymbol{\lambda}_2^{(1,2,0)} & \boldsymbol{\lambda}_3^{(1,2,0)} \\ \boldsymbol{\lambda}_1^{(1,1,1)} & \boldsymbol{\lambda}_2^{(1,1,1)} & \boldsymbol{\lambda}_3^{(1,1,1)} \\ \boldsymbol{\lambda}_1^{(1,0,2)} & \boldsymbol{\lambda}_2^{(1,0,2)} & \boldsymbol{\lambda}_3^{(1,0,2)} \\ \boldsymbol{\lambda}_1^{(0,3,0)} & \boldsymbol{\lambda}_2^{(0,3,0)} & \boldsymbol{\lambda}_3^{(0,3,0)} \\ \boldsymbol{\lambda}_1^{(0,2,1)} & \boldsymbol{\lambda}_2^{(0,2,1)} & \boldsymbol{\lambda}_3^{(0,2,1)} \\ \boldsymbol{\lambda}_1^{(0,1,2)} & \boldsymbol{\lambda}_2^{(0,1,2)} & \boldsymbol{\lambda}_3^{(0,1,2)} \\ \boldsymbol{\lambda}_1^{(0,0,3)} & \boldsymbol{\lambda}_2^{(0,0,3)} & \boldsymbol{\lambda}_3^{(0,0,3)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} (3,0,0) \\ (2,1,0) \\ (2,0,1) \\ (1,2,0) \\ (1,1,1) \\ (1,0,2) \\ (0,3,0) \\ (0,2,1) \\ (0,1,2) \\ (0,0,3) \end{matrix}.$$

Of course, any basis $\boldsymbol{\Gamma}$ for null $\boldsymbol{M}_{\boldsymbol{r}}$ works just as well to represent the set of model-compliant signals: By linearity, any sequence $\hat{\boldsymbol{y}} = \boldsymbol{\Gamma}\boldsymbol{x}$ where $\boldsymbol{x} \in \mathbb{C}^n$ is model-compliant. This particular choice of basis only reveals the intricate connection between the projective eigentuples and the basis signals. In practice,

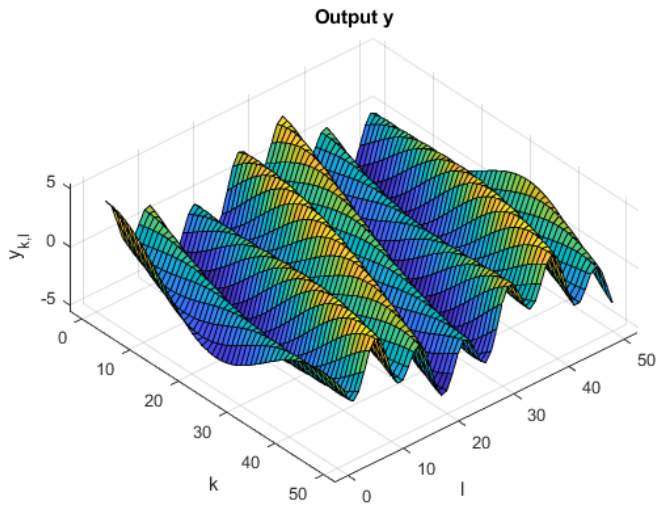


Figure 2: Output data of a model with eigentuples $(e^{i\frac{\pi}{4}}, e^{i\frac{\pi}{20}}), (e^{-i\frac{\pi}{4}}, e^{-i\frac{\pi}{20}}), (e^{i\frac{\pi}{4}}, e^{i\frac{\pi}{15}})$ $(e^{-i\frac{\pi}{4}}, e^{-i\frac{\pi}{15}})$. More precisely, $y_{(\kappa_1,\kappa_2)} = 2\cos\big(\kappa_1\frac{\pi}{4}\big)\cos\big(\kappa_2\frac{\pi}{20}\big) + 2\cos\big(\kappa_1\frac{\pi}{15}\big)\cos\big(\kappa_2\frac{\pi}{30}\big)$.

these roots are not known, however, so that one cannot directly construct this basis. On the contrary, the Macaulay framework uses a numerical basis to find the common roots of multivariate polynomial equations [3, 14, 59].

3.2 Back to a kernel representation

Due to the duality of the null and row space of a (Macaulay) matrix, returning to a kernel representation is then as simple as taking the left kernel of the basis matrix. However, this naive approach generally results in a set of difference equations of a large degree, which cannot be straightforwardly applied to smaller grids.

Alternatively, the (projective) Buchberger-Möller algorithm can be employed, which leads to a reduced Gröbner basis for the ideal, which minimizes the degree of the difference equations for the considered graded ordering of the grid [1, 4, 5, 34, 38]. Generally, this algorithm eliminates the solutions at infinity. However, a simple change of coordinates suffices to move these solutions from 'infinity' to affine space, where they can be treated like as normal common roots, as illustrated in Example 3.2 below [34]. Note that for any finite set of projective points, almost all linear coordinates transformations are such that none of the points lie at infinity in the new coordinate system[4]

[4]Let $\boldsymbol{\xi}_0 = \boldsymbol{p}\begin{bmatrix}\boldsymbol{\lambda}_1 & \cdots & \boldsymbol{\lambda}_n\end{bmatrix}$. Since $\boldsymbol{\lambda}_i \neq \mathbf{0}$, if this vector contains zero elements, there always exists an arbitrarily small perturbation on $\boldsymbol{p}$ such that these elements become nonzero.

Example 3.2. Reconsider the basis $\mathbf{\Gamma}$ from Example 3.1. Performing the invertible change of coordinates $(\lambda_0, \lambda_1, \lambda_2) = (\xi_0 + \xi_1, \xi_1 + \xi_2, \xi_0 + \xi_2) = \boldsymbol{P}\boldsymbol{\xi}$ such that $\boldsymbol{P}^{-1}\boldsymbol{\lambda}_i = \boldsymbol{\xi}_i$, leads to $\boldsymbol{\xi}_1 = (\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}), \boldsymbol{\xi}_2 = (-\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ and $\boldsymbol{\xi}_3 = (\frac{1}{2}, -\frac{1}{2}, \frac{1}{2})$. This allows us to rewrite $\mathbf{\Gamma}$ in as follows:

$$\mathbf{\Gamma} = \begin{bmatrix} (\boldsymbol{P}\boldsymbol{\xi}_1)^{(3,0,0)} & (\boldsymbol{P}\boldsymbol{\xi}_2)^{(3,0,0)} & (\boldsymbol{P}\boldsymbol{\xi}_3)^{(3,0,0)} \\ (\boldsymbol{P}\boldsymbol{\xi}_1)^{(2,1,0)} & (\boldsymbol{P}\boldsymbol{\xi}_2)^{(2,1,0)} & (\boldsymbol{P}\boldsymbol{\xi}_3)^{(2,1,0)} \\ (\boldsymbol{P}\boldsymbol{\xi}_1)^{(2,0,1)} & (\boldsymbol{P}\boldsymbol{\xi}_2)^{(2,0,1)} & (\boldsymbol{P}\boldsymbol{\xi}_3)^{(2,0,1)} \\ (\boldsymbol{P}\boldsymbol{\xi}_1)^{(1,2,0)} & (\boldsymbol{P}\boldsymbol{\xi}_2)^{(1,2,0)} & (\boldsymbol{P}\boldsymbol{\xi}_3)^{(1,2,0)} \\ (\boldsymbol{P}\boldsymbol{\xi}_1)^{(1,1,1)} & (\boldsymbol{P}\boldsymbol{\xi}_2)^{(1,1,1)} & (\boldsymbol{P}\boldsymbol{\xi}_3)^{(1,1,1)} \\ (\boldsymbol{P}\boldsymbol{\xi}_1)^{(1,0,2)} & (\boldsymbol{P}\boldsymbol{\xi}_2)^{(1,0,2)} & (\boldsymbol{P}\boldsymbol{\xi}_3)^{(1,0,2)} \\ (\boldsymbol{P}\boldsymbol{\xi}_1)^{(0,3,0)} & (\boldsymbol{P}\boldsymbol{\xi}_2)^{(0,3,0)} & (\boldsymbol{P}\boldsymbol{\xi}_3)^{(0,3,0)} \\ (\boldsymbol{P}\boldsymbol{\xi}_1)^{(0,2,1)} & (\boldsymbol{P}\boldsymbol{\xi}_2)^{(0,2,1)} & (\boldsymbol{P}\boldsymbol{\xi}_3)^{(0,2,1)} \\ (\boldsymbol{P}\boldsymbol{\xi}_1)^{(0,1,2)} & (\boldsymbol{P}\boldsymbol{\xi}_2)^{(0,1,2)} & (\boldsymbol{P}\boldsymbol{\xi}_3)^{(0,1,2)} \\ (\boldsymbol{P}\boldsymbol{\xi}_1)^{(0,0,3)} & (\boldsymbol{P}\boldsymbol{\xi}_2)^{(0,0,3)} & (\boldsymbol{P}\boldsymbol{\xi}_3)^{(0,0,3)} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 3 & & 3 & & & 1 & & & \\ & 1 & 1 & 2 & 2 & & 1 & 1 & & \\ 1 & 2 & 1 & 1 & 2 & & & 1 & & \\ & & & 1 & 2 & 1 & 1 & 2 & 1 & \\ & 1 & 1 & 1 & 2 & 1 & & 1 & 1 & \\ 1 & 1 & 2 & & 2 & 1 & & & 1 & \\ & & & & & & 1 & 3 & 3 & 1 \\ & & & 1 & 2 & 1 & & 1 & 2 & 1 \\ & 1 & 1 & & 2 & 2 & & & 1 & 1 \\ 1 & & 3 & & & 3 & & & & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}_1^{(3,0,0)} & \boldsymbol{\xi}_2^{(3,0,0)} & \boldsymbol{\xi}_3^{(3,0,0)} \\ \boldsymbol{\xi}_1^{(2,1,0)} & \boldsymbol{\xi}_2^{(2,1,0)} & \boldsymbol{\xi}_3^{(2,1,0)} \\ \boldsymbol{\xi}_1^{(2,0,1)} & \boldsymbol{\xi}_2^{(2,0,1)} & \boldsymbol{\xi}_3^{(2,0,1)} \\ \boldsymbol{\xi}_1^{(1,2,0)} & \boldsymbol{\xi}_2^{(1,2,0)} & \boldsymbol{\xi}_3^{(1,2,0)} \\ \boldsymbol{\xi}_1^{(1,1,1)} & \boldsymbol{\xi}_2^{(1,1,1)} & \boldsymbol{\xi}_3^{(1,1,1)} \\ \boldsymbol{\xi}_1^{(1,0,2)} & \boldsymbol{\xi}_2^{(1,0,2)} & \boldsymbol{\xi}_3^{(1,0,2)} \\ \boldsymbol{\xi}_1^{(0,3,0)} & \boldsymbol{\xi}_2^{(0,3,0)} & \boldsymbol{\xi}_3^{(0,3,0)} \\ \boldsymbol{\xi}_1^{(0,2,1)} & \boldsymbol{\xi}_2^{(0,2,1)} & \boldsymbol{\xi}_3^{(0,2,1)} \\ \boldsymbol{\xi}_1^{(0,1,2)} & \boldsymbol{\xi}_2^{(0,1,2)} & \boldsymbol{\xi}_3^{(0,1,2)} \\ \boldsymbol{\xi}_1^{(0,0,3)} & \boldsymbol{\xi}_2^{(0,0,3)} & \boldsymbol{\xi}_3^{(0,0,3)} \end{bmatrix} = \boldsymbol{R}\mathbf{\Xi}$$

Keep in mind that any basis for range $\mathbf{\Gamma}$ suffices to perform the algorithm. Notice the similar structure of $\mathbf{\Gamma}$ and $\mathbf{\Xi}$. Indeed, both these matrices are special cases of *observability matrices*, to which we will get back in the next section. The latter of both observability matrices can then be used to construct a set of homogeneous polynomial equations in the new variables $\boldsymbol{\xi}$, after which the coordinate transform is reverted and a set of polynomials in terms of $\boldsymbol{\lambda}$ is obtained. Also notice that, since the coordinate transform is invertible, so is $\boldsymbol{R}$, and, as a consequence, $\mathbf{\Gamma}$ and $\mathbf{\Xi}$ are of the same rank.

Let us now briefly illustrate the main ideas behind the Buchberger-Möller algorithm: First, an appropriate ordering for the multi-exponents, i.e. the grid points, is required. Here the graded inverse lexicographic ordering is used [12]. Then, through any (numerical) linear-algebra technique, the first $n = 3$ linearly independent rows of $\mathbf{\Xi}$ are determined. In this case, the rows associated with

the multi-indices $(3,0,0), (2,1,0)$ and $(2,0,1)$ are independent:

$$\boldsymbol{\Xi} = \frac{1}{8}\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \\ -1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix} \begin{matrix} (3,0,0) \\ (2,1,0) \\ (2,0,1) \\ (1,2,0) \\ (1,1,1) \\ (1,0,2) \\ (0,3,0) \\ (0,2,1) \\ (0,1,2) \\ (0,0,3) \end{matrix}.$$

Observe that any of the subsequent rows can then be written as linear combinations of these first three rows. These linear combinations lead to the following relations that hold in all $\boldsymbol{\xi}_i$: $\boldsymbol{\xi}^{(1,2,0)} - \boldsymbol{\xi}^{(3,0,0)} = 0$; $\boldsymbol{\xi}^{(1,1,1)} + \boldsymbol{\xi}^{(3,0,0)} + \boldsymbol{\xi}^{(2,1,0)} + \boldsymbol{\xi}^{(2,0,1)} = 0$ and $\boldsymbol{\xi}^{(1,0,2)} - \boldsymbol{\xi}^{(3,0,0)} = 0$. Now, since none of the eigentuples $\boldsymbol{\xi}_i$ lie in at infinity, i.e. $\xi_0 \neq 0$, we can safely divide away the ξ_0 factor in each of these equations:

$$\begin{aligned} -\boldsymbol{\xi}^{(2,0,0)} + \boldsymbol{\xi}^{(0,2,0)} &= 0, \\ \boldsymbol{\xi}^{(2,0,0)} + \boldsymbol{\xi}^{(1,1,0)} + \boldsymbol{\xi}^{(1,0,1)} + \boldsymbol{\xi}^{(0,1,1)} &= 0, \\ -\boldsymbol{\xi}^{(2,0,0)} + \boldsymbol{\xi}^{(0,0,2)} &= 0. \end{aligned}$$

This essentially serves to dehomogenize the equations, as the classical form of the algorithm constructs non-homogeneous bases.

Now, these linear expressions allow us to define a model-compliant signal at $(0,2,0), (0,1,1)$ and $(0,0,2)$ based on the preceding points $(2,0,0), (1,1,0), (1,0,1)$. This situation is illustrated in Figure 3.2. Note that by shifting these equations, all multiples of the dependent monomials can be expressed in terms of their predecessors as well. Indeed, the shifts of these polynomials with either ξ_0, ξ_1 or ξ_2, provide us with linear relations between the last four rows of $\boldsymbol{\Xi}$ and their predecessors, all of which could already be rewritten in terms of the first three rows. As such, a set of equations is obtained that recursively describe all rows in terms of first three, providing a full description of $\boldsymbol{\Xi}$. This is what makes reduced Gröbner bases so practical: for the considered grid ordering, they essentially provide us with 'initial conditions' an recursively applying the equations straightforwardly leads to the set of compatible signals.

