SIAM J. Matrix Anal. Appl.
Vol. 42, No. 2, pp. 913–953

SYSTEMS OF POLYNOMIAL EQUATIONS, HIGHER-ORDER TENSOR DECOMPOSITIONS, AND MULTIDIMENSIONAL HARMONIC RETRIEVAL: A UNIFYING FRAMEWORK. PART II: THE BLOCK TERM DECOMPOSITION*

JEROEN VANDERSTUKKEN†, PATRICK KÜRSCHNER‡, IGNAT DOMANOV†, AND LIEVEN DE LATHAUWER†

Abstract. In Part I we proposed a multilinear algebra framework to solve 0-dimensional systems of polynomial equations with simple roots. We extend this framework to incorporate multiple roots: a block term decomposition (BTD) of the null space of the Macaulay matrix reveals the dual (sub)space of a disjoint root in each term. The BTD is the joint triangularization of multiplication tables and a three-way generalization of the Jordan canonical form in the matrix case, intimately related to the border rank of a tensor. We hint at and illustrate flexible numerical optimization-based algorithms.

Key words. system of polynomial equations, multilinear algebra, block term decomposition, border rank, Macaulay matrix, multiplication table

AMS subject classifications. 13P15, 15A69, 54B05, 65H04

DOI. 10.1137/17M1150062

1. Introduction.

Systems of polynomial equations arise often in science and engineering. Solving such systems means finding all the common roots of the polynomials. Many methods have become available to solve systems of polynomial equations: algebraic geometry-based computer algebra methods, e.g., [5]; polynomial homotopy continuation (PHC), e.g., [39, 4]; and (Macaulay) resultant- and linear algebra–based methods [21, 37, 36], including, e.g., numerical polynomial algebra (NPA) [28, 34], and polynomial numerical linear algebra (PNLA) [1, 14], to name just a few.

A higher-order tensor in multilinear algebra is a multiway generalization of a one-way vector and a two-way matrix in linear algebra. Tensor decompositions like the canonical polyadic decomposition (CPD) and the block term decomposition (BTD) are then generalizations of matrix decompositions. Despite the natural generalization, multilinear algebra exhibits striking differences with linear algebra. First, a tensor that has rank greater than R is said to have border rank R if it can be approximated arbitrarily well by a (sequence of) rank-R tensor(s) [13]. This phenomenon, as shown in [32], can be seen as a multiway generalization of approximate diagonalization of a nondiagonalizable matrix, and the limit point of the approximating rank-R sequence can be seen as a multiway generalization of the Jordan canonical form. Second, the

*Received by the editors October 2, 2017; accepted for publication (in revised form) March 26, 2021; published electronically June 17, 2021.
https://doi.org/10.1137/17M1150062
Funding: This work was supported by the KU Leuven Research Council, C1 project c16/15/059-nD, the FWO under EOS project G0F6718N (SeLMA), the Flemish Government under "Onderzoeksprogramma Artificiele Intelligentie (AI) Vlaanderen," and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Advanced grant BIOTENSORS 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.
†Group Science, Engineering and Technology, KU Leuven Kulak, E. Sabbelaan 53, 8500 Kortrijk, Belgium, and Department of Electrical Engineering, ESAT/STADIUS, KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium (vanderstukken.jeroen@gmail.com, ignat.domanov@kuleuven.be, Lieven.DeLathauwer@kuleuven.be).
‡Centre for Mathematics and Natural Sciences, Leipzig University of Applied Sciences (HTWK Leipzig), PF 30 11 66, D–04251 Leipzig, Germany (patrick.kuerschner@htwk-leipzig.de).

rank of a tensor depends on the field considered for the factor entries.

For a tensor in $\mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ chosen at random according to continuous distributions (e.g., i.i.d. Gaussian entries), more than one distinct value of the rank occurs with positive probability. These rank values are called typical.

In [38] we presented a multilinear algebra framework to formulate and solve 0-dimensional polynomial root-finding problems, the solutions of which are isolated and finite in number. This discussion was limited to systems with only simple roots. For such systems we derived a connection between the null space of the Macaulay matrix and multidimensional harmonic retrieval (MHR). By jointly exploiting the multiplicative shift invariance in the different variables, we obtained a third-order tensor CPD that reveals the common roots.

In this companion paper we discuss systems of polynomial equations that are allowed to have roots with multiplicity greater than 1. Rather than just a single integer for the multiplicity, the multiplicity structure (dual space) of a multiple root is an essential means of providing characteristics of the root [6]. The dual spaces manifest themselves in the null space of the Macaulay matrix. If a system has roots with multiplicity greater than 1, the basis of the null space of the Macaulay matrix does not fully exhibit multiplicative shift invariance anymore. Consequently, we cannot derive a third-order tensor CPD that reveals the roots. Instead, we will derive a third-order tensor BTD that reveals the dual (sub)spaces of the disjoint roots.

In [38] we explained that the multiplicative shift invariance–expressing CPD can be seen in terms of the joint diagonalization of NPA multiplication tables. In this companion paper we will explain that the BTD generalization can be seen in terms of the joint block diagonalization/triangularization of the multiplication tables. Further, BTD offers a three-way generalization of the Jordan canonical form of the eigenvalue decomposition (EVD) in NPA. Such connections emphasize the unifying power of the multilinear algebra framework and its ability to help us understand the "roots" of polynomial systems and multilinear algebra more profoundly. Including BTD, our approach is able to (recursively) detect various (nested) structures in the null space of the Macaulay matrix. The multilinear approach opens a whole new range of numerical optimization techniques to solve systems of polynomial equations.

The paper is organized as follows. Section 2 will review our notation and introduce some necessary definitions. Section 3 will introduce the CPD and BTD as important tensor decompositions for this study, present a new uniqueness result for a BTD with special structure, and update the structure of the null space of the Macaulay matrix from the "simple root case" to the "case of roots with multiplicities." In section 4 we will then establish that the formerly resulting third-order tensor CPD needs to be understood as a special case of a third-order tensor BTD that also covers the more general case of roots with multiplicities. To develop insight, the emphasis is on the affine case, but the results can easily be extended to the projective case. Section 5 will further make connections between the BTD and the border rank of the higher-order tensor and between the BTD and the possible difference between the tensor's rank over the complex field and its rank over the real field. In section 6 we propose polynomial root-finding algorithms based on the insights from the previous sections. Section 7 presents the results of numerical experiments, and section 8 will summarize our findings.

2. Notation.

We give a quick summary of our notation. For more details the reader is referred to [38].

2.1. Higher-order tensors. Scalars, vectors, matrices, and tensors are denoted by italic, boldface lowercase, boldface uppercase, and calligraphic letters, respectively: $a \in \mathbb{C}$, $\mathbf{a} \in \mathbb{C}^{I_1}$, $\mathbf{A} \in \mathbb{C}^{I_1 \times I_2}$, and the Nth-order tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$. This paper will not surpass the third-order case. $a_{i_1} = \mathbf{a}(i_1) = (\mathbf{a})_{i_1}$ is the i_1th entry of vector $\mathbf{a}$. $a_{i_1,i_2} = \mathbf{A}(i_1, i_2) = (\mathbf{A})_{i_1,i_2}$ is equal to the entry of matrix $\mathbf{A}$ with row index i_1 and column index i_2. $\mathbf{a}_{i_2} = \mathbf{A}(:, i_2) = (\mathbf{A})_{i_2}$ denotes the i_2th column of $\mathbf{A}$. Likewise for the entries (a_{i_1,i_2,i_3}) and fibers $(\mathcal{A}(i_1, :, :)$, $\mathcal{A}(:, i_2, :)$, $\mathcal{A}(:, :, i_3))$ of a tensor $\mathcal{A}$; the vector obtained when all but the nth index of $\mathcal{A}$ are kept fixed is called a mode-n fiber of $\mathcal{A}$. The i_3th matrix slice $\mathcal{A}(:, :, i_3)$ of $\mathcal{A}$ is denoted by $\mathbf{A}_{i_3}$. We use $\cdot^*$, $\cdot^T$, $\cdot^H$, $\cdot^{-1}$, and $\cdot^\dagger$ to denote the complex conjugate, transpose, Hermitian transpose, inverse, and Moore–Penrose pseudoinverse, respectively.

$\mathbf{D} = \mathrm{diag}(\mathbf{d})$ represents a diagonal matrix with the vector $\mathbf{d}$ on its diagonal, and $\mathbf{D}_i(\mathbf{C}) = \mathrm{diag}(\mathbf{C}(i, :))$ holds the ith row of the matrix $\mathbf{C}$. $\mathbf{I}_I$ is the identity matrix of order $I \times I$. $\mathrm{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_I\})$ is the span of the vectors $\mathbf{a}_1$ through $\mathbf{a}_I$. $\mathrm{col}(\mathbf{A})$, $\mathrm{row}(\mathbf{A})$, and $\mathrm{null}(\mathbf{A})$ are used to denote the column, row, and right null space of $\mathbf{A}$, respectively. $r_{\mathbf{A}}$ denotes the rank of $\mathbf{A}$. Lastly, the Kronecker and Khatri–Rao products are denoted by $\otimes$ and $\odot$, respectively, and $\dot{+}$ is used to denote the direct sum of subspaces.

A third-order tensor $\mathcal{A}$ is vectorized to $\mathrm{vec}(\mathcal{A})$ by vertically stacking all entries a_{i_1,i_2,i_3} such that i_3 varies slowest and i_1 fastest: $a_{i_1,i_2,i_3} = (\mathrm{vec}(\mathcal{A}))_{(i_3-1)I_2I_1+(i_2-1)I_1+i_1}$. The matrix representation $\mathbf{A}_{[1;3,2]}$ is obtained by stacking the mode-1 fibers of $\mathcal{A}$ as columns into a matrix, in such a way that i_2 varies fastest along the second dimension: $a_{i_1,i_2,i_3} = \left(\mathbf{A}_{[1;3,2]}\right)_{i_1,(i_3-1)I_2+i_2}$. The mode-1 product $\mathcal{C} = \mathcal{A} \cdot_1 \mathbf{B} \in \mathbb{C}^{J \times I_2 \times I_3}$ of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ and a matrix $\mathbf{B} \in \mathbb{C}^{J \times I_1}$ then has the matrix representation $\mathbf{C}_{[1;3,2]} = \mathbf{B}\mathbf{A}_{[1;3,2]}$, i.e., it is the result of multiplying all mode-1 fibers of $\mathcal{A}$ from the left with $\mathbf{B}$. Other matrix representations and related products are defined analogously.

The mode-n rank $R_n = \mathrm{rank}_n(\mathcal{A})$ is the dimension of the mode-n fiber space, i.e., $R_n = r_{\mathbf{A}_{[n;\bullet]}}$, in which $\bullet$ indicates that the order of the indices different from n does not matter. The tuple $\mathrm{rank}_{\boxplus}(\mathcal{A}) = (R_1, R_2, R_3)$ is called the multilinear rank of $\mathcal{A}$. The outer product $\mathcal{T} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}$ of nonzero vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ yields a rank-1 tensor with entries $t_{i_1,i_2,i_3} = a_{i_1} b_{i_2} c_{i_3}$. The minimal number of rank-1 terms that sum to a particular tensor $\mathcal{A}$ is called the rank of $\mathcal{A}$ and denoted by $r_{\mathcal{A}}$.

2.2. Polynomial equations. Let us consider the system of polynomial equations

$$\begin{cases} f_1(x_1, \ldots, x_n) = 0, \\ \qquad \vdots \\ f_s(x_1, \ldots, x_n) = 0 \end{cases} \tag{1}$$

in n complex variables x_j, stacked in the vector $\mathbf{x} \in \mathbb{C}^n$. A monomial $\mathbf{x}^{\boldsymbol{\alpha}} = \prod_{j=1}^n x_j^{\alpha_j}$ is defined by an exponent vector $\boldsymbol{\alpha}$. The degree of a monomial is defined as $\deg(\mathbf{x}^{\boldsymbol{\alpha}}) = \sum_{j=1}^n \alpha_j$. There exist several schemes for ordering monomials by their exponent vector. As in the companion paper [38], we will adopt the degree negative lexicographic order. The monomials $\mathbf{x}^{\boldsymbol{\alpha}} < \mathbf{x}^{\boldsymbol{\beta}}$ are ordered by the degree negative lexicographic order if one of the following two conditions is satisfied: (i) $\deg(\mathbf{x}^{\boldsymbol{\alpha}}) < \deg(\mathbf{x}^{\boldsymbol{\beta}})$; or (ii) $\deg(\mathbf{x}^{\boldsymbol{\alpha}}) = \deg(\mathbf{x}^{\boldsymbol{\beta}})$ and the leftmost nonzero entry of $\boldsymbol{\beta} - \boldsymbol{\alpha}$ is negative.

A polynomial $f(x_1, \ldots, x_n) = \sum_{l=1}^p f_l \mathbf{x}_l^{\boldsymbol{\alpha}_l}$ is characterized by a coefficient vector $\mathbf{f}$. The degree d_i of a polynomial f_i in (1) is the degree of the monomial with the highest degree in f_i. The ring of all polynomials in n variables is denoted by $\mathcal{C}^n$. The vector space $\mathcal{C}^n_d$ is the subset of $\mathcal{C}^n$ that contains all polynomials up to degree d. Its dimension is given by

$$q(d) = \dim \mathcal{C}^n_d = \binom{n+d}{n}.$$

A polynomial is said to be homogeneous if all its monomials have the same degree. A polynomial f can be homogenized to a polynomial f^h by multiplying each monomial $\mathbf{x}_l^{\boldsymbol{\alpha}_l}$ in f with a power β_l of x_0, such that $\deg(x_0^{\beta_l}\mathbf{x}_l^{\boldsymbol{\alpha}_l}) = d$ for all l. The ring (vector space) of all homogeneous polynomials in $n+1$ variables (up to degree d) is denoted by $\mathcal{P}^n$ ($\mathcal{P}^n_d$). The projective space $\mathbb{P}^n$ is the set of equivalence classes on $\mathbb{C}^{n+1} \setminus \{\mathbf{0}\}$: $\begin{pmatrix} x'_0 & x'_1 & \dots & x'_n \end{pmatrix}^T \sim \begin{pmatrix} x_0 & x_1 & \dots & x_n \end{pmatrix}^T$ if there exists a $\lambda \in \mathbb{C}$ such that $\begin{pmatrix} x'_0 & x'_1 & \dots & x'_n \end{pmatrix}^T = \lambda \begin{pmatrix} x_0 & x_1 & \dots & x_n \end{pmatrix}^T$. Points with $x_0 = 0$ cannot be normalized to their affine counterpart $\begin{pmatrix} 1 & \frac{x_1}{x_0} & \dots & \frac{x_n}{x_0} \end{pmatrix}^T$: they are points at infinity.

The degree of (1) is $d_0 = \max_{i=1}^s d_i$. The set of all roots of (1) is called the solution set. Under the same assumptions as in [38] that (1) is a square system ($n = s$) with a 0-dimensional solution set, the number of roots in the projective space, counting multiplicities, is given by the Bézout number

$$m = \prod_{i=1}^n d_i.$$

If (1) has multiple roots, $m_0 < m$ denotes the number of disjoint roots. The m_0 distinct roots of (1) will be denoted by $\begin{pmatrix} x_0^{(k)} & x_1^{(k)} & x_2^{(k)} & \cdots & x_n^{(k)} \end{pmatrix}^T \in \mathbb{P}^n$, $k = 1 : m_0$.

3. Tensor decompositions, Macaulay null space, and harmonic structure: From simple roots to roots with multiplicities. Similar to the way [38] was organized, in this section we display the ingredients from the study of tensor decompositions, sets of polynomial equations, and harmonic retrieval that we will combine in our derivation. To allow roots with multiplicities, we will not only need CPD, as in [38], but also a particular type of BTD (section 3.1). We also need to discuss the multiplicity structure of a root (section 3.2). For handling roots with multiplicities, we need to take the step from the multivariate Vandermonde structure in [38] to a "confluent" extension (section 3.3).

3.1. Tensor decompositions.

3.1.1. CPD. An R-term polyadic decomposition (PD) expresses a tensor $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ as a sum of R rank-1 terms

$$\mathcal{T} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \stackrel{\text{def}}{=} \sum_{r=1}^R \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r. \tag{2}$$

The matrices $\mathbf{A} \in \mathbb{C}^{I_1 \times R}$, $\mathbf{B} \in \mathbb{C}^{I_2 \times R}$, and $\mathbf{C} \in \mathbb{C}^{I_3 \times R}$ are called factor matrices. If R is minimal, then the PD is a canonical polyadic decomposition (CPD) and $R = r_{\mathcal{T}}$ is the rank of $\mathcal{T}$. Equation (2) can be expressed in an entrywise manner as

$$t_{i_1 i_2 i_3} = \sum_{r=1}^R a_{i_1 r} b_{i_2 r} c_{i_3 r}, \qquad i_1 = 1 : I_1,\ i_2 = 1 : I_2,\ i_3 = 1 : I_3.$$

In a slicewise manner, (2) can be written as

$$\mathbf{T}_{i_3} = \mathbf{A}\mathbf{D}_{i_3}(\mathbf{C})\mathbf{B}^T, \qquad i_3 = 1 : I_3.$$

In matricized format, (2) can be written as

$$\mathbf{T}_{[1,2;3]} = \sum_{r=1}^{R} (\mathbf{a}_r \otimes \mathbf{b}_r)\mathbf{c}_r^T = (\mathbf{A} \odot \mathbf{B})\,\mathbf{C}^T.$$

A CPD can only be unique up to permutation of the rank-1 terms and scaling/counterscaling of the vectors within the same term (i.e., we can allow $\mathbf{a}_r \leftarrow \mathbf{a}_r\alpha_r$, $\mathbf{b}_r \leftarrow \mathbf{b}_r\beta_r$, $\mathbf{c}_r \leftarrow \mathbf{c}_r\gamma_r$ with $\alpha_r\beta_r\gamma_r = 1$).

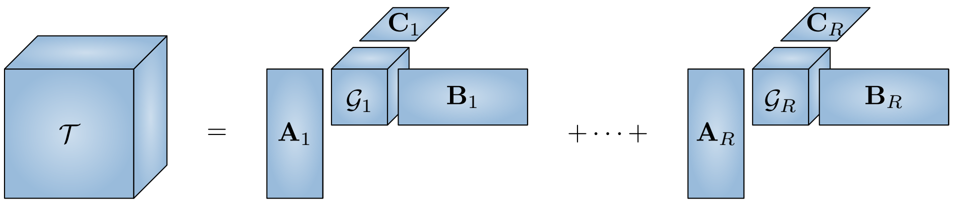


Fig. 1. *BTD of a tensor $\mathcal{T}$ is a decomposition in terms that have low multilinear rank.*

3.1.2. BTD. Block term decomposition (BTD) generalizes PD in the sense that the terms do not need to be rank-1 (i.e., have multilinear rank $(1, 1, 1)$) but only need to have low multilinear rank [8, 9, 12]. Specifically, in this paper, we will deal with the BTD

$$\mathcal{T} = \sum_{r=1}^{R} [\![\mathcal{G}_r; \mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r]\!] \stackrel{\text{def}}{=} \sum_{r=1}^{R} \mathcal{G}_r \cdot_1 \mathbf{A}_r \cdot_2 \mathbf{B}_r \cdot_3 \mathbf{C}_r, \tag{3}$$

in which $\mathcal{G}_r \in \mathbb{C}^{\mu_r\times\mu_r\times\mu_r}$ is multilinear rank-(μ_r, μ_r, μ_r) and the matrices $\mathbf{A}_r \in \mathbb{C}^{I_1\times\mu_r}$, $\mathbf{B}_r \in \mathbb{C}^{I_2\times\mu_r}$, and $\mathbf{C}_r \in \mathbb{C}^{I_3\times\mu_r}$ have full column rank, $r = 1 : R$, implying that (3) is a decomposition into a sum of multilinear rank-(μ_r, μ_r, μ_r) terms. Figure 1 depicts such a BTD. Throughout the paper we will consider only those decompositions of the form (3) for which the matrices

$$\mathbf{B} \stackrel{\text{def}}{=} \begin{pmatrix}\mathbf{B}_1 & \dots & \mathbf{B}_R\end{pmatrix} \in \mathbb{C}^{I_2\times\sum_{r=1}^{R}\mu_r} \quad\text{and}\quad \mathbf{C} \stackrel{\text{def}}{=} \begin{pmatrix}\mathbf{C}_1 & \dots & \mathbf{C}_R\end{pmatrix} \in \mathbb{C}^{I_3\times\sum_{r=1}^{R}\mu_r} \tag{4}$$

have full column rank. We say that $\mathcal{T}$ is indecomposable if $\mathcal{T}$ does not admit a decomposition of the form (3) with $R \geq 2$ terms and such that condition (4) holds. We say that decomposition (3) of $\mathcal{T}$ into a sum of R indecomposable multilinear rank-(μ_r, μ_r, μ_r) terms is unique if any other decomposition of $\mathcal{T}$ into a sum of $\tilde{R}$ indecomposable multilinear rank-$(\tilde{\mu}_r, \tilde{\mu}_r, \tilde{\mu}_r)$ terms necessarily coincides with (3) up to permutation of the terms, provided that $\sum_{r=1}^{\tilde{R}} \tilde{\mu}_r = \sum_{r=1}^{R} \mu_r$. The counterpart of the CPD scaling/counterscaling ambiguity is that we can allow $\mathbf{A}_r \leftarrow \mathbf{A}_r\mathbf{M}_r^{(1)}$, $\mathbf{B}_r \leftarrow \mathbf{B}_r\mathbf{M}_r^{(2)}$, $\mathbf{C}_r \leftarrow \mathbf{C}_r\mathbf{M}_r^{(3)}$, in which $\mathbf{M}_r^{(1)} \in \mathbb{C}^{\mu_r\times\mu_r}$, $\mathbf{M}_r^{(2)} \in \mathbb{C}^{\mu_r\times\mu_r}$, $\mathbf{M}_r^{(3)} \in \mathbb{C}^{\mu_r\times\mu_r}$ are invertible, if the transformation is compensated by $\mathcal{G}_r \leftarrow \mathcal{G}_r \cdot_1 \left(\mathbf{M}_r^{(1)}\right)^{-1} \cdot_2 \left(\mathbf{M}_r^{(2)}\right)^{-1} \cdot_3 \left(\mathbf{M}_r^{(3)}\right)^{-1}$ [9].

The following theorem presents a sufficient condition for uniqueness of BTD (3). If $\mu_1 = \cdots = \mu_R = 1$, that is, in the case of the CPD, Theorem 3.1 reduces to [38, Theorem 3.1].

THEOREM 3.1. *Let $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ admit decomposition* (3) *into a sum of multilinear rank-*(μ_r, μ_r, μ_r) *terms. Assume that*

(5) *the matrices* $\mathbf{B}$ *and* $\mathbf{C}$ *defined in* (4) *have full column rank,*

(6) *the matrix* $[\mathbf{A}_1(:,1) \ \dots \ \mathbf{A}_R(:,1)]$ *does not have proportional columns,*

and that the core tensors $\mathcal{G}_r \in \mathbb{C}^{\mu_r \times \mu_r \times \mu_r}$ *have slices* $\mathcal{G}_r(l+1,:,:) = \mathcal{G}_r(:,l+1,:) \in \mathbb{C}^{\mu_r \times \mu_r}, l = 0 : \mu_r - 1$, *which are upper-triangular or, if* $l = 0$, *equal to* $\mathbf{I}_{\mu_r}$. *Then BTD* (3) *is unique.*

Proof. The proof is given in Appendix A. ∎

Moreover, if the assumptions of Theorem 3.1 hold, the argumentation in Appendix A gives a way to compute the BTD and its factor matrices algebraically by means of a block-diagonalization by a similarity transform. As in Appendix A we consider without loss of generalization (w.l.o.g.) a tensor $\mathcal{T}$ where $\mathbf{B}$, $\mathbf{C}$ are square, i.e., the second mode dimension of $\mathcal{T}$ is equal to the third one: $I_2 = I_3 = m = \mu_1 + \cdots + \mu_R$. For $\mathcal{T}$ with larger second and third mode dimensions, this can be achieved by, e.g., a compression using the multilinear singular value decomposition (MLSVD)[1] [11]. Define two "slice mixtures" $\mathbf{T}_1 \stackrel{\text{def}}{=} \mathcal{T} \cdot_1 \mathbf{f}^T$ and $\mathbf{T}_2 \stackrel{\text{def}}{=} \mathcal{T} \cdot_1 \mathbf{g}^T \in \mathbb{C}^{m \times m}$, where $\mathbf{f}$, $\mathbf{g} \in \mathbb{C}^{I_1}$ are two generic vectors. Because

$$\mathcal{T} \cdot_1 \mathbf{h}^T = \mathbf{B} \cdot \mathrm{Blockdiag}(\mathcal{G}_1 \cdot_1 (\mathbf{h}^T \mathbf{A}_1), \dots, \mathcal{G}_{m_0} \cdot_1 (\mathbf{h}^T \mathbf{A}_R)) \cdot \mathbf{C}^T \tag{7}$$

for any vector $\mathbf{h} \in \mathbb{C}^{I_1}$, the factor matrix $\mathbf{B}$ is, up to the intrinsic indeterminacies mentioned above, given by the block-diagonal decomposition[2]

$$\mathbf{T}_2 \mathbf{T}_1^{-1} = \mathbf{B} \begin{pmatrix} \mathbf{D}_1 & & \\ & \ddots & \\ & & \mathbf{D}_R \end{pmatrix} \mathbf{B}^{-1}, \quad \mathbf{D}_r \stackrel{\text{def}}{=} (\mathcal{G}_r \cdot_1 \mathbf{g}^T \mathbf{A}_r)(\mathcal{G}_r \cdot_1 \mathbf{f}^T \mathbf{A}_r)^{-1} \in \mathbb{C}^{\mu_r \times \mu_r}.$$

The factor matrix $\mathbf{C}$ can be obtained as $\mathbf{C} = \mathbf{T}_1 \mathbf{B}^{-T}$ (this follows easily from (7)), again up to the intrinsic indeterminacies. The above block-diagonalization of $\mathbf{T}_2 \mathbf{T}_1^{-1}$ can in practice be computed, e.g., from a Schur decomposition of $\mathbf{T}_2 \mathbf{T}_1^{-1}$; see [19, section 7.6.3]. It also returns the partition of $\mathbf{B}$ into the blocks $\mathbf{B}_r \in \mathbb{C}^{m \times \mu_r}$, and consequently also the partitioning of $\mathbf{C}$ into blocks $\mathbf{C}_r \in \mathbb{C}^{m \times \mu_r}$, with correct column sizes. We have

$$\mathcal{T} \cdot_2 \mathbf{B}^{-1} \cdot_3 \mathbf{C}^{-1} = \sum_{r=1}^{R} \mathcal{G}_r \cdot_1 \mathbf{A}_r \cdot_2 (\mathbf{B}^{-1} \mathbf{B}_r) \cdot_3 (\mathbf{C}^{-1} \mathbf{C}_r) = \sum_{r=1}^{R} \mathcal{G}_r \cdot_1 \mathbf{A}_r \cdot_2 \begin{pmatrix} 0 \\ \mathbf{I}_{\mu_r} \\ 0 \end{pmatrix} \cdot_3 \begin{pmatrix} 0 \\ \mathbf{I}_{\mu_r} \\ 0 \end{pmatrix},$$

so we obtain the tensors $\tilde{\mathcal{G}}_r \stackrel{\text{def}}{=} \mathcal{G}_r \cdot_1 \mathbf{A}_r$ (indeed, the horizontal slices of $\mathcal{T} \cdot_2 \mathbf{B}^{-1} \cdot_3 \mathbf{C}^{-1}$ are block-diagonal matrices and the kth horizontal slice of $\tilde{\mathcal{G}}_r$ is just the rth block of the kth horizontal slice of $\mathcal{T} \cdot_2 \mathbf{B}^{-1} \cdot_3 \mathbf{C}^{-1}$). It is clear that $\mathcal{G}_r$ and $\mathbf{A}_r$ can be recovered from $\tilde{\mathcal{G}}_r$, again up to the intrinsic indeterminacies. For example, one can compute the SVD $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H = \tilde{\mathbf{G}}_{r[2,3:1]}$, take $\mathbf{A}_r = \mathbf{U}(:, 1 : \mu_r)$, and set $\mathcal{G}_r = \tilde{\mathcal{G}}_r \cdot_1 \mathbf{A}_r^H$. Consequently, by doing this for all R terms we obtain the BTD (3).

