21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Decomposing multivariate polynomials with structured low-rank matrix completion

Konstantin Usevich[1]

***Abstract*— We are focused on numerical methods for decomposing a multivariate polynomial as a sum of univariate polynomials in linear forms. The main tool is the recent result on correspondence between the Waring rank of a homogeneous polynomial and the rank of a partially known quasi-Hankel matrix constructed from the coefficients of the polynomial. Based on this correspondence, we show that the original decomposition problem can be reformulated as structured low-rank matrix completion (or as structured low-rank approximation in the case of approximate decomposition). We construct algorithms for the polynomial decomposition problem. In the case of bivariate polynomials, we provide an extension of the well-known Sylvester algorithm for binary forms.**

I. INTRODUCTION

A. Problem statement and motivation

In this paper, we consider the following polynomial decomposition problem. Given a multivariate complex (resp. real) polynomial $p(z_1, \ldots, z_n)$ of degree d in n variables, decompose it into a sum of r univariate polynomials of linear forms, i.e. find a decomposition (possibly minimal) of the form

$$\begin{aligned} p(z_1, \ldots, z_n) = q_1 \left(s_{1,1}z_1 + \cdots + s_{1,n}z_n\right) + \cdots \\ \cdots + q_r \left(s_{r,1}z_1 + \cdots + s_{r,n}z_n\right), \end{aligned} \tag{1}$$

where $q_1, \ldots, q_r$ are univariate complex (resp. real) polynomials of degree d and $\{s_{k,j}\}_{k,j=1}^{r,n}$ are the complex (resp. real) coefficients of the linear combinations.

The decomposition (1) is motivated by a problem arising in system identification of block structured models [1]. Often, it is possible to estimate (for example, using frequency domain methods [1]) a block-structured model, consisting of a single-input-multiple-output (SIMO) linear time-invariant (LTI) block, followed by a multivariate polynomial nonlinearity (Fig. 1, top diagram). However, this representation is coupled, and it is more desirable to find a *parallel Wiener* representation (Fig. 1, bottom diagram). This can be accomplished as shown in Fig. 1. By decomposing a polynomial nonlinearity as (1), the linear transformation can be merged with the SIMO block, which yields a parallel Wiener representation.

One also may be interested (for example, in system identification) in approximation of the form (1) (rather than in decomposition). In approximation theory, this problem is related to approximation of a multivariate nonlinear function as a sum of nonlinear functions of the form $\widetilde{f}(s_1 z_1 + \ldots + s_n z_n)$ (called *ridge functions* [2] or *planar waves* [3]). Approximation by ridge functions appears in many topics of statistics and data analysis (feed-forward neural networks, density estimation, independent component analysis, projection pursuit, etc.), see [2], [4] and references therein.

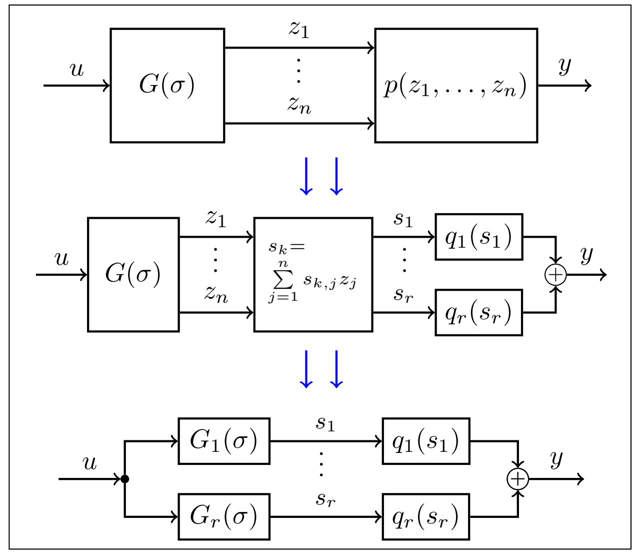

Fig. 1. Decoupling a nonlinearity in identification of parallel Wiener systems

B. Previous works

In the case when the polynomial p is homogeneous (the total degree of all its monomials is the same), the decomposition (1) is known as *Waring decomposition*, and is equivalent to symmetric tensor decomposition [5], [6], [7, Ch. 5]. There are many recent results on the *Waring rank* of homogeneous polynomials (i.e., r in the minimal decomposition (1)): maximal rank (r needed to decompose any polynomial) and generic rank (r needed to decompose a generic polynomial). There are results on uniqueness of Waring decomposition.

For $n = 2$, there is a classic Sylvester algorithm (see [5], [8]), which provides a complete solution to the Waring decomposition problem. For $n > 2$, there exist symbolic and numeric algorithms that provide partial solutions to the Waring decomposition problem [8], [9].

For the general (non-homogeneous case), there are just a few works [10], [11]. In [11], it was shown that any real or complex polynomial can be decomposed with at most

$$r_{max} \leq \binom{n+d-2}{d-1}$$

*This work is supported by ERC Grant agreement no. 258581 "Structured low-rank approximation: Theory, algorithms, and applications".

[1]Konstantin Usevich is with the Department ELEC, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium kusevich@vub.ac.be

ISBN: 978-90-367-6321-9

terms in (1), where $\binom{n}{k}$ is the binomial coefficient. There are currently no results on the generic r, to the extent of the author's knowledge.

As of numerical algorithms, probably the only available paper is [1] on decomposition of degree-3 polynomials. The approach of [1] is based on non-symmetric tensor decompositions, and treats the homogeneous components of $p(\mathbf{z})$ separately, which leads to r higher than r_{max}.

C. Contribution of this paper

We are focused on developing practical algorithms for decomposition (1). We use the connection of (1) to simultaneous Waring decomposition of homogeneous polynomials [6]. We extend existing algorithms using a well-known idea in signal processing of stacking Hankel-like matrices next to each other. This idea was also used in [12] for simultaneous Waring decomposition of a quartic and cubic (in the case $n = 2$), in the context of independent component analysis.

First, for the case $n = 2$, we extend the Sylvester algorithm. The proposed algorithm is guaranteed to find a minimal decomposition of the form (1). Second, for $n > 2$, we extend the recent result [8] on correspondence between the Waring decomposition and low-rank completion of a quasi-Hankel matrix. We reformulate the decomposition problem (1) of a non-homogeneous polynomial as a structured low-rank matrix completion problem (or structured low-rank approximation in the case of approximate decomposition).

In this paper, we also summarize the results of [8] in the language of linear recurrent arrays, which is more familiar to researchers in signal processing and system theory. Also, compared to [8], where systems of polynomial equations are solved, we propose to use general-purpose methods for matrix completion.

The paper is organized as follows. Section II is devoted to preliminaries, such as notation, Waring decomposition, connection to linear recurrent arrays and multidimensional Hankel matrices (and multidimensional systems). In Section III, the polynomial decomposition (1) is reformulated as a simultaneous Waring decomposition. Next, we state the main result of the paper: a correspondence between the polynomial decomposition problem (1) and a structured low-rank matrix completion problem. Using this correspondence, we present a generic algorithm for polynomial decomposition. We provide a simpler algorithm for bivariate polynomials (an extension of the Sylvester algorithm), which uses only standard linear algebra operations. Section IV contains some numerical examples for the algorithms.