Now, the coordinate transform can be undone through substituting in $\boldsymbol{P}^{-1}\boldsymbol{\lambda} = (\lambda_0 - \lambda_1 + \lambda_2, \lambda_0 + \lambda_1 - \lambda_2, -\lambda_0 + \lambda_1 + \lambda_2) = \boldsymbol{\xi}$ to obtain the sought-after equations in terms of $\boldsymbol{\lambda}_i$:

$$\begin{aligned} \boldsymbol{\lambda}^{(1,1,0)} - \boldsymbol{\lambda}^{(1,0,1)} &= 0, \\ \boldsymbol{\lambda}^{(1,0,1)} &= 0, \\ -\boldsymbol{\lambda}^{(1,0,1)} + \boldsymbol{\lambda}^{(0,1,1)} &= 0. \end{aligned}$$

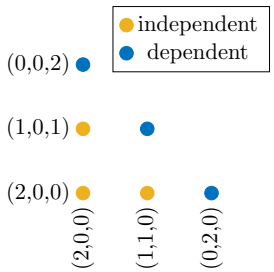


Figure 3: Visual representation of the dependent and independent grid points in Example 3.2.

Notice that linear combinations of these equations readily span the homogenization of the original equations (2) and vice verse, so that they are equivalent representations of the same behavior, that is, the projective common roots of both sets of equations coincide. These equations do not necessarily form a reduced Gröbner basis, however.

3.3 Multiplicities

In the previous subsections, the assumption that $\boldsymbol{r}(\boldsymbol{\sigma})$ has simple common roots was made, that is, roots without multiplicities. But which type of basis signals do arise in the presence of multiplicities? In case of a multiple root, there will be as many linearly independent basis signals associated with said root [43, 44, 45, 53]. As such let us substitute a signal of the form[5] $\frac{\partial \boldsymbol{\lambda}^{\boldsymbol{\kappa}}}{\partial \lambda_k} = \kappa_k \boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k}$ into the difference equations. To work this out us take a look at the action of σ_k^d on such a signal: $\sigma_k^d \kappa_k \boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k} = (\kappa_k + d)\boldsymbol{\lambda}^{\boldsymbol{\kappa}+(d-1)\boldsymbol{e}_k} = \kappa_k \lambda_k^d \boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k} + d\lambda_k^{d-1}\boldsymbol{\lambda}^{\boldsymbol{\kappa}}$, so that:

$$\boldsymbol{r}(\boldsymbol{\sigma})\kappa_k \boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k} = \boldsymbol{r}(\boldsymbol{\lambda})\kappa_i \boldsymbol{\lambda}^{\boldsymbol{\kappa}-\boldsymbol{e}_k} + \frac{\partial \boldsymbol{r}(\boldsymbol{\lambda})}{\partial \lambda_k}\boldsymbol{\lambda}^{\boldsymbol{\kappa}},$$

which vanishes for all $\boldsymbol{\kappa} \in \mathbb{N}_0^m$ if and only if both $\boldsymbol{r}(\boldsymbol{\lambda}) = 0$ and $\frac{\partial \boldsymbol{r}(\boldsymbol{\lambda})}{\partial \lambda_i} = 0$, that is, the common root $\boldsymbol{\lambda}$ has a multiplicity of at least two.

In general, multiplicity structure of an eigentuple is closely tied to the number of linearly independent differential operators that vanish in said eigentuple while defining a 'closed' vector space w.r.t. polynomial multiplication [13, 41, 51, 63]. More complex multiplicity structures give rise to different signals in the kernel [3]. These are increasingly cumbersome to work out, however. Let us illustrate such a more complicated multiplicity structure through an example instead.

Example 3.3. Consider the difference equations below:

$$\boldsymbol{r}(\boldsymbol{\sigma})y_{\boldsymbol{\kappa}} = \begin{bmatrix} \sigma_1^2 - 4\sigma_0\sigma_1 + 4\sigma_0^2 \\ \sigma_2^2 - 6\sigma_0\sigma_2 + 9\sigma_0^2 \end{bmatrix} y_{\boldsymbol{\kappa}} = \boldsymbol{0}$$

[5] For $\lambda_k = 0$ this signal is undefined at $\kappa_k = \boldsymbol{0}$. In this case, let it be zero by convention.

It can be verified that $\boldsymbol{r}(2,3) = \frac{\partial \boldsymbol{r}(\boldsymbol{\lambda})}{\partial \lambda_1}|_{\boldsymbol{\lambda}=(2,3)} = \frac{\partial \boldsymbol{r}(\boldsymbol{\lambda})}{\partial \lambda_2}|_{\boldsymbol{\lambda}=(2,3)} = \frac{\partial^2 \boldsymbol{r}(\boldsymbol{\lambda})}{\partial \lambda_1 \partial \lambda_2}|_{\boldsymbol{\lambda}=(2,3)} = 0$. Since there are four linearly independent evaluation and differential operators associated with $(2,3)$, it is a common root with multiplicity four. For any finite degree, there is a finite number of monomials such that these operators can be easily cast into a vector format by applying them to said monomial basis. The resulting sequences can then also be interpreted as a basis signal, e.g. $\forall \boldsymbol{\kappa} : \hat{y}_{\boldsymbol{\kappa}} = \frac{\partial^2 \boldsymbol{\lambda}^{\boldsymbol{\kappa}}}{\partial \lambda_1 \partial \lambda_2}|_{\boldsymbol{\lambda}=(2,3)}$ is model-compliant. Over a third-degree grid, this gives us the following basis signals:

$$\begin{array}{cccc|c} \boldsymbol{\lambda}^{\boldsymbol{\kappa}}|_{\boldsymbol{\lambda}=(2,3)} & \frac{\partial \boldsymbol{\lambda}^{\boldsymbol{\kappa}}}{\partial \lambda_1}|_{\boldsymbol{\lambda}=(2,3)} & \frac{\partial \boldsymbol{\lambda}^{\boldsymbol{\kappa}}}{\partial \lambda_2}|_{\boldsymbol{\lambda}=(2,3)} & \frac{\partial^2 \boldsymbol{\lambda}^{\boldsymbol{\kappa}}}{\partial \lambda_1 \partial \lambda_2}|_{\boldsymbol{\lambda}=(2,3)} & \boldsymbol{\kappa} \\ \hline 1 & 0 & 0 & 0 & (3,0,0) \\ 2 & 1 & 0 & 0 & (2,1,0) \\ 3 & 0 & 1 & 0 & (2,0,1) \\ 4 & 4 & 0 & 0 & (1,2,0) \\ 6 & 3 & 2 & 1 & (1,1,1) \\ 9 & 0 & 6 & 0 & (1,0,2) \\ 8 & 12 & 0 & 0 & (0,3,0) \\ 12 & 12 & 4 & 4 & (0,2,1) \\ 18 & 9 & 12 & 6 & (0,1,2) \\ 27 & 0 & 27 & 0 & (0,0,3) \end{array}.$$

For this example, these vector representations then span the kernel of any associated Macaulay matrix[6]. Note that this can be readily generalize to more complex multiplicity structures [3, 14].

4 State-space representation

4.1 Affine models

Another possible representation for autonomous, scalar, linear, shift-invariant models is a *state-space representation*. Indeed, this representation is well-established for affine mD systems [16]. Since the affine case is of the greatest importance to the remainder of this thesis and the most accessible sources with this regard concern the regular case, where there are no eigenvalues at infinity [3, 14], we focus on deriving a suitable descriptor state-space format through the well-known affine techniques. This construction relies stems from homogeneous multiplication matrices [6, 54]. For treatment of more general mD descriptor (also singular or implicit), linear state-space models, we refer the reader to [28, 31, 36, 65].

As demonstrated in Section 3.2, cases with roots at infinity can always be reduced through a coordinate transform $\boldsymbol{P}\boldsymbol{\xi} = \boldsymbol{\lambda}$, where $\boldsymbol{P} \in \mathbb{C}^{m+1 \times m+1}$ is invertible, such that $\xi_0 \neq 0$ [1, 34]. This allows us to define alternative difference

[6] In this simple case, the nullity immediately stabilizes to the number of common roots counting multiplicities, i.e. four. In general, one may need a larger degree than that of the equations themselves.

equations as $\boldsymbol{r}(\boldsymbol{\lambda}) = \boldsymbol{r}(\boldsymbol{P}\boldsymbol{\xi}) = \boldsymbol{w}(\boldsymbol{\xi})$, the common roots of which can be associated one to one with those of $\boldsymbol{r}(\boldsymbol{\lambda})$, as $\boldsymbol{P}$ is invertible. Moreover, the multiplicity of associated solutions is also preserved by the coordinate transform $\boldsymbol{P}$: By the chain rule, $\boldsymbol{0} = \frac{\partial \boldsymbol{R}(\boldsymbol{\lambda})}{\partial \lambda_i} = \frac{\partial \boldsymbol{w}(\boldsymbol{\xi}(\boldsymbol{\lambda}))}{\partial \lambda_i} = \sum_{j=0}^{m} \frac{\partial \boldsymbol{w}(\boldsymbol{\xi})}{\partial \xi_j} \frac{\partial \xi_j}{\partial \lambda_i}$, where $\frac{\partial \xi_j}{\partial \lambda_i} = \frac{\partial (\boldsymbol{P}^{-1}\boldsymbol{\lambda})_j}{\partial \lambda_i}$ are constants, implying that $\boldsymbol{w}$ also has a multiplicity in the solution $\boldsymbol{P}^{-1}\boldsymbol{\xi}$ whenever $\boldsymbol{\lambda}$ is a multiple root of $\boldsymbol{r}(\boldsymbol{\lambda})$. This argument can be extended to higher order derivatives.

Let us have a closer look at these affine techniques first. Models without eigentuples at infinity allow a state-space representation of the following form [3, 14, 16, 22]:

$$\begin{aligned} \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_1} &= \boldsymbol{A}_1 \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_0} \\ &\vdots \\ \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_m} &= \boldsymbol{A}_m \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_0} \\ \hat{y}_{\boldsymbol{\kappa}} &= \boldsymbol{c}\boldsymbol{x}_{\boldsymbol{\kappa}} \end{aligned} \tag{3}$$

where $\boldsymbol{A}_k\boldsymbol{A}_l = \boldsymbol{A}_l\boldsymbol{A}_k \ \ \forall k, l \in \{1, \dots, m\}$; $\boldsymbol{A}_k \in \mathbb{C}^{n\times n}, \boldsymbol{c} \in \mathbb{C}^{1\times n}, \boldsymbol{x}_{\boldsymbol{\kappa}} \in \mathbb{C}^n$.

Given an initial state $\boldsymbol{x}_{d\boldsymbol{e}_0} = \boldsymbol{x}$, the output of such a model can then be written as $\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}\boldsymbol{x}$ for $\|\boldsymbol{\kappa}\|_1 = d$, with $\boldsymbol{A}^{\boldsymbol{\kappa}} = \prod_{k=0}^{m} \boldsymbol{A}_k^{\kappa_k}$ and $\boldsymbol{A}_0 = \boldsymbol{I}$. Note that the set of model-compliant outputs is invariant under a joint similarity transform on the *state-transition matrices*, that is, the model $\boldsymbol{A}_k^{\text{new}} = \boldsymbol{V}^{-1}\boldsymbol{A}_k\boldsymbol{V}, \boldsymbol{c}^{\text{new}} = \boldsymbol{c}\boldsymbol{V}, \boldsymbol{x}_{\boldsymbol{\kappa}}^{\text{new}} = \boldsymbol{V}^{-1}\boldsymbol{x}_{\boldsymbol{\kappa}}$ results in the same set of outputs, where $\boldsymbol{V} \in \mathbb{C}^{n\times n}$ is invertible.

These state-transition matrices satisfy any set of difference equations $\boldsymbol{w}(\boldsymbol{\sigma})$ that represent an equivalent model: $\boldsymbol{w}(\boldsymbol{A}_0, \dots \boldsymbol{A}_m) = \boldsymbol{0}$ [14]. Furthermore, the eigenvalues of $\boldsymbol{A}_k$ correspond to the kth coordinates of each of the eigentuples $\boldsymbol{\lambda}_k$. This observation is the foundation of eigenvalue-based polynomial root finding, where these matrices are referred to as *multiplication maps* [51, 55, 59]. The commutativity of the state-transition matrices ensures that they can all be put into a upper triangular form using the same Schur transformation, grouping together the eigenvalues of the different state-transition matrices into eigentuples and revealing their multiplicity [11, 25, 48]. We refer the interested reader to Appendix A.2 for a proof of this property, as well as an analysis of the structure of the eigenspaces of commuting matrices.

We now illustrate the above representation and show how it can be obtained in practice through applying *multidimensional realization theory* from an image representation [14].

Example 4.1. Reconsider the linear coordinate transform $\boldsymbol{P}$ from Example 3.2. As previously stated, this defines another set of polynomials $\boldsymbol{r}(\boldsymbol{P}\boldsymbol{\xi}) = \boldsymbol{w}(\boldsymbol{\xi})$, and thus an associated Macaulay matrix $\boldsymbol{M}_{\boldsymbol{w}} = \boldsymbol{M}_{\boldsymbol{r}}\boldsymbol{R}$. Indeed, it was already derived that $\boldsymbol{\Xi} = \boldsymbol{R}^{-1}\boldsymbol{\Gamma}$ forms a basis for all model-compliant trajectories with $\|\boldsymbol{\kappa}\|_1 = 3$, such that it naturally forms a basis for null $\boldsymbol{M}_{\boldsymbol{w}}$.

Due to the close connection with the affine eigentuples, a representation of the form (3) is most easily derived from normalized eigentuples, i.e. with

$\xi_0 = 1$. Notice that scaling the eigentuples by a factor c scales the associated signal by a constant factor as well: $(c\boldsymbol{\xi})^{\boldsymbol{\kappa}} = c^{\|\boldsymbol{\kappa}\|_1}\boldsymbol{\xi}^{\boldsymbol{\kappa}}$, so that the signals $(c\boldsymbol{\xi})^{\boldsymbol{\kappa}}$ span the same set of model-compliant trajectories, i.e. range $\boldsymbol{\Xi}$. As such, consider the normalization of the eigentuples $\boldsymbol{\xi}_1 \cong (1,1,-1), \boldsymbol{\xi}_2 \cong (1,-1,-1)$ and $\boldsymbol{\xi}_3 \cong (1,-1,1)$. The model compliant signals can then be written as $\hat{y}_{\boldsymbol{\kappa}} = x_1(1,1,-1)^{\boldsymbol{\kappa}} + x_2(1,-1,-1)^{\boldsymbol{\kappa}} + x_1(1,-1,1)^{\boldsymbol{\kappa}}$. Since each of the ξ_0-values is 1, its contributions in the exponentials can then essentially be ignored. By assigning the multivariate exponential terms to the state vector, a state-space model in the form of (3) can be defined as follows:

$$\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_1} = \begin{bmatrix} x_1(1,1,-1)^{\boldsymbol{\kappa}+\boldsymbol{e}_1} \\ x_2(1,-1,-1)^{\boldsymbol{\kappa}+\boldsymbol{e}_1} \\ x_3(1,-1,1)^{\boldsymbol{\kappa}+\boldsymbol{e}_1} \end{bmatrix} = \begin{bmatrix} 1 & & \\ & -1 & \\ & & -1 \end{bmatrix} \begin{bmatrix} x_1(1,1,-1)^{\boldsymbol{\kappa}+\boldsymbol{e}_0} \\ x_2(1,-1,-1)^{\boldsymbol{\kappa}+\boldsymbol{e}_0} \\ x_3(1,-1,1)^{\boldsymbol{\kappa}+\boldsymbol{e}_0} \end{bmatrix} = \boldsymbol{A}_{\xi_1}\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_0},$$

$$\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_2} = \begin{bmatrix} x_1(1,1,-1)^{\boldsymbol{\kappa}+\boldsymbol{e}_1} \\ x_2(1,-1,-1)^{\boldsymbol{\kappa}+\boldsymbol{e}_1} \\ x_3(1,-1,1)^{\boldsymbol{\kappa}+\boldsymbol{e}_1} \end{bmatrix} = \begin{bmatrix} -1 & & \\ & -1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} x_1(1,1,-1)^{\boldsymbol{\kappa}+\boldsymbol{e}_0} \\ x_2(1,-1,-1)^{\boldsymbol{\kappa}+\boldsymbol{e}_0} \\ x_3(1,-1,1)^{\boldsymbol{\kappa}+\boldsymbol{e}_0} \end{bmatrix} = \boldsymbol{A}_{\xi_2}\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_0},$$

$$\boldsymbol{c} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}.$$

To clearly distinguish between the state-transition matrices of this affine model and the ones of the model characterized by $\boldsymbol{r}(\boldsymbol{\sigma})$, we added the subscripts ξ_i.