[1]In the following, we use the term "compression" to refer to the MLSVD-based compression.

[2]Step 1 in the proof of Theorem 4.4 in Appendix A ensures that a generic $\mathbf{f}$ will yield a nonsingular matrix $\mathbf{T}_1$.

We conclude by mentioning that, instead of working with the above block-diagonalization of $\mathbf{T}_2\mathbf{T}_1^{-1}$, one can also use a block-diagonalization of the matrix pencil $(\mathbf{T}_1, \mathbf{T}_2)$, which is to be preferred numerically as it avoids the inverse of $\mathbf{T}_1$. The algebraic computation discussed here generalizes the generalized eigenvalue decomposition (GEVD)-based computation of the CPD used in [38]. Just as the CPD in [38] may be seen as an extension of GEVD to more than two matrices, the considered BTD here may be seen as an extension of block-diagonalization to more than two matrices. Furthermore, one can use optimization-based approaches [29] to compute the BTD or, if necessary, refine the results obtained from algebraic methods. This situation is, again, similar to that for the CPD in [38].

3.2. The Macaulay null space. Our approach exploits the Vandermonde structure in the null space of a Macaulay matrix of sufficiently high degree.

3.2.1. Simple roots.

DEFINITION 3.2 ([15, p. 263]). *Let $f_i \in \mathcal{C}^n_{d_i}$, $i = 1 : s$, be s polynomials of degree d_i in n variables $x_1, \ldots, x_n$. Then the Macaulay matrix $\mathbf{M}(d)$ of degree d contains as its rows the coefficients of*

$$\mathbf{M}(d) = \begin{pmatrix} f_1 \\ x_1 f_1 \\ \vdots \\ x_n^{d-d^{(1)}} f_1 \\ f_2 \\ x_1 f_2 \\ \vdots \\ x_n^{d-d_s} f_s \end{pmatrix} \in \mathbb{C}^{\sum_{i=1}^{s} q(d-d_s) \times q(d)},$$

where each polynomial $f_i, i = 1 : s$, is multiplied with all possible monomials $\mathbf{x}^{\boldsymbol{\alpha}}$, $\deg(\mathbf{x}^{\boldsymbol{\alpha}}) = 0 : d - d_i \in \mathbb{N}$.

If the system (1) has only simple roots, the null space of $\mathbf{M}(d)$ constructed at a degree d greater than or equal to the so-called degree of regularity d^* is m-dimensional; it is generated by m multivariate Vandermonde vectors

(8)
$$\mathbf{v}_k(d) = \begin{pmatrix} 1 & x_1^{(k)} & x_2^{(k)} & \ldots & x_1^{(k)2} & x_1^{(k)} x_2^{(k)} & \ldots & x_{n-1}^{(k)} x_n^{(k)d-1} & x_n^{(k)d} \end{pmatrix}^T \in \mathbb{C}^{q(d)},$$

where $x_j^{(k)}$ denotes the jth coordinate of the kth root, $k = 1 : m$, $j = 1 : n$. For more background, see [38].

3.2.2. The multiplicity structure of a root. Let the fixed set of m points $\mathcal{Z} = \{\mathbf{z}_k\}_{k=1}^m \subset \mathbb{C}^n$ represent the solution set of the system (1). The system is then defined by a basis $\mathcal{F}$ for the polynomial ideal $\mathcal{I} \subset \mathcal{C}^n$ of all polynomials that attain zero on the set $\mathcal{Z}$. The set of residue classes $[r] = \{r' \in \mathcal{C}^n | \, r - r' \in \mathcal{I}\}$ is a quotient ring $\mathcal{C}^n/\mathcal{I}$ induced by the polynomial ideal $\mathcal{I}$.

If all elements of $\mathcal{Z}$ occur with multiplicity 1, i.e., if the system defined by $\mathcal{F}$ has only simple roots, then the characterization of the residue classes is straightforward. We have that a polynomial $g \in \mathcal{I} \Leftrightarrow g(\mathbf{z}_k) = 0$ for all k. Further, $g \in [r] \Leftrightarrow g - r \in \mathcal{I} \Leftrightarrow (g - r)(\mathbf{z}_k) = 0$ for all k. Any residue class is completely characterized by the value evaluations of its members on the set of m points $\mathcal{Z}$, and $\dim \mathcal{C}^n/\mathcal{I} = m$.

However, if one or more of the elements of $\mathcal{Z}$ occur with multiplicity greater than 1, i.e., if the system defined by $\mathcal{F}$ has coinciding roots, things become more subtle. Say there are $m_0 < m$ disjoint roots $\mathcal{Z}_0 = \{\mathbf{z}_k\}_{k=1}^{m_0} \subset \mathcal{Z}$, occurring with multiplicity μ_k in $\mathcal{Z}$, such that $\sum_{k=1}^{m_0} \mu_k = m$. One can show that the dimension of $\mathcal{C}^n/\mathcal{I}$ remains m but that $g(\mathbf{z}_k) = 0$ for all $k = 1 : m_0$ is no longer sufficient for $g \in \mathcal{I}$ [35, pp. 91–92]. For a concise characterization of the residue classes, we introduce differential functionals. Differential functionals act on a polynomial $f \in \mathcal{C}^n$ first by differentiation $(\cdot)$ and then by evaluation $[\cdot]$.

Definition 3.3 (differential functional [35, p. 90]). *Let $\mathbf{z} \in \mathbb{C}^n$ and $f \in \mathcal{C}^n$. Then a differential functional monomial is defined by*

$$\partial_{\mathbf{j}}[\mathbf{z}](f) = \partial_{j_1 \dots j_n}[\mathbf{z}](f) = \frac{1}{j_1! \dots j_n!} \left(\frac{\partial^{\sum_{l=1}^n j_l}}{\partial x_1^{j_1} \dots \partial x_n^{j_n}} f \right)(\mathbf{z}),$$

where $\mathbf{j} = \begin{pmatrix} j_1 & \dots & j_n \end{pmatrix}^T \in \mathbb{N}^n$. Any linear combination $\sum_{\mathbf{j}} \beta_{\mathbf{j}} \partial_{\mathbf{j}}[\mathbf{z}](f)$ with $\beta_j \in \mathbb{C}$ of differential functional monomials $\partial_{\mathbf{j}}[\mathbf{z}](f)$ is a differential functional.

The order of the differential functional monomial $\partial_{\mathbf{j}}$ is defined as $o(\partial_{\mathbf{j}}) = |\mathbf{j}| = \sum_{l=1}^n j_l$ [6, p. 2145]. The order of a linear combination is the order of the highest-order differential functional monomial in that linear combination.

Let us turn back to the characterization of the residue classes. Gröbner duality formulates a sufficient condition for $g \in \mathcal{I}$ in terms of differential functionals.

Definition 3.4 (Gröbner duality [20, pp. 174–178]). *Let the system of polynomials defined by a basis $\mathcal{F}$ for the ideal $\mathcal{I}$ have $m_0 \leq m$ disjoint roots. Then $\mathbf{z}_k$ is a root of the system with multiplicity μ_k iff μ_k linearly independent differential functionals $\sum_{\mathbf{j}} \beta_{\mathbf{j}} \partial_{\mathbf{j}}[\mathbf{z}_k](g)$ vanish for $g \in \mathcal{I}$.*

Hence, given the fixed set $\mathcal{Z}$, Gröbner duality states that a sufficient condition for $g \in \mathcal{I}$ is that $c_{kl}(g) = 0$ for all $k = 1 : m_0$, where, for the kth root (with multiplicity μ_k), we need to consider $c_{k0} = \partial_{\mathbf{0}}[\mathbf{z}_k]$ of order 0 and $\mu_k - 1$ differential functionals c_{kl} of order greater than 0. The collection $\mathcal{D}[\mathbf{z}_k](\mathcal{F}) = \{c_{kl} \,|\, \forall f \in \mathcal{F} : c_{kl}(f) = 0\}$ containing these differential functionals is referred to as the multiplicity structure of the root $\mathbf{z}_k$. The dimension of $\mathcal{D}$ equals μ_k, and the depth δ_k of $\mathcal{D}$ is defined as the highest order of the differential functionals in $\mathcal{D}$.[3] Summarizing, a residue class is now completely characterized by value and derivative evaluations contained in all the $\mathcal{D}[\mathbf{z}_k]$ together, $k = 1 : m_0$.

Several algorithms to compute the multiplicity structure have been proposed in the literature [26, 7, 41, 6]. One such algorithm is Macaulay's algorithm [26]. The idea of Macaulay's approach is to compute $\mathcal{D}$ by computing the null space of Macaulay-like matrices at increasing degrees. Indeed, as already mentioned in [38], the m-dimensional null space of $\mathbf{M}(d)$ at a degree $d \geq d^*$ is isomorphic with the set of all residue classes $\mathcal{C}_d^n/\mathcal{I}$.

In the remainder of this paper, we will write $\partial_{\mathbf{j}}[\mathbf{v}]$ or, more generally, $c[\mathbf{v}]$ for a differential functional that acts on a multivariate Vandermonde vector $\mathbf{v}$ first by differentiation and then by evaluation of its entries.

Example 3.5 ([16, Example 7]). Consider the system of $s = 2$ polynomial equa-

[3]The differential functionals constitute a basis for the so-called dual space of the ideal $\mathcal{I}$, and the dimension of $\mathcal{D}$ is the dimension of the dual subspace spanned by the elements of $\mathcal{D}$—see also Definition B.1.

tions in $n = 2$ variables

$$\begin{cases} f_1(x_1, x_2) = (x_2 - 2)^2 = 0, \\ f_2(x_1, x_2) = (x_1 - x_2 + 1)^2 = 0, \end{cases}$$

where $d^{(1)} = d^{(2)} = 2$, $d^* = 2+2-2 = 2$, and $m = 2 \cdot 2 = 4$, but $m_0 = 1$. The system has $m_0 = 1$ disjoint root $\mathbf{x}^{(1)} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} \end{pmatrix}^T = \begin{pmatrix} 1 & 2 \end{pmatrix}^T$ with multiplicity $\mu_1 = 4$. It can be verified that a basis for the $(m = 4)$-dimensional null space of

$$\mathbf{M}(d) = \begin{pmatrix} 4 & 0 & -4 & 0 & 0 & 1 \\ 1 & 2 & -2 & 1 & -2 & 1 \end{pmatrix}$$

at $d = d^*$ is given by the multivariate Vandermonde vector $c_{10}[\mathbf{v}(d)] = \partial_{\mathbf{0}}[\mathbf{v}(2)] = \mathbf{v}(2)$, the "first-order derivative vectors" $c_{11}[\mathbf{v}(2)] = \partial_{10}[\mathbf{v}(2)]$ and $c_{12}[\mathbf{v}(2)] = \partial_{01}[\mathbf{v}(2)]$, and the linear combination of "second-order derivative vectors" $c_{13}[\mathbf{v}(2)] = (2\partial_{20} + \partial_{11})\,[\mathbf{v}(2)]$. (In the notation of Definition 3.3, we have $\beta_{00} = \beta_{10} = \beta_{01} = \beta_{11} = 1$ and $\beta_{20} = 2$.) This basis[4] is stacked in a matrix that will be called confluent multivariate Vandermonde in subsection 3.3.2:

$$\begin{aligned} \tilde{\mathbf{V}}(2) &\overset{\text{def}}{=} \begin{pmatrix} c_{10}[\mathbf{v}(2)] & c_{11}[\mathbf{v}(2)] & c_{12}[\mathbf{v}(2)] & c_{13}[\mathbf{v}(2)] \end{pmatrix} \\ &= \begin{pmatrix} \partial_{00}[\mathbf{v}(2)] & \partial_{10}[\mathbf{v}(2)] & \partial_{01}[\mathbf{v}(2)] & (2\partial_{20} + \partial_{11})\,[\mathbf{v}(2)] \end{pmatrix} \\ &= \left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ \hline x_1^{(1)} & 1 & 0 & 0 \\ x_2^{(1)} & 0 & 1 & 0 \\ \hline x_1^{(1)2} & 2x_1^{(1)} & 0 & 2 \\ x_1^{(1)}x_2^{(1)} & x_2^{(1)} & x_1^{(1)} & 1 \\ x_2^{(1)2} & 0 & 2x_2^{(1)} & 0 \end{array}\right). \end{aligned} \tag{9}$$

The depth δ_1 of $\mathcal{D}[\mathbf{x}^{(1)}]$ is equal to the order of $c_{13}[\mathbf{v}(2)]$: $\delta_1 = 2$.

3.3. Vandermonde matrices.

In what follows, matrices having Vandermonde structure will play an important role, so we shall recall some properties here for both uni- and multivariate Vandermonde matrices.

3.3.1. Vandermonde matrices with distinct generators.

We consider univariate Vandermonde matrices $\mathbf{V}^{(j)}(d) \in \mathbb{C}^{(d+1)\times m}$ generated by the jth coordinate of the m roots of (1), denoted by $\{x_j^{(k)}\}$, $k = 1 : m$, $j = 1 : n$:

$$\mathbf{V}^{(j)}(d) = \left(\mathbf{v}_1^{(j)}(d), \ldots, \mathbf{v}_m^{(j)}(d)\right), \quad \mathbf{v}_k^{(j)}(d) = \left(1, x_j^{(k)}, x_j^{(k)2}, \ldots, x_j^{(k)d}\right)^T.$$

The univariate Vandermonde matrix $\mathbf{V}^{(j)}(d)$ has full column rank if all *generators* $x_j^{(k)}$ are distinct, $k = 1 : m$. We will make use of *spatial smoothing* [30]. This means that if we take the outer product of subvectors $\mathbf{v}_k^{(j)}(1 : L) \cdot \mathbf{v}_k^{(j)}(1 : d - L + 2)^T$, the

[4]Like the multivariate Vandermonde basis in the case of simple roots, this confluent multivariate Vandermonde basis is only one possible basis for the Macaulay null space. In practice, it is a numerical basis that will be computed. Both are related by an a priori unknown basis transformation—see (17).

result is a rank-1 Hankel matrix:

$$\mathbf{v}_k^{(j)}(1:L)\otimes\mathbf{v}_k^{(j)}(1:d-L+2) = \begin{pmatrix} 1 \\ x_j^{(k)} \\ \vdots \\ x_j^{(k)(L-1)} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ x_j^{(k)} \\ x_j^{(k)2} \\ \vdots \\ x_j^{(k)(d-L+1)} \end{pmatrix} \tag{10}$$

$$= \operatorname{vec}\left(\begin{pmatrix} 1 \\ x_j^{(k)} \\ \vdots \\ x_j^{(k)(L-1)} \end{pmatrix} \begin{pmatrix} 1 \\ x_j^{(k)} \\ x_j^{(k)2} \\ \vdots \\ x_j^{(k)(d-L+1)} \end{pmatrix}^T \right)$$

$$= \operatorname{vec} \underbrace{\begin{pmatrix} 1 & x_j^{(k)} & \dots & x_j^{(k)(d-L+1)} \\ x_j^{(k)} & x_j^{(k)2} & \dots & x_j^{(k)(d-L+2)} \\ \vdots & \vdots & & \vdots \\ x_j^{(k)(L-1)} & x_j^{(k)L} & \dots & x_j^{(k)d} \end{pmatrix}}_{=\mathbf{H}_k}.$$

The structure is called *(multiplicative) shift invariance*, referring to the shifting of entries when the power of $x_j^{(k)}$ is raised. In [38] we have used the variant for $L = 2$. In Part II we will use the variant for $L > 2$.

For multivariate generators $\{(x_1^{(k)}, \dots, x_n^{(k)})\}$, $k = 1 : m$, we define multivariate Vandermonde matrices of degree d as

$$\mathbf{V}(d) = \begin{pmatrix} \mathbf{v}_1(d) & \dots & \mathbf{v}_m(d) \end{pmatrix} \in \mathbb{C}^{q(d)\times m}, \tag{11}$$

where each column $\mathbf{v}_k(d)$ is in the multivariate Vandermonde form of (8). Multivariate Vandermonde matrices exhibit a multiplicative shift structure in each variable x_j. More precisely, a multivariate Vandermonde matrix consists of the rows of the Khatri–Rao product of the n univariate Vandermonde matrices $\mathbf{V}^{(j)}(d)$ that are associated with the monomials up to degree d. Formally, we have $\mathbf{V}(d) = \mathbf{S}_{(d+1)^n\to q(d)}\left(\mathbf{V}^{(1)}(d)\odot\cdots\odot\mathbf{V}^{(n)}(d)\right)$, where $\mathbf{V}^{(j)}(d) \in \mathbb{C}^{(d+1)\times m}$, $j = 1 : n$, are univariate Vandermonde matrices of degree d constructed from the jth coordinate of the m roots and $\mathbf{S}_{(d+1)^n\to q(d)} \in \mathbb{R}^{q(d)\times(d+1)^n}$ eliminates all duplicate rows in the Khatri–Rao products, truncates the monomials of degree higher than d, and reorders the remaining $q(d)$ monomials according to the chosen monomial order. The matrix $\mathbf{S}_{(d+1)^n\to q(d)}$ can be constructed by n-fold composition of the "elimination matrices" in [27]. See [38] for more details, where the n-fold multiplicative shift structure was used to connect the null space of the Macaulay matrix to CPD.

3.3.2. Confluent Vandermonde matrices. If m_0 distinct univariate generators $x_j^{(k)}$ occur each with multiplicities $\mu_k \geq 1$, and $m = \sum_{k=1}^{m_0}\mu_k$ is the total number of generators, the associated univariate Vandermonde matrix $\mathbf{V}^{(j)}(d)$ set up in a naive way would have identical columns and, hence, be rank deficient. Confluent univariate Vandermonde matrices

$$\tilde{\mathbf{V}}^{(j)}(d) = \left(\tilde{\mathbf{V}}_1^{(j)}(d), \dots, \tilde{\mathbf{V}}_{m_0}^{(j)}(d)\right)$$

capture the multiplicities by including "derivative vectors" in submatrices of the form

$$\tilde{\mathbf{V}}_k^{(j)}(d) = \begin{pmatrix} \mathbf{v}_k^{(j)}(d) & \frac{d}{dx_j}[\mathbf{v}_k^{(j)}(d)] & \dots & \frac{1}{(\mu_k-1)!}\frac{d^{\mu_k-1}}{dx_j^{\mu_k-1}}[\mathbf{v}_k^{(j)}(d)] \end{pmatrix} \in \mathbb{C}^{(d+1)\times\mu_k}, \quad k = 1 : m_0,$$

with Vandermonde vectors $\mathbf{v}_k^{(j)}(d)$ as in subsection 3.3.1; see, e.g., [22, 10]. Only the first column $\mathbf{v}_k^{(j)}(d)$ of $\tilde{\mathbf{V}}_k^{(j)}(d)$ enjoys the multiplicative shift invariance mentioned in subsection 3.3.1. The submatrices $\tilde{\mathbf{V}}_k^{(j)}(d)$ are for $I, I-L+1 \geq \mu_k$ related to a rank-μ_k Hankel matrix via $\tilde{\mathbf{H}}_k = \tilde{\mathbf{V}}_k^{(j)}(d)(1:L,:) \cdot \mathbf{D}_k^{(j)} \cdot \tilde{\mathbf{V}}_k^{(j)}(d)(1:I-L,:)$, where

$$\mathbf{D}_k^{(j)} = \begin{pmatrix} 1 & x_j^{(k)} & x_j^{(k)2} & \cdots & x_j^{(k)(\mu_k-1)} \\ x_j^{(k)} & x_j^{(k)2} & x_j^{(k)3} & \cdots & 0 \\ x_j^{(k)2} & x_j^{(k)3} & x_j^{(k)4} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_j^{(k)(\mu_k-1)} & 0 & \cdots & \cdots & 0 \end{pmatrix} \in \mathbb{C}^{\mu_k\times\mu_k}$$

is nonsingular and Hankel; see, e.g., [3, 10]. This can be seen as block generalization of the spatial smoothing structure in (10).

For the multivariate case, the multiplicity structure of a multiple root defined in subsection 3.2.2 gives rise to a generalization of multivariate Vandermonde matrices of the form

$$\tilde{\mathbf{V}}(d) = \begin{pmatrix} \tilde{\mathbf{V}}_1(d) & \dots & \tilde{\mathbf{V}}_{m_0}(d) \end{pmatrix} \in \mathbb{C}^{q(d)\times m}, \tag{12}$$

in which

$$\begin{aligned} \tilde{\mathbf{V}}_k(d) &= \begin{pmatrix} \tilde{\mathbf{v}}_{k,0}(d) & \tilde{\mathbf{v}}_{k,1}(d) & \dots & \tilde{\mathbf{v}}_{k,\mu_k-1}(d) \end{pmatrix} \\ &= \begin{pmatrix} c_{k0}[\mathbf{v}_k(d)] & c_{k1}[\mathbf{v}_k(d)] & \dots & c_{k,\mu_k-1}[\mathbf{v}_k(d)] \end{pmatrix} \in \mathbb{C}^{q(d)\times\mu_k} \end{aligned}$$

for $k = 1 : m_0$, where $c_{k,l}$ are the differential functionals from the multiplicity structure $\mathcal{D}[\mathbf{z}_k](\mathcal{F})$. We shall refer to (12) as confluent multivariate Vandermonde matrices; see also [17]. Each submatrix $\tilde{\mathbf{V}}_k(d) \in \mathbb{C}^{q(d)\times\mu_k}$ reflects the multiplicity structure $\mathcal{D}[\mathbf{z}_k]$ of the kth root. The depth δ_k of $\mathcal{D}[\mathbf{z}_k]$ is the highest order of the corresponding $c_{k,\cdot}$ in $\tilde{\mathbf{V}}_k(d)$. Only the first column $c_{k0}[\mathbf{v}_k(d)] = \tilde{\mathbf{v}}_{k,0}(d) = \mathbf{v}_k(d)$ in each submatrix has the shift-invariance property. The confluent multivariate Vandermonde matrix $\tilde{\mathbf{V}}(d)$ is of full column rank m and constitutes a basis for the m-dimensional null space of $\mathbf{M}(d)$ for $d \geq d^*$.

4. From the Macaulay null space to BTD.

Here we unravel the BTD structure in the Macaulay null space $\mathbf{K}(d)$, $d \geq d^*$. For the sake of presentation and simplicity, we mainly restrict ourselves to the affine case, but generalizations to the projective case follow by interpreting Vandermonde vectors $\mathbf{v}(d)$ as

$$\mathbf{v}^h(d) = \begin{pmatrix} x_0^d & x_0^{d-1}x_1 & \dots & x_0^{d-2}x_1^2 & x_0^{d-2}x_1x_2 & \dots & x_n^d \end{pmatrix}^T \in \mathbb{C}^{q(d)} \tag{13}$$

and consequently using $\mathbf{j} \in \mathbb{N}^{n+1}$ in the differential functionals (i.e., also include partial derivatives in x_0); see [15]. Details on a special treatment of roots at infinity ($x_0 = 0$) are given when necessary.