II. PRELIMINARIES

A. Notation

We denote by $\mathbb{Z}_+$ the set of nonnegative integers, and by $\mathbb{C}$ and $\mathbb{R}$ the sets of complex and real numbers respectively. We denote by $\mathbb{R}[\mathbf{z}]$ (resp. $\mathbb{C}[\mathbf{z}]$), where $\mathbf{z} = \begin{bmatrix} z_1 & \cdots & z_n \end{bmatrix}^\top$, the vector space of n-variate polynomials with real (resp. complex) coefficients; $\mathbb{R}_d[\mathbf{z}]$ and $\mathbb{C}_d[\mathbf{z}]$ stand for spaces of homogeneous polynomials of degree d, and $\mathbb{R}_{\le d}[\mathbf{z}]$ and $\mathbb{C}_{\le d}[\mathbf{z}]$ the spaces of multivariate polynomials with degree at most d.

For vectors $x = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top \in \mathbb{C}^n$ and $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^\top \in \mathbb{Z}_+^n$, we define in a standard way the power operation:

$$x^\alpha := x_1^{\alpha_1} \cdot \cdots \cdot x_n^{\alpha_n}.$$

For a multi-index $\alpha \in \mathbb{Z}_+^n$, we denote by $|\alpha| = \alpha_1 + \cdots + \alpha_n$ the sum of its elements. By $\blacktriangle^{(n,d)}$ we denote the set of multi-indices $\{\alpha \in \mathbb{Z}_+^n \,|\, |\alpha| \le d\}$ (the degrees of monomials in the space $\mathbb{K}_{\le d}[\mathbf{z}]$). By $\triangle^{(n,d)}$ we denote the set $\{\alpha \in \mathbb{Z}_+^n \,|\, |\alpha| = d\}$ (the degrees of monomials in the space $\mathbb{K}_d[\mathbf{z}]$).

B. Multidimensional arrays and Hankel matrices

Let $\mathbb{A}_n = \{\mathbb{Z}_+^n \to \mathbb{C}\}$ denote the space of n-dimensional infinite sequences[1], i.e., $\mathrm{A} = (\mathrm{A}_\alpha)_{\alpha \in \mathbb{Z}_+^n} \in \mathbb{A}_n$ is a collection of elements indexed by the multi-indices $\alpha \in \mathbb{Z}_+^n$.

For an $\mathrm{A} \in \mathbb{A}_n$ we consider the *infinite Hankel matrix*, with row and columns indexed by multi-indices,

$$\mathcal{H}(\mathrm{A}) := [h_{\alpha,\beta}]_{\alpha,\beta \in \mathbb{Z}_+^n}, \quad h_{\alpha,\beta} = \mathrm{A}_{\alpha+\beta}.$$

Note 1: The rows and columns of $\mathcal{H}(\mathrm{A})$ can be also conveniently indexed by integers as

$$\mathcal{H}(\mathrm{A}) = [h_{i,j}]_{i,j=0}^{\infty,\infty}, \quad h_{i,j} = \mathrm{A}_{\alpha_i+\alpha_j}.$$

if a multi-index order $0 = \alpha_0 \prec \alpha_1 \prec \cdots$ is chosen.
It is easy to see that the rows (and the columns) of $\mathcal{H}(\mathrm{A})$ are shifts of the original infinite array

$$\mathcal{H}(\mathrm{A})_{\beta,:} = \mathcal{H}(\mathrm{A})_{:,\beta} = \sigma_\beta(\mathrm{A}).$$

where the array shift $\sigma_\beta(\mathrm{A})$ is defined as

$$\sigma_\beta(\mathrm{A}) := \mathrm{B} \in \mathbb{A}_n, \quad \text{where} \quad \mathrm{B}_\alpha = \mathrm{A}_{\alpha+\beta}.$$

In our exposition, we also define an action of the polynomial $p(\mathbf{z}) = \sum\limits_{\alpha \in \blacktriangle^{(n,d)}} p_\alpha \mathbf{z}^\alpha \in \mathbb{C}_{\le d}[\mathbf{z}]$ on an array as

$$p * \mathrm{A} := \sum_{\beta \in \blacktriangle^{(n,d)}} p_\beta \sigma_\beta(\mathrm{A}).$$

Thus a β-shift is the action of $\mathbf{z}^\beta$ on A, i.e., $\mathbf{z}^\beta * \mathrm{A} = \sigma_\beta(\mathrm{A})$.

Note 2: The action $p * \mathrm{A}$ can be interpreted as filtering or convolution with an "FIR-filter" p (because a polynomial, by definition, has a finite number of nonzero coefficients).

We say that the polynomial $p(\mathbf{z})$ annihilates A if $p * \mathrm{A} \equiv 0$. Note that $p * \mathrm{A} \equiv 0$ if and only if

$$\sum_{\beta \in \blacktriangle^{(n,d)}} p_\beta \mathrm{A}_{\alpha+\beta} = 0, \quad \text{for all } \alpha \in \mathbb{Z}_+^n,$$

i.e., the infinite array A satisfies a linear recurrence relation with coefficients of $p(\mathbf{z})$. We denote by

$$\mathcal{I}(\mathrm{A}) := \{p \in \mathbb{C}[z] \,|\, p * \mathrm{A} \equiv 0\} \tag{2}$$

[1]The space $\mathbb{A}_n$ can be identified with the dual space of $\mathbb{C}[z]$, and also with the set of formal power series [8].

the set of all polynomials that annihilate an array (i.e., the set of all possible linear recurrences). The set $\mathcal{I}(\mathrm{A})$ is a polynomial ideal [13] in $\mathbb{C}[\mathbf{z}]$.

Note 3: The matrix $\mathcal{H}(\mathrm{A})$ is the matrix representation of the linear (Hankel) operator

$$\begin{aligned}\mathscr{H}^{(\mathrm{A})} :&\mathbb{C}[\mathbf{z}] \to \mathbb{A}_n\\ &p(\mathbf{z}) \mapsto p * \mathrm{A}.\end{aligned}$$

Moreover, the ideal $\mathcal{I}(\mathrm{A})$ is the kernel of the operator $\mathscr{H}^{(\mathrm{A})}$. The following theorem gives a characterization for Hankel matrices (and operators) of finite rank.

Theorem 1: The following three statements are equivalent

- $\operatorname{rank} \mathcal{H}(\mathrm{A}) = r < \infty$
- $\mathcal{I}(\mathrm{A})$ is a 0-dimensional ideal with variety $\{\boldsymbol{\lambda}_k\}_{k=1}^m \subset \mathbb{C}^n$, $m \le r$, and $\dim \mathbb{C}[z]/\mathcal{I}(\mathrm{A}) = r$.
- the array A has a representation

$$\mathrm{A}_\alpha = \sum_{k=1}^{m} h_k(\alpha)\boldsymbol{\lambda}_k^\alpha, \quad \text{for all } \alpha \tag{3}$$

where $h_k \in \mathbb{C}[\mathbf{z}]$ are polynomials and $\boldsymbol{\lambda}_k \in \mathbb{C}^n$.