Remark that not all choices of $\boldsymbol{c}$ are sufficient to fully describe all model-compliant trajectories: indeed choosing $\begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$ would not output the first exponential component, for example. However, there always exists a $\boldsymbol{c}$ that does lead to an n-dimensional space of trajectories.

When the eigentuples are unknown the state-transition matrices can be obtained from any basis for the trajectories of degree d, given that d is large enough. In practice, when the dimension of the null space basis for the associated Macaulay matrix stabilizes[7] with increasing degree d [3, 59]. In our example, it can be verified that $d = 3$ satisfies these conditions.

Now notice that the state-transition matrices satisfy the following shift equations, obtained by selecting the appropriate rows of $\boldsymbol{\Xi}$:

$$\begin{matrix} (3,0,0) \\ (2,1,0) \\ (2,0,1) \\ (1,2,0) \\ (1,1,1) \\ (1,0,2) \end{matrix} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \\ -1 & -1 & -1 \\ 1 & -1 & 1 \end{bmatrix} \boldsymbol{T}\boldsymbol{T}^{-1}\boldsymbol{A}_{\xi_1}\boldsymbol{T} = \frac{1}{8} \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{matrix} (2,1,0) \\ (1,2,0) \\ (1,1,1) \\ (0,3,0) \\ (0,2,1) \\ (0,1,2) \end{matrix} \boldsymbol{T},$$

$$\begin{matrix} (3,0,0) \\ (2,1,0) \\ (2,0,1) \\ (1,2,0) \\ (1,1,1) \\ (1,0,2) \end{matrix} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \\ -1 & -1 & -1 \\ 1 & -1 & 1 \end{bmatrix} \boldsymbol{T}\boldsymbol{T}^{-1}\boldsymbol{A}_{\xi_2}\boldsymbol{T} = \frac{1}{8} \begin{bmatrix} 1 & 1 & -1 \\ -1 & -1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix} \begin{matrix} (2,1,0) \\ (1,1,1) \\ (1,0,2) \\ (0,2,1) \\ (0,1,2) \\ (0,0,3) \end{matrix} \boldsymbol{T}.$$

[7] When there are roots at infinity, a so called 'gap' appears, which can be used to eliminate these roots using a column compression [14].

While $\boldsymbol{T}$ may be any (unknown) invertible matrix, any basis $\boldsymbol{\Xi T}$ suffices to determine $\boldsymbol{A}_{\xi_1}^{\text{new}} = \boldsymbol{T}^{-1}\boldsymbol{A}_{\xi_2}\boldsymbol{T}$ and $\boldsymbol{A}_{\xi_2}^{\text{new}} = \boldsymbol{T}^{-1}\boldsymbol{A}_{\xi_2}\boldsymbol{T}$, which can be used to build an equivalent state-space model, as previously indicated. Since all eligible state transition matrices are joint similarity transforms of $\boldsymbol{A}_{\xi_1}$ and $\boldsymbol{A}_{\xi_2}$, respectively, their eigentuples always correspond to the common roots of $\boldsymbol{w}(\boldsymbol{\xi})$. As such, it holds that $\boldsymbol{w}(\boldsymbol{A}_{\xi_1}^{\text{new}}, \boldsymbol{A}_{\xi_2}^{\text{new}}) = \boldsymbol{T}^{-1}\boldsymbol{w}(\boldsymbol{A}_{\xi_1}, \boldsymbol{A}_{\xi_2})\boldsymbol{T} = 0$. Remark that equations of this format also apply to systems with a more complex multiplicity structure.

Now that we built some intuition on the inner workings of the state-space format, we briefly discuss how to return to an image representation starting from a state-space model. Recall that an image representation simply requires a basis for all the linearly independent signals over the considered grid. As such, to obtain all possible model outputs, one could consider the set of signals $\boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}\boldsymbol{x}$ generated by n linearly independent initial state vectors $\boldsymbol{x}$, e.g. the columns of the identity matrix. It can be worked out that the model-compliant output associated with any other initial state can then be written as a linear combination of these basis signals. Capturing these trajectories in a matrix format then results in the *observability matrix*:

$$\forall \boldsymbol{\kappa} \in \mathbb{N}^{m+1}, \|\boldsymbol{\kappa}\|_1 = d : \boldsymbol{\Gamma}_{\boldsymbol{\kappa},:} = \boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}.$$

The columns of this matrix do not necessarily form a basis, however. In general, this observability matrix only guarantees to span the exact set of trajectories found in the null space of a Macaulay matrix once the degree of regularity is considered, an upper bound for which can be found in [3]. Computing such a matrix for the model considered in Example 4.1 results in a multiple of the basis $\boldsymbol{\Xi}$.

Note that the *shift-invariant structure* of $\boldsymbol{\Gamma}$ helps to shed some light on the shift equations that were set up in Example 4.1. This allows us to easily obtain a general form for these equations, which can be solved for $\boldsymbol{A}_k$ whenever the left-hand-side coefficient matrix is of full column rank:

$$\forall k \in \{1, \ldots, m\}, \|\boldsymbol{\kappa}\|_1 = d - 1 : \boldsymbol{\Gamma}_{\boldsymbol{\kappa}+\boldsymbol{e}_0,:}\boldsymbol{A}_k = \boldsymbol{\Gamma}_{\boldsymbol{\kappa}+\boldsymbol{e}_k,:}. \tag{4}$$

Since $\boldsymbol{A}_0 = \boldsymbol{I}$, one can choose $\boldsymbol{c} = \boldsymbol{\Gamma}_{d\boldsymbol{e}_0,:}$ to obtain a full state-space description.

4.2 Descriptor systems

To build a state-space model for models with eigentuples at infinity, the change of coordinates $\boldsymbol{P}$ can now be applied to $\boldsymbol{A}_{\boldsymbol{\xi}_j}$. This results in another set of state-transition matrices: $\forall k \in \{0, \ldots, m\} : \boldsymbol{A}_{\lambda_k} = (\sum_{l=0}^{m} p_{k,l}\boldsymbol{A}_{\xi_l})$. Now observe that trajectories of the form $\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{\boldsymbol{\kappa}}\boldsymbol{x}$, where $\boldsymbol{A}_{\boldsymbol{\lambda}} = (\boldsymbol{A}_{\lambda_0}, \ldots, \boldsymbol{A}_{\lambda_m})$, comply with $\boldsymbol{r}(\boldsymbol{\sigma})$:

$$\forall \|\boldsymbol{\kappa}\|_1 = d - \deg(\boldsymbol{r}(\boldsymbol{\sigma})) : \boldsymbol{r}(\boldsymbol{\sigma})\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{c}\boldsymbol{r}(\boldsymbol{A}_{\boldsymbol{\lambda}})\boldsymbol{A}_{\boldsymbol{\lambda}}^{\boldsymbol{\kappa}}\boldsymbol{x} = \boldsymbol{c}\boldsymbol{w}(\boldsymbol{A}_{\boldsymbol{\xi}})\boldsymbol{A}_{\boldsymbol{\lambda}}^{\boldsymbol{\kappa}}\boldsymbol{x} = \boldsymbol{0},$$

since $\boldsymbol{w}(\boldsymbol{A}_{\boldsymbol{\xi}}) = \boldsymbol{0}$. This definition of $\boldsymbol{A}_{\boldsymbol{\lambda}}$ also allows us to associate any signal of the form $\boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{\boldsymbol{\kappa}}\boldsymbol{x}$ with a signal $\boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\xi}}^{\boldsymbol{\gamma}}\boldsymbol{x}$ by an invertible linear combination over

the multi-indices $\boldsymbol{\gamma}$, where $\|\boldsymbol{\gamma}\|_1 = \|\boldsymbol{\kappa}\|_1$, much like in Example 3.2. As such, the former signals span an n-dimensional set of trajectories, forming a basis for the $\boldsymbol{r}(\boldsymbol{\sigma})$-compliant signals of degree d. Let us illustrate this using the running example.

Example 4.2. Applying the linear transformation defined by $\boldsymbol{P}$ from Example 3.2 to the state-transition matrices $\boldsymbol{A}_{\xi_i}$ obtained in Example 4.1 we obtain:

$$\boldsymbol{A}_0 = \boldsymbol{A}_{\xi_0} + \boldsymbol{A}_{\xi_1} = \begin{bmatrix} 2 & & \\ & 0 & \\ & & 0 \end{bmatrix}$$
$$\boldsymbol{A}_1 = \boldsymbol{A}_{\xi_1} + \boldsymbol{A}_{\xi_2} = \begin{bmatrix} 0 & & \\ & 2 & \\ & & 0 \end{bmatrix}$$
$$\boldsymbol{A}_2 = \boldsymbol{A}_{\xi_0} + \boldsymbol{A}_{\xi_2} = \begin{bmatrix} 0 & & \\ & 0 & \\ & & 2 \end{bmatrix}.$$

As one would expect, the eigentuples of this model capture the common roots $\boldsymbol{\lambda}_1 \cong (2,0,0), \boldsymbol{\lambda}_2 \cong (0,2,0)$ and $\boldsymbol{\lambda}_2 \cong (0,0,2)$ up to scaling, as is allowed for projective points. Choosing $\boldsymbol{c}$ as before, an observability matrix for this model can be set up, which forms a basis for the model-compliant signals, and thus for the null space of $\boldsymbol{M}_{\boldsymbol{r}}$:

$$\begin{bmatrix} \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{(3,0,0)} \\ \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{(2,1,0)} \\ \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{(2,0,1)} \\ \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{(1,2,0)} \\ \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{(1,1,1)} \\ \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{(1,0,2)} \\ \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{(0,3,0)} \\ \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{(0,2,1)} \\ \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{(0,1,2)} \\ \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{(0,0,3)} \end{bmatrix} = \begin{array}{c} \begin{matrix} (2,0,0)^{\boldsymbol{\kappa}} & (0,2,0)^{\boldsymbol{\kappa}} & (0,0,2)^{\boldsymbol{\kappa}} \end{matrix} \\ \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix} \end{array}.$$

Notice that the range of this matrix corresponds exactly to the range of $\boldsymbol{\Gamma}$ in Example 3.1 and thus describes the same model.

Now that the model-compliant trajectories are known, a suitable state-space description remains to be found. Since there is no need anymore to distinguish between the descriptor system coordinate-transformed system, we drop the λ-subscripts and denote $\boldsymbol{A}_k = \boldsymbol{A}_{\lambda_k}$ from hereon. To find the projective roots of set of polynomials, a set of 'rational' shift equations has been suggested in the literature [3, 7]. In the affine case, $\boldsymbol{A}_0$ could be assumed to be $\boldsymbol{I}$, so that its effects could easily be undone to decrease the 0th index. For descriptor

systems, $\boldsymbol{A}_0$ is not invertible, however. Defining a state sequence $\boldsymbol{x}_{\boldsymbol{\kappa}} = \boldsymbol{A}_{\boldsymbol{\lambda}}^{\boldsymbol{\kappa}} \boldsymbol{x}$ the following equality can also be derived $\boldsymbol{A}_k \boldsymbol{A}_l \boldsymbol{x}_{\boldsymbol{\kappa}} = \boldsymbol{A}_k \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_l} = \boldsymbol{A}_l \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_k}$, leading to the set of state-space equations:

$$\begin{aligned} \forall k, l \in \{0, \ldots, m\}, \boldsymbol{\kappa} \in \mathbb{N}_0^{m+1}, \|\boldsymbol{\kappa}\|_1 = d-1 : \boldsymbol{A}_l \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_k} &= \boldsymbol{A}_k \boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_l} \qquad (5) \\ \text{where } \boldsymbol{A}_k \boldsymbol{A}_l &= \boldsymbol{A}_l \boldsymbol{A}_k. \\ \hat{y}_{\boldsymbol{\kappa}} &= \boldsymbol{c} \boldsymbol{x}_{\boldsymbol{\kappa}}. \end{aligned}$$

Notice that by their earlier construction, the eigentuples of the $\boldsymbol{A}_k$-matrices are nonzero $m+1$-tuples[8], as they correspond to the solutions of $\boldsymbol{r}(\boldsymbol{\sigma})$ over $\mathbb{P}^m$.

The solution trajectory $\boldsymbol{x}_{\boldsymbol{\kappa}}$ of these state-transition equations remains a solution under right-multiplication of one of the equations by some invertible $\boldsymbol{T}$. Moreover, the matrices $\boldsymbol{T}\boldsymbol{A}_k$ then do not necessarily commute, providing an alternative model description to 5. However, keep in mind that constructing the state-trajectory is not as straightforward for such a representation. However, as these are a set of linear equations and may thus be solved as such.

A state-space model as in (5) is always possible, though. Indeed, it should be clear that all $\boldsymbol{r}(\boldsymbol{\sigma})$-compliant trajectories $\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\lambda}}^{\boldsymbol{\kappa}} \boldsymbol{x}$ satisfy (5). We now show the converse: any state sequence $\boldsymbol{x}_{\boldsymbol{\kappa}}$, where $\|\boldsymbol{\kappa}\|_1 = d$, that complies with (5) can be written in the form $\boldsymbol{A}_{\boldsymbol{\lambda}}^{\boldsymbol{\kappa}} \boldsymbol{x}$.

First, since the matrices $\boldsymbol{A}_k$ commute, there exists a basis $\boldsymbol{V}$ such that $\forall k : \boldsymbol{A}_k = \boldsymbol{V}\boldsymbol{J}_k\boldsymbol{V}^{-1}$, where the $\boldsymbol{J}_k$ are pairwise commutative, each $\boldsymbol{J}_i$ is block-diagonal

$$\boldsymbol{J}_k = \begin{bmatrix} \boldsymbol{J}_k^{(1)} & & \\ & \boldsymbol{J}_k^{(2)} & \\ & & \ddots \end{bmatrix},$$

each set of blocks $(\boldsymbol{J}_1^{(s)}, \ldots, \boldsymbol{J}_m^{(s)})$ is associated with precisely one eigentuple of $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m)$, and its dimension reflects the multiplicity of this eigentuple, as discussed in Appendix A.3 [25, 42]. Since these eigentuples correspond to projective points, there is always at least one nonzero element in each eigentuple. This implies that for each tuple $(\boldsymbol{J}_1^{(s)}, \ldots, \boldsymbol{J}_m^{(s)})$ there exists an index i such that $\boldsymbol{J}_i^{(s)}$ has nonzero eigenvalues and is thus invertible.