4.1. CPD and simple roots. Part I [38] jointly exploits the multiplicative shift invariance in each variable x_j in the null space of the Macaulay matrix of a system with only simple roots. The null space admits a multivariate Vandermonde basis, corresponding to the columns of $\mathbf{V}(d) \in \mathbb{C}^{q(d)\times m}$. This multivariate Vandermonde basis is not readily available. What we can find is a numerical basis, which we stack in $\mathbf{K}(d) \in \mathbb{C}^{q(d)\times m}$. Obviously, we have $\mathbf{K}(d) = \mathbf{V}(d)\mathbf{C}(d)^T$ with an invertible basis transformation matrix $\mathbf{C}(d) \in \mathbb{C}^{m\times m}$. Exploiting the structure results in the following third-order tensor CPD [38]:

$$
\begin{aligned}
\mathbf{Y}_{[1,2;3]} &\overset{\text{def}}{=} \begin{pmatrix} \overline{\mathbf{S}}^{(0)}(d-1)\cdot\mathbf{K}(d) \\ \overline{\mathbf{S}}^{(1)}(d-1)\cdot\mathbf{K}(d) \\ \vdots \\ \overline{\mathbf{S}}^{(n)}(d-1)\cdot\mathbf{K}(d) \end{pmatrix} \\
&= \left(\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(m)} \\ \vdots & \vdots & & \vdots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(m)} \end{pmatrix} \odot \mathbf{B}(d-1) \right) \mathbf{C}(d)^T \qquad (14) \\
&= (\mathbf{V}(1)\odot\mathbf{V}(d-1))\cdot\mathbf{C}(d)^T \in \mathbb{C}^{((n+1)\cdot q(d-1))\times m}, \\
\text{or } \mathcal{Y} &= [\![\mathbf{V}(1), \mathbf{V}(d-1), \mathbf{C}(d)]\!] \in \mathbb{C}^{(n+1)\times q(d-1)\times m}, \qquad (15)
\end{aligned}
$$

where $\overline{\mathbf{S}}^{(j)}(d-1)$ selects all rows of $\mathbf{K}(d)$ onto which the rows of $\mathbf{K}(d)$, associated with monomials of degree at most $d-1$ in x_j, are mapped after multiplication with x_j. In the projective case the CPD in (14) is constructed using multivariate Vandermonde matrices $\mathbf{V}^h(1)$, $\mathbf{V}^h(d-1)$ of the form $\mathbf{V}^h(d) = \begin{pmatrix}\mathbf{v}_1^h(d) & \dots & \mathbf{v}_m^h(d)\end{pmatrix} \in \mathbb{C}^{q(d)\times m}$ with $\mathbf{v}_k^h(d)$ as in (13) and containing the kth root $\begin{pmatrix}x_0^{(k)} & x_1^{(k)} & \dots & x_n^{(k)}\end{pmatrix}^T$ in the projective interpretation.

4.2. BTD and multiple roots. Now let $\tilde{\mathbf{V}}(d)$ as in (12) denote a confluent multivariate Vandermonde ("multivariate Vandermonde plus derivative") basis for the null space of the Macaulay matrix of a system with multiple roots:

$$
\mathbf{M}(d)\cdot\tilde{\mathbf{V}}_k(d) = \mathbf{M}(d)\cdot\begin{pmatrix}c_{k0}[\mathbf{v}(d)] & \dots & c_{k,\mu_k-1}[\mathbf{v}(d)]\end{pmatrix} = \mathbf{0}, \qquad k = 1:m_0. \qquad (16)
$$

The multiplicity structure in (16) is not unique [15] (unless $\mu_k = 1$ for all k). Indeed, multiplying both sides in (16) with a nonsingular transformation matrix $\mathbf{T} \in \mathbb{C}^{\mu_k\times\mu_k}$ yields the equally valid relation

$$
\mathbf{M}(d)\tilde{\mathbf{V}}_k(d)\mathbf{T} = \mathbf{M}(d)\left(\tilde{\mathbf{V}}_k(d)\mathbf{T}\right) = \mathbf{0}. \qquad (17)
$$

In the following we partition the invertible transformation matrix $\mathbf{C}(d)$ so that it matches the partition in (12):

$$
\mathbf{C}(d) = \begin{pmatrix}\mathbf{C}_1(d) & \dots & \mathbf{C}_{m_0}(d)\end{pmatrix} \in \mathbb{C}^{m\times m}.
$$

We emphasize that $\tilde{\mathbf{V}}_k(d)$ ($\tilde{\mathbf{V}}(d)$) is *not* multivariate Vandermonde and that the newly introduced columns in $\tilde{\mathbf{V}}_k(d)$ (in $\tilde{\mathbf{V}}(d)$) do *not* exhibit shift invariance as discussed in [38, section 3.3]. Hence, we cannot implement *simple* spatial smoothing to exploit this shift invariance, and we do not obtain the CPD in (2) anymore.

Example 4.1. Consider again the system in Example 3.5. Since it has $m_0 = 1$ distinct roots, we omit the subscript indicating the numbering of the distinct roots in (12) and use $\tilde{\mathbf{V}}(2) = \tilde{\mathbf{V}}_1(2)$ as in (9). The first column of $\tilde{\mathbf{V}}(2)$ enjoys shift invariance:

$$\tilde{\mathbf{V}}([1\,2\,3],1)\cdot x_1^{(1)} = \begin{pmatrix} 1 \\ \hline x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} \cdot x_1^{(1)} = \begin{pmatrix} x_1^{(1)} \\ \hline x_1^{(1)2} \\ x_1^{(1)} x_2^{(1)} \end{pmatrix} = \tilde{\mathbf{V}}([2\,4\,5],1).$$

Similarly, $\tilde{\mathbf{V}}([1\,2\,3],1)\cdot x_2^{(1)} = \tilde{\mathbf{V}}([3\,5\,6],1)$. However, the other columns do not exhibit this shift-invariance property. For instance, for the second column $\left(\tilde{\mathbf{V}}(2)\right)_2 = \partial_{10}[\mathbf{v}(2)]$ we have

$$\tilde{\mathbf{V}}([1\,2\,3],2)\cdot x_1^{(1)} = \begin{pmatrix} 0 \\ \hline 1 \\ 0 \end{pmatrix} \cdot x_1^{(1)} = \begin{pmatrix} 0 \\ \hline x_1^{(1)} \\ 0 \end{pmatrix} \neq \begin{pmatrix} 1 \\ \hline 2x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \tilde{\mathbf{V}}([2\,4\,5],2).$$

Nonetheless, we can formulate a BTD for $\mathcal{Y}$ using a more general row selection in the confluent multivariate Vandermonde null space of the Macaulay matrix. Theorem 4.2 gives this decomposition, and its derivation is given in Appendix B.

Let us already give that Example 4.7 at the end of this section clarifies the upcoming insights on the well-known univariate playground.

THEOREM 4.2. *Let the system of polynomials $\mathcal{F}$ in n (affine) variables $x_1, \ldots, x_n$ have $m_0 \leq m$ disjoint roots with multiplicity $\mu_k, k = 1 : m_0$. Assume $d = d^{(1)} + d^{(2)} \geq d^*$ with $1 \leq d^{(1)} < d$. Consider the third-order tensor with matrix representation*

$$\text{(18)}\quad \mathbf{Y}_{[1,2;3]}(d^{(1)}, d^{(2)}) = \begin{pmatrix} \overline{\mathbf{S}}^{(0)}(d^{(2)}) \cdot \mathbf{K}(d^{(1)} + d^{(2)}) \\ \overline{\mathbf{S}}^{(1)}(d^{(2)}) \cdot \mathbf{K}(d^{(1)} + d^{(2)}) \\ \vdots \\ \overline{\mathbf{S}}^{(q(d^{(1)})-1)}(d^{(2)}) \cdot \mathbf{K}(d^{(1)} + d^{(2)}) \end{pmatrix} \in \mathbb{C}^{\left(q(d^{(1)})\cdot q(d^{(2)})\right)\times m},$$

where $\mathbf{K}(d^{(1)} + d^{(2)})$ is a basis for the null space of $\mathbf{M}(d^{(1)} + d^{(2)})$. Moreover, $\overline{\mathbf{S}}^{(l)}(d^{(2)}) \in \mathbb{R}^{q(d^{(2)})\times q(d)}$, $l = 0 : q(d^{(1)})$, denote the row selection matrices that select the rows of $\mathbf{K}(d^{(1)} + d^{(2)})$ onto which the monomials of degree 0 up to $d^{(2)}$ are mapped after multiplication with the $(l+1)$th monomial of degree at most $d^{(1)}$ in the degree negative lexicographic order. Then $\mathbf{Y}_{[1,2;3]}$ admits the BTD

$$\text{(19)}\quad \mathcal{Y}(d^{(1)}, d^{(2)}) = \sum_{k=1}^{m_0} \mathcal{G}_k(d^{(1)}, d^{(2)}) \cdot_1 \mathbf{A}_k(d^{(1)}) \cdot_2 \mathbf{B}_k(d^{(2)}) \cdot_3 \mathbf{C}_k(d) \in \mathbb{C}^{q(d^{(1)})\times q(d^{(2)})\times m}$$

with factor matrices $\mathbf{A}_k(d^{(1)}) = \tilde{\mathbf{V}}_k(d^{(1)}) \in \mathbb{C}^{q(d^{(1)})\times \mu_k}$, $\mathbf{B}_k(d^{(2)}) = \tilde{\mathbf{V}}_k(d^{(2)}) \in \mathbb{C}^{q(d^{(2)})\times \mu_k}$, and $\mathbf{C}_k(d) \in \mathbb{C}^{m\times \mu_k}$. The core tensors $\mathcal{G}_k(d^{(1)}, d^{(2)}) \in \mathbb{C}^{\mu_k\times\mu_k\times\mu_k}$ have slices $\mathcal{G}_k(l+1,:,:) = \mathcal{G}_k(:,l+1,:) \in \mathbb{C}^{\mu_k\times\mu_k}$, $l = 0 : \mu_k - 1$, which are upper-triangular or, if $l = 0$, equal to $\mathbf{I}_{\mu_k}$.

In other words, Theorem 4.2 states that if we choose $d^{(1)}$ and $d^{(2)}$ appropriately, then the third-order tensor $\mathcal{Y}$ admits the BTD in (19). See Figure 2 for an illustration. Each of the m_0 terms in Figure 2 reveals in its first and second factor matrices

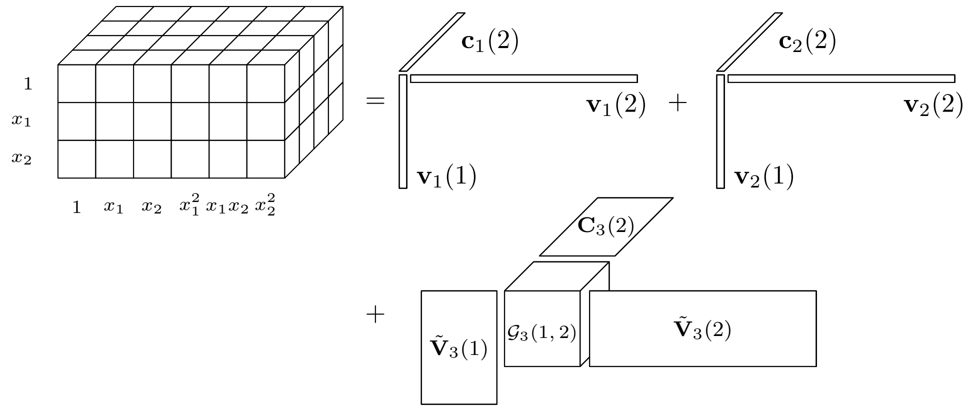


FIG. 2. *Schematic of the BTD* (19) *for* $\mathcal{Y} \in \mathbb{C}^{q(d^{(1)})\times q(d^{(2)})\times m}$ *for a system of* $s = 2$ *polynomial equations in* $n = 2$ *unknowns. Counting multiplicities, the number of roots* $m = 4$*. The number of distinct roots* $m_0 = 3$*. The first two roots are isolated* ($\mu_1 = \mu_2 = 1$)*. The third root has multiplicity* $\mu_3 = 2$ *with depth* $\delta_3 = 1$*. The degrees in Theorem* 4.2 *are chosen as* $d^{(1)} = 1$ *and* $d^{(2)} = 2$ *such that* $d^{(1)} + d^{(2)} = 3 \geq d^* = 2$.

a disjoint root and its multiplicity structure. The dimensions of the core tensors correspond to the multiplicities μ_k. Recall from subsection 3.1.2 that BTD is subject to basic linear transformation indeterminacies. This is consistent with the multiplicity structure of a root being determined up to an invertible basis transformation matrix, as shown in (17).

If all roots are distinct, i.e., if $m_0 = m$, the BTD simplifies to a CPD. In other words, the CPD in [38, eqn. (31)] is the special case of the BTD (19) for which $d^{(1)} = 1$, $d^{(2)} = d - 1$, and $\tilde{\mathbf{V}}_k = (\mathbf{V})_k = \mathbf{v}_k = c_{k0}[\mathbf{v}]$. Note that if $d^{(1)} > 1$, $\mathcal{Y}(d^{(1)}, d^{(2)})$ holds more than $n + 1$ horizontal slices.

Example 4.3. Consider again the system in Example 3.5. Take $d^{(1)} = d^{(2)} = 2$, such that $2 + 2 = 4 \geq 2 = d^*$ and the assumptions of Theorem 4.2 are satisfied. Following the reasoning in Appendix B, it can be verified that $\mathcal{Y}(2, 2)$ in (19) admits the single-term BTD

$$\mathcal{Y}(2,2) = \mathcal{G}(2,2) \cdot_1 \tilde{\mathbf{V}}(2) \cdot_2 \tilde{\mathbf{V}}(2) \cdot_3 \mathbf{C}(4) \in \mathbb{C}^{6\times 6\times 4},$$

with $\tilde{\mathbf{V}}(2)$ as given in (9), and with

$$\mathbf{G}(2,2)_{[1;2,3]} = \mathbf{G}(2,2)_{[2;1,3]} = \left(\begin{array}{cccc|cccc|cccc|cccc} 1 & & & & & 1 & & & & & 1 & & & & & 1 \\ & 1 & & & & & & 2 & & & & 1 & & & & \\ & & 1 & & & & & 1 & & & & & & & & \\ & & & 1 & & & & & & & & & & & & \end{array}\right)$$

comprising identity and upper-triangular matrix slices.

Theorem 4.2 gives only the BTD (19) but not its uniqueness or a way to compute it algebraically. However, if it is unique, it could already be computed by means of optimization-based algorithms [29].

4.3. Uniqueness and algebraic computation of the BTD. Theorem 4.4 below gives conditions that ensure uniqueness of (19) and, furthermore, enable an algebraic computation of the factor matrices using block-diagonalization of certain matrices. While (19) may typically be constructed using lower values for d, $d^{(1)}$, and $d^{(2)}$ than the lower bounds presented by Theorem 4.4, the lower bounds in Theorem 4.4 guarantee uniqueness of the decomposition. Moreover, Theorem 4.4 forms the counterpart to [38, Theorem 6.1], which established the uniqueness of the CPD (15) and the ability to compute it via eigenvalue decompositions in the case of only simple roots.

THEOREM 4.4. *Define* $\mathbf{A} \stackrel{\text{def}}{=} \begin{pmatrix}\mathbf{A}_1 & \dots & \mathbf{A}_{m_0}\end{pmatrix} \in \mathbb{C}^{q(d^{(1)})\times m}$, $\mathbf{B} \stackrel{\text{def}}{=} \begin{pmatrix}\mathbf{B}_1 & \dots & \mathbf{B}_{m_0}\end{pmatrix} \in \mathbb{C}^{q(d^{(2)})\times m}$ *and let* $\mathbf{C} \in \mathbb{C}^{m\times m}$ *be the invertible basis transformation from above. Let* $d = d^{(1)} + d^{(2)}$, *where* $d^{(1)}, d^{(2)}$ *satisfy*

1. $d^{(2)} \geq d^*$,
2. $d^{(1)} \geq \max\{1, \max_k \delta_k\}$.

Then the BTD (19) *is unique.*

Proof. The condition $d^{(1)} \geq \max\{1, \max_k \delta_k\}$ ensures that all individual blocks $\mathbf{A}_r = \tilde{\mathbf{V}}_r(d^{(1)})$, $r = 1 : m_0$, have full column rank, so (19) is a decomposition into a sum of multilinear rank-(μ_r, μ_r, μ_r) terms. To prove uniqueness we show that the assumptions in Theorem 3.1 hold for $R = m_0$, $I_1 = q(d^{(1)})$, $I_2 = q(d^{(2)})$, and $I_3 = m$. By Theorem 4.2, it is sufficient to show that assumptions (5) and (6) hold. Note that both conditions always imply $d \geq d^* + 1$. For $d \geq d^*$ we have that $\dim \operatorname{null}(\mathbf{M}(d)) = m$ and that the numerical basis $\mathbf{K}(d) \in \mathbb{C}^{q(d)\times m}$ has full column rank $r_{\mathbf{K}(d)} = m$. Thus, $\mathbf{C}$ has also full column rank. Since $\mathbf{B} = \tilde{\mathbf{V}}(d^{(2)})$ and $r_{\tilde{\mathbf{V}}(d^{(2)})} = m$ for $d^{(2)} \geq d^*$ [15], the second condition ensures full column rank of $\mathbf{B}$. Finally, since the first columns of the $\mathbf{A}_k$, $k = 1 : m_0$, are genuine multivariate Vandermonde vectors associated to the m_0 distinct roots, (6) is always satisfied for $d^{(1)} \geq 1$. ∎

Example 4.5. We revisit Example 3.5 (see also Example 4.3) with $n = s = 2$, initial degree $d_0 = 2$ so that $d_* = d_0 \cdot n - n = 2$, $m_0 = 1 < m = d_0^2 = 4$, $\mu_1 = 4$, $\delta_1 = 2$. Taking $d^{(2)} = 2$ and $d^{(1)} = 2$ as we did before satisfies the conditions 1 and 2 of Theorem 4.4.

Under the conditions of Theorem 4.4, the BTD of $\mathcal{Y}$ and its factor matrices can be computed algebraically by following the steps outlined in subsection 3.1.2. Similarly as in [38, Algorithm 1], we start from a compressed version $\mathcal{Y}_c \in \mathbb{C}^{q(d^{(1)})\times m\times m}$ of $\mathcal{Y}$.

The algebraic method in subsection 3.1.2 requires a block-diagonal decomposition of $\mathbf{T}_2\mathbf{T}_1^{-1}$, where $\mathbf{T}_1 \stackrel{\text{def}}{=} \mathcal{Y}_c \cdot_1 \mathbf{f}^T$, $\mathbf{T}_2 \stackrel{\text{def}}{=} \mathcal{Y}_c \cdot_1 \mathbf{g}^T \in \mathbb{C}^{m\times m}$ are generic linear combinations of the horizontal slices $\mathcal{Y}_c(i,:,:)$ with $\mathbf{f}, \mathbf{g} \in \mathbb{C}^{I_1}$. In practice, one would compute this block-diagonal decomposition of $\mathbf{T}_2\mathbf{T}_1^{-1}$ from a Schur decomposition (see [19, section 7.6.3]), resulting in factor matrices $\mathbf{A}$, $\mathbf{B}$ that are not in confluent multivariate Vandermonde form, but rather in the form $\mathbf{A} = \tilde{\mathbf{V}}(d^{(1)})\mathbf{R}^{(1)}$, $\mathbf{B} = \tilde{\mathbf{V}}(d^{(2)})\mathbf{R}^{(2)}$ with some unknown invertible transformations $\mathbf{R}^{(1)}$, $\mathbf{R}^{(2)} \in \mathbb{C}^{m\times m}$. This does not immediately reveal the roots, but we will see later in section 6 how the roots and their multiplicities can nevertheless be retrieved.

4.4. Connection with NPA. Let the system of polynomials $\mathcal{F}$ have $m_0 \leq m$ disjoint roots. Consider the family of multiplication tables $\{\mathbf{A}_{x_j}\}_{j=1}^n$, where $\mathbf{A}_h \in \mathbb{C}^{m\times m}$ represents a multiplication with the residue class $[h]$ in the m-dimensional quotient ring $\mathcal{C}^n/\mathcal{I} = \mathcal{C}^n/\langle\mathcal{F}\rangle$ associated to an arbitrary basis, e.g., the standard monomials.[5] Then the central theorem of NPA [34, Theorem 2.27] states that a μ_k-

[5]Standard monomials refer to the monomials in the normal set basis, which relate to the Macaulay matrix as follows. If we flip the columns of $\mathbf{M}(d)$ from left to right, then the standard monomials

fold root $\mathbf{x}^{(k)}$ of $\mathcal{F}$ yields eigenvalues $x_j^{(k)}$ of $\mathbf{A}_{x_j}$ with algebraic multiplicity μ_k. There is also an associated joint invariant subspace span$(\mathbf{X}_k)$, $\mathbf{X}_k \in \mathbb{C}^{m\times\mu_k}$ such that

$$\mathbf{A}_{x_j}\begin{pmatrix}\mathbf{X}_1 & \dots & \mathbf{X}_{m_0}\end{pmatrix} = \begin{pmatrix}\mathbf{X}_1 & \dots & \mathbf{X}_{m_0}\end{pmatrix}\begin{pmatrix}\mathbf{T}_{x_{j,1}} & & \\ & \ddots & \\ & & \mathbf{T}_{x_{j,m_0}}\end{pmatrix}, \tag{20}$$

with $\mathbf{T}_{x_{j,k}} \in \mathbb{C}^{\mu_k\times\mu_k}$ upper-triangular and $x_j^{(k)}$ on the diagonal. Note that only the first columns of $\mathbf{X}_k$ are joint eigenvectors. In the case of only simple roots ($m = m_0$), this reduces to a joint diagonalization of the multiplication tables $\left\{\mathbf{A}_{x_j}\right\}_{j=1}^{n}$. Briefly, [38, Corollary 6.3] showed that if a tensor $\mathcal{H}(d) \in \mathbb{C}^{n\times m\times m}$ is constructed as in (14), (15) but using a column echelon basis $\mathbf{H}(d)$ of null$(\mathbf{M}(d))$ as well as n proper selection matrices, associated to the m standard monomials, then the n slices of $\mathcal{H}$ are equal to the n multiplication tables w.r.t. the normal set basis for $\mathcal{C}^n/\langle\mathcal{F}\rangle$, i.e., $\mathcal{Y}(j,:,:) = \mathbf{A}_{x_j}$, $j = 1:n$. Corollary 4.6 extends this result to roots with multiplicities using the BTD from Theorem 4.2. The tensors $\mathcal{H}$ in Corollary 4.6 and [38, Corollary 6.3] are constructed in the same manner, but in the case of roots with multiplicities, the expressions are more involved.

Corollary 4.6. *Let the polynomial system $\mathcal{F}$ have $m_0 \leq m$ disjoint affine roots with multiplicity $\mu_k, k = 1 : m_0$, and let $\mathbf{H}(d)$ hold the column echelon basis of* null$(\mathbf{M}(d))$. *For $d \geq d^* + 1$ let $d^{(1)}$, $d^{(2)}$ satisfy the conditions of Theorem* 4.4. *Consider the third-order tensor $\mathcal{H}(d)$ with matrix representation*

$$\mathbf{H}_{[1,2;3]} = \begin{pmatrix}\hat{\tilde{\mathbf{S}}}^{(1)}(d-1)\mathbf{H}(d) \\ \vdots \\ \hat{\tilde{\mathbf{S}}}^{(n)}(d-1)\mathbf{H}(d)\end{pmatrix} \in \mathbb{C}^{(n\cdot m)\times m},$$

where $\hat{\tilde{\mathbf{S}}}^{(j)}(d-1)$ denotes the row selection matrix that selects the rows of $\mathbf{H}(d)$ onto which the m standard monomials are mapped after multiplication with x_j. Then the n slices $\{\mathcal{H}(j,:,:)\}_{j=1}^{n}$ of $\mathcal{H}(d)$ are equal to the n multiplication tables $\left\{\mathbf{A}_{x_j}\right\}_{j=1}^{n}$ w.r.t. the normal set basis for the quotient ring $\mathcal{C}^n/\langle\mathcal{F}\rangle$.

Proof. The structure in (19) does not depend on the specific choice $\mathbf{K}(d) = \tilde{\mathbf{V}}(d)\mathbf{C}(d)^T$ that is made for the basis of null$(\mathbf{M}(d))$, so the BTD (19) holds for $\mathbf{K}(d) = \mathbf{H}(d)$ as well. For a slice of $\mathcal{H}(d)$ we have

$$\operatorname{vec}\left(\mathcal{H}(j,:,:)\right)^T = (\mathbf{I}_{n+1})_{j+1}^T \sum_{k=1}^{m_0} \mathbf{A}_k(1)\cdot(\mathbf{G}_k(d))_{[1;3,2]}\cdot\Big(\mathbf{C}_k(d)\otimes\hat{\mathbf{B}}_k(d-1)\Big)^T,$$

where $\hat{\mathbf{B}}_k \in \mathbb{C}^{m\times\mu_k}$ contains the m rows of $\mathbf{B}_k(d-1) \in \mathbb{C}^{q(d-1)\times\mu_k}$ that correspond to the m standard monomials. At least one standard monomial has exactly degree d^*, meaning that $d = d^* + 1$ is needed for $\mathbf{B}_k(d-1)$ to contain all the rows that

are those monomials that correspond to the linearly dependent columns of the row echelon form of the flipped matrix [1, p. 97]. Equivalently, they correspond to the first m linearly independent rows of a multivariate Vandermonde basis for null$(\mathbf{M}(d))$ [14].

correspond to the standard monomials. Multiplication with $(\mathbf{I}_{n+1})^T_{j+1}$ reveals

$$\text{(21)} \qquad \operatorname{vec}(\mathcal{H}(j,:,:))^T = \sum_{k=1}^{m_0} \underbrace{\mathbf{A}_k(1)(j+1,:) \cdot \mathbf{G}_{k[1;3,2]}}_{= x_j^{(k)} \cdot \mathbf{G}_{k[1;3,2]}(1,:) + 1_j \cdot \mathbf{G}_{k[1;3,2]}(l+1,:)} \cdot \left(\mathbf{C}_k \otimes \hat{\mathbf{B}}_k\right)^T,$$

where $1_j = 1$ if $\partial_{0\ldots j\ldots 0} = c_{kl} \in \mathcal{D}[\mathbf{x}^{(k)}]$ and 0 otherwise. Let $\tilde{\mathbf{V}}(d) = \mathbf{H}(d)\mathbf{U}$, where $\mathbf{U} \in \mathbb{C}^{m\times m}$ is an invertible transformation matrix and $\mathbf{C}^T = \mathbf{U}^{-1}$. Reference [16, Proposition 1] shows that $\begin{pmatrix}\hat{\mathbf{B}}_1 & \ldots & \hat{\mathbf{B}}_{m_0}\end{pmatrix} = \mathbf{U}$, which, together with a matricization of (21), yields

$$\begin{aligned}
\mathcal{H}(j,:,:) &= \sum_{k=1}^{m_0} \hat{\mathbf{B}}_k \left(x_j^{(k)} \cdot \mathcal{G}_k(1,:,:) + 1_j \cdot \mathcal{G}_k(l+1,:,:)\right) \mathbf{C}_k^T \\
&= \sum_{k=1}^{m_0} \hat{\mathbf{B}}_k \underbrace{\left(x_j^{(k)} \cdot \mathbf{I}_{\mu_k} + 1_j \cdot \mathcal{G}_k(l+1,:,:)\right)}_{= \mathbf{T}_{x_{j,k}}} \mathbf{C}_k^T = \mathbf{U} \begin{pmatrix} \mathbf{T}_{x_{j,1}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{T}_{x_{j,m_0}} \end{pmatrix} \mathbf{U}^{-1},
\end{aligned}$$

where the right-hand side equals $\mathbf{A}_{x_j}$ per [34, Theorem 2.27]. ∎

We give an example that connects the insights that have emerged for multivariate polynomial equations with multiple roots to the basic univariate case.