If one of the statements holds true, then

$$\mathcal{I}(\mathrm{A}) \text{ is radical} \iff \mathrm{A}_\alpha = \sum_{k=1}^{m} c_k\boldsymbol{\lambda}_k^\alpha \iff m = r.$$

The case when $\mathcal{I}(\mathrm{A})$ is not radical, corresponds to the case of multiple roots for univariate polynomials.

Theorem 1 (or its variations) can be found in different papers and formulated with different terminology: in [14], [15] it is formulated in terms of linear recurrent sequences (for example [14, §20.20]); in [8] it is formulated in terms of linear functionals on $\mathbb{C}[\mathbf{z}]$ (see also [16]); in [17], [18] similar results were formulated in terms of multidimensional systems theory.

C. Waring decomposition and polynomial decomposition

Problem 1 (Waring decomposition): Given a homogeneous polynomial $a(\mathbf{z}) \in \mathbb{K}_d[\mathbf{z}]$, $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$, find the minimal r and vectors $\mathbf{s}_1, \ldots, \mathbf{s}_r \in \mathbb{K}^n$ such that

$$a(\mathbf{z}) = c_1 \cdot (\mathbf{s}_1^\top \mathbf{z})^d + \cdots + c_r \cdot (\mathbf{s}_r^\top \mathbf{z})^d, \tag{4}$$

with $c_j \in \mathbb{K}^n \setminus \{0\}$. The minimal such r is called the *Waring rank* of $a(\mathbf{z})$.

Note 4: Waring decomposition problem can be considered as a restriction of the decomposition problem (1) to homogeneous polynomials.

Waring problem is a classic problem in algebra (see [5] for a historical overview), and recently received much attention due to a remarkable result of [19] on the generic ranks, and also due to many important applications (see [6] for a list of references).

Waring problem is also known as a tensor decomposition problem. Consider a symmetric d-way tensor

$$\mathcal{A} \in \mathbb{K}^{\overbrace{n \times \cdots \times n}^{d \text{ times}}},$$

where the term *symmetric* means that

$$\mathcal{A}_{(j_1,\ldots,j_d)} = \mathcal{A}_{\Pi(j_1,\ldots,j_d)}$$

for any permutation of indeces Π. For any symmetric tensor $\mathcal{A}$ we can construct a homogeneous polynomial

$$a(\mathbf{z}) = \mathcal{A}\times_1\mathbf{z}\times_2\mathbf{z}\cdots\times_d\mathbf{z} = \sum_{j_1,\ldots,j_d=1}^{n} (\mathcal{A}_{(j_1,\ldots,j_d)})\cdot z_{j_1}\cdots z_{j_d},$$

and vice versa, for any homogeneous polynomial there exist a unique symmetric tensor $\mathcal{A}$.

The powers of linear forms correspond to rank-1 symmetric tensors, i.e., for $\mathbf{s} \in \mathbb{K}^n$ we have that

$$(\mathbf{s} \otimes \cdots \otimes \mathbf{s}) \times_1 \mathbf{z} \times_2 \mathbf{z} \cdots \times_d \mathbf{z} = (\mathbf{s}^\top \mathbf{z})^d.$$

Hence, (4) is equivalent to symmetric tensor decomposition

$$\mathcal{A} = \sum_{i=1}^{r} c_i \cdot \overbrace{\mathbf{s}_i \otimes \cdots \otimes \mathbf{s}_i}^{d}, \quad \text{where } s_i \in \mathbb{K}^n. \tag{5}$$

Example 1: Consider the case $d = 2$, i.e., $a(z)$ is a quadratic form. In linear algebra, quadratic forms are represented by symmetric matrices $\mathcal{A} \in \mathbb{K}^{n\times n}$ such that $a(\mathbf{z}) = \mathbf{z}^\top \mathcal{A}\mathbf{z}$. For the polynomial

$$a(\mathbf{z}) = a_1z_1^2 + a_2z_2^2 + a_3z_3^2 + a_4z_1z_2 + a_5z_1z_3 + a_6z_2z_3,$$

the corresponding symmetric matrix is

$$\mathcal{A} = \begin{bmatrix} a_1 & \frac{a_4}{2} & \frac{a_5}{2} \\ \frac{a_4}{2} & a_2 & \frac{a_6}{2} \\ \frac{a_5}{2} & \frac{a_6}{2} & a_3 \end{bmatrix}.$$

Then (5) is equivalent to the factorization

$$\mathcal{A} = S \operatorname{diag}(c_1, \ldots, c_r) S^\top, \quad S := \begin{bmatrix}\mathbf{s}_1 & \cdots & \mathbf{s}_r\end{bmatrix}.$$

One of the possible such factorizations is given by the eigenvalue decomposition (where the vectors $\mathbf{s}_k$ are orthogonal). Thus, the Waring decomposition can be considered as a generalization of the symmetric factorization of symmetric matrices.

Note 5: Note that decompositions (4) and (5) contain an obvious overparameterization: the vectors $\mathbf{s}_k$ can be scaled and permuted, without changing the decomposition. Thus in decomposition (4) we aim at finding $\mathbf{s}_k$ up to permutation and scaling.

Note 6: The decomposition (4) can be still nonunique, even if we take into account scaling and permutation and consider a minimal decomposition (for the minimal r).

For more background on Waring and symmetric tensor decompositions, we refer the reader to [6].

D. From Waring decomposition to sums of exponentials

This subsection contains a brief description of the principle of *apolar duality*, widely used in the theory of Waring decompositions.

For a homogeneous polynomial $a(\mathbf{z}) = \sum\limits_{\alpha\in\triangle^{(n,d)}} a_\alpha\mathbf{z}^\alpha$, we define its *normalized coefficients* $\{a_\alpha^{(N)}\}_{\alpha\in\triangle^{(n,d)}}$ as

$$a_\alpha^{(N)} = a_\alpha \cdot \binom{d}{\alpha}^{-1},$$

where $\binom{d}{\alpha} := \frac{d!}{\alpha_1!\cdots\alpha_n!}$ is the *multinomial coefficient*. Hence, we have that

$$a(\mathbf{z}) = \sum_{\alpha\in\triangle^{(n,d)}} \binom{d}{\alpha} a_\alpha^{(N)} \mathbf{z}^\alpha.$$

Due to the properties of the multinomial coefficient, we have

$$(\mathbf{s}^\top \mathbf{z})^d = \sum_{\alpha\in\triangle^{(n,d)}} \binom{d}{\alpha} \mathbf{s}^\alpha \mathbf{z}^\alpha.$$

Hence, the Waring decomposition (4) is equivalent to

$$a_\alpha^{(N)} = \sum_{k=1}^{r} c_k \mathbf{s}_k^\alpha, \quad \text{for all } \alpha \in \triangle^{(n,d)}, \tag{6}$$

which is a decomposition of the array of normalized coefficients $\{a_\alpha^{(N)}\}_{\alpha\in\triangle^{(n,d)}}$ as a sum of projective exponentials.