Since $\boldsymbol{V}$ is invertible, we can associate each $\boldsymbol{x}_{\boldsymbol{\kappa}}$ one-to-one with another state sequence as follows $\boldsymbol{x}_{\boldsymbol{\kappa}} = \boldsymbol{V}\boldsymbol{v}_{\boldsymbol{\kappa}}$. Substituting $\boldsymbol{A}_k = \boldsymbol{V}\boldsymbol{J}_k\boldsymbol{V}^{-1}$ and left-multiplying both sides by $\boldsymbol{V}^{-1}$, (5) results in an equivalent model that describes

[8] Note that this condition of nonzero eigentuples is equivalent with the condition that $\begin{bmatrix}\boldsymbol{A}_1^\mathsf{T} & \ldots & \boldsymbol{A}_m^\mathsf{T}\end{bmatrix}^\mathsf{T}$ is of full column rank and $\begin{bmatrix}\boldsymbol{A}_1 & \ldots & \boldsymbol{A}_m\end{bmatrix}$ is of full row rank.

$\boldsymbol{v}_{\boldsymbol{\kappa}}$:

$$\forall k,l \in \{0,\dots,n\}, \|\boldsymbol{\kappa}\|_1 = d-1: \quad (6)$$
$$\begin{bmatrix} \boldsymbol{J}_k^{(1)} & & \\ & \boldsymbol{J}_k^{(2)} & \\ & & \ddots \end{bmatrix} \boldsymbol{v}_{\boldsymbol{\kappa}+\boldsymbol{e}_l} = \begin{bmatrix} \boldsymbol{J}_l^{(1)} & & \\ & \boldsymbol{J}_l^{(2)} & \\ & & \ddots \end{bmatrix} \boldsymbol{v}_{\boldsymbol{\kappa}+\boldsymbol{e}_k}.$$

This essentially decouples (5) into a set of similar, lower-dimensional equations. Choosing $\boldsymbol{c}^{\text{new}} = \boldsymbol{cV}$ then outputs $\hat{y}_{\boldsymbol{\kappa}}$ compliant with (5).

As just shown, for each of these smaller models, there then exists an invertible state-transition matrix $\boldsymbol{J}_k^{(s)}$. Defining $\boldsymbol{J}^{(s)} = \left(\boldsymbol{J}_1^{(s)},\dots,\boldsymbol{J}_m^{(s)}\right)$ and assuming $\left(\boldsymbol{J}^{(s)}\right)^{d\boldsymbol{e}_k}$ is invertible, we can define $\boldsymbol{v}$ as the unique solution of $\boldsymbol{v}_{d\boldsymbol{e}_k} = \left(\boldsymbol{J}^{(s)}\right)^{d\boldsymbol{e}_k}\boldsymbol{v}$. By repeated application of (6), it holds that[9] $\left(\boldsymbol{J}^{(s)}\right)^{\boldsymbol{\gamma}}\boldsymbol{v}_{\boldsymbol{\kappa}} = \left(\boldsymbol{J}^{(s)}\right)^{\boldsymbol{\gamma}-\boldsymbol{e}_k}\left(\boldsymbol{J}_l^{(s)}\right)\boldsymbol{v}_{\boldsymbol{\kappa}-\boldsymbol{e}_l+\boldsymbol{e}_k} = \ldots = \left(\boldsymbol{J}^{(s)}\right)^{\boldsymbol{\kappa}}\boldsymbol{v}_{\boldsymbol{\gamma}}, \forall \boldsymbol{\gamma},\boldsymbol{\kappa} \in \mathbb{N}_0^m$ where $\|\boldsymbol{\gamma}\|_1 = \|\boldsymbol{\kappa}\|_1 = d$. Choosing $\boldsymbol{\gamma} = d\boldsymbol{e}_k$ then ensures that $\left(\boldsymbol{J}^{(s)}\right)^{\boldsymbol{\gamma}} = \left(\boldsymbol{J}^{(s)}\right)^{d\boldsymbol{e}_k}$ is invertible, such that[10]:

$$\left(\boldsymbol{J}^{(s)}\right)^{d\boldsymbol{e}_k}\boldsymbol{v}_{\boldsymbol{\kappa}} = \left(\boldsymbol{J}^{(s)}\right)^{\boldsymbol{\kappa}}\boldsymbol{v}_{d\boldsymbol{e}_k} \leftrightarrow \boldsymbol{v}_{\boldsymbol{\kappa}} = \left(\boldsymbol{J}^{(s)}\right)^{\boldsymbol{\kappa}}\left(\boldsymbol{J}^{(s)}\right)^{-d\boldsymbol{e}_k}\boldsymbol{v}_{d\boldsymbol{e}_k} = \left(\boldsymbol{J}^{(s)}\right)^{\boldsymbol{\kappa}}\boldsymbol{v}.$$

We now illustrate this model description for the running example.

Example 4.3. Continuing from example 4.2 and the equality $\boldsymbol{A}_{\lambda_i}\boldsymbol{A}_{\lambda_j}\boldsymbol{x}_{\boldsymbol{\kappa}} = \boldsymbol{A}_{\lambda_i}\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_l} = \boldsymbol{A}_{\lambda_j}\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_k}$ leads to the following state-space model:

$$\begin{aligned}
\boldsymbol{A}_0\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_1} &= \begin{bmatrix}1 & & \\ & 0 & \\ & & 0\end{bmatrix}\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_1} = \begin{bmatrix}0 & & \\ & 1 & \\ & & 0\end{bmatrix}\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_0} = \boldsymbol{A}_1\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_0} \\
\boldsymbol{A}_0\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_2} &= \begin{bmatrix}1 & & \\ & 0 & \\ & & 0\end{bmatrix}\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_2} = \begin{bmatrix}0 & & \\ & 0 & \\ & & 1\end{bmatrix}\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_0} = \boldsymbol{A}_2\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_0} \\
\boldsymbol{A}_2\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_1} &= \begin{bmatrix}0 & & \\ & 0 & \\ & & 1\end{bmatrix}\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_1} = \begin{bmatrix}0 & & \\ & 1 & \\ & & 0\end{bmatrix}\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_2} = \boldsymbol{A}_1\boldsymbol{x}_{\boldsymbol{\kappa}+\boldsymbol{e}_2} \\
\hat{y}_{\boldsymbol{\kappa}} &= \begin{bmatrix}1 & 1 & 1\end{bmatrix}\boldsymbol{x}_{\boldsymbol{\kappa}}.
\end{aligned}$$

Finally, notice that the equivalence between this state-space description and the signals spanned by multivariate the 'exponential modes' associated with the

[9] Remark that, for this argument in the general case, all combinations of indices (k,l) should be 'exchangeable' with one another by some equation of the form (5). Indeed, eliminating some of the equations for our running example would result in more viable trajectories. When one of the $\boldsymbol{A}_k$ is invertible, e.g. $\boldsymbol{A}_0$ in the affine case, it can be verified that a subset of these equations implies the rest as well.

[10] note that $\boldsymbol{AB} = \boldsymbol{BA} \rightarrow \boldsymbol{BA}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{ABA}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{BAA}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{B}$

eigentuples only holds when all grid points of a given degree are considered. For example, let us consider a subset of the points $\|\boldsymbol{\kappa}\|_1 = 2$. Writing out the equations by which we defined $(x_2)_{\boldsymbol{\kappa}}$ above over a subset of grid points, e.g. $(2,0,0), (1,1,0), (1,0,1), (0,1,1)$ results in the following linear system of equations:

$$\begin{bmatrix} 1 & 0 & & \\ & & 1 & 0 \\ 0 & & 0 & \\ & 0 & & 0 \\ & 0 & 1 & \end{bmatrix} \begin{bmatrix} (x_2)_{(2,0,0)} \\ (x_2)_{(1,1,0)} \\ (x_2)_{(1,0,1)} \\ (x_2)_{(0,1,1)} \end{bmatrix} = \mathbf{0}$$

Clearly, the sequence $\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\mathsf{T}$ satisfies this set of equations, while this does not comply with the associated multi-expontial $(x_2)_{\boldsymbol{\kappa}} = (0,1,0)^{\boldsymbol{\kappa}}$.

4.3 Realization theory for descriptor systems

The techniques of affine realization theory have to slightly adapted to successfully deal with state-transition matrices with eigentuples at infinity. Generally, to account for $\boldsymbol{A}_0$ it has to be added to the right-hand side of equation (4) [3]. This results in an equation with two unknown shift matrices. One may try to consider a different pair of shifts and to solve for the single unknown $\boldsymbol{A}_k^{-1}\boldsymbol{A}_l$, however, in general, all the $\boldsymbol{A}_k$ may be singular, e.g., in the running example. Moreover, a singular state-transition matrix also causes the coefficient matrix on the other side of the equation to be not of full rank. Nevertheless, as long as there are a finite number of projective roots, these issues can be circumvented using the techniques from [6, 7, 54]. From the descriptor observability matrix, the shift equations (4) can be easily adjusted, from which a set of invertible shift matrices and corresponding coefficient matrices can be made by linear combinations:

$$\begin{aligned} \forall k,l \in \{1,\ldots,m\}, \|\boldsymbol{\kappa}\|_1 = d-1 : \boldsymbol{\Gamma}_{\boldsymbol{\kappa}+\boldsymbol{e}_l,:}\boldsymbol{A}_k &= \boldsymbol{\Gamma}_{\boldsymbol{\kappa}+\boldsymbol{e}_k,:}\boldsymbol{A}_l \qquad (7) \\ \rightarrow \boldsymbol{\Gamma}_{\boldsymbol{\kappa}+\boldsymbol{e}_l,:}\sum_{k=0}^{m} f_k\boldsymbol{A}_k &= \sum_{k=0}^{m} f_k\boldsymbol{\Gamma}_{\boldsymbol{\kappa}+\boldsymbol{e}_k,:}\boldsymbol{A}_l \\ \leftrightarrow \left(\sum_{k=0}^{m} f_k\boldsymbol{\Gamma}_{\boldsymbol{\kappa}+\boldsymbol{e}_k,:}\right)^{-1} \boldsymbol{\Gamma}_{\boldsymbol{\kappa}+\boldsymbol{e}_l,:} &= \boldsymbol{A}_l \left(\sum_{k=0}^{m} f_k\boldsymbol{A}_k\right)^{-1} \end{aligned}$$

Again, almost all linear combinations suffice to ensure $\boldsymbol{A}_f$ is invertible. While Example 4.4 below illustrates these techniques for a null space in the generalized Vandermonde format, using a similarity transform on $\boldsymbol{A}_k\boldsymbol{A}_f^{-1}$, these procedures can be readily generalized to general bases for the trajectories, as already highlighted in the affine case [3].

Example 4.4. Reconsider $\boldsymbol{\Gamma}$ from Example (3.1). Choosing $f(\boldsymbol{\lambda}) = \boldsymbol{f}^\mathsf{T}\boldsymbol{\lambda}$ as a linear function of $\boldsymbol{\lambda}$, with $\boldsymbol{f} \in \mathbb{R}^{m+1}$ suffices to ensure that $f(\boldsymbol{\lambda})$ is nonzero in all

the eigentuples $\boldsymbol{\lambda}_i$. This causes the associated shift matrix $\boldsymbol{A}_f = f(\boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{A}_2)$ to be nonsingular as well. Writing out (7) in matrix format for $i = 1, \|\boldsymbol{\kappa}\|_1 = 2$ and $\boldsymbol{f} = \begin{bmatrix}3 & 2 & 1\end{bmatrix}^\mathsf{T}$ then results in:

$$\begin{matrix}3(3,0,0)+2(2,1,0)+(2,0,1)\\3(2,1,0)+2(1,2,0)+(1,1,1)\\3(1,0,1)+2(1,1,1)+(1,0,2)\\3(1,2,0)+2(0,3,0)+(0,2,1)\\3(1,1,1)+2(0,2,1)+(0,1,2)\\3(1,0,2)+2(0,1,2)+(0,0,3)\end{matrix}\begin{bmatrix}3&0&0\\0&0&0\\0&0&0\\0&2&0\\0&0&0\\0&0&1\end{bmatrix}\boldsymbol{A}_1 = \begin{matrix}(2,1,0)\\(1,2,0)\\(1,1,1)\\(0,3,0)\\(0,2,1)\\(0,1,2)\end{matrix}\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\\0&1&0\\0&0&0\\0&0&0\end{bmatrix}\boldsymbol{A}_f$$

Notice that this linear combination causes the left coefficient matrix to be left invertible and $\boldsymbol{A}_f = \mathrm{diag}(3, 2, 1)$ invertible as well, such that one can solve this system for $\boldsymbol{A}_1\boldsymbol{A}_f^{-1}$:

$$\boldsymbol{A}_1\boldsymbol{A}_f^{-1} = \begin{bmatrix}0&&\\&\frac{1}{2}&\\&&0\end{bmatrix}$$

A similar analysis applies to $\boldsymbol{A}_0\boldsymbol{A}_f^{-1}$ and $\boldsymbol{A}_2\boldsymbol{A}_f^{-1}$.

It can be seen that the eigentuples of the set of commuting matrices $\boldsymbol{A}_k\boldsymbol{A}_f^{-1}, i = 0, 1, 2$ then correspond to the values of the rational function $\frac{\lambda_i}{f(\boldsymbol{\lambda})}$ in the eigentuples $(1, 0, 0), (0, 1, 0)$ and $(0, 0, 1)$. Notice that these rational functions, where the numerator is of the same degree as the denominator, are well-defined on the projective eigentuples. That is, evaluating these functions in another representative for the projective points $\boldsymbol{\lambda}_i$, that is, a scaling $\boldsymbol{\lambda}$, does not affect the resulting value. Finally, notice that for nonzero $f(\boldsymbol{\lambda})$, $\frac{\boldsymbol{\lambda}}{f(\boldsymbol{\lambda})} \cong \boldsymbol{\lambda}$, so that the eigentuples can be obtained from these matrices in similar fashion as in the affine case.

In general, the matrices $\boldsymbol{A}_i\boldsymbol{A}_f^{-1}$ can be used directly as the state-transition matrices in (5). Indeed, multiplying both sides of the equations by $\boldsymbol{A}_f^{-1}$ does not alter the solutions $\boldsymbol{x}_{\boldsymbol{\kappa}}$. However, constructing trajectories as $\boldsymbol{c}(\boldsymbol{A}_1\boldsymbol{A}_f^{-1}, \ldots, \boldsymbol{A}_m\boldsymbol{A}_f^{-1})^{\boldsymbol{\kappa}}\boldsymbol{x}$ leads to $\boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}\boldsymbol{A}_f^{-d}\boldsymbol{x}$, so that the corresponding observability matrix differs from the original one by the invertible factor $\boldsymbol{A}_f^{-d}$.

5 Data-based realization

Now that we have discussed all the state-space representation, we possess the necessary tools to discuss a special case of the image representation: the data-based representation. Indeed, for LTI systems, it is well-known that, under some conditions, a single trajectory can be used to infer a basis for all possible trajectories using the shift-invariance property [35, 46]. By commutativity of the shift operators, this also applies to the model class under consideration: $\boldsymbol{r}(\boldsymbol{\sigma})\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{0} \to \boldsymbol{r}(\boldsymbol{\sigma})y_{\boldsymbol{\kappa}+\boldsymbol{\gamma}} = \boldsymbol{\sigma}^{\boldsymbol{\gamma}}\boldsymbol{r}(\boldsymbol{\sigma})\hat{y}_{\boldsymbol{\kappa}} = \boldsymbol{0}$. This observation has been used in

the mD case as well to define a generalized Hankel matrix and build a state-space model using multidimensional realization theory [2, 22]. Assuming a state-space representation of the form 5, and given a row and column degree d_r and d_c, respectively, the elements of such a matrix can be compactly defined using multi-indices:

$$\forall \|\boldsymbol{\rho}\|_1 = d_r, \|\boldsymbol{\gamma}\|_1 = d_c : \boldsymbol{H}_{\boldsymbol{\rho},\boldsymbol{\gamma}} = y_{\boldsymbol{\rho}+\boldsymbol{\gamma}} = \boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\rho}+\boldsymbol{\gamma}}\boldsymbol{x} = \boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\rho}}\boldsymbol{x}_{\boldsymbol{\gamma}}.$$

This matrix can then be decomposed as

$$\boldsymbol{H} = \boldsymbol{\Gamma}\boldsymbol{X}, \tag{8}$$

where $\boldsymbol{\Gamma}$ is a degree d_r observability matrix and the columns of $\boldsymbol{X}$ are $\boldsymbol{x}_{\boldsymbol{\gamma}}$. As such, the rank of $\boldsymbol{H}$ is at most n. Moreover, we know that, when $\operatorname{rank}\boldsymbol{H} = n$, it holds that $\operatorname{range}\boldsymbol{H} = \operatorname{range}\boldsymbol{\Gamma}$, so that a basis for its column space can be used as a basis for all model-compliant trajectories of degree d_r. This basis can then be employed to find a kernel representation or a state-space model using the techniques previously described.