Example 4.7. Consider the univariate polynomial equation

$$f(x) = (x-\alpha)^2 = x^2 - 2\alpha x + \alpha^2 = 0$$

of degree $d = 2$ and with a total number of $m = 2$ roots. The polynomial f has only $m_0 = 1$ disjoint root $x^{(1)} = \alpha$, with multiplicity $\mu_1 = 2$.

The Frobenius companion matrix of f,

$$\mathbf{A}_x = \begin{pmatrix} 0 & 1 \\ -\alpha^2 & 2\alpha \end{pmatrix},$$

is the matrix that describes the effect of multiplying the normal set $\{1, x\}$ with $h = x$ in terms of $\{1, x\}$; i.e., in terms of [34, Theorem 2.27] it is a multiplication table. The matrix $\mathbf{A}_x$ has the eigenvalue $x^{(1)} = \alpha$ with algebraic multiplicity $\mu_1 = 2$ but with geometric multiplicity 1. Consequently, $\mathbf{A}_x$ cannot be diagonalized, but it admits a Jordan canonical form, $\mathbf{A}_x = \mathbf{U}\mathbf{T}\mathbf{U}^{-1}$, in which

$$\mathbf{T} = \begin{pmatrix} x^{(1)} & 1 \\ 0 & x^{(1)} \end{pmatrix} = \begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix} \quad \text{and} \quad \mathbf{U} = \begin{pmatrix} -\alpha & 1 \\ -\alpha^2 & 0 \end{pmatrix}$$

are an upper-triangular matrix with both diagonal elements equal to $x^{(1)} = \alpha$ and a matrix whose columns span the invariant subspace of dimension $\mu_1 = 2$, respectively.

In the univariate case, the multiplicity structure is of the form $\mathcal{D}[x^{(1)}] = \left\{\partial_l[x^{(1)}]\right\}_{l=0}^{\mu_1 - 1}$. A confluent Vandermonde basis for the $(m = 2)$-dimensional null space of $\mathbf{f}^T = \begin{pmatrix} \alpha^2 & -2\alpha & 1 \end{pmatrix}$ is thus given by

$$\tilde{\mathbf{V}}_1 = \begin{pmatrix} \partial_0[\mathbf{v}_1] & \partial_1[\mathbf{v}_1] \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \alpha & 1 \\ \alpha^2 & 2\alpha \end{pmatrix},$$

with $\mathbf{v}_1(2) = \begin{pmatrix}1 & \alpha & \alpha^2\end{pmatrix}^T$. Take $d^{(1)} = d^{(2)} = 1$, such that the conditions in Theorem 4.2 are satisfied: $d^{(1)} + d^{(2)} = 1 + 1 = 2 \geq 2 = d^* + 1 > d^*$.

Next, as mentioned in the proof of Corollary 4.6, $\mathbf{Y}(1,1)_{[1,2;3]}$ in (18) may be constructed from $\mathbf{H}(2) = \tilde{\mathbf{V}}_1\mathbf{C}^T$ as a special case of $\mathbf{K}(2) = \mathbf{V}(2)\mathbf{C}(2)^T$:

$$\begin{aligned}\mathbf{Y}_{[1,2;3]}(1,1) &= \left(\begin{array}{c}\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}_{2\times 1}\end{pmatrix}\cdot \mathbf{H}(2)\\ \hline \begin{pmatrix}\mathbf{0}_{2\times 1} & \mathbf{I}_2\end{pmatrix}\cdot \mathbf{H}(2)\end{array}\right) = \left(\begin{array}{c}\underline{\mathbf{H}}\\ \hline \overline{\mathbf{H}}\end{array}\right) = \left(\begin{array}{cc}1 & 0\\ 0 & 1\\ \hline 0 & 1\\ -\alpha^2 & 2\alpha\end{array}\right) = \left(\begin{array}{c}\mathbf{I}_2\\ \hline \mathbf{A}_x\end{array}\right)\\ &= \left(\begin{array}{c}\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}_{2\times 1}\end{pmatrix}\cdot \tilde{\mathbf{V}}_1(2)\\ \hline \begin{pmatrix}\mathbf{0}_{2\times 1} & \mathbf{I}_2\end{pmatrix}\cdot \tilde{\mathbf{V}}_1(2)\end{array}\right)\mathbf{C}(2)^T = \left(\begin{array}{cc}\underline{\partial_0[\mathbf{v}_1(2)]} & \underline{\partial_1[\mathbf{v}_1(2)]}\\ \hline \overline{\partial_0[\mathbf{v}_1(2)]} & \overline{\partial_1[\mathbf{v}_1(2)]}\end{array}\right)\mathbf{C}(2)^T\\ &= \left(\begin{array}{cc}1 & 0\\ \alpha & 1\\ \hline \alpha & 1\\ \alpha^2 & 2\alpha\end{array}\right)\mathbf{C}(2)^T,\end{aligned}$$

in which the basis transformation matrix

$$\mathbf{C}(2)^T = \begin{pmatrix}1 & 0\\ \alpha & 1\end{pmatrix}^{-1}.$$

It can be verified that $\mathcal{Y}(1,1)$ admits the single-term BTD

$$\mathcal{Y}(1,1) = \mathcal{G} \cdot_1 \tilde{\mathbf{V}}_1(1) \cdot_2 \tilde{\mathbf{V}}_1(1) \cdot_3 \mathbf{C}(2) \in \mathbb{C}^{2\times 2\times 2} \tag{22}$$

in which the core tensor, given by

$$\mathbf{G}_{[1;2,3]} = \mathbf{G}_{[2;1,3]} = \left(\begin{array}{cc|cc}1 & 0 & 0 & 1\\ 0 & 1 & 0 & 0\end{array}\right), \tag{23}$$

can be seen as a three-way variant of a (2×2) Jordan cell. Given that $\partial_0[\mathbf{v}_1] = \mathbf{v}_1$, (22) becomes

$$\mathcal{Y}(1,1) = \mathbf{v}_1(1)\otimes\mathbf{v}_1(1)\otimes\mathbf{c}_{1,1} + \underbrace{\partial_1[\mathbf{v}_1(1)]\otimes\mathbf{v}_1(1)\otimes\mathbf{c}_{1,2} + \mathbf{v}_1(1)\otimes\partial_1[\mathbf{v}_1(1)]\otimes\mathbf{c}_{1,2}}_{\partial_1[\mathbf{v}_1(1)\otimes\mathbf{v}_1(1)]\otimes\mathbf{c}_{1,2}}. \tag{24}$$

5. Connection with border rank and typical rank.

The concepts of border and typical ranks belong to the striking differences between linear (matrix) algebra and multilinear (tensor) algebra. Subsection 5.1 and 5.2 will discuss border rank and typical rank of a tensor, respectively, and establish a connection with the BTD in Theorem 4.2. Next to novel fundamental insights, the conclusions at the end of each subsection will be used to design algorithms in section 6.

5.1. Border rank.

The set of tensors that have rank at most R,

$$\begin{aligned}S_R(I_1,I_2,I_3) &= \{\mathcal{T}\in\mathbb{C}^{I_1\times I_2\times I_3} \mid r_{\mathcal{T}} \leq R\}\\ &= \{\mathcal{T}\in\mathbb{C}^{I_1\times I_2\times I_3} \mid \exists\, \mathbf{A}\in\mathbb{C}^{I_1\times R}, \mathbf{B}\in\mathbb{C}^{I_2\times R}, \mathbf{C}\in\mathbb{C}^{I_3\times R} : \mathcal{T} = [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!]\},\end{aligned}$$

is not closed for $R \geq 2$ [13]. A consequence is that the computation of the best rank-R approximation of $\mathcal{T}\in\mathbb{C}^{I_1\times I_2\times I_3}$ may result in a sequence of rank-R estimates $\mathcal{T}_n$

that converge to a boundary point $\hat{\mathcal{T}}$ of $S_R(I_1, I_2, I_3)$ which itself has rank $r_{\hat{\mathcal{T}}} > R$. In such a case, the best rank-R approximation does not exist; the cost function has an infimum but not a minimum. If a tensor $\mathcal{T}$ can be approximated arbitrarily well by rank-R tensors, and R is minimal in this sense, then $\mathcal{T}$ is said to have border rank R. Numerically, it is observed that the convergence towards $\hat{\mathcal{T}}$ is slow and that some of the rank-1 terms "diverge" in the sense that they become increasingly linearly dependent, while their norms grow without bound [25, 24]. The columns of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ that correspond to the diverging rank-1 terms necessarily become more and more linearly dependent as well.

Example 5.1 ([13, Proposition 4.6]). Consider the third-order tensor

$$\mathcal{T} = \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u}, \tag{25}$$

with $\mathbf{u}$ and $\mathbf{v}$ linearly independent. The tensor $\mathcal{T}$ is known to have rank $r_{\mathcal{T}} = 3 > 2$ and border rank 2 [25]. It is approximated arbitrarily well, for $n \to \infty$, by a sequence of two diverging rank-1 terms:

$$\begin{aligned} \mathcal{T}_n &= n\left(\mathbf{u} + \frac{1}{n}\mathbf{v}\right) \otimes \left(\mathbf{u} + \frac{1}{n}\mathbf{v}\right) \otimes \left(\mathbf{u} + \frac{1}{n}\mathbf{v}\right) - n\mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u} \\ &= \mathcal{T} + \frac{1}{n}\left(\mathbf{v} \otimes \mathbf{v} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{v} + \frac{1}{n}\mathbf{v} \otimes \mathbf{v} \otimes \mathbf{v}\right) = \mathcal{T} + \mathcal{O}\left(\frac{1}{n}\right). \end{aligned} \tag{26}$$

Theorem 5.2 shows that if $\mathcal{T}$ is the limit sum of two diverging rank-1 terms, it has multilinear rank $(2, 2, 2)$ and the core tensor admits a third-order variant of the Jordan canonical form of (2×2) matrices.

THEOREM 5.2 ([13, Lemma 4.7]). *For a group of $R = 2$ diverging rank-1 terms, $\mathcal{T}$ can be written as*

$$\mathcal{T} = \mathcal{G} \cdot_1 \mathbf{A} \cdot_2 \mathbf{B} \cdot_3 \mathbf{C}, \tag{27}$$

where $r_{\mathbf{A}} = r_{\mathbf{B}} = r_{\mathbf{C}} = 2$ and where $\mathcal{G} \in \mathbb{C}^{2\times2\times2}$ is given by

$$\mathbf{G}_{[2;1,3]} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{array}\right). \tag{28}$$

Moreover, $r_{\mathcal{G}} = r_{\mathcal{T}} = 3$.

More generally, divergence can happen in several groups of rank-1 terms, and groups can involve more than two terms [33]. Divergence can be avoided by decomposing the tensor in block terms of proper multilinear rank, rather than rank-1 terms. The multilinear rank of a block term matches the cardinality of the group of diverging rank-1 terms that it represents. In [32] third-order variants of the Jordan canonical form are derived for groups up to four diverging rank-1 terms. In [31, section 2] a procedure is proposed to estimate the multilinear rank of the block terms and to obtain an initialization for the BTD algorithm from a "naively fitted" CPD.

Recall from [38] that in the case of simple roots, $\mathcal{Y}(d)$ has rank m. The CPD of $\mathcal{Y}(d)$ can be related to a matrix EVD in which all eigenvalues are distinct. Example 5.3 illustrates that $\mathcal{Y}(d^{(1)}, d^{(2)})$ in Theorem 4.2 has border rank m in the case of multiple roots. Indeed, roots with multiplicity greater than 1 may be seen as the limit case of simple roots that get closer and closer. In Theorem 4.2 the m_0 groups of μ_k

diverging rank-1 terms are collected in m_0 block terms of multilinear rank (μ_k, μ_k, μ_k), $k = 1 : m_0$. While the CPD is related to an EVD in the case of only distinct roots, the BTD in (19) may be seen as a third-order generalization of the Jordan canonical form when there are eigenvalues that have an algebraic multiplicity greater than the geometric multiplicity.

Example 5.3. Consider again the polynomial equation in Example 4.7. Recall that we built $\mathcal{Y}(1,1)$ from the slices $\mathbf{I}_2$ and $\mathbf{A}_x$. The matrix $(\mathbf{I}_2)^{-1}\mathbf{A}_x = \mathbf{A}_x$ has a double eigenvalue α with geometric multiplicity 1. The matrix $\mathbf{A}_x$ cannot be diagonalized, but it does admit a Jordan canonical form. Further, $\mathcal{Y}(1,1)$ itself admits the third-order variant of the Jordan canonical form in Theorem 5.2; i.e., (22) is an instance of (27) and (23) matches (28). One can show that $r_{\mathcal{Y}} = 3$ but that $\mathcal{Y}(1,1)$ has border rank $m = 2$. Trying to compute a rank-2 PD of $\mathcal{Y}(1,1)$ results in a sequence of $m = 2$ diverging rank-1 terms as in Example 5.1.

On the other hand, Example 4.3 exhibited in fact the third-order variant of a (4×4) Jordan cell in the form of the core tensor $\mathcal{G}(2,2)$. The root with multiplicity 4 led to a block term of border rank 4. Fitting a rank-4 PD results in a sequence of $m = 4$ diverging rank-1 terms.

We can conclude that if we proceed in the multiple root case as we have done for simple roots in [38], i.e., by fitting a rank-m CPD to $\mathcal{Y}(1, d-1)$, this will result in m_0 groups of diverging rank-1 terms, with μ_k rank-1 terms in the kth group. Such divergence does not occur if we fit the BTD (19) to $\mathcal{Y}(d^{(1)}, d^{(2)})$. The crucial point is not to split a multilinear rank-(μ_k, μ_k, μ_k) term into terms of lower multilinear rank, such as rank-1 terms. As in [31, section 2], estimates of the multiplicities μ_k and an initialization for the BTD algorithm may nevertheless be obtained from a "naive" use of the algorithm for simple roots in [38] (see section 6 for an illustration).

5.2. Rank over the real or the complex field. The rank of a tensor depends on the field of the entries. Consider, for instance, $\mathcal{T} \in \mathbb{R}^{2\times2\times2}$, whose entries are sampled randomly from a continuous probability distribution. If $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are constrained to be real, then $r_{\mathcal{T}} = 2$ and $r_{\mathcal{T}} = 3$ both occur with nonzero probability—whereas if $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ can be complex, $r_{\mathcal{T}} = 2$ occurs with probability 1 [23, 2]. When the rank takes more than one value with nonzero possibility, the values that occur are called typical. A rank value that occurs with probability 1 is called generic.

The roots of a system of polynomial equations with real-valued coefficients are real-valued or appear in complex conjugated pairs. Example 5.4 shows that a simple pair of complex conjugated roots yields a real-valued block term of multilinear rank $(2,2,2)$ that takes rank 2 over $\mathbb{C}$ but rank 3 over $\mathbb{R}$. In general, the computation of the roots of a system of polynomial equations with real-valued coefficients can be done in $\mathbb{R}$, provided we allow block terms, where block terms that take rank 2 over $\mathbb{C}$ but rank 3 over $\mathbb{R}$ capture simple pairs of complex conjugated roots. Block terms that capture a pair of real-valued simple roots have rank 2 over *both* $\mathbb{C}$ and $\mathbb{R}$; such terms can be further decomposed into two real-valued rank-1 terms that correspond to the individual roots.

Example 5.4. Consider the univariate polynomial equation

$$f(x) = x^2 - 2x + 2 = 0$$

of degree $d = m = 2$. There are $m = 2$ complex conjugated roots: $x^{(1)} = 1 + i$ and $x^{(2)} = 1 - i$. The degree of regularity $d^* = 1$. At $d = d^* + 1 = 2$, $\mathcal{Y}(1, d-1) =$

$\mathcal{Y}(1,1) \in \mathbb{R}^{2\times 2\times 2}$ is constructed from $\mathbf{K}(2)(= \mathbf{V}(2)\mathbf{C}(2)^T) \in \mathbb{R}^{3\times 2}$ as follows:

$$\mathbf{Y}_{[1,2;3]}(1,1) = \left(\frac{\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}_{2\times 1}\end{pmatrix} \cdot \mathbf{K}(2)}{\begin{pmatrix}\mathbf{0}_{2\times 1} & \mathbf{I}_2\end{pmatrix} \cdot \mathbf{K}(2)} \right) = \left(\begin{array}{cc} 1 & 1 \\ 1+i & 1-i \\ \hline 1+i & 1-i \\ (1+i)^2 & (1-i)^2 \end{array} \right) \mathbf{C}(2)^T \in \mathbb{R}^{(2\cdot 2)\times 2}.$$

Since both roots are simple, $\mathcal{Y}$ admits the CPD $\mathcal{Y}(1,1) = [\![\mathbf{V}(1), \mathbf{V}(1), \mathbf{C}(2)]\!]$ with

$$\mathbf{V}(1) = \begin{pmatrix} 1 & 1 \\ 1+i & 1-i \end{pmatrix}.$$

We can rewrite the CPD as a single-term BTD:

$$\mathcal{Y}(1,1) = \mathcal{G}(1,1) \cdot_1 \mathbf{A}(1) \cdot_2 \mathbf{B}(1) \cdot_3 \mathbf{C}(2),$$

in which

$$\mathbf{G}_{[1;3,2]}(1,1) = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right)$$

and in which the (2×2) factor matrices $\mathbf{A}(1) = \mathbf{B}(1) = \mathbf{V}(1)$ and $\mathbf{C}(2)$ are complex-valued. From the sparsity pattern of $\mathcal{G}$ it is obvious that $r_{\mathcal{G}} = r_{\mathcal{Y}} = m = 2$.

The tensor $\mathcal{Y}(1,1)$ can equally well be decomposed as

$$\mathcal{Y}(1,1) = \tilde{\mathcal{G}}(1,1) \cdot_1 \tilde{\mathbf{A}}(1) \cdot_2 \tilde{\mathbf{B}}(1) \cdot_3 \tilde{\mathbf{C}}(2),$$

in which

$$\tilde{\mathcal{G}}(1,1) = \mathcal{G}(1,1) \cdot_1 \left(\mathbf{M}^{(1)}\right)^{-1} \cdot_2 \left(\mathbf{M}^{(2)}\right)^{-1} \cdot_3 \left(\mathbf{M}^{(3)}\right)^{-1},$$

and $\tilde{\mathbf{A}}(1) = \mathbf{A}(1)\mathbf{M}^{(1)}$, $\tilde{\mathbf{B}}(1) = \mathbf{B}(1)\mathbf{M}^{(2)}$, $\tilde{\mathbf{C}}(1) = \mathbf{C}(1)\mathbf{M}^{(3)} \in \mathbb{C}^{2\times 2}$, where $\mathbf{M}^{(1)}, \mathbf{M}^{(2)}$, $\mathbf{M}^{(3)} \in \mathbb{C}^{2\times 2}$ are invertible basis transformation matrices. If we take

$$\mathbf{M}^{(1)} = \mathbf{M}^{(2)} = \mathbf{M}^{(3)} = \mathbf{M} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2i} \\ \frac{1}{2} & -\frac{1}{2i} \end{pmatrix},$$

then $\tilde{\mathbf{A}}(1), \tilde{\mathbf{B}}(1), \tilde{\mathbf{C}}(1)$ are real-valued and

$$\tilde{\mathbf{G}}_{[1;3,2]}(1,1) = \left(\begin{array}{cc|cc} 2 & 0 & 0 & -2 \\ 0 & -2 & -2 & 0 \end{array} \right).$$

The core tensor $\tilde{\mathcal{G}}(1,1) \in \mathbb{R}^{2\times 2\times 2}$ has rank 3 over $\mathbb{R}$. (On the other hand, like $\mathcal{Y}(1,1)$, it has rank 2 over $\mathbb{C}$.)

6. Algorithm.

The goal of this section is to use the fundamental insights from the previous sections to design numerical methods for the multivariate root-finding problem.

6.1. A BTD based root-finding method.

Theorem 4.4 hints at an algebraic BTD-based algorithm illustrated in Algorithm 1 for finding the roots of a polynomial system that can handle multiple roots. It generalizes the algebraic method in [38, Algorithm 1]. For roots with multiplicities, the algorithm first finds the column spaces of the BTD factor matrices $\mathbf{B} \stackrel{\text{def}}{=} \begin{pmatrix}\mathbf{B}_1 & \dots & \mathbf{B}_{m_0}\end{pmatrix} \in \mathbb{C}^{q(d^{(2)})\times m}$. These correspond to the μ_k-dimensional multivariate confluent Vandermonde subspaces associated with the dual spaces of the m_0 disjoint roots.

We comment on the main steps of Algorithm 1.

Algorithm 1 BTD for multivariate polynomial root-finding.

Input: A system $f_i \in \mathcal{C}^n_{d_i}, i = 1 : n$, in $n+1$ projective unknowns $x_j \in \mathbb{C}, j = 0 : n$.
Output: Roots $x_1^{(k)}, \ldots, x_n^{(k)}$ and multiplicities μ_k, $k = 1 : m_0$.
1: Choose $d^{(1)}$, $d^{(2)}$ such that $d = d^{(1)} + d^{(2)} \geq d^* + 1$ and $d^{(1)}, d^{(2)}$ satisfy the conditions of Theorem 4.4.
2: Construct Macaulay matrix $\mathbf{M}(d)$.
3: Compute null space basis $\mathbf{K}(d) \leftarrow \text{null}\,(\mathbf{M}(d))$.
4: **for** $j = 0 : q(d^{(1)}) - 1$ **do**
5: $\quad \mathcal{Y}(j+1, :, :) \leftarrow \overline{\mathbf{S}}^{(j)}(d^{(2)}) \cdot \mathbf{K}(d)$.
6: Compute the SVD $\mathbf{Y}_{[2;1,3]} = \mathbf{U}^{(2)} \cdot \mathbf{\Sigma}^{(2)} \cdot \mathbf{U}^{(1,3)H}$.
7: Orthogonal compression: $\mathcal{Y}_c \leftarrow \mathcal{Y} \cdot_2 \mathbf{U}^{(2)H}$.
8: Compute the BTD

$$\mathcal{Y}_c = \sum_{k=1}^{m_0} \mathcal{G}_k \cdot_1 \mathbf{A}_k \cdot_2 \tilde{\mathbf{B}}_k \cdot_3 \mathbf{C}_k, \tag{29}$$

with $\mathcal{G}_k \in \mathbb{C}^{\mu_k \times \mu_k \times \mu_k}$, $\mathbf{A}_k \in \mathbb{C}^{q(d^{(1)}) \times \mu_k}$, and $\tilde{\mathbf{B}}_k$, $\mathbf{C}_k \in \mathbb{C}^{m \times \mu_k}$, $k = 1 : m_0$.
9: Expand $\mathbf{B}_k = \mathbf{U}^{(2)} \tilde{\mathbf{B}}_k \in \mathbb{C}^{q(d^{(2)}) \times \mu_k}$ and retrieve the roots via generalized ESPRIT approach, $k = 1 : m_0$.
10: **return** $x_1^{(k)}, \ldots, x_n^{(k)}$ and μ_k, $k = 1 : m_0$.

Step 1. The degrees $d^{(1)}$, $d^{(2)}$ have to be chosen sufficiently large according to the conditions of Theorem 4.4 to ensure uniqueness of the BTD and to allow its algebraic computation. The condition $d^{(1)} \geq \max\{1, \max_k \delta_k\}$ leads to the obstacle that the depths δ_k of the roots are generally unknown beforehand. It holds that $\delta_k \leq \mu_k - 1$, but also the multiplicities μ_k are generally not known either. However, if the degree $d^{(1)}$ is chosen large enough, the number m_0 of distinct roots and the individual multiplicities μ_k are directly obtained in the course of the algebraic computation of the BTD in step 8, where m_0 is the number of detected terms and the μ_k appear as the sizes of the individual blocks in the factor matrices. One obvious possibility is to use the upper bound $\delta_k \leq \max_i d_i$ and set $d^{(1)} = \max_i d_i$. However, such an increase in d would lead to a larger Macaulay matrix and make the computation of basis for the null space more expensive.

Steps 2–5. These are the same calculations as in [38, Algorithm 1] for simple roots. The only difference is that in step 5 more than $n+1$ selections $\overline{\mathbf{S}}^{(j)}(d^{(2)})$ are applied if $d^{(1)} > 1$. These execute a generalized spatial smoothing with monomials of degree greater than one.

Steps 6 *and* 7. As in the root-finding procedure for simple roots [38, Algorithm 1], compression of $\mathcal{Y}$ is carried out. This reduces the computational load in the later steps.

Step 8. Here the factor matrices and cores of the BTD (29) are obtained using the algebraic computation outlined in subsection 3.1.2. The main computational step is the block-diagonalization by similarity of an $m \times m$ matrix. This block-diagonalization returns $\tilde{\mathbf{B}}_k$, $\mathbf{C}_k \in \mathbb{C}^{m \times \mu_k}$, $k = 1 : m_0$, where the column dimensions match the multiplicity μ_k of the kth root (provided $d^{(1)}, d^{(2)}$ have been chosen appropriately). The blocks $\mathbf{B}_k = \mathbf{U}^{(2)} \tilde{\mathbf{B}}_k$, $\mathbf{C}_k$ are the blocks of the second and third factor matrix $\mathbf{B}$, $\mathbf{C}$ of the BTD (19). With $\mathbf{B}_k$, $\mathbf{C}_k$ the blocks $\mathbf{A}_k$ of the first factor matrix and cores

$\mathcal{G}_k$ can be obtained. In step 9 we will see that for obtaining the roots, only $\mathbf{A}_k$ or $\mathbf{B}_k$ are required.