E. Affine decomposition

Assume that we deal only with the case $(\mathbf{s}_k)_1 \neq 0$ for all k. As noted in [8, Def. 2.1], we can always achieve $(\mathbf{s}_k)_1 \neq 0$ by a generic rotation of the variables $\mathbf{z}$. Then, due to Note 5, we may assume that $\mathbf{s}_k$ have the form

$$\mathbf{s}_k = \begin{bmatrix} 1 \\ \boldsymbol{\lambda}_k \end{bmatrix}, \quad \boldsymbol{\lambda}_k = \begin{bmatrix} \lambda_{k,1} & \cdots & \lambda_{k,n-1} \end{bmatrix}^\top \in \mathbb{K}^{n-1}. \tag{7}$$

Then the decomposition (6) is equivalent to

$$a_\alpha^{(D)} = \sum_{k=1}^{r} c_k \boldsymbol{\lambda}_k^\alpha, \quad \text{for all } \alpha \in \blacktriangle^{(n-1,d)}, \tag{8}$$

where $\{a_\alpha^{(D)}\}_{\alpha\in\blacktriangle^{(n-1,d)}}$ are the *dehomogenized coefficients*

$$a_\alpha^{(D)} := a_{d-|\alpha|,\alpha_1,\ldots,\alpha_{n-1}}^{(N)}, \tag{9}$$

obtained from the normalized coefficients. The decomposition (8) is called *affine decomposition* [8, Def. 2.1] (as opposed to its projective counterpart (6)).

Note that the operation of dehomogenization drops the dimension of the array of coefficients by 1.

F. Affine decompositions and Hankel matrices

In what follows, we deal with complex decompositions ($\mathbb{K} = \mathbb{C}$). In [8] it was proposed to relax the problem from finding the Waring decomposition, to finding the decomposition of the dehomogenized coefficients of the form

$$a_\alpha^{(D)} = \sum_{k=1}^{r} h_k(\alpha) \boldsymbol{\lambda}_k^\alpha, \quad \text{for all } \alpha \in \blacktriangle^{(n-1,d)}. \tag{10}$$

where $h_k \in \mathbb{C}[z_1, \ldots, z_{n-1}]$. Note that this decomposition is exactly of the form (3). We will refer to (10) as a *generalized affine decomposition*.

By Theorem 1, the problem of finding minimal decomposition (10) is equivalent to the following problem.

Problem 2:

$$\begin{aligned} &\underset{\mathrm{A}\in\mathbb{A}_{n-1}}{\text{minimize}} \;\; \operatorname{rank} \mathcal{H}(\mathrm{A}) \\ &\text{subject to} \;\; \mathrm{A}_\alpha = a_\alpha^{(D)} \text{ for all } \alpha \in \blacktriangle^{(n-1,d)}. \end{aligned}$$

This is a problem of finding the minimal low-rank completion of the infinite Hankel matrix $\mathcal{H}(\mathrm{A})$, given the initial entries $\{\mathrm{A}_\alpha\}_{\alpha\in\blacktriangle^{(n-1,d)}}$.

Note 7: It is well known (see Note 6), that a symmetric tensor may have a nonunique symmetric decomposition. The same holds for the generalized affine decomposition (10). However, after the completion in the Problem 2, the $\boldsymbol{\lambda}_k$ in the decomposition (10) are determined uniquely. Therefore, the nonuniqueness of symmetric tensor decompositions corresponds to nonuniqueness of matrix completion in Problem 2.

G. Reduction to quasi-Hankel matrix completion

Since the rank of an infinite Hankel matrix is finite, it is possible to construct a submatrix that will completely define the elements of the whole Hankel matrix $\mathscr{H}^{(\mathrm{A})}$ [8].

Consider two ordered sets of multi-indices, described by matrices $\mathcal{A} \in \mathbb{Z}_+^{n\times m_a}$ and $\mathcal{B} \in \mathbb{Z}_+^{n\times m_b}$, where

$$\begin{aligned} \mathcal{A} &= \begin{bmatrix} \alpha^{(1)} & \ldots & \alpha^{(m_a)} \end{bmatrix} \in \mathbb{Z}_+^{n\times m_a}, \\ \mathcal{B} &= \begin{bmatrix} \beta^{(1)} & \ldots & \beta^{(m_b)} \end{bmatrix} \in \mathbb{Z}_+^{n\times m_b}. \end{aligned}$$

Then the quasi-Hankel [20] matrix $\mathscr{H}_{\mathcal{A},\mathcal{B}}(\mathrm{A}) \in \mathbb{C}^{m_a\times m_b}$ is defined as

$$(\mathscr{H}_{\mathcal{A},\mathcal{B}}(\mathrm{A}))_{i,j} := \mathrm{A}_{\alpha^{(i)}+\beta^{(j)}}.$$

The quasi-Hankel matrix is a submatrix of $\mathcal{H}(\mathrm{A})$ corresponding to the sets $\{\alpha^{(1)}, \ldots, \alpha^{(m_a)}\}$ and $\{\beta^{(1)}, \ldots, \beta^{(m_b)}\}$ of row and column indices. $\mathscr{H}_{\mathcal{A},\mathcal{B}}$ takes into account only the multi-indices from the Minkowski sum $\mathcal{A} + \mathcal{B}$.

Now we go back to Problem 2. In [8], it was shown that it is sufficient to consider a submatrix $\mathscr{H}_{\mathcal{A},\mathcal{A}}$ with $\mathcal{A} = \blacktriangle^{(n-1,d)}$. This results in the following algorithm (we present it in a simplified form).

Algorithm 1: Input: the homogeneous polynomial $a(\mathbf{z}) \in \mathbb{C}_d[\mathbf{z}]$. Output: the factors in decomposition (10).

1) Construct the normalized dehomogenized coefficients $\{a_\alpha^{(D)}\}_{\alpha\in\blacktriangle^{(n-1,d)}}$.
2) Construct the quasi-Hankel matrix $\mathscr{H}_{\blacktriangle^{(n-1,d)},\blacktriangle^{(n-1,d)}}(\mathrm{A})$, where $\mathrm{A}_\alpha = a_\alpha^{(D)}$ for $\alpha \in \blacktriangle^{(n-1,d)}$, and A_α are unknown for $\alpha \in \blacktriangle^{(n-1,2d)} \setminus \blacktriangle^{(n-1,d)}$.
3) Find the minimal rank completion for the unknown entries A_α for $\alpha \in \blacktriangle^{(n-1,2d)} \setminus \blacktriangle^{(n-1,d)}$.
4) Reconstruct the exponents $\boldsymbol{\lambda}_k$ in the expansion (10) from the image of the matrix $\mathscr{H}_{\mathcal{A},\mathcal{A}}(\mathrm{A})$.
5) Obtain the coefficients of h_k by solving the linear system (10) (e.g., via least squares fitting).

Algorithm 1 is a simplification of [8, Alg. 7.1] (in particular, more details on step 4 can be found in [8]).

The minimal decomposition (10) does not always coincide with (8), thus Algorithm 1 does not solve Waring decomposition problem in general. Unfortunately, for a generic homogeneous polynomial the rank of the minimal completion in Problem 2 may be smaller than the generic Waring rank.

Note 8: In the original algorithm [8, Alg. 7.1], it is proposed to increase the number of exponents until the polynomials h_k in (10) are constant (which yields a decomposition (8)). However, this approach requires solving polynomial systems of equations, which is not always feasible.