Conditions do apply for the rank of $\boldsymbol{H}$ to become n though: first and foremost $\boldsymbol{X}$ should also be of full rank, which is not the case for all $\boldsymbol{x}$, much like some output matrices $\boldsymbol{c}$ are not suitable either, as mentioned in Example 4.2. This will be discussed in-depth in the next subsection. In practice, the model order can be detected by building such Hankel-like matrices of increasing row degrees[11], until the rank stabilizes, similarly to building the null space of a Macaulay matrix [14, 59].

Example 5.1. Consider a data sequence of degree $\|\boldsymbol{\kappa}\|_1 = 6$, e.g.:

$$y_{(6,0,0)} = 1,\ y_{(0,6,0)} = 2,\ y_{(0,0,6)} = 3,$$
$$y_{\boldsymbol{\kappa}} = 0, \text{ for other } \boldsymbol{\kappa}, \text{with } \|\boldsymbol{\kappa}\|_1 = 6.$$

A generalized Hankel matrix can then be built with column and row indices up

[11] Note that decreasing the column degree while increasing the row degree may hamper the growth in rank. However, as will be detailed in the next subsection, the lowest of both degrees is the 'bottleneck' for the rank, so keeping as $d_c \geq d_r$ suffices.

to degree three:

$$\boldsymbol{H} = \begin{matrix} \\ (6,0,0) \\ (5,1,0) \\ \vdots \\ (3,0,3) \end{matrix} \overset{\begin{matrix} (3,0,0) & (2,1,0) & \dots & (0,0,3) \end{matrix}}{\begin{bmatrix} y_{(6,0,0)} & y_{(5,1,0)} & \cdots & y_{(3,0,3)} \\ y_{(5,1,0)} & y_{(4,2,0)} & \cdots & y_{(2,1,3)} \\ \vdots & \vdots & \vdots & \vdots \\ y_{(3,0,3)} & y_{(2,1,3)} & \cdots & y_{(0,0,6)} \end{bmatrix}} \begin{matrix} (3,0,0) \\ (2,1,0) \\ \\ (0,0,3) \end{matrix}$$

$$= \begin{bmatrix} 1 & & & & & & & & & \\ & 0 & & & & & & & & \\ & & 0 & & & & & & & \\ & & & 0 & & & & & & \\ & & & & 0 & & & & & \\ & & & & & 0 & & & & \\ & & & & & & 2 & & & \\ & & & & & & & 0 & & \\ & & & & & & & & 0 & \\ & & & & & & & & & 3 \end{bmatrix}.$$

Clearly, this matrix is of rank 3, as was expected from the model order.

Next, computing a basis $\boldsymbol{\Gamma}$ for range $\boldsymbol{H}$ results in the observability matrix that was previously encountered for the running example:

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} (3,0,0) \\ (2,1,0) \\ (2,0,1) \\ (1,2,0) \\ (1,1,1) \\ (1,0,2) \\ (0,3,0) \\ (0,2,1) \\ (0,1,2) \\ (0,0,3) \end{matrix}.$$

The techniques of section 4.3 can then be applied to construct a state-space model, as done in Example 4.4.

5.1 Order of the identified model

It was already mentioned that some values of $\boldsymbol{c}$ may lead to the observability matrix to be rank-deficient. Indeed, while an n system characterized by difference equations always allows a representation of the form (5), the converse is not guaranteed: In similar fashion as for LTI systems, a state-space model of this form may describe a lower-order model due to some *unobservable modes*. The model can also be split in an observable and an unobservable part, only the former of which can be reconstructed through realization theory [2]. To provide more insight in the algebraic structure of such models as well as a a succinct

check for such modes, we now present a multidimensional generalization of the Popov-Belevitch-Hautus (PBH) test for observability [29]. There exist multiple notions of controllability and observability in the multidimensional behavioral systems theory [43], which we do not explicitly explore. Instead, we focus on the rank of $\boldsymbol{\Gamma}$. A more general, highly algebraic account using the behavioral framework for systems theory can be found in [33], though our presentation strongly differs from this. A very similar analysis applies to $\boldsymbol{X}$ in (8), though with left instead of right eigenvectors.

To state the result, we first note that for any set of m commuting matrices $\boldsymbol{A}_0, \ldots, \boldsymbol{A}_m \in \mathbb{C}^{n\times n}$, there exists some finite degree d^* so that for all $\boldsymbol{\gamma} : \boldsymbol{A}^{\gamma} = \sum_{\|\boldsymbol{\kappa}\|_1 \leq d^*} l_{\boldsymbol{\kappa}} \boldsymbol{A}^{\boldsymbol{\kappa}}$, i.e. all monomials in these matrices are linear combinations the monomials up to degree d^*. Indeed, by the Cayley-Hamilton theorem, all monomials in $\boldsymbol{A}_0, \ldots, \boldsymbol{A}_m \in \mathbb{C}^{n\times n}$ can be written in terms of the finite set $\{\boldsymbol{A}^{\boldsymbol{\kappa}} \mid \|\boldsymbol{\kappa}\|_\infty \leq n-1\}$, a superset of all linearly independent $\boldsymbol{A}$-monomials. Smaller sets of monomials exist, however. For example, [58] provides a smaller superset[12] which includes all degree monomials of $n-1$. To understand the PBH-like property, only the existence of this degree is of importance.

Theorem 5.1. *Let $\boldsymbol{\Gamma}$ be the degree $d_r \geq d^*$ observability matrix associated with a state-space model (5). It then holds that* $\operatorname{rank} \boldsymbol{\Gamma} < n$ *iff there exists a joint eigenvector $\boldsymbol{x}$ of the commuting family $\boldsymbol{A}_k, k = 0, \ldots, m$ such that $\boldsymbol{c}\boldsymbol{x} = 0$.*

Proof. Let us start with the 'if' part. If there is such a joint eigenvector $\boldsymbol{x}$, associated with some eigentuple $\boldsymbol{\lambda}$, then $\boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}\boldsymbol{x} = \boldsymbol{c}\boldsymbol{\lambda}^{\boldsymbol{\kappa}}\boldsymbol{x} = 0, \forall \boldsymbol{\kappa}$. By its definition, it then holds that $\boldsymbol{\Gamma}$ has a nontrivial right null space, since $\boldsymbol{\Gamma}_{\boldsymbol{\kappa}}\boldsymbol{x} = 0, \forall \boldsymbol{\kappa}$.

Now, let us consider the 'only if' part. As discussed in Section 4.2 we can transform $\boldsymbol{\Gamma}$ into the observability matrix associated with a model with $\boldsymbol{A}_{\boldsymbol{\xi}} = \sum_{l=0}^{m} p_{k,l} \boldsymbol{A}_l$, which has no eigentuples at infinity. This transformation then captured by the invertible matrix $\boldsymbol{R}$: $\boldsymbol{\Xi} = \boldsymbol{R}\boldsymbol{\Gamma}$. Due to the invertibility of $\boldsymbol{R}$, both these observability matrices have the same right null space.

As $\boldsymbol{A}_{\xi_0} = \boldsymbol{I}$, the rows of $\boldsymbol{\Xi}$ consist of all vectors of the form $\boldsymbol{c}\prod_{k=1}^{m} \boldsymbol{A}_{\xi_1}^{\boldsymbol{\kappa}_k}$, with $\sum_{k=1}^{m} \boldsymbol{\kappa}_k \leq d_r$. Since all $\boldsymbol{A}_{\boldsymbol{\xi}}$-monomials can be written in terms of the degree $\leq d^*$ $\boldsymbol{A}_{\boldsymbol{\xi}}$-monomials[13], any vector of the form $\boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\xi}}^{\boldsymbol{\kappa}}$ can be written as a linear combination of the rows of $\boldsymbol{\Xi}$. Consequently, if $\boldsymbol{\Xi}\boldsymbol{x} = \boldsymbol{0}$, then for any $\boldsymbol{\kappa}$ and k there exists a vector $\boldsymbol{l} : \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\xi}}^{\boldsymbol{\kappa}}(\boldsymbol{A}_{\xi_k}\boldsymbol{x}) = \boldsymbol{c}\boldsymbol{A}_{\boldsymbol{\xi}}^{\boldsymbol{\kappa}+\boldsymbol{e}_k}\boldsymbol{x} = \boldsymbol{l}\boldsymbol{\Xi}\boldsymbol{x} = 0$. We conclude that $\forall k : \boldsymbol{A}_{\xi_k}(\operatorname{null}\boldsymbol{\Xi}) \subseteq \operatorname{null}\boldsymbol{\Xi}$, that is, $\operatorname{null}\boldsymbol{\Xi}$ is an invariant subspace of all $\boldsymbol{A}_{\xi_k}$, and by linearity, of the original state-transition matrices $\boldsymbol{A}_k$.

Now, every joint invariant subspace of a set of commuting matrices contains a joint eigenvector [25], which we briefly prove for didactical purposes: Let $\mathcal{X}$ be an $\boldsymbol{P}$- and $\boldsymbol{T}$-invariant subspaces of $\mathbb{C}^n$, where $\boldsymbol{P}$ and $\boldsymbol{T}$ commute. As

[12] In general, any normal set of a border basis for the ideal $I = \{p(\boldsymbol{\lambda}) \in \mathbb{C}[\boldsymbol{\lambda}] \mid p(\boldsymbol{A}) = \boldsymbol{0}\}$ provides a set of linearly independent matrix monomials. In practice, such a set can be found through what is essentially a 'block Buchberger-Möller' algorithm.

[13] It can be seen that d^* is the same for monomials in $\boldsymbol{A}_{\boldsymbol{\xi}}$ and $\boldsymbol{A}$: Using the coordinate transform, write the $\boldsymbol{A}_{\boldsymbol{\xi}}$-monomial in terms of $\boldsymbol{A}$-monomials of the same degree. Next, write those in terms of the lower degree $\boldsymbol{A}$-monomials. Finally express the result in terms of lower degree $\boldsymbol{A}_{\boldsymbol{\xi}}$-monomials again.

discussed in Appendix A.2, each $\boldsymbol{P}$-invariant subspace contains an eigenvector of $\boldsymbol{P}$ belonging to some eigenspace $\mathcal{V}_\lambda$. Due to commutativity $\boldsymbol{T}\mathcal{V}_\lambda \subseteq \mathcal{V}_\lambda$ and $\boldsymbol{T}\mathcal{X} \subseteq \mathcal{X}$, so that $\boldsymbol{T}(\mathcal{V}_\lambda \cap \mathcal{X}) \subseteq \mathcal{V}_\lambda \cap \mathcal{X} \neq \{0\}$. That is, the intersection is $\boldsymbol{T}$-invariant as well, and thus, there lies a $\boldsymbol{T}$-eigenvector in $\mathcal{V}_\lambda \cap \mathcal{X}$.

Consequently, there exists some joint eigenvector $\boldsymbol{x} \in \operatorname{null} \boldsymbol{\Xi}$ of all $\boldsymbol{A}_k$ such that $\boldsymbol{c}\boldsymbol{A}^{\boldsymbol{\kappa}}\boldsymbol{x} = \boldsymbol{c}\boldsymbol{\lambda}^{\boldsymbol{\kappa}}\boldsymbol{v} = 0, \|\boldsymbol{\kappa}\|_1 = d_r$. Since the eigentuples are nonzero, $\boldsymbol{\lambda}^{\boldsymbol{\kappa}}$ is also nonzero for some $\boldsymbol{\kappa}$ and thus $\boldsymbol{c}\boldsymbol{x} = 0$. □

6 Applications

The preceding sections showed that autonomous, scalar, linear, shift-invariant models can output finite sums of multivariate exponential signals. Such signals are common solution components of linear, homogeneous partial differential equations (PDE's) with constant coefficients. Of course, there exist many PDE's that that are incompatible with such separable solution components as well. Nonetheless, the considered model class may still serve as approximation tools for PDE's [8, 23, 24, 62]. The formal study of PDE's, the approximation thereof and partial difference equations is a vast field of research in and of itself, outside the scope of this thesis. For an overview, we refer the reader to [26, 39, 52]. Let illustrate how the considered model class can model a finite set of solution components of a sampled solution to the wave equation.

Example 6.1. Consider the 1D wave equation with Dirichlet boundary conditions over the x-domain:

$$\begin{aligned} \frac{\partial y(x,t)}{\partial t^2} &= c^2 \frac{\partial y(x,t)}{\partial x^2} \text{ for } 0 < x < L, \\ y(0,t) &= 0 = y(0,t). \end{aligned}$$

For an appropriate set of initial conditions $y(x,0) = \phi(x), \frac{\partial y(x,0)}{\partial t} = \psi(x)$, the solution can be found as [52]:

$$y(x,t) = \sum_{i=1}^{\infty} \left(a_i \cos\left(\frac{n\pi ct}{L}\right) + b_i \sin\left(\frac{n\pi ct}{L}\right) \right) \sin\left(\frac{n\pi x}{L}\right),$$

where a_n and b_n are the coefficients of the Fourier series expansion of $y(x,0)$.

Now, we sample this solution at discrete grid points $(k\Delta x, l\Delta t)$ in time and space and rewrite it as follows:

$$\begin{aligned} y_{\boldsymbol{\kappa}} = \frac{1}{2} \sum_{i=1}^{\infty} & a_i \sin\left(\frac{i\pi(\kappa_1 \Delta x + c\kappa_2 \Delta t)}{L}\right) + a_i \sin\left(\frac{i\pi(\kappa_1 \Delta x - c\kappa_2 \Delta t)}{L}\right) \\ & + b_i \cos\left(\frac{i\pi(\kappa_1 \Delta x + c\kappa_2 \Delta t)}{L}\right) + b_i \cos\left(\frac{i\pi(\kappa_1 \Delta x - c\kappa_2 \Delta t)}{L}\right). \end{aligned}$$

As such, choosing r_i and θ_i such that $b_i + a_i = 2r_i \cos(\theta_i)$ and $b_i - a_i = 2r_i \sin(\theta_i)$, we conclude that the contribution of the ith *solution mode* corresponds to the output of the following fourth-order system:

$$\boldsymbol{A}_1 = \begin{bmatrix} \cos\left(\frac{i\pi\Delta x}{L}\right) & \sin\left(\frac{i\pi\Delta x}{L}\right) & & \\ -\sin\left(\frac{i\pi\Delta x}{L}\right) & \cos\left(\frac{i\pi\Delta x}{L}\right) & & \\ & & \cos\left(\frac{i\pi\Delta x}{L}\right) & \sin\left(\frac{i\pi\Delta x}{L}\right) \\ & & -\sin\left(\frac{i\pi\Delta x}{L}\right) & \cos\left(\frac{i\pi\Delta x}{L}\right) \end{bmatrix},$$

$$\boldsymbol{A}_2 = \begin{bmatrix} \cos\left(\frac{i\pi c\Delta t}{L}\right) & \sin\left(\frac{i\pi c\Delta t}{L}\right) & & \\ -\sin\left(\frac{i\pi c\Delta t}{L}\right) & \cos\left(\frac{i\pi c\Delta t}{L}\right) & & \\ & & \cos\left(\frac{i\pi c\Delta t}{L}\right) & -\sin\left(\frac{i\pi c\Delta t}{L}\right) \\ & & \sin\left(\frac{i\pi c\Delta t}{L}\right) & \cos\left(\frac{i\pi c\Delta t}{L}\right) \end{bmatrix},$$

$$\boldsymbol{c} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} r_i \sin\left(\frac{i\pi(\kappa_1\Delta x + c\kappa_2\Delta t)}{L} + \theta_i\right) \\ r_i \cos\left(\frac{i\pi(\kappa_1\Delta x + c\kappa_2\Delta t)}{L} + \theta_i\right) \\ r_i \sin\left(\frac{i\pi(\kappa_1\Delta x - c\kappa_2\Delta t)}{L} + \theta_i\right) \\ r_i \cos\left(\frac{i\pi(\kappa_1\Delta x - c\kappa_2\Delta t)}{L} + \theta_i\right) \end{bmatrix}.$$

By extension, any solutions to this wave equation consisting of $s \in \mathbb{N}$ nonzero *modes* can be described by a $4s$th order state-space model of the form (3), by joining s models of the form above. As the general solutions can have an infinite number of modes, these models are almost always an approximation. Then again, good approximations are often the best one can hope for when dealing with PDE's. Note that the sampling can cause aliasing to arise, resulting in some *modes* becoming indistinguishable from one another, so that $y_{\boldsymbol{\kappa}}$ becomes a sum of a finite number of modes.