As an alternative one could, similarly to the CPD root-finding method in [38], compute the BTD (19) in step 8 by, e.g., nonlinear Schrödinger (NLS)-type methods [29]. Although this requires in theory less stringent conditions on $d^{(1)}, d^{(2)}$, in practice the performance of such NLS methods is highly dependent on good initial guesses. Thus, the outcome of the algebraic method can be used as an initial guess for NLS methods, which would then refine the quality of the result.

Step 9. The decomposition of $\mathcal{Y}$ obtained in step 8 yields a splitting of contributions of the m_0 different roots. Rank-1 terms are given by vectors $\mathbf{a}_k = \mathbf{A}_k \in \mathbb{C}^{q(d^{(1)})}$, $\mathbf{b}_k = \mathbf{B}_k \in \mathbb{C}^{q(d^{(2)})}$ and belong to simple roots ($\mu_k = 1$) which can be readily retrieved from $\mathbf{A}_k$ or $\mathbf{B}_k$ by means of a simple scaling (e.g., dividing $\mathbf{A}_k$ by its first entry) as discussed in [38]. Alternatively, the multiplicative shift structure of multivariate Vandermonde vectors and matrices can be used: $\overline{\mathbf{S}}^{(i)}\mathbf{A}_k = \overline{\mathbf{S}}^{(0)}\mathbf{A}_k \cdot x_i^{(k)}$, $i = 1 : n$, where $\overline{\mathbf{S}}^{(0)}$, $\overline{\mathbf{S}}^{(i)}$ select the rows associated to monomials of degree 0 to $d^{(1)} - 1$ and, respectively, the rows associated to monomials up to degree $d^{(1)}$, where x_i is of degree at least one. Using the $\mathbf{b}_k$ vectors works in the same way.

Retrieving the roots with multiplicities requires some additional work because, due to the (multi)linear transformation indeterminacies, the computed block matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ do not directly reveal the roots. The roots can be found from $\mathbf{A}_k$ or $\mathbf{B}_k$ by using the generalized multiplicative shift structure of confluent multivariate Vandermonde matrices; see Lemma B.4. We will illustrate this using the $\mathbf{A}_k$ blocks here, but the variant using the $\mathbf{B}_k$ works in the same way. Note that we originally used this multiplicative shift structure to derive the BTD (19) in Theorem 4.2. Recall that $\mathbf{A}_k = \tilde{\mathbf{V}}_k(d^{(1)})\tilde{\mathbf{M}}_k$ for some invertible $\tilde{\mathbf{M}}_k \in \mathbb{C}^{\mu_k \times \mu_k}$, $k = 1 : m_0$. For an affine root $\mathbf{x}_k$ with multiplicity $\mu_k > 1$ and depth $\delta_k \leq \mu_k - 1$, we have for the corresponding confluent multivariate Vandermonde matrix $\tilde{\mathbf{V}}_k(d^{(2)})$

$$\tilde{\mathbf{S}}^{(i)}\tilde{\mathbf{V}}_k(d^{(1)}) = \tilde{\mathbf{S}}^{(0)}\tilde{\mathbf{V}}_k(d^{(1)})\mathbf{J}_k^{(i)}, \quad i = 1 : n,$$

where $\tilde{\mathbf{S}}^{(0)}$ selects the first $I_k \geq \mu_k$ rows of $\tilde{\mathbf{V}}_k(d^{(1)})$ such that $\tilde{\mathbf{S}}^{(0)}\tilde{\mathbf{V}}_k(d^{(1)}) \in \mathbb{C}^{I_k \times \mu_k}$ has full column rank, $\tilde{\mathbf{S}}^{(i)}$ selects the rows of $\tilde{\mathbf{V}}_k(d^{(1)})$ onto which these $I_k \geq \mu_k$ monomials are mapped after a multiplication with the ith variable x_i, and $\mathbf{J}_k^{(i)} \in \mathbb{C}^{\mu_k \times \mu_k}$ is upper-triangular with $x_i^{(k)}$ (the value of the ith variable of the kth distinct root) on the diagonal; see Lemma B.4 in Appendix B.1 or [15, section 4.4], [14, section 6.1] for details. Using $\tilde{\mathbf{V}}_k(d^{(1)}) = \mathbf{A}_k\tilde{\mathbf{M}}_k^{-1}$ yields

$$\left(\tilde{\mathbf{S}}^{(0)}\mathbf{A}_k\right)^{\dagger}\tilde{\mathbf{S}}^{(i)}\mathbf{A}_k = \tilde{\mathbf{M}}_k^{-1}\mathbf{J}_k^{(i)}\tilde{\mathbf{M}}_k \stackrel{\text{def}}{=} \tilde{\mathbf{J}}_k^{(i)}, \quad i = 1 : n.$$

In other words, $\tilde{\mathbf{J}}_k^{(i)}$ can be obtained by solving the linear system $\big(\tilde{\mathbf{S}}^{(0)}\mathbf{A}_k\big)\tilde{\mathbf{J}}_k^{(i)} = \tilde{\mathbf{S}}^{(i)}\mathbf{A}_k$ and it has a single distinct eigenvalue $x_i^{(k)}$ with algebraic multiplicity μ_k. This eigenvalue can be retrieved by $x_i^{(k)} = \text{trace}(\tilde{\mathbf{J}}_k^{(i)})/\mu_k$ or from a Schur decomposition $\tilde{\mathbf{J}}_k^{(i)} = \mathbf{Q}_{k,i}^H\mathbf{R}_{k,i}\mathbf{Q}_{k,i}$ with $\mathbf{Q}_{k,i}$ unitary and $\mathbf{R}_{k,i}$ upper-triangular with $x_i^{(k)}$ on the diagonal.

Step 9 is the only part of Algorithm 1 that needs to be slightly adapted in case of roots at infinity. If $x_0^{(k)} = 0, x_1^{(k)}, \ldots, x_n^{(k)}$ is a root in the $n+1$ projective coordinates, $\tilde{\mathbf{S}}^{(0)}\mathbf{A}_k$ will not have full column rank because $\tilde{\mathbf{V}}_k(d^{(1)})$ will have zero columns and zero top rows. Thus, we use a rank test on $\tilde{\mathbf{S}}^{(0)}\mathbf{A}_k$ to decide whether the kth root is projective or not. If $r_{\tilde{\mathbf{S}}^{(0)}\mathbf{A}_k} < \mu_k$, then the kth root is at infinity and we set $x_0^{(k)} = 0$. Otherwise, we are in the affine situation and set $x_0^{(k)} = 1$ and proceed as outlined above. For a root at infinity, recall that the components $x_i^{(k)}$, $i = 1 : n$, are only determined up to scalar factor $\lambda \neq 0$. We continue in this case by testing if $\tilde{\mathbf{S}}^{(i)}\mathbf{A}_k$ has full column rank for $i = 1 : n$. If $r_{\tilde{\mathbf{S}}^{(i)}\mathbf{A}_k} < \mu_k$, we set $x_i^{(k)} = 0$; otherwise we continue as in the affine case to retrieve the component $x_i^{(k)}$. Note that at least one component $x_i^{(k)}$, $i = 1 : n$, has to be nonzero.

In the form presented in Algorithm 1, the method will return the roots and the individual multiplicities, but not their complete multiplicity structures. One possibility to get the multiplicity structure for a known root with known multiplicity $\mu_k > 1$ is to find the differential functionals $c_{kl} = \sum_{\mathbf{j}} \beta_{\mathbf{j}} \partial_{\mathbf{j}}$, $l = 0 : \mu_k - 1$, from all possible differential functional monomials (Definition 3.3) up to order $\mu_k - 1$: $\partial_{\mathbf{0}}, \partial_{1,0\ldots,0}, \ldots, \partial_{\mathbf{h}}$, $|\mathbf{h}| = \mu_k - 1$. It holds that

$$\tilde{\mathbf{V}}_k(d) = \begin{pmatrix} c_{k0}[\mathbf{v}_k] & \ldots & c_{k\mu_k-1}[\mathbf{v}_k] \end{pmatrix} = \underbrace{\begin{pmatrix} \partial_{\mathbf{0}}[\mathbf{v}_k] & \partial_{1,0\ldots,0}[\mathbf{v}_k] & \ldots & \partial_{\mathbf{h}}[\mathbf{v}_k] \end{pmatrix}}_{\stackrel{\text{def}}{=}\mathbf{U}_k} \mathbf{P}_k,$$

where $\mathbf{P}_k \in \mathbb{C}^{q(\mu_k-1)\times\mu_k}$ holds the coefficients β of the functional c_{kl}. Only $\mathbf{U}_k \in \mathbb{C}^{q(d)\times q(\mu_k-1)}$ is explicitly known in the above equality. Since $\mathbf{M}(d)\tilde{\mathbf{V}}_k(d) = \mathbf{0}$, the matrix $\mathbf{P}_k$ can be computed from the null space problem

$$(\mathbf{M}(d)\mathbf{U}_k)\,\mathbf{P}_k = \mathbf{0};$$

see also [1, section 3.6.2] for similar approaches. Alternatively, one could resort to algorithms for computing the multiplicity structure [26, 28, 7, 41, 6].

6.2. A recursive root-finding method. The BTD in Algorithm 1 and section 5 prompt the unconstrained *recursive* polynomial root-finding Algorithm 2. The algorithm allows us to (recursively) detect various (nested) structures in the null space of the Macaulay matrix. We give this algorithm as an illustration of the remarkable new possibilities in our framework.

Some explanation is in order. In Example 5.4 we combined a pair(s) of rank-1 terms, which per definition are pairs of multilinear rank-$(1,1,1)$ terms, to rewrite the CPD of $\mathcal{Y}(1, d-1)$ as a BTD. That is, we expressed $\mathcal{Y}(1, d-1)$ as a BTD with (one) multilinear rank-$(2,2,2)$ term(s). There is no reason why we should refrain from further combining pairs of multilinear rank-$(2,2,2)$ terms to obtain a BTD in multilinear rank-$(4,4,4)$ term(s), and so on. The converse of this bottom-up reasoning is the top-down schematic in Figure 3; Algorithm 2 is the implied recursive root-finding algorithm. It proceeds as follows. Take the initial input $\hat{\mathcal{Y}} = \hat{\mathcal{Y}}(1, d-1)$ embodying all $R = m$ roots. Next, compute the BTD in step 7 with, for instance, $R_1 = \lfloor m/2 \rfloor$ and $R_2 = \lceil m/2 \rceil$. Then descend to the next level of the tree in Figure 3. Recursively run the same procedure on $\hat{\mathcal{Y}}_1$ embodying $R = R_1 = \lfloor m/2 \rfloor$ roots and on $\hat{\mathcal{Y}}_2$ embodying $R = R_2 = \lceil m/2 \rceil$ roots. After having repeated this procedure $\mathcal{O}(\log_2 m)$ times, each CPD in step 2 in Algorithm 2 (at the leave nodes in Figure 3) reveals the minimum possible $R = 2$ roots left. The columns of the obtained factor matrices $\hat{\mathbf{A}}_n$, $\tilde{\mathbf{B}}_n$, and $\tilde{\mathbf{C}}_n$ could thereby serve as an initialization for computing the BTD or the CPD at a lower level.

The root node in Figure 3 embodies (a full basis for) the $(R = m)$-dimensional null space of the Macaulay matrix. The lower-level nodes embody increasingly lower-dimensional nested subspaces $\subseteq \mathbb{C}^n$. They provide an increasingly finer-grained view

Algorithm 2 Recursive multivariate polynomial root-finding.

Input: A compressed $\hat{\mathcal{Y}} \in \mathbb{C}^{(n+1)\times R\times R}$ $(R \leq m)$ for the system $f_i \in \mathcal{C}^n_{d_i}, i = 1:n$, in the $n+1$ projective unknowns $x_j \in \mathbb{C}, j = 0:n$, with $m_0 = m$ simple roots.

Output: $\left\{\mathbf{x}^{(k)}\right\}_{k=1}^{R}$

if $R \leq 2$ **then** ▷ termination

 Compute the R-term CPD $\hat{\mathcal{Y}} = \left[\!\left[\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}\right]\!\right]$.

 $\mathbf{X} \leftarrow \sim \hat{\mathbf{A}}$.

 return X

else ▷ divide

 $R_1 \leftarrow \lfloor R/2 \rfloor$ and $R_2 \leftarrow \lceil R/2 \rceil$.

 Compute the BTD

$$\hat{\mathcal{Y}} = \underbrace{\hat{\mathcal{G}}_1 \cdot_1 \hat{\mathbf{A}}_1 \cdot_2 \hat{\mathbf{B}}_1 \cdot_3 \hat{\mathbf{C}}_1}_{=\hat{\mathcal{Y}}_1 \in \mathbb{C}^{n+1\times R_1\times R_1}} + \underbrace{\hat{\mathcal{G}}_2 \cdot_1 \hat{\mathbf{A}}_2 \cdot_2 \hat{\mathbf{B}}_2 \cdot_3 \hat{\mathbf{C}}_2}_{=\hat{\mathcal{Y}}_2 \in \mathbb{C}^{n+1\times R_2\times R_2}}$$

 in which $\hat{\mathcal{G}}_1 \in \mathbb{C}^{R_1\times R_1\times R_1}$ and $\hat{\mathcal{G}}_2 \in \mathbb{C}^{R_2\times R_2\times R_2}$.

 Compress $\hat{\mathcal{Y}}_1$ and $\hat{\mathcal{Y}}_2$ using the MLSVD.

 return { Algorithm 2($\hat{\mathcal{Y}}_1$), Algorithm 2($\hat{\mathcal{Y}}_2$) } ▷ conquer

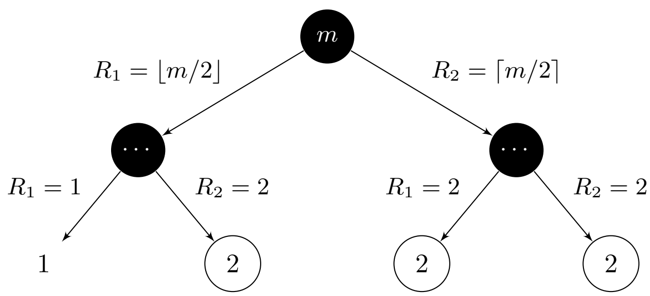


Fig. 3. *Tree-like schematic of a complete run of Algorithm 2 for $\hat{\mathcal{Y}} = \hat{\mathcal{Y}}(1, d-1) \in \mathbb{C}^{(n+1)\times m\times m}$. BTDs at the top levels (second and third mode dimensions $R > 2$) are indicated in black, and CPDs in the leaves (with $R \leq 2$) are indicated in white. The rank values $r_{\hat{\mathcal{Y}}} = R$ are also depicted in each node.*

on the roots $\mathbf{x}^{(k)} \in \mathbb{C}^n$ of the system. One could alternatively terminate the recursion over $\mathbb{R}$ at a multilinear rank-$(2, 2, 2)$, rank-3 term that corresponds to a pair of complex conjugated roots, or at a multilinear rank-(μ_k, μ_k, μ_k) term. In the latter case the leaf node would embody the μ_k-dimensional dual space $\mathcal{D}[\mathbf{x}^{(k)}]$. Owing to many NLS runs, the recursive procedure does not in the case of simple roots compete with [38, Algorithm 1] in terms of computational cost, but it is extremely flexible and interesting conceptually. One could, for instance, decide to "zoom in" on a select cluster of roots in one block term. Example 6.1 sketches the idea.

Example 6.1. Consider first the univariate case. Say that we are only interested in the roots of a univariate polynomial $f(x)$ within a Δ-neighborhood of a given x, i.e., roots $x + \delta, |\delta| \leq \Delta$. For

$$\mathbf{v}_x = \begin{pmatrix} 1 & x & x^2 & \ldots & x^d \end{pmatrix}^T \quad \text{and} \quad \mathbf{v}_{x+\delta} = \begin{pmatrix} 1 & x+\delta & (x+\delta)^2 & \ldots & (x+\delta)^d \end{pmatrix}^T$$

we have

$$\cos(\mathbf{v}_x \sphericalangle \mathbf{v}_{x+\delta}) = \frac{\langle \mathbf{v}_x, \mathbf{v}_{x+\delta}\rangle}{\|\mathbf{v}_x\|\|\mathbf{v}_{x+\delta}\|} = \frac{\frac{1-[x(x+\delta)]^{d+1}}{1-x(x+\delta)}}{\sqrt{\frac{1-[x^2]^{d+1}}{1-x^2}}\sqrt{\frac{1-[(x+\delta)^2]^{d+1}}{1-(x+\delta)^2}}}. \tag{30}$$

Evidently, $\lim_{|\delta|\le\Delta\to 0}\cos(\mathbf{v}_x \sphericalangle \mathbf{v}_{x+\delta}) = 1$. To assess whether a candidate root y is sufficiently close to x to be of further interest, we will consider $|x-y|$ if both values are available. If the Vandermonde vectors $\mathbf{v}_x$ and $\mathbf{v}_y$ are available, we may obviously also compare the latter, as is clear from (30). However, the block terms in step 7 of Algorithm 2 are characterized by confluent Vandermonde subspaces rather than individual Vandermonde vectors. The subspaces may be generated by several roots, which can themselves be simple or have multiplicity greater than 1. Here, we can assess the angle between a subspace (say $\mathcal{S}$) and Vandermonde vector $\mathbf{v}_x$ of matching size. For a block term that captures (possibly among other roots) a root y that is close to x, $\cos(\mathbf{v}_x \sphericalangle \mathcal{S})$ is bounded from below by (30) for a given tolerance Δ. Conversely, we can discard the block terms for which $\cos(\mathbf{v}_x \sphericalangle \mathcal{S})$ is not large enough, since their subspaces cannot contain a Vandermonde vector with a generator sufficiently close to x.

In the multivariate case it is possible to assess the proximity for all variables together. Let us consider the bivariate case by way of example. Let $\boldsymbol{\Delta} = \begin{pmatrix}\delta_1 & \delta_2\end{pmatrix}^T$ demarcate a region around $\mathbf{x} = \begin{pmatrix}x_1 & x_2\end{pmatrix}^T$. For assessing the proximity of $\mathbf{v}_\mathbf{x} = \mathbf{v}_{x_1} \otimes \mathbf{v}_{x_2}$ and $\mathbf{v}_{\mathbf{x}+\boldsymbol{\delta}} = \mathbf{v}_{x_1+\delta_1} \otimes \mathbf{v}_{x_2+\delta_2}$, note that

$$\begin{aligned}\langle \mathbf{v}_{x_1} \otimes \mathbf{v}_{x_2}, \mathbf{v}_{x_1+\delta_1} \otimes \mathbf{v}_{x_2+\delta_2}\rangle &= (\mathbf{v}_{x_1} \otimes \mathbf{v}_{x_2})^H (\mathbf{v}_{x_1+\delta_1} \otimes \mathbf{v}_{x_2+\delta_2})\\ &= \left(\mathbf{v}_{x_1}^H \mathbf{v}_{x_1+\delta_1}\right)\cdot\left(\mathbf{v}_{x_2}^H \mathbf{v}_{x_2+\delta_2}\right)\\ &= \langle \mathbf{v}_{x_1}, \mathbf{v}_{x_1+\delta_1}\rangle \cdot \langle \mathbf{v}_{x_2}, \mathbf{v}_{x_2+\delta_2}\rangle\end{aligned}$$

and that $\|\mathbf{v}_{x_1} \otimes \mathbf{v}_{x_2}\| = \|\mathbf{v}_{x_1}\| \cdot \|\mathbf{v}_{x_2}\|$. This allows the threshold (30) to be replaced by a product of such thresholds.

7. Experimental results.

This section contains the results of some numerical experiments that illustrate the potential of our approach.

7.1. BTD-based root-finding.

As an illustration of the discussion in subsection 5.1 we compare fitting of the m_0-term BTD (19) and the m-term CPD (2) in the multiple root case, and we showcase the divergence of rank-1 terms when fitting the CPD. By way of example, we consider the system [35, Example 1.3.1]

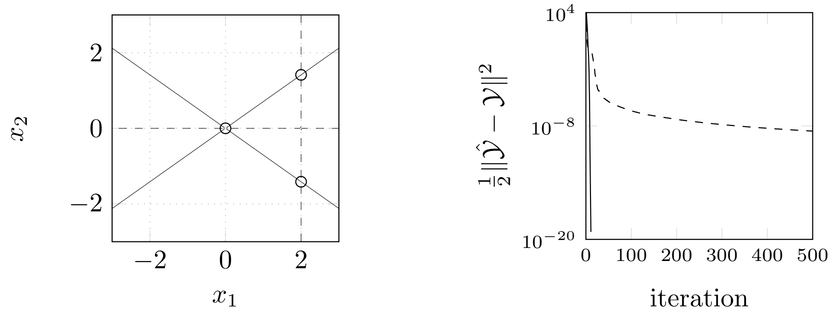
$$\begin{cases} f_1(x_1, x_2) = x_1x_2 - 2x_2 = 0,\\ f_2(x_1, x_2) = 2x_2^2 - x_1^2 = 0\end{cases} \tag{31}$$

shown in Figure 4(a). We have $s = n = 2$, $d_0 = 2$, $d^* = 2+2-2 = 2$, and $m = 2\cdot 2 = 4$, but $m_0 = 3$. The system has $m_0 = 3 < 4 = m$ disjoint (and affine) roots

$$\mathbf{x}^{(1)} = \begin{pmatrix}x_1^{(1)} & x_2^{(1)}\end{pmatrix}^T = \begin{pmatrix}0 & 0\end{pmatrix}^T \quad\text{and}\quad \begin{pmatrix}x_1^{(2,3)} & x_2^{(2,3)}\end{pmatrix}^T = \begin{pmatrix}2 & \pm\sqrt{2}\end{pmatrix}^T,$$

with multiplicity $\mu_1 = 2$ and $\mu_2 = \mu_3 = 1$, respectively. The confluent multivariate Vandermonde basis $\tilde{\mathbf{V}}(2)$ for $\mathrm{null}(\mathbf{M}(d^*)) = \mathrm{null}(\mathbf{M}(2))$ is given by

$$\tilde{\mathbf{V}}(2) = \left(\begin{array}{c|c|c}\tilde{\mathbf{V}}_1(2) & \tilde{\mathbf{v}}_2(2) & \tilde{\mathbf{v}}_3(2)\end{array}\right) = \left(\begin{array}{cc|c|c}\partial_{00}[\mathbf{v}_1(2)] & \partial_{10}[\mathbf{v}_1(2)] & \partial_{00}[\mathbf{v}_2(2)] & \partial_{00}[\mathbf{v}_3(2)]\end{array}\right) \in \mathbb{C}^{q(2)\times m},$$



(a) (b)

FIG. 4. (a) *Zero-level curves of* f_1 *(- - -) and* f_2 *(——) in* (31). *The roots are marked with "o".* (b) *Convergence of an optimization-based NLS-type algorithm to fit a CPD (- - -) and a BTD (——) to* $\mathcal{Y}(1,1)$ *in* (32) *as a function of the iteration step.*

where

$$\tilde{\mathbf{V}}_1(2) = \begin{pmatrix} \partial_{00}[\mathbf{v}_1(2)] & \partial_{10}[\mathbf{v}_1(2)] \end{pmatrix} = \left(\begin{array}{cc} 1 & 0 \\ \hline x_1^{(1)} & 1 \\ x_2^{(1)} & 0 \\ \hline x_1^{(1)2} & 2x_1^{(1)} \\ x_1^{(1)}x_2^{(1)} & x_2^{(1)} \\ x_2^{(1)2} & 0 \end{array}\right) = \left(\begin{array}{cc} 1 & 0 \\ \hline 0 & 1 \\ 0 & 0 \\ \hline 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{array}\right) \in \mathbb{C}^{q(2)\times\mu_1}.$$

The depth δ_1 of $\mathcal{D}[\mathbf{x}^{(1)}]$ equals $o\left(\partial_{10}[\mathbf{x}^{(1)}]\right) = 1$. Take $d^{(1)} = d^{(2)} = 1$ such that $d^{(1)} + d^{(2)} = 2 \geq 2 = d^*$. The tensor $\mathcal{Y}(1,1) \in \mathbb{C}^{q(1)\times q(1)\times m}$, constructed as shown in (18), admits the BTD

$$\begin{aligned}\mathcal{Y}(1,1) = {} & \mathcal{G}_1(1,1) \cdot_1 \tilde{\mathbf{V}}_1(1) \cdot_2 \tilde{\mathbf{V}}_1(1) \cdot_3 \mathbf{C}_1(2) \\ & + \mathbf{v}_2(1) \otimes \mathbf{v}_2(1) \otimes \mathbf{c}_{2,1}(2) + \mathbf{v}_3(1) \otimes \mathbf{v}_3(1) \otimes \mathbf{c}_{3,1}(2),\end{aligned} \tag{32}$$

in which

$$(\mathbf{G}_1(1,1))_{[2;1,3]} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{array}\right).$$

First we fit an $(m = 4)$-term CPD using the randomly initialized NLS algorithm in Tensorlab [40], until the relative change in objective function drops below 10^{-9} or a maximum of 500 iterations is reached. Figure 4(b) shows the convergence: it is slow. A collinearity criterion [31, eqn. (2.2)] identifies a group of $\mu_1 = 2$ diverging rank-1 terms and two linearly independent nondiverging rank-1 terms ($\mu_2 = \mu_3 = 1$).[6]

Next we fit a BTD with $m_0 = 3$ and the identified, correct multiplicities μ_k using NLS with the same stopping criterion. We use the CPD results to initialize the BTD fitting by means of the SGSD-based procedure in [31, p. 299],

$$\tilde{\mathbf{A}}_1(1) = \begin{pmatrix} 1 & 1 \\ 0.0138 & 0.0003 \\ 0 & 0 \end{pmatrix},$$

[6]When the algorithm terminates, the cosine between the vector representations of the two diverging rank-1 terms has become 0.9998 in absolute value.

which satisfies $\tilde{\mathbf{A}}_1(1) = \tilde{\mathbf{V}}_1(1)\mathbf{M}_1^{(1)}$ for some nonsingular matrix $\mathbf{M}_1^{(1)}$. From the last row of $\tilde{\mathbf{A}}_1(1)$ it follows that $x_2^{(1)} = 0$. The value of $x_1^{(1)}$ may be recovered from $f_2(x_1, x_2) = 0$.