H. The case $n = 2$, Sylvester algorithm

For $n = 2$, the set $\blacktriangle^{(n-1,d)}$ is equal to $\{0, 1, \ldots, d\}$. Therefore, the quasi-Hankel matrix is just a Hankel matrix

$$\mathscr{H}_{\blacktriangle^{(1,d)},\blacktriangle^{(1,d)}}(\mathrm{A}) = \mathscr{H}_d(\mathrm{A}) := \begin{bmatrix} \mathrm{A}_0 & \mathrm{A}_1 & \cdots & \mathrm{A}_d \\ \mathrm{A}_1 & \iddots & \iddots & \vdots \\ \vdots & \iddots & & \mathrm{A}_{2d-1} \\ \mathrm{A}_d & \cdots & \mathrm{A}_{2d-1} & \mathrm{A}_{2d} \end{bmatrix}.$$

Note 9: With some abuse of notation, for a vector $A \in \mathbb{C}^T$ we will denote by $\mathscr{H}_L(A) \in \mathbb{C}^{L\times(T-L+1)}$ the Hankel matrix with L rows constructed from A.

Since, the values $\{\mathrm{A}_{d+1}, \ldots, \mathrm{A}_{2d}\}$ are unknown, it is a classic minimal realization problem [21] for scalar Hankel matrices. For $n = 2$, the minimal realization problem is completely solved [22]. The solution is essentially determined by the "maximally square" submatrix of $\mathscr{H}_d(\mathrm{A})$.

Moreover, for $n = 2$ it is possible to construct an algorithm which yields the Waring decomposition (8) (compared to Algorithm 1).

Algorithm 2: Input: homogeneous polynomial $a(z_1, z_2)$. Output: $\mathbf{s}_k$ and c_k in (4).

1) Compute the dehomogenized coefficients $A = [a_j^{(D)}]_{j=0}^d$.
2) Set $r = 1$.
3) If rank $\mathscr{H}_{r+1}(A) = r + 1$, increase r by 1, repeat;
4) Consider a vector $\theta = \begin{bmatrix}\theta_0 & \cdots & \theta_r\end{bmatrix}^\top$, such that $\theta^\top \mathscr{H}_{r+1}(A) = 0$.
5) Define $\{\lambda_k\}_{k=1}^r$ as the roots of $\theta(z) := \sum\limits_{j=0}^{r} \theta_r z^r$.
6) If all the roots are simple, goto step 8.
7) Try another specialization of θ at step 4, and go to step 5. If all specializations for current r were exhausted, increase r and go to step 4.
8) Define $\mathbf{s}_k = \begin{bmatrix}1 & \lambda_k\end{bmatrix}^\top$. Determine the coefficients c_k by solving the linear system (4).

Algorithm 2 is known as Sylvester algorithm [5], [8].

Note 10: Algorithm 2 is suitable not only for the $\mathbf{s}_k$ of the form $\mathbf{s}_k = \begin{bmatrix}1 & \lambda_k\end{bmatrix}^\top$, but also for $\mathbf{s}_k = \begin{bmatrix}0 & 1\end{bmatrix}^\top$. This corresponds to the case $\theta_r = 0$ (the root at ∞ of $\theta(z)$ [22]).

III. MAIN RESULTS

A. Problem (re)statement

It is easy to relate decomposition (1) to Waring decomposition (4). Denote by $a^{(k)}(\mathbf{z})$ the degree-k homogeneous part of $p(\mathbf{z})$ in (1). Then (1) can be rewritten as a system of equations

$$\begin{aligned} a^{(0)}(\mathbf{z}) &= c_{1,0} + \cdots + c_{r,0}, \\ a^{(1)}(\mathbf{z}) &= c_{1,1} \cdot (\mathbf{s}_1^\top \mathbf{z}) + \cdots + c_{r,1} \cdot (\mathbf{s}_r^\top \mathbf{z}), \\ &\ \ \vdots \\ a^{(d)}(\mathbf{z}) &= c_{1,d} \cdot (\mathbf{s}_1^\top \mathbf{z})^d + \cdots + c_{r,d} \cdot (\mathbf{s}_r^\top \mathbf{z})^d, \end{aligned} \tag{11}$$

where $c_{j,k}$ are the coefficients of the polynomials $q_j(t)$, i.e. $q_j(t) = c_{j,0} + c_{j,1}t + \ldots + c_{j,d}t^d$. In other words, decompostion (1) is the problem of simultaneous Waring decomposition of several homogeneous polynomials. Equivalently, it is a simultaneous symmetric tensor decomposition problem.

As in Section II-E, we construct the dehomogenized normalized coefficient arrays for $j = 1, \ldots, d$: $\{a_\alpha^{(j,D)}\}_{\alpha\in\blacktriangle^{(n-1,j)}}$. We also assume that the vectors $\mathbf{s}_k$ are given as in (7), by the vectors $\boldsymbol{\lambda}_k$. Then the system of equations (11) becomes

$$\begin{aligned} a_{\alpha_1}^{(1,D)} &= \sum_{k=1}^{r} c_{1,k} \boldsymbol{\lambda}_k^{\alpha_1}, \quad \text{for all } \alpha_1 \in \mathcal{A}_1, \\ a_{\alpha_2}^{(2,D)} &= \sum_{k=1}^{r} c_{2,k} \boldsymbol{\lambda}_k^{\alpha_2}, \quad \text{for all } \alpha_2 \in \mathcal{A}_2, \\ &\ \ \vdots \\ a_{\alpha_d}^{(d,D)} &= \sum_{k=1}^{r} c_{d,k} \boldsymbol{\lambda}_k^{\alpha_d}, \quad \text{for all } \alpha_d \in \mathcal{A}_d, \end{aligned} \tag{12}$$

where $\mathcal{A}_j := \blacktriangle^{(n-1,j)}$. Therefore, the problem of polynomial decomposition is equivalent to simultaneous sums-of-exponentials fitting. For this problem (at least, for one-dimensional exponents), there is a well-known approach in signal processing [23] to stack the Hankel matrices next to each other. This idea is exploited in the next two subsections.

B. Case $n = 2$: extended Sylvester's algorithm

Let $\{a_i^{(j,D)}\}_{i=0}^j$ be the sequence dehomogenized normalized coefficients. Define the vectors $A^{(j)} \in \mathbb{C}^{j+1}$ as

$$A^{(j)} := \begin{bmatrix} a_0^{(j,D)} & \cdots & a_j^{(j,D)} \end{bmatrix}^\top.$$

Consider the following family of structured matrices

$$\mathscr{S}_r(p) := \begin{bmatrix} \mathscr{H}_{r+1}(A^{(r)}) & \cdots & \mathscr{H}_{r+1}(A^{(d)}) \end{bmatrix}, \tag{13}$$

where $\mathscr{H}_{r+1}(\cdot)$ is a Hankel matrix with r rows (see Note 9). A matrix $\mathscr{S}_r(p)$ has $r + 1$ rows, and contains only homogeneous terms of p from the r-th degree to the d-th degree. Then the decomposition of a bivariate polynomial can be obtained by the following algorithm.

Algorithm 3: Input: polynomial $p(z_1, z_2)$. Output: $\mathbf{s}_k$ and $c_{j,k}$ in (11).