Since an array physical phenomena is governed by such linear PDE's, these complex exponentials are a useful tool for analyzing multidimensional signals. Most importantly, the field of multidimensional harmonic retrieval often aims to decompose signals into such components [15, 27, 47, 50, 56, 57]. Many signal processing applications on the rely on these techniques, for example: analysis of wireless communications channels [21, 37], nuclear magnetic resonance spectroscopy [19, 32, 61], blind source separation [49] and direction of arrival [20], among others.

7 Conclusion

We briefly introduced the model class under consideration in this thesis, that is, a straightforward multidimensional generalization of autonomous LTI descriptor systems. More precisely, its intricate connection with multivariate polynomials with a zero-dimensional set of common roots, and, by extension, the Macaulay framework, was explored, resulting in a kernel, image and state-space representation for the model-compliant trajectories. The difference linear-algebra based

transformations between these representations were also briefly highlighted. Finally, a data-based variation of the image representation using of a generalized Hankel matrix was mentioned, as well as potential applications for this class of models.

A Jordan canonical form, commuting matrices and invariant subspaces

In this section, some useful properties of commuting matrices are outlined. As previously shown, commuting matrices arise in state-space models associated with linear, shift-invariant mD models and, similarly, as multiplication matrices in polynomial coordinate rings.

Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ be a pair of commuting matrices. Commutativity implies some interesting interactions between the (generalized) eigenvectors of $\boldsymbol{A}, \boldsymbol{B}$. Before we get to these properties, we first take take a closer look at the generalized eigenvectors and the Jordan canonical form (JCF).

A.1 Jordan canonical form and invariant subspaces

For a detailed treatment of the JCF, we refer the reader to [10, 17, 25].

A vector $\boldsymbol{v}$ is a rank k generalized eigenvector of $\boldsymbol{A}$ associated with the eigenvalue λ if:

$$(\boldsymbol{A} - \lambda \boldsymbol{I})^k \boldsymbol{v} = \boldsymbol{0},$$
$$(\boldsymbol{A} - \lambda \boldsymbol{I})^{k-1} \boldsymbol{v} \neq \boldsymbol{0}.$$

Note that generalized eigenvectors associated with different eigenvalues linearly independent, as will become evident from their construction in the derivations below, and by the linear independence of eigenvectors associated with different eigenvalues. The generalized eigenspaces thus span a set of disjoint vector spaces.

A.1.1 Construction of generalized eigenvectors

These vectors can be found though the computation of the Jordan Canonical form. Moreover, the generalized eigenvectors associated with some eigenvalue λ form a chain (or multiple chains): the rank k generalized eigenvectors are mapped onto the rank $k-1$ eigenvectors by $(\boldsymbol{A} - \lambda \boldsymbol{I})$, and so on. Finally, the rank 1 generalized eigenvectors are just eigenvectors associated with λ, such that they lie in the null space of $(\boldsymbol{A} - \lambda \boldsymbol{I})$. The total dimension of the space spanned by the generalized eigenvectors is equal to the algebraic multiplicity of λ. We formalize this with the following derivation.

Let λ be an eigenvalue of $\boldsymbol{A}$ with algebraic multiplicity m_a and geometric multiplicity m_g (each matrix has at least one eigenvalue and associated eigenvector, since $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$ has a solution). Now let $\boldsymbol{V}$ be an orthogonal basis

for the m_g-dimensional eigenspace associated with λ and let $\boldsymbol{Q}$ be an orthogonal basis for the orthogonal complement of this space. We then perform the similarity transformation:

$$\begin{bmatrix}\boldsymbol{V}^\mathsf{T}\\ \boldsymbol{Q}^\mathsf{T}\end{bmatrix}(\boldsymbol{A}-\lambda\boldsymbol{I})\begin{bmatrix}\boldsymbol{V} & \boldsymbol{Q}\end{bmatrix} = \begin{bmatrix}\boldsymbol{0} & \boldsymbol{V}^\mathsf{T}(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{Q}\\ \boldsymbol{0} & (\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}-\lambda\boldsymbol{I})\end{bmatrix}.$$

Now, since this matrix is block triangular, its characteristic equation is the product of the characteristic equations of the diagonal blocks: $\det(\boldsymbol{A}-\lambda\boldsymbol{I}) = (x-\lambda)^{m_g}\det\big(\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}-\lambda\boldsymbol{I}\big)$. It is then clear that $\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}$ has the eigenvalue λ with an algebraic multiplicity $m_a - m_g$.

Let $\boldsymbol{V}_2$ be the eigenspace of $\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}$ w.r.t. λ and note that:

$$\begin{aligned}\begin{bmatrix}\boldsymbol{V}^\mathsf{T}(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{Q}\\ (\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}-\lambda\boldsymbol{I})\end{bmatrix}\boldsymbol{V}_2 &= \begin{bmatrix}\boldsymbol{V}^\mathsf{T}(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{Q}\boldsymbol{V}_2\\ \boldsymbol{0}\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{V}^\mathsf{T}\\ \boldsymbol{Q}^\mathsf{T}\end{bmatrix}(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{Q}\boldsymbol{V}_2.\end{aligned}$$

Because $\begin{bmatrix}\boldsymbol{V} & \boldsymbol{Q}\end{bmatrix}$ is an orthogonal basis for $\mathbb{R}^n$, it is clear that $(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{Q}\boldsymbol{V}_2$ lies fully in the subspace spanned by $\boldsymbol{V}$, i.e. the eigenspace w.r.t. λ. So under the map $(\boldsymbol{A}-\lambda\boldsymbol{I})$, the vectors in $\operatorname{range}\boldsymbol{Q}\boldsymbol{V}_2$ get mapped onto $\operatorname{range}\boldsymbol{V}$, which get mapped onto $\boldsymbol{0}$. As such $\boldsymbol{Q}\boldsymbol{V}_2$ clearly span the rank 2 generalized eigenvectors. If the full algebraic multiplicity is not accounted for yet, this procedure can be continued for $\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}\boldsymbol{Q}$ to find the higher rank eigenvectors.

A.1.2 Jordan chains and elementary invariant subspaces

Next, let $\boldsymbol{p}, \boldsymbol{q}$ be two linearly independent rank $k > 1$ generalized eigenvectors and let them be orthogonal to all rank $< k$ generalized eigenvectors, as per the construction above. Notice that these vectors cannot be mapped to the same rank $k-1$ generalized eigenvector $\boldsymbol{v}$ under $(\boldsymbol{A}-\lambda\boldsymbol{I})$. Indeed, it would then hold that $(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{p} = (\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{q} = \boldsymbol{v}$, which implies that $\boldsymbol{q}-\boldsymbol{p}$ lies in the null space of $(\boldsymbol{A}-\lambda\boldsymbol{I})$. However, this is impossible since they are orthogonal to all eigenvectors associated with λ. By taking linear combinations, this argument can be easily extended to show that any number of rank-k vectors is to be mapped to the same number of rank $k-1$ vectors, implying that each linearly independent rank k eigenvector defines its own chain of lower-rank eigenvectors, disjoint from the space spanned by other rank k chains.

These generalized eigenvectors thus define chains of vectors that are mapped onto one another by $(\boldsymbol{A}-\lambda\boldsymbol{I})$ from a higher rank to a lower rank. The space spanned by the vectors in such a chain then forms an *invariant subspace* of the matrix $\boldsymbol{A}$, that is, a subspace $\operatorname{range}\boldsymbol{V}$ such that $\operatorname{range}\boldsymbol{A}\boldsymbol{V} \subseteq \operatorname{range}\boldsymbol{V}$. A well-known example of such an invariant subspace is the space spanned by an eigenvector $\boldsymbol{v}$ of $\boldsymbol{A}$: $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \in \operatorname{range}\boldsymbol{v}$. Indeed, by the observations above, one can choose a basis $\boldsymbol{V}_\lambda \in \mathbb{R}^{n\times m_a}$ with as its columns generalized eigenvectors, such that at each rank $k > 1$ column of $\boldsymbol{V}_\lambda$ is mapped onto exactly one rank $k-1$ column by $(\boldsymbol{A}-\lambda\boldsymbol{I})$. In this basis, the map defined by $\boldsymbol{A}$ is described by

the Jordan block $\boldsymbol{J}_\lambda$ associated with λ, defined as: $\boldsymbol{J}_\lambda = \boldsymbol{V}_\lambda{}^\dagger \boldsymbol{A} \boldsymbol{V}_\lambda$. For example, let the ith column be mapped onto the $(i-1)$th column:

$$\begin{aligned} \boldsymbol{V}_\lambda \boldsymbol{e}_{i-1} &= (\boldsymbol{A} - \lambda \boldsymbol{I}) \boldsymbol{V}_\lambda \boldsymbol{e}_i \\ \boldsymbol{e}_{i-1} &= \boldsymbol{V}_\lambda{}^\dagger \boldsymbol{V}_\lambda (\boldsymbol{J}_\lambda - \lambda \boldsymbol{I}) \boldsymbol{V}_\lambda{}^\dagger \boldsymbol{V}_\lambda \boldsymbol{e}_i \\ \boldsymbol{e}_{i-1} &= (\boldsymbol{J}_\lambda - \lambda \boldsymbol{I}) \boldsymbol{e}_i. \end{aligned}$$

Similar bases can be constructed for all q eigenvalues of $\boldsymbol{A}$, leading to the full JCF, using the basis $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_{\lambda_1} & \cdots & \boldsymbol{V}_{\lambda_q} \end{bmatrix}$.

Finally, notice that, since no other generalized eigenvector maps onto $\boldsymbol{V}_\lambda \boldsymbol{e}_{i-1}$, this also implies that $\boldsymbol{e}_{i-1}{}^\mathsf{T} (\boldsymbol{J}_\lambda - \lambda \boldsymbol{I}) = \boldsymbol{e}_i{}^\mathsf{T}$, such $\boldsymbol{A}$ maps the $(i-1)$th row to the ith row of $\boldsymbol{V}_\lambda{}^\dagger$. As such, any chain of right generalized eigenvectors,

$$\boldsymbol{V}_\lambda \boldsymbol{e}_i \xrightarrow{(\boldsymbol{A} - \lambda \boldsymbol{I})} \boldsymbol{V}_\lambda \boldsymbol{e}_{i-1} \xrightarrow{(\boldsymbol{A} - \lambda \boldsymbol{I})} \dots \xrightarrow{(\boldsymbol{A} - \lambda \boldsymbol{I})} \boldsymbol{V}_\lambda \boldsymbol{e}_{i-l}$$

implies the existence of the following "reversed" chain of left generalized eigenvectors:

$$\boldsymbol{e}_{i-l}{}^\mathsf{T} \boldsymbol{V}_\lambda{}^\dagger \xrightarrow{(\boldsymbol{A} - \lambda \boldsymbol{I})} \boldsymbol{e}_{i-l+1}{}^\mathsf{T} \boldsymbol{V}_\lambda{}^\dagger \xrightarrow{(\boldsymbol{A} - \lambda \boldsymbol{I})} \dots \xrightarrow{(\boldsymbol{A} - \lambda \boldsymbol{I})} \boldsymbol{e}_i{}^\mathsf{T} \boldsymbol{V}_\lambda{}^\dagger. \tag{9}$$

Example A.1. To illustrate these rather abstract observations, let us take a closer look at an example JCF of some matrix with an eigenvalue λ, with an algebraic multiplicity of six, involving two Jordan chains: one of length four and one of length two. As was just derived, there exists a basis $\boldsymbol{V}_\lambda$ that leads to the following form:

$$\boldsymbol{J}_\lambda = \boldsymbol{V}_\lambda{}^\dagger \boldsymbol{A} \boldsymbol{V}_\lambda = \left[\begin{array}{cccccc} \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & \lambda & 1 & & \\ & & & \lambda & 0 & \\ \hline & & & & \lambda & 1 \\ & & & & & \lambda \end{array}\right].$$

Notice how, for example, $\boldsymbol{J}_\lambda$ maps $\boldsymbol{e}_4$ as: $\boldsymbol{J}_\lambda \boldsymbol{e}_4 = \lambda \boldsymbol{e}_4 + \boldsymbol{e}_3$, or, in the original coordinates, that $\boldsymbol{A} \boldsymbol{V}_\lambda \boldsymbol{e}_4 = \boldsymbol{V}_\lambda (\boldsymbol{e}_4 + \boldsymbol{e}_3)$, since range $\boldsymbol{A} \boldsymbol{V}_\lambda \subseteq$ range $\boldsymbol{V}_\lambda$. However, $\boldsymbol{J}_\lambda \boldsymbol{e}_5 = \lambda \boldsymbol{e}_5$, is not mapped to $\boldsymbol{e}_4$, breaking the Jordan form into two blocks, each corresponding to one Jordan chain. As such, $\boldsymbol{V}_\lambda \boldsymbol{e}_1, \boldsymbol{V}_\lambda \boldsymbol{e}_5$ are the eigenvectors of $\boldsymbol{A}$ associated with λ. Each block then corresponds to an invariant subspace of $\boldsymbol{A}$ range $\boldsymbol{A} \boldsymbol{V}_\lambda \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 & \boldsymbol{e}_4 \end{bmatrix} \subseteq$ range $\boldsymbol{V}_\lambda \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 & \boldsymbol{e}_4 \end{bmatrix}$ and, similarly, range $\boldsymbol{A} \boldsymbol{V}_\lambda \begin{bmatrix} \boldsymbol{e}_5 & \boldsymbol{e}_6 \end{bmatrix} \subseteq$ range $\boldsymbol{V}_\lambda \begin{bmatrix} \boldsymbol{e}_5 & \boldsymbol{e}_6 \end{bmatrix}$. Similarly, $\boldsymbol{e}_4{}^\mathsf{T} \boldsymbol{V}_\lambda{}^\dagger$ and $\boldsymbol{e}_6{}^\mathsf{T} \boldsymbol{V}_\lambda{}^\dagger$ are left eigenvectors of $\boldsymbol{A}$ and the Jordan chains now correspond to ascending rows of $\boldsymbol{V}_\lambda{}^\dagger$, e.g. $\boldsymbol{e}_3{}^\mathsf{T} \boldsymbol{V}_\lambda{}^\dagger \boldsymbol{A} = (\lambda \boldsymbol{e}_3{}^\mathsf{T} + \boldsymbol{e}_4{}^\mathsf{T}) \boldsymbol{V}_\lambda{}^\dagger$.