Now we repeat the above experiment using Algorithm 1 with the algebraic BTD computation to find the roots and multiplicities for the system (31).

Setting $d^{(1)} = 1$ and $d^{(2)} = 2$ ensures that the prerequisites of Theorem 4.4 are met and, consequently, the matrices $\mathbf{A}_1(2) \in \mathbb{C}^{3\times 2}$, $\mathbf{A}_{2,3}(2) \in \mathbb{C}^{3\times 1}$, $\mathbf{B}_1(2) \in \mathbb{C}^{6\times 2}$, and $\mathbf{B}_{2,3}(2) \in \mathbb{C}^{6\times 1}$ can be readily computed algebraically via a block-diagonalization. The block-diagonalization already reveals the correct multiplicities $\mu_1 = 2$, $\mu_{2,3} = 1$. From $\mathbf{B}_1(2)$ the twofold root $\mathbf{x}^{(1)} = (0, 0)^T$ is retrieved using the generalized ESPRIT approach (step 9 of subsection 6.1; see also Appendix B.1). The simple roots $\mathbf{x}^{(2,3)}$ are retrieved from scaling the factor vectors of the rank-1 terms of the BTD as in [38, Algorithm 1]. Let $\mathbf{V}(d) = \begin{pmatrix}\mathbf{v}_1(d) & \mathbf{v}_2(d) & \mathbf{v}_3(d)\end{pmatrix} \in \mathbb{C}^{q(d)\times m_0}$ be the multivariate Vandermonde matrix of degree $d \geq 1$ associated to the true solutions of the polynomial system and $\hat{\mathbf{V}}(d)$ the estimated counterpart computed by Algorithm 1. Note that we do not add derivative columns corresponding to the roots with multiplicities here. The algebraic BTD-based procedure achieves a relative forward error[7]

$$\epsilon_{\hat{\mathbf{V}}(1)} = \frac{\|\hat{\mathbf{V}}(1) - \mathbf{V}(1)\|}{\|\mathbf{V}(1)\|}$$

of $\mathcal{O}(10^{-14})$ and a residual norm $\|\mathbf{M}(d_0)\mathbf{V}(d_0)\| = \mathcal{O}(10^{-13})$. Not only are these results significantly more accurate compared to the ones obtained with the NLS-based BTD computation that we executed before, but also the algebraic computation is carried out without the need for iterative procedures and initial guesses (obtained, e.g., by a preliminary CPD fit). This indicates that the algebraic BTD computation is more reliable compared to a BTD computation using optimization-based methods. Nevertheless, optimization-based methods can still be used in cases where some refinement of the algebraic results is needed, such as for noisy equations (see [38] for an illustration).

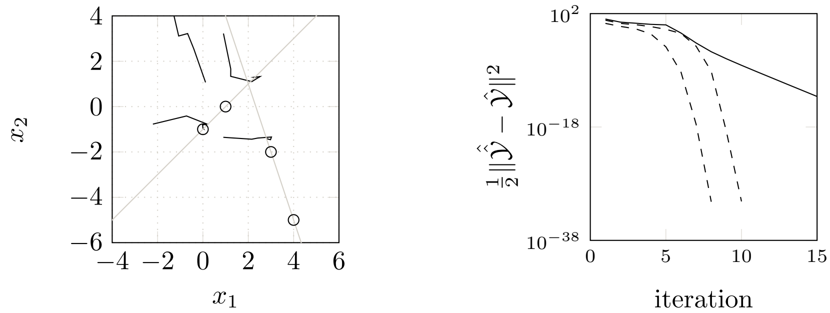
7.2. A recursive polynomial root-finding algorithm. As a numerical illustration of Algorithm 2, consider again the system of $s = 2$ polynomial equations in $n = 2$ variables [38, Example 3.2]:

$$\begin{cases} f_1(x_1, x_2) = -x_1^2 + 2x_1x_2 + x_2^2 + 5x_1 - 3x_2 - 4 = 0, \\ f_2(x_1, x_2) = x_1^2 + 2x_1x_2 + x_2^2 - 1 = 0, \end{cases} \tag{33}$$

with $d^{(1)} = d^{(2)} = 2$ and $d^* = 2+2-2 = 2$. The system has $m = 2\cdot 2 = 4$ simple roots $\begin{pmatrix}x_1 & x_2\end{pmatrix}^T = \begin{pmatrix}0 & -1\end{pmatrix}^T$, $\begin{pmatrix}1 & 0\end{pmatrix}^T$, $\begin{pmatrix}3 & -2\end{pmatrix}^T$, and $\begin{pmatrix}4 & -5\end{pmatrix}^T$ ("$\circ$" in Figure 5(a)).

From the numerical basis $\mathbf{K}(d) = \mathbf{K}(d^* + 1) = \mathbf{K}(2 + 1)$ for the null space of $\mathbf{M}(d)$ we construct the tensor $\mathcal{Y}(1, 2) \in \mathbb{C}^{3\times 6\times 4}$ which has multilinear rank-$(3, 4, 4)$; an MLSVD compression yields $\hat{\mathcal{Y}} \in \mathbb{C}^{3\times 4\times 4}$. We run Algorithm 2 using NLS and convergence criterion 10^{-6} for both the CPD in step 2 and the BTD in step 7. As the initial $\hat{\mathcal{Y}}$ has $R = m = 4$, the BTD (top level in Figure 3) directly uses the minimum sizes $R_1 = R_2 = m/2 = 2$ for the core tensors. To fit the BTD, we randomly initialized the first factor matrices $\hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2 \in \mathbb{C}^{3\times 2}$ for the optimization algorithm (alternatively, it is also possible to employ an algebraic BTD algorithm as in section 7.1) . Figure 5(a) illustrates how the $R_1 = 2$ columns of $\hat{\mathbf{A}}_1$ (first normalized so that $x_0 = 1$ and then

[7]Computed using the `cpderr` routine of Tensorlab [40].



(a) (b)

FIG. 5. (a) *Convergence of the projected terms in the BTD at the top level in Figure* 3 *for* (33) *from a random initialization to subspaces (——) spanned by two roots "○" each.* (b) *Convergence of an optimization-based NLS-type algorithm to fit the BTD (——) and two CPDs in the leaves (- - -) as a function of the iteration step.*

projected as points on the (x_1, x_2)-plane $\mathbb{C}^2$) converge from their random initialization to the lower-dimensional subspace (plotted as a gray line (——) in $\mathbb{C}^2$) spanned by the columns of

$$\left(\hat{\mathbf{V}}\right)_{1,2} = \begin{pmatrix} 1 & 1 \\ \hline x_1^{(1)} & x_1^{(2)} \\ x_2^{(1)} & x_2^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \hline 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Likewise, the $R_2 = 2$ columns of $\hat{\mathbf{A}}_2$ converge to the subspace (drawn as a gray line (——)) spanned by the two columns of

$$\left(\hat{\mathbf{V}}\right)_{3,4} = \begin{pmatrix} 1 & 1 \\ \hline x_1^{(3)} & x_1^{(4)} \\ x_2^{(3)} & x_2^{(4)} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \hline 3 & 4 \\ -2 & -5 \end{pmatrix}.$$

Note that one converged column of $\hat{\mathbf{A}}_2$ is kept outside Figure 5(a) for visibility. Next, each CPD in a recursive call of Algorithm 2 (leaf nodes in Figure 3) will converge within these subspaces to the sought after roots. Figure 5(b) shows the convergence. Because there are no multiple roots, there are no diverging rank-1 terms, and convergence is fast.

8. Conclusions. In [38] we have attempted to show that multilinear algebra is a convincing framework to formulate and solve 0-dimensional polynomial root-finding problems. This paper has taken the multilinear algebra framework to the next level. The third-order tensor BTD proposed in Theorem 4.2 is the most general decomposition in our framework. It incorporates multiple roots, reducing to the CPD if all roots happen to be simple; it coincides with the triangularization in NPA's central theorem; and it is a three-way generalization of the Jordan canonical form, intimately related to border rank. Furthermore, Theorem 4.4 establishes uniqueness properties for the BTD and enables its algebraic computation by means of a block-diagonalization. Future work might use our findings to formulate a three-way Jordan form for groups of many diverging rank-1 terms, which has so far only been done for relatively simple cases [31, 32]; general expressions are still elusive. We have illustrated how our BTD-based framework is able to retrieve the roots and their multiplicities from the null space of the Macaulay matrix. Moreover, we proposed a recursive method to detect nested structures in the nullspace. This essentially amounts to splitting a tensor that captures all roots into smaller tensors that capture subsets of roots, and iterating over such splittings. Future work might also investigate the use of constrained optimization techniques or prior knowledge to improve the accuracy with which the roots are found. It may also be interesting to see whether, e.g., clusters of roots of no interest can be discarded early in the polynomial root-finding procedures.

Appendix A. Proof of Theorem 3.1. We will need the following lemma.

LEMMA A.1. *Let* $M_1, \dots, M_K$ *be linear transformations on* $\mathbb{C}^m$ *and let*

$$\mathbb{C}^m = V_1 \dotplus \cdots \dotplus V_R, \qquad \dim V_r = \mu_r, \tag{34}$$

be a direct sum decomposition of $\mathbb{C}^m$ *into subspaces that are invariant for all* $M_1, \dots, M_K$*,*

$$M_k V_r \subseteq V_r, \quad r = 1, \dots, R, \quad k = 1, \dots, K.$$

Let also

$$\mathbf{M}_k = \operatorname{Blockdiag}(\mathbf{M}_k^{(1)}, \dots, \mathbf{M}_k^{(R)}), \quad \mathbf{M}_k^{(r)} \in \mathbb{C}^{\mu_r \times \mu_r}, \quad r = 1, \dots, R, \quad k = 1, \dots, K, \tag{35}$$

be the block-diagonal forms of $M_1, \dots, M_K$ *in a basis derived from decomposition* (34)*. Assume that*

1. *there exists a linear combination of* $M_1, \dots, M_K$ *with matrix representation* $\mathbf{M} = \operatorname{Blockdiag}(\mathbf{M}^{(1)}, \dots, \mathbf{M}^{(R)})$ *such that the spectra of any two blocks do not intersect;*
2. *none of the subspaces* V_r *can be further decomposed into a direct sum of subspaces that are invariant for all transformations* $M_1, \dots, M_K$*.*

Then any other decomposition of $\mathbb{C}^m$ *into a direct sum of* $\tilde{R} \geq R$ *subspaces that are invariant for all transformations* $M_1, \dots, M_K$*,*

$$\mathbb{C}^m = \tilde{V}_1 \dotplus \cdots \dotplus \tilde{V}_{\tilde{R}}, \qquad \dim \tilde{V}_r = \tilde{\mu}_r, \tag{36}$$

coincides with decomposition (34) *up to permutation of terms, that is,* $\tilde{V}_1 = V_{\pi(1)}, \dots, \tilde{V}_R = V_{\pi(R)}$ *for some permutation* π *of* $\{1, \dots, R\}$*. In particular, it necessarily holds that* $\tilde{R} = R$ *and that* $\tilde{\mu}_1 = \mu_{\pi(1)}, \dots, \tilde{\mu}_R = \mu_{\pi(R)}$*.*

Proof. Let subspace W be invariant for all transformations $M_1, \dots, M_K$. Then W is also invariant for the transformation M. Hence, by assumption 1 and [18, Theorem 2.1.5], $W = W_1 \dotplus \cdots \dotplus W_R$, where the subspaces $W_1 \subseteq V_1, \dots, W_R \subseteq V_R$ are invariant for M. Moreover, since W is invariant for all $M_1, \dots, M_K$ and (34) is a direct sum decomposition, it follows that the subspaces $W_1, \dots, W_R$ are also invariant for all transformations $M_1, \dots, M_K$. Applying this result to the subspaces $\tilde{V}_1, \dots, \tilde{V}_{\tilde{R}}$ in decomposition (36), we obtain that

$$\tilde{V}_1 = W_{11} \dotplus \cdots \dotplus W_{1R}, \dots, \tilde{V}_{\tilde{R}} = W_{\tilde{R}1} \dotplus \cdots \dotplus W_{\tilde{R}R}, \tag{37}$$

where the subspaces

$$W_{11}, W_{21}, \dots, W_{\tilde{R}1} \subseteq V_1, \dots, W_{1R}, W_{2R}, \dots, W_{\tilde{R}R} \subseteq V_R \tag{38}$$

are invariant for all transformations $M_1, \dots, M_K$. Now from (34), (36), (37), and (38) we obtain that

$$
\begin{aligned}
V_1 \dotplus \cdots \dotplus V_R = \mathbb{C}^I = \tilde{V}_1 \dotplus \cdots \dotplus \tilde{V}_{\tilde{R}} &= (W_{11} \dotplus \cdots \dotplus W_{1R}) \dotplus \cdots \dotplus (W_{\tilde{R}1} \dotplus \cdots \dotplus W_{\tilde{R}R}) \\
&= (W_{11} \dotplus W_{21} \dotplus \cdots \dotplus W_{\tilde{R}1}) \dotplus \cdots \dotplus (W_{1R} \dotplus W_{2R} \dotplus \cdots \dotplus W_{\tilde{R}R}) \subseteq V_1 \dotplus \cdots \dotplus V_R.
\end{aligned}
\tag{39}
$$

Hence $V_r = W_{1r} \dotplus W_{2r} \dotplus \cdots \dotplus W_{\tilde{R}r}$, $r = 1, \dots, R$. By assumption 2, this is possible only if one of the subspaces $W_{1r}, W_{2r}, \dots, W_{\tilde{R}r}$ coincides with V_r and the other subspaces are zero. This easily implies the statement of the lemma. ∎

Proof of Theorem 3.1. Since the matrix $\mathbf{B}$ has full column rank, it is sufficient to prove that for any decomposition of $\mathcal{T}$ into a sum of indecomposable tensors the blocks of the matrix in the second mode can be permuted so that their column spaces coincide with the column spaces of the blocks $\mathbf{B}_1, \dots, \mathbf{B}_R$. To prove the uniqueness of the column spaces $\operatorname{col}(\mathbf{B}_1), \dots, \operatorname{col}(\mathbf{B}_R)$ we will use Lemma A.1. In our derivation we assume without loss of generality that the matrix $\mathbf{B}$ is square, so $\mu_1 + \cdots + \mu_R = m$ and $\mathbf{B} \in \mathbb{C}^{m \times m}$.

Step 1*: Reduction to Lemma* A.1. For any $\mathbf{f} \in \mathbb{C}^{I_1}$ we have that

$$
\mathcal{T} \cdot_1 \mathbf{f}^T = \mathbf{B} \cdot \operatorname{Blockdiag}(\mathcal{G}_1 \cdot_1 (\mathbf{f}^T \mathbf{A}_1), \dots, \mathcal{G}_R \cdot_1 (\mathbf{f}^T \mathbf{A}_R)) \cdot \mathbf{C}^T,
\tag{40}
$$

where we identify the one-slice tensors $\mathcal{T} \cdot_1 \mathbf{f}^T \in \mathbb{C}^{1 \times m \times m}$ and $\mathcal{G}_1 \cdot_1 (\mathbf{f}^T \mathbf{A}_1) \in \mathbb{C}^{1 \times \mu_1 \times \mu_1}, \dots, \mathcal{G}_R \cdot_1 (\mathbf{f}^T \mathbf{A}_R) \in \mathbb{C}^{1 \times \mu_R \times \mu_R}$ with matrices. Since the first horizontal slice of $\mathcal{G}_r$ is the identity matrix and the other frontal slices are strictly upper-triangular, we have that

$$
\mathcal{G}_r \cdot_1 (\mathbf{f}^T \mathbf{A}_r) \ \text{ is the sum of } \mathbf{f}^T \mathbf{A}_r(:,1) \mathbf{I}_{\mu_r} \text{ and a strictly upper triangular matrix.}
\tag{41}
$$

Since, by (6), the first columns of the matrices $\mathbf{A}_1, \dots, \mathbf{A}_R$ are nonzero, it easily follows that for generic $\mathbf{f} \in \mathbb{C}^{I_1}$ all values $\mathbf{f}^T \mathbf{A}_1(:,1), \dots, \mathbf{f}^T \mathbf{A}_R(:,1)$ are nonzero. Hence, by (40) and (41), the $m \times m$ matrix $\mathcal{T} \cdot_1 \mathbf{f}^T$ is nonsingular for generic $\mathbf{f} \in \mathbb{C}^{I_1}$. Hence for $k = 1, \dots, I_1$ we have that

$$
\begin{aligned}
\mathcal{T}(k,:,:)(\mathcal{T} \cdot_1 \mathbf{f}^T)^{-1} = \mathbf{B} \cdot \operatorname{Blockdiag}(&(\mathcal{G}_1 \cdot_1 (\mathbf{A}_1(k,:)))(\mathcal{G}_1 \cdot_1 (\mathbf{f}^T \mathbf{A}_1))^{-1}, \dots, \\
&(\mathcal{G}_1 \cdot_1 (\mathbf{A}_R(k,:)))(\mathcal{G}_1 \cdot_1 (\mathbf{f}^T \mathbf{A}_R))^{-1}) \cdot \mathbf{B}^{-1}.
\end{aligned}
\tag{42}
$$

Thus, the matrices $\mathcal{T}(k,:,:)(\mathcal{T} \cdot_1 \mathbf{f}^T)^{-1}$ can be simultaneously reduced to block-diagonal form by a similarity transform. This means that the column spaces of the blocks $\mathbf{B}_1, \dots, \mathbf{B}_{m_0}$ are invariant for all matrices $\mathcal{T}(1,:,:)(\mathcal{T} \cdot_1 \mathbf{f}^T)^{-1}, \dots, \mathcal{T}(I_1, :,:)(\mathcal{T} \cdot_1 \mathbf{f}^T)^{-1}$ and that the whole space $\mathbb{C}^m$ can be decomposed into the direct sum of $\operatorname{col}(\mathbf{B}_1), \dots, \operatorname{col}(\mathbf{B}_R)$: $\mathbb{C}^m = \operatorname{col}(\mathbf{B}_1) \dotplus \cdots \dotplus \operatorname{col}(\mathbf{B}_R)$.

Step 2. By Step 1, any BTD $\mathcal{T} = \sum_{r=1}^{\tilde{R}} [\![\tilde{\mathcal{G}}_r; \tilde{\mathbf{A}}_r, \tilde{\mathbf{B}}_r, \tilde{\mathbf{C}}_r]\!]$ with nonsingular $\tilde{\mathbf{B}} \stackrel{\text{def}}{=} \begin{pmatrix}\tilde{\mathbf{B}}_1 & \dots & \tilde{\mathbf{B}}_{\tilde{R}}\end{pmatrix}$ and $\tilde{\mathbf{C}} \stackrel{\text{def}}{=} \begin{pmatrix}\tilde{\mathbf{C}}_1 & \dots & \tilde{\mathbf{C}}_{\tilde{R}}\end{pmatrix}$ generates a decomposition of $\mathbb{C}^m$ into a direct sum of $\operatorname{col}(\tilde{\mathbf{B}}_1), \dots, \operatorname{col}(\tilde{\mathbf{B}}_{\tilde{R}})$. To show that all such decompositions coincide up to permutation of the terms with the decomposition $\mathbb{C}^m = \operatorname{col}(\mathbf{B}_1) \dotplus \cdots \dotplus \operatorname{col}(\mathbf{B}_R)$, we show that the assumptions in Lemma A.1 hold for $K = I_1$, $V_r = \operatorname{col}(\mathbf{B}_r)$, and

$$
\mathbf{M}_k = \operatorname{Blockdiag}(\mathbf{M}_k^{(1)}, \dots, \mathbf{M}_k^{(R)}) \text{ with } \mathbf{M}_k^{(r)} = (\mathcal{G}_r \cdot_1 (\mathbf{A}_r(k,:)))(\mathcal{G}_r \cdot_1 (\mathbf{f}^T \mathbf{A}_r))^{-1}.
\tag{43}
$$

Assumption 1. Let $\mathbf{h} \in \mathbb{C}^K$ and $\mathbf{M} \overset{\text{def}}{=} h_1\mathbf{M}_1 + \cdots + h_K\mathbf{M}_K$. Then by (43), the rth diagonal block of $\mathbf{M}$ is the sum of $(\mathbf{h}^T\mathbf{A}_r(:,1))(\mathbf{f}^T\mathbf{A}_r(:,1))^{-1}\mathbf{I}_{\mu_r}$ and a strictly upper-triangular matrix. Hence, the diagonal blocks of $\mathbf{M}$ have one-point spectra $(\mathbf{h}^T\mathbf{A}_1(:,1))(\mathbf{f}^T\mathbf{A}_1(:,1))^{-1},\dots,(\mathbf{h}^T\mathbf{A}_R(:,1))(\mathbf{f}^T\mathbf{A}_R(:,1))^{-1}$. We show that there exists a vector $\mathbf{h}$ such that the values $(\mathbf{h}^T\mathbf{A}_1(:,1))(\mathbf{f}^T\mathbf{A}_1(:,1))^{-1},\dots,$ $(\mathbf{h}^T\mathbf{A}_R(:,1))(\mathbf{f}^T\mathbf{A}_R(:,1))^{-1}$ are distinct. Indeed, if $(\mathbf{h}^T\mathbf{A}_{r_1}(:,1))(\mathbf{f}^T\mathbf{A}_{r_1}(:,1))^{-1} = (\mathbf{h}^T\mathbf{A}_{r_2}(:,1))(\mathbf{f}^T\mathbf{A}_{r_2}(:,1))^{-1}$, then easy algebraic manipulations imply that

$$\mathbf{h}^T(\mathbf{f}^T\mathbf{A}_{r_2}(:,1))\mathbf{A}_{r_1}(:,1) = \mathbf{h}^T(\mathbf{f}^T\mathbf{A}_{r_1}(:,1))\mathbf{A}_{r_2}(:,1). \tag{44}$$

Thus, equation (44) holds only for vectors $\mathbf{h}$ that are orthogonal to the vector $(\mathbf{f}^T\mathbf{A}_{r_2}(:,1)\mathbf{A}_{r_1}(:,1) - (\mathbf{f}^T\mathbf{A}_{r_1}(:,1))\mathbf{A}_{r_2}(:,1))^*$, which, due to the generic choice of $\mathbf{f}$ in Step 1 and by assumption (6), is nonzero. Hence, the values $(\mathbf{h}^T\mathbf{A}_1(:,1))(\mathbf{f}^T\mathbf{A}_1(:,1))^{-1}$, $\dots,(\mathbf{h}^T\mathbf{A}_R(:,1))(\mathbf{f}^T\mathbf{A}_R(:,1))^{-1}$ are distinct for any vector $\mathbf{h}$ that is not orthogonal to any of the $\frac{R(R-1)}{2}$ vectors $(\mathbf{f}^T\mathbf{A}_{r_2}(:,1)\mathbf{A}_{r_1}(:,1) - (\mathbf{f}^T\mathbf{A}_{r_1}(:,1))\mathbf{A}_{r_2}(:,1))^*$, $1 \le r_1 < r_2 \le R$.

Assumption 2. Since the matrix $\mathbf{A}_r$ has full column rank, its row space is equal to $\mathbb{C}^{\mu_r}$. Hence the subspace spanned by the matrices $\mathbf{M}_1^{(r)},\dots,\mathbf{M}_{\mu_r}^{(r)}$ coincides with the subspace spanned by the nonsingular upper triangular matrix $\mathbf{S}_1 \overset{\text{def}}{=} (\mathcal{G}_r \cdot_1 (\mathbf{I}_{\mu_r}(1,:)))(\mathcal{G}_r \cdot_1 (\mathbf{f}^T\mathbf{A}_r))^{-1} = (\mathcal{G}_r(1,:,:))(\mathcal{G}_r \cdot_1 (\mathbf{f}^T\mathbf{A}_r))^{-1}$ and the $\mu_r - 1$ strictly upper-triangular matrices $\mathbf{S}_{l+1} \overset{\text{def}}{=} (\mathcal{G}_r \cdot_1 (\mathbf{I}_{\mu_r}(l+1,:)))(\mathcal{G}_r \cdot_1 (\mathbf{f}^T\mathbf{A}_r))^{-1} = (\mathcal{G}_r(l+1,:,:))(\mathcal{G}_r \cdot_1 (\mathbf{f}^T\mathbf{A}_r))^{-1}$, $l = 1,\dots,\mu_r - 1$. To prove that the subspace $\mathbb{C}^{\mu_r}$ cannot be decomposed into a direct sum of subspaces that are invariant for all matrices $\mathbf{M}_1^{(r)},\dots,\mathbf{M}_{\mu_r}^{(r)}$ we prove a stronger statement: the subspace $\mathbb{C}^{\mu_r}$ cannot be decomposed into a direct sum of subspaces that are invariant for all matrices $\mathbf{S}_2,\dots,\mathbf{S}_{\mu_R}$. Since $\mathbf{S}_2,\dots,\mathbf{S}_{\mu_R}$ are nilpotent matrices, it is sufficient to prove that the common null space of $\mathbf{S}_2,\dots,\mathbf{S}_{\mu_R}$ is trivial, i.e., is spanned by the vector $\mathbf{I}_{\mu_r}(:,1)$. Let $\mathbf{u}$ be a nonzero vector such that $\mathbf{S}_2\mathbf{u} = \cdots = \mathbf{S}_{\mu_R}\mathbf{u} = \mathbf{0}$. Since $\mathcal{G}_r(:,1,:) = \mathbf{I}_{\mu_r}$, it follows that the first rows of the matrices $\mathbf{S}_2,\dots,\mathbf{S}_{\mu_r}$ are proportional, respectively, to the second, third, $\dots$, μ_rth row of the matrix $(\mathcal{G}_r \cdot_1 (\mathbf{f}^T\mathbf{A}_r))^{-1}$. Hence, the identities $\mathbf{S}_2\mathbf{u} = \cdots = \mathbf{S}_{\mu_R}\mathbf{u} = \mathbf{0}$ imply that the last $\mu_r - 1$ entries of the vector $(\mathcal{G}_r \cdot_1 (\mathbf{f}^T\mathbf{A}_r))^{-1}\mathbf{u}$ are zero. Since the matrix $(\mathcal{G}_r \cdot_1 (\mathbf{f}^T\mathbf{A}_r))^{-1}$ is nonsingular and upper-triangular, it follows that the last $\mu_r - 1$ entries of the vector $\mathbf{u}$ are zero as well. ∎

Appendix B. Derivation of Theorem 4.2. In this section we derive the BTD structure in Theorem 4.2. Throughout this derivation we will make frequent use of the following Definition B.1 and Lemma B.3.