1) Compute the vectors $A^{(j)}$.
2) Set $r = 1$.
3) If rank $\mathscr{S}_r(p) = r + 1$, increase r by 1, repeat;
4) Consider a vector $\theta = \begin{bmatrix}\theta_0 & \cdots & \theta_r\end{bmatrix}^\top$, such that $\theta^\top \mathscr{S}_r(p) = 0$.
5) Define $\{\lambda_k\}_{k=1}^r$ as the roots of $\theta(z) := \sum\limits_{j=0}^{r} \theta_r z^r$.
6) If all the roots are simple, goto step 8.
7) Try another specialization of θ at step 4, and go to step 5. If all specializations for current r were exhausted, increase r and go to step 4.
8) Define $\mathbf{s}_k = \begin{bmatrix}1 & \lambda_k\end{bmatrix}^\top$. Determine the coefficients $c_{j,k}$ by solving the linear system (11).

Algorithm 3 yields the minimal decomposition of the form (11). Since all the vectors $A^{(j)}$, $j \geq r$ satisfy linear recurrent relations with coefficients θ_k (such that the corresponding polynomial $\theta(z)$ has only simple roots), they should be representable as sums of the corresponding complex exponents (12) (from the algebraic theory of Hankel matrices [22]). Moreover, by construction, there cannot be a decomposition of smaller order $r' < r$.

C. Case $n > 2$: a matrix completion algorithm

For $n > 2$, we propose to extend Algorithm 1.

Algorithm 4: Input: a polynomial $p(\mathbf{z}) \in \mathbb{C}_{\le d}[\mathbf{z}]$. Output: the vectors $\mathbf{s}_k$, and coefficients $c_{j,k}$ in the expansion (11)

1) Split the polynomial into homogeneous parts $a^{(j)}(\mathbf{z})$.
2) Construct the normalized dehomogenized coefficients $\{a^{(j,D)}_{\alpha}\}_{\alpha\in\mathcal{A}_j}$.
3) Construct the following matrix
$$\left[\mathscr{H}_{\mathcal{A}_d,\mathcal{A}_1}(\mathrm{A}^{(1)}) \quad \dots \quad \mathscr{H}_{\mathcal{A}_d,\mathcal{A}_d}(\mathrm{A}^{(d)})\right], \qquad (14)$$
where $\mathrm{A}^{(j)}_{\alpha} = a^{(j,D)}_{\alpha}$ for $\alpha \in \mathcal{A}_j$, and other elements are missing.
4) Find the minimal rank completion of (14).
5) Compute $U \in \mathbb{R}^{|\mathcal{A}_d|\times r}$ — the basis of the column space of the matrix (14).
6) Determine the multidimensional exponents $\{\boldsymbol{\lambda}_k\}_{k=1}^{r}$.
7) Obtain the coefficients $c_{j,k}$ by solving the linear system (11) (e.g., via least squares fitting).

Note 11: We use a different from [8] approach for obtaining the exponents $\boldsymbol{\lambda}_k$. We use multidimensional ESPRIT-type method [24] (more precisely, an ESPRIT version for shaped arrays [25]).

We do not prove the correctness of the algorithm, but the main idea is along the same lines with Algorithm 1. The key fact is that an analogue of Theorem 1 holds for a collection of infinite arrays $\mathrm{A}^{(1)},\dots,\mathrm{A}^{(d)}$ which are annihilated by the same set of polynomials (see [17], [18], [26]). In general, the result of the algorithm is not the decomposition (12), but rather an analogous simultaneous generalized affine decomposition (10) of the homogeneous parts $a^{(j)}(\mathbf{z})$.

We also provide an example in Section IV, for $n = 3$.

IV. Numerical Examples

We present several examples, for the algorithms proposed in the paper. All the algorithms were implemented in Matlab; the code is available on request, and is planned to be released publicly.

A. Decomposition of bivariate polynomials ($n = 2$)

1) Example 1: We construct a polynomial

$$p(\mathbf{z}) = q_1(\mathbf{s}_1^{\top}\mathbf{z}) + q_2(\mathbf{s}_2^{\top}\mathbf{z}) + q_3(\mathbf{s}_3^{\top}\mathbf{z}),$$

where the linear forms are given by

$$\mathbf{s}_1 = \begin{bmatrix}1 & 1\end{bmatrix}, \mathbf{s}_2 = \begin{bmatrix}1 & -0.35\end{bmatrix}, \mathbf{s}_3 = \begin{bmatrix}-0.35 & 1\end{bmatrix},$$

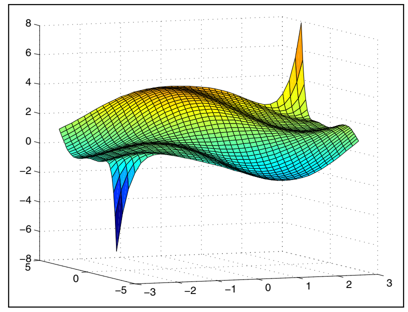

Fig. 2. The first test bivariate polynomial

and q_1, q_2 and q_3 are degree-9 Chebyshev approximations on $[-5;5]$ of the functions

$$f_1(t) = \tanh\left(\frac{\pi t}{2}\right) + \frac{t}{20},$$
$$f_2(t) = \cos(t+1), \quad f_3(t) = \exp(0.2t).$$

respectively. The resulting polynomial $p(\mathbf{z})$ is plotted in Fig. 2.

We consider the sequence of matrices $\mathscr{S}_k(p)$, defined in (13). In Fig. 3 we plot the rank of $\mathscr{S}_r(p)$ depending on r. As we can see in Fig. 3 the rank is full until the point $r = 3$,

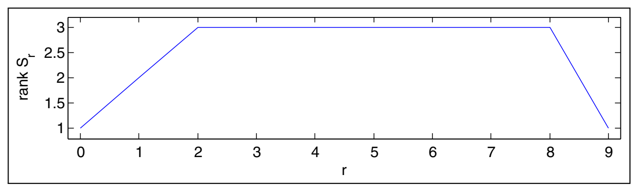

Fig. 3. Ranks of the matrices $\mathscr{S}_r(p)$ depending on r, example 1

after which it stays constant (and equal to 3). The rank drops again at $r = 9$, where the number of columns becomes too small.

From the kernel of $\mathscr{S}_3(p)$ (as well as the kernels of other matrices), we can extract the unique (up to a scaling factor) vector $\theta \approx \begin{bmatrix}-1 & -2.2071 & 2.2071 & 1\end{bmatrix}^{\top}$. The roots of the characteristic polynomial $\theta(z)$ are

$$\lambda_1 = -2.8571, \quad \lambda_2 = 1, \quad \lambda_3 = -0.35,$$

and the directions $\widehat{\mathbf{s}}_1 = \begin{bmatrix}1 & \lambda_1\end{bmatrix}$, $\widehat{\mathbf{s}}_2 = \begin{bmatrix}1 & \lambda_2\end{bmatrix}$, and $\widehat{\mathbf{s}}_3 = \begin{bmatrix}1 & \lambda_3\end{bmatrix}$ exactly correspond to the original directions $\mathbf{s}_1$, $\mathbf{s}_2$ and $\mathbf{s}_3$ (up to scaling and permutation).