A.2 Joint Schur triangularization

For more in-depth depositions on joint triangularization, we refer the reader to [25, 42].

Now that we have some more intuition about the JCF, which describes the elementary invariant subspaces of a matrix, i.e. a single Jordan chain, let us consider the case with two commuting matrices $\boldsymbol{A}, \boldsymbol{B}$. Notice that for any eigenvector $\boldsymbol{v}$ of $\boldsymbol{A}$, it holds that:

$$\boldsymbol{ABv} = \boldsymbol{BAv} = \lambda \boldsymbol{Bv},$$

implying that $\boldsymbol{Bv}$ is again an eigenvector of $\boldsymbol{A}$ associated with λ. As such, the eigenvectors of $\boldsymbol{A}$ associated with λ form a $\boldsymbol{B}-$invariant subspace. Formally, let $\boldsymbol{V}$ be an orthogonal basis for this eigenspace of $\boldsymbol{A}$, it holds that range $\boldsymbol{BV} \subseteq$ range $\boldsymbol{V}$.

Now, any $\boldsymbol{B}$-invariant space range $\boldsymbol{V}$ contains an eigenvector of $\boldsymbol{B}$. Indeed, take an eigenvector x: $\boldsymbol{V}^\mathsf{T}\boldsymbol{BV}\boldsymbol{x} = \mu \boldsymbol{x}$ and note that $\boldsymbol{VV}^\mathsf{T}\boldsymbol{p}$ is an identity operation for any vector $\boldsymbol{p} \in$ range $\boldsymbol{V}$. By range $\boldsymbol{BV} \subseteq$ range $\boldsymbol{V}$, it thus holds that $\boldsymbol{B}(\boldsymbol{Vx}) = \boldsymbol{VV}^\mathsf{T}\boldsymbol{B}(\boldsymbol{Vx}) = \mu \boldsymbol{Vx}$, i.e. $\boldsymbol{Vx}$ is an eigenvector of $\boldsymbol{B}$ and clearly lies in range $\boldsymbol{V}$.

Since the $\boldsymbol{B}$-invariant space $\boldsymbol{V}$ is the eigenspace of $\boldsymbol{A}$ in this case, this then leads to the existence of a common Schur transformation: Take $\boldsymbol{v}$ to be a common eigenvector of $\boldsymbol{A}, \boldsymbol{B}$ and let $\boldsymbol{Q}$ be an orthogonal basis for the orthogonal complement of $\boldsymbol{v}$. Now take the following common, orthogonal similarity transformation:

$$\begin{aligned}
\boldsymbol{A}_\text{sim} &= \begin{bmatrix} \boldsymbol{v}^\mathsf{T} \\ \boldsymbol{Q}^\mathsf{T} \end{bmatrix} \boldsymbol{A} \begin{bmatrix} \boldsymbol{v} & \boldsymbol{Q} \end{bmatrix} = \begin{bmatrix} \lambda & \boldsymbol{v}^\mathsf{T}\boldsymbol{AQ} \\ \boldsymbol{0} & \boldsymbol{Q}^\mathsf{T}\boldsymbol{AQ} \end{bmatrix} \\
\boldsymbol{B}_\text{sim} &= \begin{bmatrix} \boldsymbol{v}^\mathsf{T} \\ \boldsymbol{Q}^\mathsf{T} \end{bmatrix} \boldsymbol{B} \begin{bmatrix} \boldsymbol{v} & \boldsymbol{Q} \end{bmatrix} = \begin{bmatrix} \mu & \boldsymbol{v}^\mathsf{T}\boldsymbol{BQ} \\ \boldsymbol{0} & \boldsymbol{Q}^\mathsf{T}\boldsymbol{BQ} \end{bmatrix}.
\end{aligned}$$

Note that $\boldsymbol{A}_\text{sim}$ and $\boldsymbol{B}_\text{sim}$ also commute, such that from the bottom-right corner of this equality, it follows that $\boldsymbol{Q}^\mathsf{T}\boldsymbol{BQQ}^\mathsf{T}\boldsymbol{AQ} = \boldsymbol{Q}^\mathsf{T}\boldsymbol{AQQ}^\mathsf{T}\boldsymbol{BQ}$. As such, there again exists a common eigenvector for these matrices, such that the procedure can be repeated until a fully upper triangular form is achieved for both matrices.

This argument can be readily extended to any number of pairwise commuting matrices. For example, consider three commuting matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{c}$. As was previously shown, there exist eigenvalues of λ and μ $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Consequently, range $\boldsymbol{V}_\lambda \cap$ range $\boldsymbol{V}_\mu =$ range $\boldsymbol{V}_{\lambda,\mu} \neq \{\boldsymbol{0}\}$. By the $\boldsymbol{C}$-invariance of $\boldsymbol{V}_\lambda$ and $\boldsymbol{V}_\mu$, it holds that range $\boldsymbol{CV}_{\lambda,\mu} \subseteq$ range $\boldsymbol{V}_\lambda \cap$ range $\boldsymbol{V}_\mu =$ range $\boldsymbol{V}_{\lambda,\mu}$, such that this space is also $\boldsymbol{C}$-invariant and thus contains an eigenvector of $\boldsymbol{C}$.

The existence of a joint eigenvector suffices to perform a step of the joint Schur decomposition. By bringing all commuting matrices in a jointly triangular form, which reveals the eigenvalues on the diagonal, the corresponding diagonal entries can be paired into m-tuples: the eigentuples.

A.3 Jordan form and commutativity

For a more elaborate treatment of this subject we refer the reader to [17, 42] and more precisely Theorem 2.4.6.3 in [25].

A.3.1 Decomposition into joint invariant subspaces associated with eigentuples

Instead of looking at just the eigenvectors, let us take a closer look at the generalized eigenspaces of commuting matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Again, notice that:

$$(\boldsymbol{A} - \lambda \boldsymbol{I})^k \boldsymbol{B}\boldsymbol{v} = \boldsymbol{B}(\boldsymbol{A} - \lambda \boldsymbol{I})^k \boldsymbol{v} = \boldsymbol{0}$$

While it may or may not hold that $\boldsymbol{B}(\boldsymbol{A} - \lambda \boldsymbol{I})^{k-1}\boldsymbol{v} \neq \boldsymbol{0}$, it is clear that $\boldsymbol{B}\boldsymbol{v}$ is a generalized eigenvector of $\boldsymbol{A}$ rank $\leq k$ associated with λ. So, letting $\boldsymbol{V}_\lambda$ be the basis associated with the λ-Jordan block of $\boldsymbol{A}$, then range $\boldsymbol{V}_\lambda$ is $\boldsymbol{B}$-invariant. Moreover, let the columns of $\boldsymbol{V}_\lambda^{(k)}$ be the subset of rank $\leq k$ generalized eigenvectors in $\boldsymbol{V}_\lambda$, then range $\boldsymbol{V}_\lambda^{(k)}$ is $\boldsymbol{B}$-invariant for any k.

Let us look at how the generalized eigenspaces of $\boldsymbol{A}$ interact with those of $\boldsymbol{B}$. Let $\boldsymbol{V}_\mu$ be a basis for a generalized eigenspace of $\boldsymbol{B}$ associated with an eigenvalue μ, i.e. the space spanned by all generalized eigenvectors associated with μ. As was previously established, the elementary invariant subspaces of a matrix are spanned by a chain of generalized eigenvectors, i.e. the generalized eigenvectors associated with one subblock of the JCF. As such, the $\boldsymbol{B}$-invariant subspace range $\boldsymbol{V}_\lambda$ can be written as a direct sum of such elementary invariant subspaces.

Consequently, there must exist some values of λ and μ such that their intersection range $\boldsymbol{V}_\lambda \cap$ range $\boldsymbol{V}_\mu =$ range $\boldsymbol{V}_{\mu,\lambda}$ is nontrivial, i.e. range $\boldsymbol{V}_{\mu,\lambda} \neq \{\boldsymbol{0}\}$. Moreover, notice that $\boldsymbol{V}_\mu$ is $\boldsymbol{A}$-invariant by the same argument used to show that $\boldsymbol{V}_\lambda$ is $\boldsymbol{B}$-invariant, such that range $\boldsymbol{A}\boldsymbol{V}_{\mu,\lambda} \subseteq$ range $\boldsymbol{V}_\mu$. By the $\boldsymbol{A}$-invariance of range $\boldsymbol{V}_\lambda$, it also holds that range $\boldsymbol{A}\boldsymbol{V}_{\mu,\lambda} \subseteq$ range $\boldsymbol{V}_\lambda$, implying that $\boldsymbol{A}\boldsymbol{V}_{\mu,\lambda} \subseteq$ range $\boldsymbol{V}_\lambda \cap$ range $\boldsymbol{V}_\mu =$ range $\boldsymbol{V}_{\lambda,\mu}$. Range $\boldsymbol{V}_{\lambda,\mu}$ thus defines an $\boldsymbol{A}$-invariant subspace, and by a similar argument, a $\boldsymbol{B}$-invariant subspace.

Because generalized eigenspaces associated with different eigenvalues are disjoint spaces, it holds that range $\boldsymbol{V}_{\lambda_1,\mu} \cap$ range $\boldsymbol{V}_{\lambda_2,\mu} = \{\boldsymbol{0}\}$ whenever $\lambda_1 \neq \lambda_2$, and similarly for distinct values of μ. In such cases, the $\boldsymbol{B}$-invariant subspace range $\boldsymbol{V}_\mu$ thus contains at least two disjoint invariant subspaces: range $\boldsymbol{V}_{\lambda_1,\mu}$ and range $\boldsymbol{V}_{\lambda_2,\mu}$, showing that there are at least two Jordan blocks associated with μ. This shows that there is at least one joint invariant subspace associated with each distinct eigentuple (λ, μ), the dimension of which is no greater than the individual generalized eigenspaces associated with λ and μ.

Note that this argument can be readily extended to a larger number of commuting matrices to show that there always exists an invariant subspace associated with each distinct eigentuple.

A.3.2 Interaction with different Jordan chains

Remark that range $\boldsymbol{V}_\lambda^{(k)}$ being $\boldsymbol{B}$-invariant has strong implications about how generalized eigenvectors of $\boldsymbol{A}$ can be mapped under $\boldsymbol{B}$. Indeed, this implies that $\boldsymbol{B}$ can map generalized eigenvectors associated with a λ-Jordan block of $\boldsymbol{A}$ to linear combinations involving other eigenvectors associated with different λ-Jordan chains, as long as the resulting generalized eigenvector is at most of

the same rank[14]. For a rank k generalized eigenvector, this can then be written out as $\boldsymbol{B}\boldsymbol{V}_\lambda^{(k)}\boldsymbol{v} = \boldsymbol{V}_\lambda^{(k)}\boldsymbol{x}$. Let $\mathcal{N}$ denote the ordered set of column indices corresponding to the nonzero columns of $(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{V}_\lambda^{(k)}$, which form the basis $\boldsymbol{V}_\lambda^{(k-1)}$. Then the next vector in the corresponding chain is then mapped as

$$\boldsymbol{B}\boldsymbol{V}_\lambda^{(k-1)}\boldsymbol{v}_{\mathcal{N}} = \boldsymbol{B}(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{V}_\lambda^{(k)}\boldsymbol{v} = (\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{V}_\lambda^{(k)}\boldsymbol{x} = \boldsymbol{V}_\lambda^{(k-1)}\boldsymbol{x}_{\mathcal{N}}. \quad (10)$$

Notice that $(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{V}_\lambda^{(k)}\boldsymbol{v}$ is the rank $k-1$ generalized eigenvector in the chain defined by $\boldsymbol{V}_\lambda^{(k)}\boldsymbol{v}$, and similarly, the columns of $\boldsymbol{V}_\lambda^{(k-1)}$ are the next nonzero vectors in the Jordan chains defined by the columns of $\boldsymbol{V}_\lambda^{(k)}$. In a sense, $\boldsymbol{B}$ thus maps Jordan chains to linear combinations of other Jordan chains: if $\boldsymbol{V}_\lambda^{(k)}\boldsymbol{v}$ is mapped to a linear combination of generalized eigenvectors, any "shift" of $\boldsymbol{V}_\lambda^{(k)}\boldsymbol{v}$ is mapped to **the same linear combination** of the "shifts" of those generalized eigenvectors.

A similar analysis can then be carried out for the left generalized eigenvectors to show that rank k left eigenvectors of $\boldsymbol{A}$ can only be mapped to linear combinations of rank $\leq k$ left eigenvectors associated with the same eigenvalue by $\boldsymbol{B}$. Notice that, when looking at the map described by $\boldsymbol{B}$ in the JCF basis, i.e. $\boldsymbol{V}_\lambda{}^\dagger\boldsymbol{B}\boldsymbol{V}_\lambda$, then the entry at position (j,i) being nonzero, that is, $\boldsymbol{e}_j{}^\mathsf{T}\boldsymbol{V}_\lambda{}^\dagger\boldsymbol{B}\boldsymbol{V}_\lambda\boldsymbol{e}_i \neq 0$, not only implies that the generalized right eigenvector defined $\boldsymbol{V}_\lambda\boldsymbol{e}_i$ is mapped to a linear combination involving $\boldsymbol{V}_\lambda\boldsymbol{e}_j$, but it also implies that the left eigenvector $\boldsymbol{e}_j{}^\mathsf{T}\boldsymbol{V}_\lambda{}^\dagger$ is mapped onto a linear combination involving $\boldsymbol{e}_i{}^\mathsf{T}\boldsymbol{V}_\lambda{}^\dagger$. However, the rank of the right generalized eigenvector $\boldsymbol{V}\boldsymbol{e}_i$ is not equal to that of the left eigenvector $\boldsymbol{e}_i{}^\mathsf{T}\boldsymbol{V}^\dagger$. Rather, by (9), if the former is of rank k_1 and the total length of the chain l_1, then the latter is of rank l_1-k_1+1. Consequently, a vector of this first chain can only be mapped to linear combinations involving a rank k_2 vector in a different Jordan chain of length l_2, if $k_2 \leq k_1$ and $l_1-k_1 \leq l_2-k_2 \to l_1-l_2+k_2 \leq k_1$. This ensures that only the highest rank (left or right) eigenvectors in a chain can map onto vectors in shorter λ-Jordan chains under $\boldsymbol{B}$, imposing additional zero entries in $\boldsymbol{V}_\lambda{}^\dagger\boldsymbol{B}\boldsymbol{V}_\lambda$.

Example A.2. Let us illustrate all the observations above by looking at a matrix $\boldsymbol{B}$ that commutes with $\boldsymbol{A}$ from Example A.1. Consider the map $\boldsymbol{B}$ defines in the JCF basis $\boldsymbol{V}_\lambda$ used in said example:

$$\boldsymbol{V}_\lambda{}^\dagger\boldsymbol{B}\boldsymbol{V}_\lambda = \left[\begin{array}{cccc|cc} x_4 & x_3 & x_2 & x_1 & y_2 & y_1 \\ & x_4 & x_3 & x_2 & & y_2 \\ & & x_4 & x_3 & & \\ & & & x_4 & & \\ \hline & & x_6 & x_5 & y_6 & y_5 \\ & & & x_6 & & y_6 \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{C} & \boldsymbol{D} \\ \hline \boldsymbol{E} & \boldsymbol{F} \end{array}\right].$$

[14]These interactions can be better captured using the lesser-known Weyr canonical form, rather than a Jordan form. Indeed, all members of a commuting family of matrices can be put in a simultaneous upper triangular form, while one of the matrices is in Weyr from [25, 42].