DEFINITION B.1 ([6, Definition 1]). *Let the linear transformation ϕ_j be defined by*

$$\phi_j\left(\partial_{j_1\dots j_n}[\mathbf{z}](f)\right) = \begin{cases} \partial_{j_1\dots j_{j-1},j_j-1,j_{j+1}\dots j_n}[\mathbf{z}](f), & j_j \neq 0,\\ 0\text{-}\textit{functional}, & j_j = 0. \end{cases}$$

Given a system of polynomial equations $\mathcal{F}$ and a μ_k-fold root $\mathbf{z}$, the dual subspaces $\mathcal{D}_t[\mathbf{z}](\mathcal{F})$ are the strictly enlarging sets $\mathcal{D}_0[\mathbf{z}](\mathcal{F}) = \operatorname{span}(\partial_{\mathbf{0}}[\mathbf{z}])$ and

$$\mathcal{D}_t[\mathbf{z}](\mathcal{F}) = \operatorname{span}\left(\left\{c = \sum_{|\mathbf{j}|\le t}\beta_{\mathbf{j}}\partial_{\mathbf{j}}[\mathbf{z}](f) \mid c(\mathcal{F}) = \{0\} \,\&\, \forall j : \phi_j(c) \in \mathcal{D}_{t-1}[\mathbf{z}](\mathcal{F})\right\}\right).$$

If $\mathcal{D}_{\delta+1}[\mathbf{z}] = \mathcal{D}_\delta[\mathbf{z}]$, then the vector space $\mathcal{D}_\delta[\mathbf{z}] = \mathcal{D}[\mathbf{z}]$ is called the dual space of *the system $\mathcal{F}$ at $\mathbf{z}$ and δ is called its depth. The dual space reveals the multiplicity structure of the root $\mathbf{z}$; its dimension equals the multiplicity μ_k.*

Example B.2. Consider again Example 3.5 with $f \in \mathcal{C}^2$, a fourfold root $\mathbf{z} \in \mathbb{C}^2$ with $\delta = 2$, and the differential functionals $c_{10} = \partial_{00}$, $c_{11} = \partial_{10}$, $c_{12} = \partial_{01}$, $c_{13} = (2\partial_{20} + \partial_{11})$. Obviously, $c_{10} \in \mathcal{D}_0[\mathbf{z}] \subset \mathcal{D}_2[\mathbf{z}]$. Since $\phi_1(c_{11}) = \phi_1(\partial_{10}) = \partial_{00}$, $\phi_2(c_{11}) = 0$ we have $c_{11} \in \mathcal{D}_1[z] \subset \mathcal{D}_2[\mathbf{z}]$, and likewise for c_{12}. For c_{13} we have that $\phi_1(c_{13}) = 2\partial_{10} + \partial_{01} \in \mathcal{D}_1[\mathbf{z}]$ and $\phi_2(c_{13}) = \partial_{10} \in \mathcal{D}_1[\mathbf{z}]$ so that $c_{13} \in \mathcal{D}_2[\mathbf{z}]$. Due to the nested structure of $\mathcal{D}$, it also holds that $\phi_i(\phi_j(c_{kl})) \in \mathcal{D}$, $i, j = 1, 2$. Indeed, we have, e.g., $\phi_1(\phi_1(c_{13})) = 2\partial_{00} \in \mathcal{D}_2[\mathbf{z}]$ as well as $\phi_2(\phi_1(c_{13})) = \partial_{00} \in \mathcal{D}_2[\mathbf{z}]$, $\phi_2(\phi_2(c_{13})) = 0 \in \mathcal{D}_2[\mathbf{z}]$.

We use the Leibniz formula (generalization of the product rule).

Lemma B.3. *Let $p, q \in \mathcal{C}^n$. Then for $\mathbf{k} \in \mathbb{N}^n$*

$$\partial_{\mathbf{k}}[p \cdot q] = \sum_{\mathbf{0} \leq \mathbf{j} \leq \mathbf{k}} \partial_{\mathbf{j}}[p] \cdot \partial_{\mathbf{k}-\mathbf{j}}[q].$$

With these prerequisites we are now ready to establish the BTD (19) in Theorem 4.2. We will do so in two steps: First we generalize the multiplicative shift structure for multivariate Vandermonde matrices, which was used in [38] for the case of only simple roots, to confluent multivariate Vandermonde matrices and roots with multiplicities greater than one (section B.1). This result is afterwards used to establish the BTD (19) starting from the null space of the Macaulay matrix (section B.2). Throughout the whole derivation, examples will illustrate main intermediate steps.

B.1. First step: Generalization of the multiplicative shift structure. We consider the confluent multivariate Vandermonde matrix

$$\tilde{\mathbf{V}}(d) = \begin{pmatrix} \tilde{\mathbf{V}}_1(d) & \dots & \tilde{\mathbf{V}}_{m_0}(d) \end{pmatrix} \in \mathbb{C}^{q(d) \times m} \tag{45a}$$

associated to a 0-dimensional polynomial system $\mathcal{F}$ with $m_0 \leq m$ distinct roots. Each block $\tilde{\mathbf{V}}_k(d)$, $k = 1 : m_0$, is of the form

$$\tilde{\mathbf{V}}_k(d) = \Big(\overbrace{c_{k0}[\mathbf{v}(d)]}^{\text{order } 0} \; \Big| \; \overbrace{c_{k1}[\mathbf{v}(d)] \quad \dots}^{\text{order } 1} \; \Big| \; \dots \; \Big| \; \overbrace{\dots \quad c_{k,\mu_k-1}[\mathbf{v}(d)]}^{\text{order } \delta_k} \Big) \in \mathbb{C}^{q(d) \times \mu_k} \tag{45b}$$

and contains the μ_k unique differential functional columns $c_{kl}[\mathbf{v}] \in \mathcal{D}[\mathbf{z}_k]$ which we assume w.l.o.g. to be ordered increasingly regarding the differentiation order of the differential functionals.

Lemma B.4. *Let $\tilde{\mathbf{V}}(d)$ be as in* (45) *with $d = d^{(1)} + d^{(2)} \geq d^*$, $d^{(1)} \geq 1$. Further, let $\overline{\mathbf{S}}^{(0)} \in \mathbb{C}^{q(d^{(2)}) \times q(d)}$ select the rows of $\tilde{\mathbf{V}}(d)$ associated to the monomials of degree 0 to $d^{(2)}$ and let $\overline{\mathbf{S}}^{(j)} \in \mathbb{C}^{q(d^{(2)}) \times q(d)}$ select the rows onto which these monomials are mapped after a multiplication with the $(j+1)$th monomial $\mathbf{x}^{\boldsymbol{\alpha}_j} = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, $\boldsymbol{\alpha}_j \in \mathbb{N}^n$ with $|\boldsymbol{\alpha}_j| \leq d^{(1)}$.*

Then the generalized multiplicative shift structure/ESPRIT-type relation

$$\overline{\mathbf{S}}^{(j)}(d^{(2)})\tilde{\mathbf{V}}_k(d) = \overline{\mathbf{S}}^{(0)}(d^{(2)})\tilde{\mathbf{V}}_k(d)\mathbf{J}_k^{(j)}, \quad 0 \leq j \leq q(d^{(1)}) - 1, \quad k = 1 : m_0, \tag{46}$$

holds, where $\mathbf{J}_k^{(j)} = \mathbf{x}^{\boldsymbol{\alpha}_j}\mathbf{I}_{\mu_k} + \mathbf{N}_k^{(j)} \in \mathbb{C}^{\mu_k \times \mu_k}$, with $\mathbf{N}_k^{(j)}$ strictly upper-triangular. For all $0 \leq i, j$ the upper-triangular matrices $\mathbf{J}_k^{(i)}, \mathbf{J}_k^{(j)}$ commute. Moreover, for the $(j+1)$th *monomial* $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ *the associated upper-triangular matrix* $\mathbf{J}_k^{(j)}$ *in* (46) *is given by*

$$\mathbf{J}_k^{(j)} = (\mathbf{J}_k^{(1)})^{\alpha_1} \cdots (\mathbf{J}_k^{(n)})^{\alpha_n} \tag{47}$$

so that all $\mathbf{J}_k^{(j)}$ *are defined by the* n *upper-triangular matrices* $\mathbf{J}_k^{(1)}, \ldots, \mathbf{J}_k^{(n)}$ *associated to the monomials* $x_1, \ldots, x_n$ *of degree one.*

Proof. Equation (46) holds trivially for $j = 0$ with $\mathbf{J}_k^{(0)} = \mathbf{I}_{\mu_k}$. We begin the derivation with shifts by the degree-one monomials x_j, $j = 1 : n$ (i.e., $\boldsymbol{\alpha}_j = \mathbf{e}_j$, $|\boldsymbol{\alpha}_j| = 1$). Only the first columns $c_{k0}[\mathbf{v}_k(d)] = \partial_{00}[\mathbf{v}_k(d)] = \mathbf{v}_k(d) = \mathbf{v}_k$ are genuine multivariate Vandermonde vectors for which the simple multiplicative shift invariance holds:

$$\overline{\mathbf{S}}^{(j)}(d^{(2)})\mathbf{v}_k = x_j \cdot \overline{\mathbf{S}}^{(0)}(d^{(2)})\mathbf{v}_k, \quad j = 1 : n, \tag{48a}$$

whereas by linearity of c_{kl} and the multiplication by $\overline{\mathbf{S}}^{(j)}(d^{(2)})$, we have for the remaining columns

$$\overline{\mathbf{S}}^{(j)}(d^{(2)})c_{kl}[\mathbf{v}_k] = \overline{\mathbf{S}}^{(0)}(d^{(2)})c_{kl}[x_j\mathbf{v}_k], \quad j = 1 : n. \tag{48b}$$

With the help of Definition B.1 and Lemma B.3 it holds for the application of $c_{kl} = \sum_{\mathbf{r}} \beta_{\mathbf{r}} \partial_{\mathbf{r}}$ to $x_j \mathbf{v}_k$ for $l = 1 : \mu_k - 1$ that

$$\begin{aligned} c_{kl}[x_j\mathbf{v}_k] &= \sum_{\mathbf{r}} \beta_{\mathbf{r}} \partial_{\mathbf{r}}[x_j\mathbf{v}_k] = \sum_{\mathbf{r}} \beta_{\mathbf{r}} \sum_{\mathbf{0}\le\mathbf{i}\le\mathbf{r}} \partial_{\mathbf{i}}[x_j]\partial_{\mathbf{r}-\mathbf{i}}[\mathbf{v}_k] = \sum_{\mathbf{r}} \beta_{\mathbf{r}} \sum_{|\mathbf{i}|=\mathbf{0}}^{1} \partial_{\mathbf{i}}[x_j]\partial_{\mathbf{r}-\mathbf{i}}[\mathbf{v}_k] \\ &= \sum_{\mathbf{r}} \beta_{\mathbf{r}} \left(x_j \partial_{\mathbf{r}}[\mathbf{v}_k] + \partial_{\mathbf{r}-\mathbf{e}_j}[\mathbf{v}_k] \right) = x_j c_{kl}[\mathbf{v}_k] + \phi_j(c_{kl})[\mathbf{v}_k]. \end{aligned}$$

Now let $1 \le t \le \delta_k$ be the differential order of c_{kl}. Since $c_{kl} \in \mathcal{D}_t[\mathbf{z}_k] \subseteq \mathcal{D}[\mathbf{z}_k]$, it holds by Definition B.1 that $\phi_j(c_{kl}) \in \mathcal{D}_{t-1} \subset \mathcal{D}[\mathbf{z}_k]$, which means $\phi_j(c_{kl})$ can be expressed as a linear combination of differential functionals from $\mathcal{D}[\mathbf{z}_k]$ of order less than t. In other words, $\phi_j(c_{kl})[\mathbf{v}_k]$ can be expressed as linear combinations of columns of $\tilde{\mathbf{V}}_k(:, 1 : l')$, $l' < l$, and whose differential order is strictly smaller than t. Hence,

$$\begin{aligned} c_{kl}[x_j\mathbf{v}_k] &= x_j c_{kl}[\mathbf{v}_k] + \sum_{l'<l} \gamma_{l'l} c_{kl'}[\mathbf{v}_k] \quad \text{for some} \quad \gamma_{l'l} \in \mathbb{C} \\ &= \tilde{\mathbf{V}}_k \mathbf{J}_k^{(j)}(:, l+1), \quad \mathbf{J}_k^{(j)}(:, l+1) = \begin{pmatrix} \gamma_{0l} \\ \vdots \\ \gamma_{l-1,l} \\ x_j \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad l = 1 : \mu_k - 1. \end{aligned} \tag{49}$$

Together with (48a), deploying the relations (49) in all μ_k columns in (48b) yields

$$\overline{\mathbf{S}}^{(j)}(d^{(2)})\tilde{\mathbf{V}}_k(d) = \overline{\mathbf{S}}^{(0)}(d^{(2)})\tilde{\mathbf{V}}_k(d)\mathbf{J}_k^{(j)},$$

with $\mathbf{J}_k^{(j)} = x_j \mathbf{I}_{\mu_k} + \mathbf{N}_k^{(j)} \in \mathbb{C}^{\mu_k \times \mu_k}$ with the γ's in the strictly upper-triangular part $\mathbf{N}_k^{(j)}$.

This relation can be extended towards shifts with higher degree monomials, i.e., $\mathbf{x}^{\boldsymbol{\alpha}_j} \cdot \mathbf{v}_k$ with $|\boldsymbol{\alpha}_j| > 1$. It similarly holds that $\overline{\mathbf{S}}^{(j)}(d^{(2)})\mathbf{v}_k = \mathbf{x}^{\boldsymbol{\alpha}_j} \cdot \overline{\mathbf{S}}^{(0)}(d^{(2)})\mathbf{v}_k$ for the first columns. The application of the functionals yields

$$c_{kl}[\mathbf{x}^{\boldsymbol{\alpha}_j}\mathbf{v}_k] = \sum_{\mathbf{r}} \beta_{\mathbf{r}} \sum_{\mathbf{0}\le\mathbf{i}\le\mathbf{r}} \partial_{\mathbf{i}}[\mathbf{x}^{\boldsymbol{\alpha}_j}]\partial_{\mathbf{r}-\mathbf{i}}[\mathbf{v}_k] = \sum_{\mathbf{r}} \beta_{\mathbf{r}} \sum_{\mathbf{i}=0}^{\min(t,|\boldsymbol{\alpha}_j|)} \partial_{\mathbf{i}}[\mathbf{x}^{\boldsymbol{\alpha}_j}]\phi^{\mathbf{i}}(\partial_{\mathbf{r}})[\mathbf{v}_k], \tag{50}$$

where we again used Definition B.1 and Lemma B.3 and introduced the notation $\phi^{\mathbf{i}} \overset{\text{def}}{=} \phi_1^{i_1}(\phi_2^{i_2}(\dots\phi_n^{i_n}))$ and $\phi_j^{i_j} \overset{\text{def}}{=} \phi_j(\phi_j(\dots\phi_j))$ (i_j-fold application of ϕ_j). Because of the nested structure of the dual space $\mathcal{D}[\mathbf{z}_k]$ it still holds that $\phi^{\mathbf{i}}(c_{kl}) \in \mathcal{D}_{\max(0,t-|\mathbf{i}|)}[\mathbf{z}_k] \subset \mathcal{D}[\mathbf{z}_k]$. Hence, (50) can be written as

$$\begin{aligned} c_{kl}[\mathbf{x}^{\boldsymbol{\alpha}_j}\mathbf{v}_k] &= \mathbf{x}^{\boldsymbol{\alpha}_j} c_{kl}[\mathbf{v}_k] + \sum_{l'<l} \gamma_{l'l}(\mathbf{x}) c_{kl'}[\mathbf{v}_k] \quad \text{for some} \quad \gamma_{l'l}(\mathbf{x}) \in \mathcal{C}^{|\boldsymbol{\alpha}_j|-1} \\ &= \tilde{\mathbf{V}}_k \begin{pmatrix} \gamma_{0l}(\mathbf{x}) \\ \vdots \\ \gamma_{l-1,l}(\mathbf{x}) \\ \mathbf{x}^{\boldsymbol{\alpha}_j} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad l = 1 : \mu_k - 1, \end{aligned}$$

so that (46) also holds for all $j \le q(d^{(1)}) - 1$, where $j > n$ indicates a multiplicative shift with the $(j+1)$th monomial in the chosen monomial ordering. The associated upper-triangular matrices $\mathbf{J}_k^{(j)}$ will have strict upper-triangular parts that may depend on the values of $x_1^{(k)}, \dots, x_n^{(k)}$.

We now establish (47) for the sake of presentation for the shift x_j^2, i.e., $\boldsymbol{\alpha} = 2\mathbf{e}_j$. We proceed through the steps in (50) in a slightly different way (but again making use of Definition B.1 and Lemma B.3):

$$\begin{aligned} c_{kl}[x_j^2\mathbf{v}_k] &= c_{kl}[x_j(x_j\mathbf{v}_k)] = \sum_{\mathbf{r}} \beta_{\mathbf{r}} \sum_{\mathbf{0}\le\mathbf{i}\le\mathbf{r}} \partial_{\mathbf{i}}[x_j]\partial_{\mathbf{r}-\mathbf{i}}[x_j\mathbf{v}_k] \\ &= \sum_{\mathbf{r}} \beta_{\mathbf{r}} \left(x_j\partial_{\mathbf{r}}[x_j\mathbf{v}_k] + \partial_{\mathbf{r}-\mathbf{e}_j}[x_j\mathbf{v}_k]\right) = x_j c_{kl}[x_j\mathbf{v}_k] + \phi_j c_{kl}[x_j\mathbf{v}_k] \\ &= x_j \left(x_j c_{kl}[\mathbf{v}_k] + \sum_{l'<l} \gamma_{l'l} c_{kl'}[\mathbf{v}_k]\right) + \phi_j \left(x_j c_{kl}[\mathbf{v}_k] + \sum_{l'<l} \gamma_{l'l} c_{kl'}[\mathbf{v}_k]\right) \\ &= x_j \tilde{\mathbf{V}}_k \mathbf{J}_k^{(j)}(:, l+1) + x_j \tilde{\mathbf{V}}_k(:, 1:l)\mathbf{J}_k^{(j)}(1:l, l+1) + \sum_{l'<l} \gamma_{l'l}\phi_j\left(c_{kl'}\right)[\mathbf{v}_k], \end{aligned} \tag{51}$$

where we used (49). For the rightmost term in (51), recall that $c_{kl'} \in \mathcal{D}_{t-1}[\mathbf{z}_k]$ if $1 \le t \le \delta_k$ is the differentiation order of c_{kl}. Thus, by the nested structure of $\mathcal{D}[\mathbf{z}_k]$, $\phi_j(c_{kl'}) \in \mathcal{D}_{\max(0,t-2)}[\mathbf{z}_k]$ so that $\phi_j\left(c_{kl'}\right) = \sum_{l''<l'} \gamma_{l''l'} c_{kl''}$. Consequently,

$$\sum_{l'<l} \gamma_{l'l}\phi_j\left(c_{kl'}\right)[\mathbf{v}_k] = \sum_{l'<l} \gamma_{l'l} \sum_{l''<l'} \gamma_{l''l'} c_{kl''}[\mathbf{v}_k] = \sum_{l'<l} \gamma_{l'l} \tilde{\mathbf{V}}_k \mathbf{J}_k^{(j)}(:, l'+1)$$

and, by recalling that $\mathbf{J}_k^{(j)}(l+1, l+1) = \gamma_{ll} = x_j$ for $l = 0 : \mu_k - 1$, we can write (51) as

$$c_{kl}[x_j^2\mathbf{v}_k] = \tilde{\mathbf{V}}_k\left(\gamma_{ll}\mathbf{J}_k^{(j)}(:,l+1)+\gamma_{ll}\mathbf{J}_k^{(j)}(1:l,l+1)+\mathbf{J}_k^{(j)}(:,1:l'+1)\mathbf{J}_k^{(j)}(:,l+1)\right) \tag{52}$$

$$= \tilde{\mathbf{V}}_k\mathbf{J}_k^{(j)}\mathbf{J}_k^{(j)}(:,l+1). \tag{53}$$

We identify $\mathbf{J}_k^{(j)}\mathbf{J}_k^{(j)}(:,l+1)$ as the $(l+1)$th column of $\mathbf{J}_k^{(j)2}$, and using (52) for $l=0:\mu_k-1$ yields (47) for quadratic shifting monomials x_j^2. The above reasoning can be extended first towards higher degree pure monomials $x_j^{\alpha_j}$, $\alpha_j>2$, and finally to general monomials $x_1^{\alpha_1}\cdots x_n^{\alpha_n}$, which establishes (47). ∎

Example B.5. Consider again Example 3.5 with the differential functionals $c_{10}=\partial_{00}$, $c_{11}=\partial_{10}$, $c_{12}=\partial_{01}$, and $c_{13}=(2\partial_{20}+\partial_{11})$ and, thus,

$$\tilde{\mathbf{V}}_1(d)=\begin{pmatrix}\mathbf{v}_1(d) & c_{11}[\mathbf{v}_1(d)] & c_{12}[\mathbf{v}_1(d)] & c_{13}[\mathbf{v}_1(d)]\end{pmatrix}\in\mathbb{C}^{q(d)\times 4}.$$

We omit the degree indications $(d, d^{(2)})$ for the rest of the example for better readability. For $j=1,2$ it clearly holds that $\overline{\mathbf{S}}^{(j)}\mathbf{v}_1=x_j\cdot\overline{\mathbf{S}}^{(0)}\mathbf{v}_1$. For the second differential functional $c_{11}=\partial_{10}$, i.e., the second column of $\tilde{\mathbf{V}}_1$, we have

$$\begin{aligned}c_{11}[x_j\mathbf{v}_1]=\partial_{10}[x_j\mathbf{v}_1]&=x_j\partial_{10}[\mathbf{v}_1]+\phi_j(\partial_{10}[\mathbf{v}_1])\\&=x_j\partial_{10}[\mathbf{v}_1]+\begin{cases}\partial_{00}[\mathbf{v}_1]=\mathbf{v}_1 & :\ j=1,\\ \mathbf{0} & :\ j=2.\end{cases}\end{aligned}$$

Thus,

$$\overline{\mathbf{S}}^{(1)}c_{11}[\mathbf{v}_1]=\overline{\mathbf{S}}^{(1)}\tilde{\mathbf{V}}_1(:,2)=\overline{\mathbf{S}}^{(0)}\tilde{\mathbf{V}}_1\begin{pmatrix}1\\x_1\\0\\0\end{pmatrix}\quad\text{and}\quad\overline{\mathbf{S}}^{(2)}\tilde{\mathbf{V}}_1(:,2)=\overline{\mathbf{S}}^{(0)}\tilde{\mathbf{V}}_2\begin{pmatrix}0\\x_2\\0\\0\end{pmatrix}.$$

Likewise, we find

$$\overline{\mathbf{S}}^{(1)}c_{12}[\mathbf{v}_1]=\overline{\mathbf{S}}^{(1)}\tilde{\mathbf{V}}_1(:,3)=\overline{\mathbf{S}}^{(0)}\tilde{\mathbf{V}}_1\begin{pmatrix}0\\0\\x_1\\0\end{pmatrix}\quad\text{and}\quad\overline{\mathbf{S}}^{(2)}\tilde{\mathbf{V}}_1(:,3)=\overline{\mathbf{S}}^{(0)}\tilde{\mathbf{V}}_1\begin{pmatrix}1\\0\\x_2\\0\end{pmatrix}.$$

For the fourth functional c_{13} we have

$$\begin{aligned}c_{13}[x_j\mathbf{v}_1]=(2\partial_{20}+\partial_{11})[x_j\mathbf{v}_1]&=x_j(2\partial_{20}+\partial_{11})[\mathbf{v}_1]+\phi_j(2\partial_{20}+\partial_{11})[\mathbf{v}_1]\\&=x_jc_{13}[\mathbf{v}_1]+\begin{cases}(2\partial_{10}+\partial_{01})[\mathbf{v}_1]=(2c_{11}+c_{12})[\mathbf{v}_1] & :\ j=1,\\ \partial_{10}[\mathbf{v}_1]=c_{11}[\mathbf{v}_1] & :\ j=2.\end{cases}\end{aligned}$$

Consequently,

$$\overline{\mathbf{S}}^{(1)}c_{13}[\mathbf{v}_1]=\overline{\mathbf{S}}^{(1)}\tilde{\mathbf{V}}_1(:,4)=\overline{\mathbf{S}}^{(0)}\tilde{\mathbf{V}}_1\begin{pmatrix}0\\2\\1\\x_1\end{pmatrix}\quad\text{and}\quad\overline{\mathbf{S}}^{(2)}\tilde{\mathbf{V}}_1(:,4)=\overline{\mathbf{S}}^{(0)}\tilde{\mathbf{V}}_1\begin{pmatrix}0\\1\\0\\x_2\end{pmatrix}.$$

Collecting all these relations yields (46) with the upper-triangular matrices

$$\mathbf{J}_1^{(1)}=\begin{pmatrix}x_1 & 1 & & \\ & x_1 & & 2\\ & & x_1 & 1\\ & & & x_1\end{pmatrix}=x_1\mathbf{I}_4+\begin{pmatrix} & 1 & & \\ & & & 2\\ & & & 1\\ & & & \end{pmatrix},\quad \mathbf{J}_1^{(2)}=\begin{pmatrix}x_2 & 0 & 1 & \\ & x_2 & & 1\\ & & x_2 & 0\\ & & & x_2\end{pmatrix}.$$

Finally let's consider as one shift with a higher degree monomial the shift with the $(j=3)$rd monomial x_1^2. It clearly holds that $\overline{\mathbf{S}}^{(3)}\mathbf{v}_1 = x_1^2 \cdot \overline{\mathbf{S}}^{(0)}\mathbf{v}_1$. For the remaining columns we get

$$\begin{aligned}
c_{11}[x_1^2\mathbf{v}_1] &= \partial_{10}[x_1^2\mathbf{v}_1] = x_1^2\partial_{10}[\mathbf{v}_1] + 2x_1\mathbf{v}_1 = x_1^2 c_{11}[\mathbf{v}_1] + 2x_1\mathbf{v}_1,\\
c_{12}[x_1^2\mathbf{v}_1] &= \partial_{01}[x_1^2\mathbf{v}_1] = x_1^2\partial_{01}[\mathbf{v}_1] = x_1^2 c_{12}[\mathbf{v}_1],\\
c_{13}[x_1^2\mathbf{v}_1] &= (2\partial_{20} + \partial_{11})[x_1^2\mathbf{v}_1]\\
&= 2(x_1^2\partial_{20}[\mathbf{v}_1] + 2x_1\partial_{10}[\mathbf{v}_1] + \mathbf{v}_1) + x_1^2\partial_{11}[\mathbf{v}_1] + 2x_1\partial_{01}[\mathbf{v}_1]\\
&= 2\mathbf{v}_1 + 4x_1c_{11}[\mathbf{v}_1] + 2x_1c_{12}[\mathbf{v}_1] + x_1^2c_{13}[\mathbf{v}_1].
\end{aligned}$$

Hence,

$$\mathbf{J}_1^{(3)} = \begin{pmatrix} x_1^2 & 2x_1 & & 2 \\ & x_1^2 & & 4x_1 \\ & & x_1^2 & 2x_1 \\ & & & x_1^2 \end{pmatrix} = x_1^2\mathbf{I}_4 + \begin{pmatrix} & 2x_1 & & 2 \\ & & & 4x_1 \\ & & & 2x_1 \\ & & & \end{pmatrix} = \mathbf{J}_1^{(1)2}.$$

B.2. Step 2. Establishing the BTD structure.