Least squares fitting of the coefficients yields estimated polynomials $\widehat{q}_1$, $\widehat{q}_2$, $\widehat{q}_3$. The coefficients of the polynomial coincide with the coefficients of the original q_k up to exponential scaling and numerical precision.

The computed norm of the difference between normalized coefficients of $p(\mathbf{z})$ and $\widehat{p}(\mathbf{z}) = \widehat{q}_1(\widehat{\mathbf{s}}_1^{\top}\mathbf{z}) + \widehat{q}_2(\widehat{\mathbf{s}}_2^{\top}\mathbf{z}) + \widehat{q}_3(\widehat{\mathbf{s}}_3^{\top}\mathbf{z})$ is equal to $\approx 1.5\cdot 10^{-14}$, which confirms that the exact decomposition was achieved up to numerical precision.

Note 12: The behavior of the rank is similar to the rank behavior of block-Hankel matrices. In fact, matrices $\mathscr{S}_r$ are *striped Hankel*, which is a special case of the more general mosaic Hankel case.

2) Example 2: We construct an example for which t he first rank drop may be larger than by one (in the sequence of the ranks of $\mathscr{S}_r(p)$). We consider the previous example, with the last polynomial replaced by $q'_3(t) = 0.2 \cdot t^2$, and everything else (including the vectors $\mathbf{s}_k$) left the same.

As in the previous example, we consider the sequence of matrices $\mathscr{S}_k(p)$, defined in (13), and we plot the ranks of $\mathscr{S}_r$ in Fig. 4. As we can see in Fig. 4, the rank drops by 2

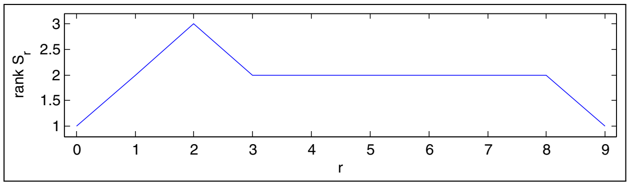

Fig. 4. Ranks of the matrices $\mathscr{S}_r(p)$ depending on r, example 2

when $r = 3$. It happens because the polynomial $q'_3(t)$ is of degree 2 and is not represented in the matrix $\mathscr{S}_3(p)$

Thus we have a nonunique vector in the left kernel of $\mathscr{S}_3(p)$ (the dimension of the kernel is 2). We pick one of the vectors, and it is different from the one in example 1, $\theta = \begin{bmatrix}-0.6947 & -1.6402 & 1.3349 & 1.0000\end{bmatrix}^\top$. The roots of the characteristic polynomial are

$$\lambda_1 = -1.9849, \quad \lambda_2 = 1, \quad \lambda_3 = -0.35.$$

Thus we see that the directions $\mathbf{s}_1$ and $\mathbf{s}_2$ are preserved and the direction $\mathbf{s}_3$ (corresponding to the polynomial q'_3) is changed.

The computed norm of the difference between normalized coefficients for the original polynomial $p(\mathbf{z})$ and recovered polynomial is equal to $\approx 10^{-15}$.

B. Decomposition of polynomials for $n = 3$

We consider the test polynomial from [1], which is a polynomial of degree 3 in 3 variables

$$\begin{aligned} p(\mathbf{z}) =& 0.9743 \\ & - 7.7940z_0 + 0.4601z_1 + 1.1426z_2 \\ & - 76.8242z_0^2 - 10.9872z_1^2 - 13.1377z_2^2 \\ & - 9.8017z_0z_1 - 8.8899z_0z_2 - 9.8928z_1z_2 \\ & - 296.1180z_0^3 + 9.2577z_1^3 + 10.2615z_2^3 \\ & - 135.2378z_0^2z_1 - 96.1882z_0^2z_2 - 95.2097z_0z_1^2 \\ & - 73.7725z_0z_2^2 + 10.4748z_1^2z_2 + 29.7747z_1z_2^2 \\ & - 58.3379z_0z_1z_2. \end{aligned}$$

The method of [1] treats homogeneous components separately, and produces a decomposition of rank 8. However, the theory in [10], [11] suggests that the maximal rank of the polynomial decomposition is 6 (for real decomposition) and 5 (for complex decomposition).

We ran Algorithm 4 with additional knowledge of bound on the maximal rank 5, since Algorithm 4 produces complex decompositions. The algorithm gives a complex rank-5 decomposition with

$$\begin{aligned} \boldsymbol{\lambda}_1 &= (-0.0593, -0.1583), \\ \boldsymbol{\lambda}_2 &= (-0.0149 + 0.2265i, -0.0659 + 0.1987i), \\ \boldsymbol{\lambda}_3 &= (-0.0149 - 0.2265i, -0.0659 - 0.1987i), \\ \boldsymbol{\lambda}_4 &= (-0.3153 + 0.0917i, 0.4166 + 0.1739i), \\ \boldsymbol{\lambda}_5 &= (-0.3153 - 0.0917i, 0.4166 - 0.1739i). \end{aligned}$$

and the basis vectors $\mathbf{s}_k = \begin{bmatrix}1 & \boldsymbol{\lambda}_k^\top\end{bmatrix}^\top$. The error of the decomposition algorithm (between the original and decomposed polynomial) is $\approx 10^{-7}$.

For matrix completion in this example we chose the following method. Consider an affine matrix structure $\mathscr{S} : \mathbb{R}^{n_p} \to \mathbb{R}^{m \times n}$, $\mathscr{S}(p) = S_0 + \sum_{i=1}^{n_p} S_i p_i$, which parametrizes all possible completions (i.e., S_0 contains all known elements, and S_i and p_i parametrize the elements to be completed). For a given r, we minimize the cost function

$$\min_{P \in \mathbb{R}^{m\times r},\, L \in \mathbb{R}^{r \times n}} \|PL - \Pi_{\mathscr{S}}(PL)\|_F^2,$$

where $\Pi_{\mathscr{S}} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is the orthogonal projection on the space of structured matrices $\{\mathscr{S}(p) \,|\, p \in \mathbb{R}^{n_p}\}$. Thus we deal with minimization of the distance from the set of low-rank matrices to the space of structured matrices. (Also, the described method can be considered as a special case of the method in [27].)

There exist other methods for matrix completion. A probably earliest algorithm of matrix completion [28], [29], is based on Cadzow-like iterations, and alternately projects on the sets of low-rank matrices and the space of matrices with given entries. The Cadzow-like algorithm is attractive due to simplicity of its steps, which allow fast implementation. Another alternative proposed recently is to minimize the nuclear norm instead of the rank [30]. We should note, however, that the so-called first-order methods for the nuclear norm heuristic are very similar to the Cadzow-like methods.

Note 13: All the methods described above do not guarantee convergence to a solution of the original rank minimization problem.

V. CONCLUDING REMARKS

We have considered a problem of decomposing a polynomial as a sum of univariate polynomials of linear forms. We proposed numerical algorithms, which are extensions of existing algorithms for Waring decomposition. In the case of bivariate polynomials, the algorithm is an extension of the Sylvester algorithm.