Notice that the rank 4 eigenvector $\boldsymbol{V}_\lambda \boldsymbol{e}_4$ is mapped onto $\boldsymbol{V}_\lambda \boldsymbol{x}$ under $\boldsymbol{B}$. Since this is the highest rank eigenvector, it can be mapped to any other eigenvector. The Toeplitz structure of $\boldsymbol{C}$ and $\boldsymbol{D}$ is explained by (10). Indeed, notice that the next element in this Jordan chain is $\boldsymbol{V}_\lambda \boldsymbol{e}_3 = (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{V}_\lambda \boldsymbol{e}_4$ and that $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{V}_\lambda = \boldsymbol{V}_\lambda \begin{bmatrix} \boldsymbol{0} & \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 & \boldsymbol{0} & \boldsymbol{e}_5 \end{bmatrix}$, such that $\boldsymbol{B}\boldsymbol{V}_\lambda \boldsymbol{e}_3 = \boldsymbol{V}_\lambda \begin{bmatrix} x_2 & x_3 & x_4 & \boldsymbol{0} & x_6 & \boldsymbol{0} \end{bmatrix}^\mathsf{T}$. The same reasoning can be applied to subsequent elements in this Jordan chain to explain the full Toeplitz structure.

The upper triangular structure in both $\boldsymbol{C}$ and $\boldsymbol{D}$ can be explained by the "rank conditions": $k_2 \leq k_1$ and $2 + k_2 \leq k_1$. Taking a closer look at $\boldsymbol{D}$, note that only the eigenvectors $\boldsymbol{V}_\lambda \boldsymbol{e}_3, \boldsymbol{V}_\lambda \boldsymbol{e}_4$, of rank $k_1 = 3$ and $k_1 = 4$, respectively, can map onto $\boldsymbol{V}_\lambda \boldsymbol{e}_5, \boldsymbol{V}_\lambda \boldsymbol{e}_6$, which are of ranks $k_1 = 1$ and $k_1 = 2$. This constrains the first two columns of $\boldsymbol{D}$ to be zero and imposes the upper triangular structure.

For the second Jordan chain, one can also consider the image of the largest rank vector: $\boldsymbol{B}\boldsymbol{V}_\lambda \boldsymbol{e}_6 \boldsymbol{V}_\lambda \boldsymbol{y}$. In this case, the first "rank conditions" explains the upper triangular structure and the fact that $y_3 = y_4 = 0$, while the Toeplitz structure is explained by (10).

Lastly, note that, since each eigentuple (λ, μ) defines its own invariant subspace, the off-diagonal blocks $\boldsymbol{D}$ and $\boldsymbol{E}$ can only be nonzero if all eigenvalues of ${\boldsymbol{V}_\lambda}^\dagger \boldsymbol{B}\boldsymbol{V}_\lambda$ are equal. Indeed, in case there are two different eigenvalues, it was established that there exist two $\boldsymbol{A}$- and $\boldsymbol{B}$- invariant subspaces $\boldsymbol{V}_{\lambda,\mu_1}$ and $\boldsymbol{V}_{\lambda,\mu_2}$. Clearly, the only two $\boldsymbol{A}$-invariant subspaces associated with λ are spanned by the bases corresponding to each Jordan block: $\operatorname{range} \boldsymbol{V}_\lambda \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 & \boldsymbol{e}_4 \end{bmatrix}$ and $\operatorname{range} \boldsymbol{V}_\lambda \begin{bmatrix} \boldsymbol{e}_5 & \boldsymbol{e}_6 \end{bmatrix}$. As these generalized eigenspaces are $\boldsymbol{B}$-invariant subspaces, it holds that $\boldsymbol{B}\boldsymbol{V}_\lambda \boldsymbol{e}_6 = \boldsymbol{V}_\lambda \begin{bmatrix} \boldsymbol{e}_5 & \boldsymbol{e}_6 \end{bmatrix} \boldsymbol{y}$, such that $y_1 = y_2 = 0$ and $\boldsymbol{B}\boldsymbol{V}_\lambda \boldsymbol{e}_4 = \boldsymbol{V}_\lambda \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 & \boldsymbol{e}_4 \end{bmatrix} \boldsymbol{x}$, such that $x_5 = x_6 = 0$.

References

[1] John Abbott et al. "Computing Ideals of Points". In: *J. Symb. Comput.* 30 (Oct. 2000), pp. 341–356.

[2] S. Attasi. "Modelling and Recursive Estimation for Double Indexed Sequences". In: *System Identification Advances and Case Studies*. Vol. 126. Elsevier, 1976, pp. 289–348.

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

[4] K. Batselier and N. Wong. "Computing the state difference equations for discrete overdetermined linear mD systems". In: *Automatica* 64 (2016), pp. 254–261.

[5] Kim Batselier and Ngai Wong. "Inverse multivariate polynomial root-finding: Numerical implementations of the affine and projective Buchberger–Möller algorithm". In: *Journal of Computational and Applied Mathematics* 320 (February 2017 Aug. 2017), pp. 15–29.

[6] Matías R. Bender and Simon Telen. "Toric Eigenvalue Methods for Solving Sparse Polynomial Systems". In: *Mathematics of Computations* 91 (2020), pp. 2397–2429.

[7] Matías R. Bender and Simon Telen. *Yet another eigenvalue algorithm for solving polynomial systems*. 2022. arXiv: 2105.08472. URL: https://arxiv.org/abs/2105.08472.

[8] Gregory Beylkin and Martin J. Mohlenkamp. "Algorithms for Numerical Analysis in High Dimensions". In: *SIAM Journal on Scientific Computing* 26.6 (2005), pp. 2133–2159.

[9] J-E Björk. *Rings of differential operators*. Vol. 21. Elsevier, 1979.

[10] Richard Bronson. *Matrix methods: An introduction*. Gulf Professional Publishing, 1991.

[11] Robert M. Corless, Patrizia M. Gianni, and Barry M. Trager. "A reordered Schur factorization method for zero-dimensional polynomial systems with multiple roots". In: *Proc. of the International Symposium on Symbolic and Algebraic Computation*. 1997, pp. 133–140.

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

[13] B. H. Dayton and Z. Zeng. "Computing the multiplicity structure in solving polynomial systems". In: *Proc. of the International Symposium on Symbolic and Algebraic Computation* (2005), pp. 116–123.

[14] P. Dreesen, K. Batselier, and B. De Moor. "Multidimensional realisation theory and polynomial system solving". In: *International Journal of Control* 91.12 (2018), pp. 2692–2704.

[15] Mohammed Nabil El Korso et al. "Statistical resolution limit for the multidimensional harmonic retrieval model: hypothesis test and Cramér-Rao Bound approaches". In: *EURASIP Journal on Advances in Signal Processing* 2011.1 (2011), p. 12.

[16] E. Fornasini, P. Rocha, and S. Zampieri. "State Space Realization of 2-D Finite-Dimensional Behaviours". In: *SIAM Journal on Control and Optimization* 31.6 (1993), pp. 1502–1517.

[17] Feliks Ruvimovich Gantmacher. *The Theory of Matrices, Volume 1*. Chelsea Publishing Company, 1977.

[18] Feliks Ruvimovich Gantmacher. *The Theory of Matrices, Volume 2*. Chelsea Publishing Company, 1974.

[19] Di Guo, Hengfa Lu, and Xiaobo Qu. "A Fast Low Rank Hankel Matrix Factorization Reconstruction Method for Non-Uniformly Sampled Magnetic Resonance Spectroscopy". In: *IEEE Access* 5 (2017), pp. 16033–16039.

[20] 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.

[21] 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*. Vol. 1. 1997, 255–258 vol.1.

[22] B. Hanzon. "Attasi nD Systems and Polynomial System Solving". In: *IFAC-PapersOnLine* 54 (9 2021), pp. 145–150.

[23] S.H. Hashemi, B. Boroomand, and B. Movahedian. "Exponential basis functions in space and time: A meshless method for 2D time dependent problems". In: *Journal of Computational Physics* 241 (2013), pp. 526–545.

[24] Ralf Hiptmair, Andrea Moiola, and Ilaria Perugia. "A survey of Trefftz methods for the Helmholtz equation". In: *Building bridges: connections and challenges in modern approaches to numerical partial differential equations*. Springer, 2016, pp. 237–279.

[25] Roger A Horn and Charles R Johnson. *Matrix Analysis*. 2nd. Cambridge University Press, 2013.

[26] P. Hydon. *Difference Equations by Differential Equation Methods*. Cambridge University Press, 2014, p. 143.

[27] Tao Jiang, Nicholas D Sidiropoulos, and Jos MF Ten Berge. "Almost-sure identifiability of multidimensional harmonic retrieval". In: *IEEE Transactions on Signal Processing* 49.9 (2001), pp. 1849–1859.

[28] T. Kaczorek. "The singular general model of 2D systems and its solution". In: *IEEE Transactions on Automatic Control* 33.11 (1988), pp. 1060–1061.

[29] T Kailath. *Linear Systems*. Series Title: Prentice-Hall information and system sciences series. Prentice-Hall International, 1998.

[30] F. L. Lewis. "A survey of linear singular systems". In: *Circuits Systems and Signal Processing* 5 (1986).

[31] F.L. Lewis. "A review of 2-D implicit systems". In: *Automatica* 28.2 (1992), pp. 345–354.

[32] Y. Li, J. Razavilar, and K.J. Ray Liu. "A high-resolution technique for multidimensional NMR spectroscopy". In: *IEEE Transactions on Biomedical Engineering* 45.1 (1998), pp. 78–86.

[33] Vakhtang Lomadze. "The PBH test for multidimensional LTID systems". In: *Automatica* 49.9 (2013), pp. 2933–2937.

[34] Maria Grazia Marinari, H. Michael Möller, and Teo Mora. "Gröbner bases of ideals defined by functionals with an application to ideals of projective points". In: *Applicable Algebra in Engineering, Communication and Computing* 4 (1993), pp. 103–145.

[35] Ivan Markovsky and Florian Dörfler. "Behavioral systems theory in data-driven analysis, signal processing, and control". In: *Annual Reviews in Control* 52 (Jan. 1, 2021), pp. 42–64.

[36] SA Miri and J Dwight Aplevich. "Modeling and realization of n-dimensional linear discrete systems". In: *Multidimensional Systems and Signal Processing* 9.3 (1998), pp. 241–253.

[37] K.N. Mokios et al. "On 3D harmonic retrieval for wireless channel sounding". In: *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 2. 2004, pp. ii–89.

[38] H. M. Möller and B. Buchberger. "The construction of multivariate polynomials with preassigned zeros". In: *Computer Algebra*. Springer Berlin Heidelberg, 1982, pp. 24–31.

[39] Keith W Morton and David Francis Mayers. *Numerical solution of partial differential equations: an introduction*. Cambridge university press, 2005.

[40] B Mourrain. "A new criterion for normal form algorithms". In: *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes. AAECC 1999* (1999), pp. 430–443.

[41] B. Mourrain. "Isolated points, duality and residues". In: *Journal of Pure and Applied Algebra* 117-118 (1997), pp. 469–493.

[42] Kevin C. O'Meara, John Clark, and Charles I. Vinsonhaler. *Advanced Topics in Linear Algebra*. Oxford University Press, 2011.

[43] U. Oberst. "Multidimensional constant linear systems". In: *Acta Applicandae Mathematicae* 20 (1-2 1990), pp. 1–175.

[44] Ulrich Oberst. "Variations on the fundamental principle for linear systems of partial differential and difference equations with constant coefficients". In: *Applicable Algebra in Engineering, Communication and Computing* 6.4 (1995), pp. 211–243.

[45] Ulrich Oberst and Franz Pauer. "The constructive solution of linear systems of partial difference and differential equations with constant coefficients". In: *Multidimensional Systems and Signal Processing* 12.3 (2001), pp. 253–308.

[46] P. Van Overschee and B. De Moor. "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems". In: *Automatica* 30 (1 1994), pp. 75–93.

[47] S. Sahnoun, K. Usevich, and P. Comon. "Multidimensional ESPRIT for Damped and Undamped Signals: Algorithm, Computations, and Perturbation Analysis". In: *IEEE Transactions on Signal Processing* 65.22 (2017), pp. 5897–5910.

[48] J. Schur. "Zur Theorie der vertauschbaren Matrizen." In: *Journal für die reine und angewandte Mathematik* 130 (1905), pp. 66–76.

[49] Mikael Sørensen and Lieven De Lathauwer. "Blind Signal Separation via Tensor Decomposition With Vandermonde Factor: Canonical Polyadic Decomposition". In: *IEEE Transactions on Signal Processing* 61.22 (2013), pp. 5507–5519.

[50] Mikael Sørensen and Lieven De Lathauwer. "Multidimensional harmonic retrieval via coupled canonical polyadic decomposition—Part I: Model and identifiability". In: *IEEE Transactions on Signal Processing* 65.2 (2016), pp. 517–527.

[51] H. J. Stetter. *Numerical Polynomial Algebra*. SIAM, 2004, p. 57.

[52] Walter A Strauss. *Partial differential equations: An introduction*. John Wiley & Sons, 2007.

[53] Bernd Sturmfels. *Solving Systems of Polynomial Equations*. American Mathematical Society, 2002.

[54] Simon Telen. "Numerical root finding via Cox rings". In: *Journal of Pure and Applied Algebra* 224.9 (2020), p. 106367.

[55] Simon Telen. "Solving Systems of Polynomial Equations". PhD thesis. KU Leuven, 2020.

[56] Jeroen Vanderstukken and Lieven De Lathauwer. "Systems of polynomial equations, higher-order tensor decompositions, and multidimensional harmonic retrieval: A unifying framework. Part I: The canonical polyadic decomposition". In: *SIAM Journal on Matrix Analysis and Applications* 42.2 (Jan. 2021), pp. 883–912.

[57] Jeroen Vanderstukken et al. "Systems of polynomial equations, higher-order tensor decompositions, and multidimensional harmonic retrieval: A unifying framework. part II: The block term decomposition". In: *SIAM Journal on Matrix Analysis and Applications* 42.2 (2021), pp. 913–953.

[58] Lukas Vanpoucke and Bart De Moor. "Constructing Multidimensional Difference Equations From a State-Space Representation Using the Generalized Cayley–Hamilton Theorem". In: *IEEE Control Systems Letters* 8 (2024), pp. 2259–2264.

[59] Christof Vermeersch. "The (Block) Macaulay Matrix: Solving Systems of Multivariate Polynomial Equations and Multiparameter Eigenvalue Problems". eng. PhD thesis. KU Leuven, 2023.

[60] J.C. Willems. "From time series to linear system - Part I. Finite dimensional linear time invariant systems". In: *Automatica* 22 (5 1986), pp. 561–580.

[61] Jiaxi Ying et al. "Hankel Matrix Nuclear Norm Regularized Tensor Completion for N-dimensional Exponential Signals". In: *IEEE Transactions on Signal Processing* 65.14 (2017), pp. 3702–3717.

[62] S.M. Zandi, B. Boroomand, and S. Soghrati. "Exponential basis functions in solution of problems with fully incompressible materials: A mesh-free method". In: *Journal of Computational Physics* 231.21 (2012), pp. 7255–7273.

[63] Zhonggang Zeng. "The closedness subspace method for computing the multiplicity structure of a polynomial system". In: *Contemporary Mathematics*. Vol. 496. Providence, Rhode Island: American Mathematical Society, 2009, pp. 347–362.

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

[65] Eva Zerz. "First–Order Representations of Discrete Linear MultidimensionalSystems". In: *Multidimensional Syst. Signal Process.* 11.4 (Oct. 2000), pp. 359–380.