Proof of Theorem 4.2. Recall that for $d \geq d^*$ the numerical basis $\mathbf{K}(d)$ of the Macaulay null space and the confluent multivariate Vandermonde matrix $\tilde{\mathbf{V}}(d)$ are linked by

$$\mathbf{K}(d) = \tilde{\mathbf{V}}(d)\mathbf{C}^T = \begin{pmatrix} \tilde{\mathbf{V}}_1(d) & \dots & \tilde{\mathbf{V}}_{m_0}(d) \end{pmatrix} \begin{pmatrix} \mathbf{C}_1^T \\ \vdots \\ \mathbf{C}_{m_0}^T \end{pmatrix}$$

and consider the matrix representation (18) of the third-order tensor $\mathcal{Y}(d^{(1)}, d^{(2)})$:

$$\mathbf{Y}_{[1,2;3]}(d^{(1)}, d^{(2)}) = \begin{pmatrix} \overline{\mathbf{S}}^{(0)}(d^{(2)}) \cdot \mathbf{K}(d) \\ \overline{\mathbf{S}}^{(1)}(d^{(2)}) \cdot \mathbf{K}(d) \\ \vdots \\ \overline{\mathbf{S}}^{(q(d^{(1)})-1)}(d^{(2)}) \cdot \mathbf{K}(d) \end{pmatrix} = \sum_{k=1}^{m_0} \begin{pmatrix} \overline{\mathbf{S}}^{(0)}(d^{(2)}) \cdot \tilde{\mathbf{V}}_k(d) \\ \overline{\mathbf{S}}^{(1)}(d^{(2)}) \cdot \tilde{\mathbf{V}}_k(d) \\ \vdots \\ \overline{\mathbf{S}}^{(q(d^{(1)})-1)}(d^{(2)}) \cdot \tilde{\mathbf{V}}_k(d) \end{pmatrix} \mathbf{C}_k^T$$

$$= \sum_{k=1}^{m_0} \begin{pmatrix} \tilde{\mathbf{V}}_k(d^{(2)}) \\ \tilde{\mathbf{V}}_k(d^{(2)})\mathbf{J}_k^{(1)} \\ \vdots \\ \tilde{\mathbf{V}}_k(d^{(2)})\mathbf{J}_k^{(q(d^{(1)})-1)} \end{pmatrix} \mathbf{C}_k^T = \sum_{k=1}^{m_0} (\mathbf{I}_{q(d^{(1)})} \otimes \tilde{\mathbf{V}}_k(d^{(2)})) \begin{pmatrix} \mathbf{I}_{\mu_k} \\ \mathbf{J}_k^{(1)} \\ \vdots \\ \mathbf{J}_k^{(q(d^{(1)})-1)} \end{pmatrix} \mathbf{C}_k^T,$$

with the upper-triangular matrices $\mathbf{J}_k^{(j)}$, $j = 1 : q(d^{(1)}) - 1$, $k = 1 : m_0$, from Lemma B.4 associated to the $q(d^{(1)})$ shifting monomials of degree 0 to $d^{(1)}$ which are assumed to be ordered consistently in the chosen monomial order. Consider the kth term in the above sum, which is associated to the kth root $\mathbf{z}_k$ with multiplicity $\mu_k \geq 1$ and depth $0 \leq \delta_k \leq \mu_k - 1$. For the strictly upper-triangular parts of $\mathbf{J}_k^{(i)} = x_i\mathbf{I}_{\mu_k} + \mathbf{N}_k^{(i)}$, $i = 1 : n$, we have the nilpotency properties

$$(\mathbf{N}_k^{(1)})^{\alpha_1} \cdots (\mathbf{N}_k^{(n)})^{\alpha_n} = \mathbf{0}_{\mu_k} \quad \forall \{\alpha_j\}_{j=1}^n \quad \text{with} \quad \sum_j \alpha_j > \delta_k, \tag{54a}$$

which include the individual properties

$$(\mathbf{N}_k^{(j)})^{\alpha} = \mathbf{0}_{\mu_k}, \quad \alpha > \delta_k, \tag{54b}$$

as a special case. Furthermore,

$$(\mathbf{N}_k^{(1)})^{\alpha_1}\cdots(\mathbf{N}_k^{(n)})^{\alpha_n} = \eta\mathbf{e}_1\mathbf{e}_{\mu_k}^T, \quad \eta\in\mathbb{C}, \quad \text{if} \quad \sum_j \alpha_j = \delta_k. \tag{54c}$$

Trivially, $(\mathbf{N}_k^{(j)})^{\mu_k} = \mathbf{0}_{\mu_k}$ since $\mu_k \geq \delta_k + 1$.

Let $\mathbf{J}_k^{(j)}$ be associated to the monomial $\mathbf{x}^{\boldsymbol{\alpha}_j}$ and express it in terms of the upper-triangular matrices $\mathbf{N}_k^{(i)}$, $i = 1 : n$, by using the multibinomial formula

$$\mathbf{J}_k^{(j)} = \sum_{\mathbf{h}\leq\boldsymbol{\alpha}_j} \mathbf{x}^{\mathbf{h}} \prod_{i=1}^{n} \binom{\alpha_i}{h_i} (\mathbf{N}_k^{(i)})^{\alpha_i - h_i} = \mathbf{x}^{\boldsymbol{\alpha}_j}\mathbf{I}_{\mu_k} + \sum_{\substack{\mathbf{h}\leq\boldsymbol{\alpha}_j \\ \mathbf{h}\neq\boldsymbol{\alpha}_j}} \mathbf{x}^{\mathbf{h}} \prod_{i=1}^{n} \binom{\alpha_i}{h_i} (\mathbf{N}_k^{(i)})^{\alpha_i - h_i}. \tag{55}$$

Using the above nilpotency properties (54) and also the property that all $\mathbf{N}_k^{(i)}$ commute indicates that at most $q(\delta_k) - 1$ different products of strictly upper-triangular matrices appear in (55) (all products of powers of the $\mathbf{N}_k^{(i)}$ with $\sum_i(\alpha_i - h_i) > \delta_k$) cancel out). The factors in front of the upper-triangular matrices can be collected to match the functional evaluations $c_{kl}[\mathbf{x}^{\boldsymbol{\alpha}_j}]$, $l = 1 : \mu_k - 1$, so that every $\mathbf{J}_k^{(j)}$ can be written as

$$\mathbf{J}_k^{(j)} = \mathbf{x}^{\boldsymbol{\alpha}_j}\mathbf{I}_{\mu_k} + c_{k1}[\mathbf{x}^{\boldsymbol{\alpha}_j}]\hat{\mathbf{N}}_k^{(1)} + \cdots + c_{k,\mu_k-1}[\mathbf{x}^{\boldsymbol{\alpha}_j}]\hat{\mathbf{N}}_k^{(\mu_k-1)}. \tag{56}$$

Here, the $\hat{\mathbf{N}}_k^{(i)}$ are linear combinations of those $\mathbf{N}_k^{(h)} = \mathbf{J}_k^{(h)} - \mathbf{x}^{\boldsymbol{\alpha}_h}\mathbf{I}_{\mu_k}$ that are associated to shifting monomials $\mathbf{x}^{\boldsymbol{\alpha}_h}$ with degrees equal to the differential order of c_{ki}, that is,

$$\hat{\mathbf{N}}_k^{(i)} = \sum_{\{h\,:\,|\boldsymbol{\alpha}_h| = o(c_{ki})\}} \omega_h \mathbf{N}_k^{(h)}, \qquad \omega_h\in\mathbb{C}.$$

Consequently, since the selection matrices $\overline{\mathbf{S}}^{(i)}$ are applied in the chosen monomial order, we find

$$\begin{aligned}\begin{pmatrix}\mathbf{I}_{\mu_k}\\ \mathbf{J}_k^{(1)}\\ \vdots\\ \mathbf{J}_k^{(q(d^{(1)})-1)}\end{pmatrix} &= \mathbf{v}_k(d^{(1)})\otimes\mathbf{I}_{\mu_k} + c_{k1}[\mathbf{v}_k(d^{(1)})]\otimes\hat{\mathbf{N}}_k^{(1)} + \cdots + c_{k,\mu_k-1}[\mathbf{v}_k(d^{(1)})]\otimes\hat{\mathbf{N}}_k^{(\mu_k-1)}\\ &= \left(\tilde{\mathbf{V}}_k(d^{(1)})\otimes\mathbf{I}_{\mu_k}\right)\mathbf{G}_{k[1,2;3]}, \quad \mathbf{G}_{k[1,2;3]} \stackrel{\text{def}}{=} \begin{pmatrix}\mathbf{I}_{\mu_k}\\ \hat{\mathbf{N}}_k^{(1)}\\ \vdots\\ \hat{\mathbf{N}}_k^{(\mu_k-1)}\end{pmatrix}.\end{aligned}$$

Hence, one term of $\mathbf{Y}_{[1,2;3]}(d^{(1)}, d^{(2)})$ can be written as

$$(\mathbf{I}_{q(d^{(1)})}\otimes\tilde{\mathbf{V}}_k(d^{(2)}))\begin{pmatrix}\mathbf{I}_{\mu_k}\\ \mathbf{J}_k^{(1)}\\ \vdots\\ \mathbf{J}_k^{(q(d^{(1)})-1)}\end{pmatrix}\mathbf{C}_k^T = (\tilde{\mathbf{V}}_k(d^{(1)})\otimes\tilde{\mathbf{V}}_k(d^{(2)}))\mathbf{G}_{k[1,2;3]}\mathbf{C}_k^T = \mathbf{Y}_{k[1,2;3]},$$

which is a matrix unfolding of one term of a BTD $\mathcal{Y}_k(d^{(1)}, d^{(2)}) = [\![\mathcal{G}_k; \tilde{\mathbf{V}}_k(d^{(1)}), \tilde{\mathbf{V}}_k(d^{(2)}), \mathbf{C}_k(d)]\!]$ of a third-order tensor $\mathcal{Y}_k(d^{(1)}, d^{(2)}) \in \mathbb{C}^{q(d^{(1)}\times q(d^{(2)}\times\mu_k}$. Since this holds for all $k = 1 : m_0$, we established the BTD (19). The equality $\mathcal{G}_k(l_1+1,:,:) = \mathcal{G}_k(:,l_1+1,:)$ follows by symmetry. ∎

We illustrate this BTD construction with an example.

Example B.6. Continuing Example B.5 with $d^{(1)} = d^{(2)} = 2$,

$$\tilde{\mathbf{V}}_1(2) = \begin{pmatrix} c_{10}[\mathbf{v}_1] & c_{11}[\mathbf{v}_1] & c_{12}[\mathbf{v}_1] & c_{13}[\mathbf{v}_1] \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \hline x_1 & 1 & 0 & 0 \\ x_2 & 0 & 1 & 0 \\ \hline x_1^2 & 2x_1 & 0 & 2 \\ x_1x_2 & x_2 & x_1 & 1 \\ x_2^2 & 0 & 2x_2 & 0 \end{pmatrix},$$

with differential functions given in Example B.2, and upper-triangular matrices $\mathbf{J}_k^{(j)}$ from the previous subsection. Now note that

$$\begin{pmatrix} \mathbf{I}_{\mu_1} \\ \mathbf{J}_1^{(1)} \\ \mathbf{J}_1^{(2)} \\ \mathbf{J}_1^{(3)} \\ \vdots \\ \mathbf{J}_1^{(5)} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{\mu_1} \\ x_1 \cdot \mathbf{I}_{\mu_1} + \mathbf{N}_1^{(1)} \\ x_2 \cdot \mathbf{I}_{\mu_1} + \mathbf{N}_1^{(2)} \\ x_1^2 \cdot \mathbf{I}_{\mu_1} + 2x_1 \cdot \mathbf{N}_1^{(1)} + \mathbf{N}_1^{(1)2} \\ \vdots \\ x_2^2 \cdot \mathbf{I}_{\mu_1} + 2x_2 \cdot \mathbf{N}_1^{(2)} + \mathbf{N}_1^{(2)2} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{\mu_1} \\ x_1 \cdot \mathbf{I}_{\mu_k} + c_{11}[x_1]\hat{\mathbf{N}}_1^{(1)} + c_{12}[x_1]\hat{\mathbf{N}}_1^{(2)} + c_{13}[x_1]\hat{\mathbf{N}}_1^{(3)} \\ x_2 \cdot \mathbf{I}_{\mu_1} + c_{11}[x_2]\hat{\mathbf{N}}_1^{(1)} + c_{12}[x_2]\hat{\mathbf{N}}_1^{(2)} + c_{13}[x_2]\hat{\mathbf{N}}_1^{(3)} \\ x_1^2 \cdot \mathbf{I}_{\mu_1} + c_{11}[x_1^2]\hat{\mathbf{N}}_1^{(1)} + c_{12}[x_1^2]\hat{\mathbf{N}}_1^{(2)} + c_{13}[x_1^2]\hat{\mathbf{N}}_1^{(3)} \\ \vdots \\ x_2^2 \cdot \mathbf{I}_{\mu_1} + c_{11}[x_2^2]\hat{\mathbf{N}}_1^{(1)} + c_{12}[x_2^2]\hat{\mathbf{N}}_1^{(2)} + c_{13}[x_2^2]\hat{\mathbf{N}}_1^{(3)} \end{pmatrix}$$
$$= c_{10}[\mathbf{v}_1] \otimes \mathbf{I}_4 + c_{11}[\mathbf{v}_1] \otimes \hat{\mathbf{N}}_1^{(1)} + c_{12}[\mathbf{v}_1] \otimes \hat{\mathbf{N}}_1^{(2)} + c_{13}[\mathbf{v}_1] \otimes \hat{\mathbf{N}}_1^{(3)},$$

which corresponds to (56) with

$$\hat{\mathbf{N}}_1^{(1)} \overset{\text{def}}{=} \mathbf{N}_1^{(1)}, \quad \hat{\mathbf{N}}_1^{(2)} \overset{\text{def}}{=} \mathbf{N}_1^{(2)}, \quad \hat{\mathbf{N}}_1^{(3)} \overset{\text{def}}{=} \tfrac{1}{2}\mathbf{N}_1^{(1)2} = \mathbf{N}_1^{(1)}\mathbf{N}_1^{(2)}.$$

Consequently,

$$(\mathbf{I}_{q(d^{(1)})} \otimes \tilde{\mathbf{V}}_1(d^{(2)})) \begin{pmatrix} \mathbf{I}_{\mu_1} \\ \mathbf{J}_1^{(1)} \\ \vdots \\ \mathbf{J}_1^{(q(d^{(1)})-1)} \end{pmatrix}$$
$$= (\mathbf{I}_{q(d^{(1)})} \otimes \tilde{\mathbf{V}}_1(d^{(2)})) \left(c_{10}[\mathbf{v}_1] \otimes \mathbf{I}_4 + c_{11}[\mathbf{v}_1] \otimes \begin{pmatrix} 0 & 1 & & \\ & 0 & & 2 \\ & & 0 & 1 \\ & & & 0 \end{pmatrix} \right.$$
$$\left. + c_{12}[\mathbf{v}_1] \otimes \begin{pmatrix} 0 & 0 & 1 & \\ & 0 & & 1 \\ & & 0 & \\ & & & 0 \end{pmatrix} + c_{13}[\mathbf{v}_1] \otimes \begin{pmatrix} 0 & 0 & 0 & 1 \\ & 0 & & 0 \\ & & 0 & \\ & & & 0 \end{pmatrix} \right)$$
$$= (\tilde{\mathbf{V}}_1(d^{(1)}) \otimes \tilde{\mathbf{V}}_1(d^{(2)})) \underbrace{\begin{pmatrix} \multicolumn{4}{c}{\mathbf{I}_4} \\ \hline 0 & 1 & & \\ & 0 & & 2 \\ & & 0 & 1 \\ & & & 0 \\ \hline 0 & 0 & 1 & \\ & 0 & & 1 \\ & & 0 & 0 \\ & & & 0 \\ \hline 0 & 0 & & 1 \\ & 0 & & 0 \\ & & 0 & 0 \\ & & & 0 \end{pmatrix}}_{=\mathbf{G}_{1[1,2;3]}}.$$

REFERENCES

[1] K. Batselier, *A Numerical Linear Algebra Framework for Solving Problems with Multivariate Polynomials*, Ph.D. thesis, ESAT/STADIUS, KU Leuven, 2013; advisor: B. De Moor.

[2] G. Bergqvist, *Exact probabilities for typical ranks of* $2 \times 2 \times 2$ *and* $3 \times 3 \times 2$ *tensors*, Linear Algebra Appl., 438 (2013), pp. 663–667, https://doi.org/10.1016/j.laa.2011.02.041.

[3] D. L. Boley, F. T. Luk, and D. Vandevoorde, *A general Vandermonde factorization of a Hankel matrix*, in ILAS Symp. on Fast Algorithms for Control, Signals and Image Processing, Winnipeg, 1997.

[4] P. Breiding and S. Timme, *HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia*, preprint, https://arxiv.org/abs/1711.10911, 2017.

[5] B. Buchberger, *Gröbner bases: A short introduction for systems theorists*, in Computer Aided Systems Theory, Lecture Notes in Comput. Sci. 2178, Springer, Berlin, 2001, pp. 1–19, https://doi.org/10.1007/354045654.

[6] B. Dayton, T.-Y. Li, and Z. Zeng, *Multiple zeros of nonlinear systems*, Math. Comp., 80 (2011), pp. 2143–2168.

[7] B. Dayton and Z. Zeng, *Computing the multiplicity structure in solving polynomial systems*, in Proceedings of the 2005 International Symposium on Symbolic and Algebraic Computation, ISSAC '05, ACM, New York, 2005, pp. 116–123.

[8] L. De Lathauwer, *Decompositions of a higher-order tensor in block terms—Part* I*: Lemmas for partitioned matrices*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1022–1032, https://doi.org/10.1137/060661685.

[9] L. De Lathauwer, *Decompositions of a higher-order tensor in block terms—Part* II*: Definitions and uniqueness*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1033–1066, https://doi.org/10.1137/070690729.

[10] L. De Lathauwer, *Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1451–1474, https://doi.org/10.1137/100805510.

[11] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278, https://doi.org/10.1137/S0895479896305696.

[12] L. De Lathauwer and D. Nion, *Decompositions of a higher-order tensor in block terms—Part* III*: Alternating least squares algorithms*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1067–1083, https://doi.org/10.1137/070690730.

[13] V. De Silva and L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127, https://doi.org/10.1137/06066518X.

[14] P. Dreesen, *Back to the Roots: Polynomial System Solving Using Linear Algebra*, Ph.D. thesis, ESAT/STADIUS, KU Leuven, 2013; advisor: B. De Moor.

[15] P. Dreesen, K. Batselier, and B. De Moor, *On the null spaces of the Macaulay matrix*, Linear Algebra Appl., 460 (2014), pp. 259–289, https://doi.org/10.1016/j.laa.2014.07.035.

[16] P. Dreesen, K. Batselier, and B. De Moor, *Multidimensional realisation theory and polynomial system solving*, Internat. J. Control, 91 (2018), pp. 2692–2704, https://doi.org/10.1080/00207179.2017.1378924.

[17] M. Gasca and J. J. Martínez, *On the computation of multivariate confluent Vandermonde determinants and its applications*, in The Mathematics of Surfaces II, Clarendon Press, New York, 1988, pp. 101–114.

[18] I. Gohberg, P. Lancaster, and L. Rodman, *Invariant Subspaces of Matrices with Applications*, SIAM, Philadelphia, 2006, https://doi.org/10.1137/1.9780898719093.

[19] G. Golub and C. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, MD, 2013.

[20] W. Gröbner, *Algebraische Geometrie* II, Manheim University Library, 1970.

[21] G. Jónsson and S. Vavasis, *Accurate solution of polynomial equations using Macaulay resultant matrices*, Math. Comp., 74 (2004), pp. 221–262, https://doi.org/10.1090/S0025-5718-04-01722-3.

[22] D. Kalman, *The generalized Vandermonde matrix*, Math. Mag., 57 (1984), pp. 15–21.

[23] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500, https://doi.org/10.1137/07070111X.

[24] W. Krijnen, T. Dijkstra, and A. Stegeman, *On the non-existence of optimal solutions and the occurrence of "degeneracy" in the Candecomp/Parafac model*, Psychometrika, 73 (2008), pp. 431–439, https://doi.org/10.1007/s11336-008-9056-1.

[25] J. Kruskal, R. Harshman, and M. Lundy, *How* 3*-MFA data can cause degenerate Parafac solutions, among other relationships*, in Multiway Data Analysis, North-Holland, 1989, pp. 115–121.

[26] F. Macaulay, *The Algebraic Theory of Modular Systems*, Cambridge University Press, Cambridge, UK, 1916.

[27] J. R. Magnus and H. Neudecker, *The elimination matrix: Some lemmas and applications*, SIAM J. Alg. Discrete Methods, 1 (1980), pp. 422–449, https://doi.org/10.1137/0601049.

[28] H. M. Möller and H. J. Stetter, *Multivariate polynomial equations with multiple zeros solved by matrix eigenproblems*, Numer. Math., 70 (1995), pp. 311–329.

[29] L. Sorber, M. Van Barel, and L. De Lathauwer, *Optimization-based algorithms for ten-*

sor decompositions: Canonical polyadic decomposition, decomposition in rank-$(L_r, L_r, 1)$ *terms, and a new generalization*, SIAM J. Optim., 23 (2013), pp. 695–720, https://doi.org/10.1137/120868323.

[30] M. Sørensen and L. De Lathauwer, *Multidimensional harmonic retrieval via coupled canonical polyadic decomposition—Part* I*: Model and identifiability*, IEEE Trans. Signal Process., 65 (2017), pp. 517–527, https://doi.org/10.1109/TSP.2016.2614796.

[31] A. Stegeman, *Candecomp/Parafac: From diverging components to a decomposition in block terms*, SIAM J. Matrix Anal. Appl., 32 (2012), pp. 291–316, https://doi.org/10.1137/110825327.

[32] A. Stegeman, *A three-way Jordan canonical form as limit of low-rank tensor approximations*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 624–650, https://doi.org/10.1137/120875806.

[33] A. Stegeman and L. De Lathauwer, *Are Diverging CP Components Always Nearly Proportional?*, Tech. Report 11-174, ESAT-SISTA, KU Leuven, 2011.

[34] H. J. Stetter, *Numerical Polynomial Algebra*, SIAM, Philadelphia, 2004, https://doi.org/10.1137/1.9780898717976.

[35] S. Telen, *Solving Systems of Polynomial Equations*, Master's thesis, KU Leuven, 2016; advisor: M. Van Barel.

[36] S. Telen, *Solving Systems of Polynomial Equations*, Ph.D. thesis, Faculty of Engineering Science, KU Leuven, 2020; advisor: M. Van Barel.

[37] S. Telen, B. Mourrain, and M. Van Barel, *Solving polynomial systems via truncated normal forms*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 1421–1447, https://doi.org/10.1137/17M1162433.

[38] J. Vanderstukken and L. De Lathauwer, *Systems of polynomial equations, higher-order tensor decompositions, and multidimensional harmonic retrieval: A unifying framework. Part* I*: The canonical polyadic decomposition*, SIAM J. Matrix Anal. Appl., 42 (2021), pp. 883–912, https://doi.org/10.1137/17M1150050.

[39] J. Verschelde, *Homotopy Continuation Methods for Solving Polynomial Systems*, Ph.D. thesis, Faculty of Engineering Science, KU Leuven, 1996; advisor: A. Haegemans.

[40] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, *Tensorlab* 3.0, 2016; available online from http://www.tensorlab.net.

[41] X. Wu and L. Zhi, *Computing the multiplicity structure from geometric involutive form*, in Proceedings of the Twenty-First International Symposium on Symbolic and Algebraic Computation, ISSAC '08, ACM, New York, 2008, pp. 325–332.