The developed ideas can be also applied for the approximation problems. This would lead to a structured low-rank approximation (rather than just matrix completion). Although it is known that the best low-rank tensor approximation may not exist, the best structured low-rank approximation always exists (for a linear structure). However, in the structured low-rank approximation approach, the best approximation may be in the form of a generalized affine decomposition (and not a Waring/polynomial decomposition).

The problem that we considered was motivated by a polynomial decomposition in identification of parallel Wiener systems. A more challenging problem is simultaneous decomposition of several coupled polynomials, which appears in identification of the more flexible parallel Wiener-Hammerstein models [31] and nonlinear state-space models [32]. As is, the methods of this paper are not applicable to these cases, due to additional constraints on polynomials. It is an interesting open question whether these cases can be treated in the framework of this paper.

There are also many open theoretical questions on the polynomial decomposition considered in this paper, such as uniqueness and generic number of terms. There are a few works on simultaneous Waring decompositions, for example [33], but mostly homogeneous polynomials of the same degree were considered.

ACKNOWLEDGMENT

I would like to thank my colleagues from the Department ELEC, especially Laurent Vanbeylen, for the discussions that triggered my interest in this topic. I also would like to thank Pierre Comon and anonymous reviewers for very valuable comments.

REFERENCES

[1] M. Schoukens and Y. Rolain, "Cross-term elimination in parallel Wiener systems using a linear input transformation," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 3, pp. 845–847, 2012.

[2] V. Lin and A. Pinkus, "Fundamentality of ridge functions," *Journal of Approximation Theory*, vol. 75, no. 3, pp. 295 – 311, 1993.

[3] K. I. Oskolkov, "On representations of algebraic polynomials as a sum of plane waves," *Serdica Mathematical Journal*, pp. 379–390, 2002.

[4] D. Donoho and I. M. Johnstone, "Projection-based approximation and a duality with kernel methods," *The Annals of Statistics*, vol. 17, no. 1, pp. 58–106, 1989.

[5] A. Iarobbino and V. Kanev, *Power sums, Gorenstein Algebras and Determinantal Loci*, ser. Lecture Notes in Mathematics. Springer, 1999, vol. 1721.

[6] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, "Symmetric tensors and symmetric tensor rank," *SIAM. J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1254–1279, 2008.

[7] J. M. Landsberg, *Tensors: Geometry and applications*. American Mathematical Soc., 2012, vol. 128.

[8] J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas, "Symmetric tensor decomposition," *Linear Algebra and its Applications*, vol. 433, no. 11–12, pp. 1851–1872, 2010.

[9] L. Oeding and G. Ottaviani, "Eigenvectors of tensors and algorithms for waring decomposition," *Journal of Symbolic Computation*, vol. 54, pp. 9–35, 2013.

[10] A. Schinzel, "On a decomposition of polynomials in several variables," *Journal de Théorie de Nombres de Bordeaux*, vol. 14, no. 2, pp. 647–666, 2002.

[11] A. Białynicki-Birula and A. Schinzel, "Representations of multivariate polynomials as sums of polynomials in linear forms," *Colloquium Mathematicum*, vol. 112, no. 2, pp. 201–233, 2008.

[12] L. De Lathauwer, P. Comon, B. De Moor, and J. Vandewalle, "ICA algorithms for 3 sources and 2 sensors," in *Higher-Order Statistics, 1999. Proceedings of the IEEE Signal Processing Workshop on*, 1999, pp. 116–120.

[13] D. Cox, J. Little, and D. O'Shea, *Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 2nd ed. Springer, 1997.

[14] V. Kurakin, A. Kuzmin, A. Mikhalev, and A. Nechaev, "Linear recurring sequences over rings and modules," *Journal of Mathematical Sciences*, vol. 76, no. 6, pp. 2793–2915, 1995.

[15] V. Kurakin, "Linear complexity of polylinear sequences," *Discrete Math. Appl.*, vol. 11, no. 1, pp. 1–51, 2001.

[16] N. Golyandina and K. Usevich, "An algebraic view on finite rank in 2D-SSA," in *Proceedings of the 6th St. Petersburg Workshop on Simulation*, S. Ermakov, V. Melas, and A. Pepelyshev, Eds., vol. 1. VVM com. Ltd., July 2009, pp. 308–313.

[17] U. Oberst, "Variations on the fundamental principle for linear systems of partial differential and difference equations with constant coefficients," *Applicable Algebra in Engineering, Communication and Computing*, vol. 6, no. 4-5, pp. 211–243, 1995.

[18] ——, "Finite dimensional systems of partial differential or difference equations," *Advances in Applied Mathematics*, vol. 17, no. 3, pp. 337 – 356, 1996.

[19] J. Alexander and A. Hirschowitz, "Polynomial interpolation in several variables," *Journal of Algebraic Geometry*, vol. 4, pp. 201–222, 1995.

[20] B. Mourrain and V. Y. Pan, "Multivariate polynomials, duality, and structured matrices," *Journal of complexity*, vol. 16, no. 1, pp. 110–180, 2000.

[21] B. Ho, "An effective construction of realizations from input/output descriptions," Ph.D. dissertation, Stanford University, 1966.

[22] G. Heinig and K. Rost, *Algebraic methods for Toeplitz-like matrices and operators*. Birkhäuser, Boston, 1984.

[23] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications*, ser. Communications and Control Engineering. Springer, 2012.

[24] S. Rouquette and M. Najim, "Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 237–245, 2001.

[25] A. Shlemov and N. Golyandina, "Shaped extension of singular spectrum analysis," in *Proceedings of the 21st International Symposium on Mathematical Theory of Networks and Systems (MTNS 2014)*, July 7-11, 2014, Groningen, The Netherlands, 2014.

[26] E. Zerz, *Topics in multidimensional linear systems theory*, ser. Springer Lecture Notes in Control and Information Sciences. Springer London, 2000, vol. 256.

[27] M. Ishteva, K. Usevich, and I. Markovsky, "Regularized structured low-rank approximation," 2013, submitted to *SIAM Journal on Matrix Analysis and Applications*. Available from: http://arxiv.org/abs/1308.1827.

[28] D. Kondrashov and M. Ghil, "Spatio-temporal filling of missing points in geophysical data sets," *Nonlinear Processes in Geophysics*, vol. 13, no. 2, pp. 151–159, 2006.

[29] P. J. Shin, P. E. Z. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, and M. Lustig, "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion," *Magnetic Resonance in Medicine*, 2013. [Online]. Available: http://dx.doi.org/10.1002/mrm.24997

[30] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Elec. Eng. Dept, Stanford University, 2002.

[31] K. Tiels and J. Schoukens, "From coupled to decoupled polynomial representations in parallel Wiener-Hammerstein models," in *Proceedings of the 52nd IEEE Conference on Decision and Control*, Florence, Italy, December 10-13, 2013, pp. 4937–4942.

[32] A. Van Mulders, L. Vanbeylen, and K. Usevich, "Identification of a block-structured model with several sources of nonlinearity," in *Proceedings of the 14th European Control Conference (ECC 2014)*, 2014.

[33] E. Carlini and J. V. Chipalkatti, "On Waring's problem for several algebraic forms," *Commentarii Mathematici Helvetici*, vol. 78, no. 3, pp. 494–517, 2003.