



**HAL**  
open science

# Decoupling multivariate polynomials: Interconnections between tensorizations

Konstantin Usevich, Philippe Dreesen, Mariya Ishteva

► **To cite this version:**

Konstantin Usevich, Philippe Dreesen, Mariya Ishteva. Decoupling multivariate polynomials: Interconnections between tensorizations. *Journal of Computational and Applied Mathematics*, 2020, 363, pp.22-34. 10.1016/j.cam.2019.03.036 . hal-02994705

**HAL Id: hal-02994705**

**<https://hal.science/hal-02994705v1>**

Submitted on 25 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



25 In this paper, we focus on the task of decoupling a set of polynomial vector functions,  
26 that is, decomposing a set of multivariate real polynomials into linear combinations of  
27 univariate polynomials in linear forms of the input variables. This task has attracted a  
28 spark of research attention over the last years, motivated by several applications, such as  
29 system identification [9, 10, 11, 12, 13, 14], approximation theory [15, 16, 17], and neural  
30 networks [18]. Restricting polynomial decoupling to a single homogeneous polynomial is  
31 equivalent to the well-known Waring decomposition [19, 20], but some generalizations to  
32 non-homogeneous polynomials or joint Waring decompositions are studied as well [21, 22]  
33 and [23, 24, 11].

34 Several tensor-based approaches were proposed for computing a decoupled represen-  
35 tation of a given function [12, 13, 25, 26, 24]. These solutions can be categorized into  
36 two classes. The methods [12, 13, 25, 26] build a tensor from the polynomial coefficients,  
37 whereas the method of [24] builds a tensor from the Jacobian matrices of the functions,  
38 evaluated at a set of sampling points. Ultimately, all methods boil down to a canonical  
39 polyadic decomposition (CP decomposition) of the constructed tensor to retrieve  
40 a decoupled representation in which the nonlinearities occur as univariate polynomial  
41 mappings.

42 The benefit of using a tensor-based approach for decoupling is twofold. First, ‘ten-  
43 sorization’ procedures often lead to (essentially) uniquely decomposable tensors [8], i.e.,  
44 ensuring that identifiable structures can be retrieved. Second, by solving the decoupling  
45 problem as a CP decomposition, one can use recent widely available and robust numerical  
46 tools, such as Tensorlab for MATLAB [27] (or alternatives [28, 29]).

47 This paper specifically focuses on the two tensorization methods [26] and [24]. Al-  
48 though both associated tensors have a particular structure, both approaches seem quite  
49 different in nature, and each of the methods has distinct advantages over the other one.  
50 For instance, the coefficient-based methods [12, 13, 25, 26] require several high-order  
51 tensors (or their matricizations) for polynomials of high degrees, whereas [24] involves a  
52 single third-order tensor only. Coefficient-based approaches can easily deal with single  
53 polynomials, whereas [24] would in that case not be able to take advantage of the unique-  
54 ness properties of the CP decomposition, as the *tensor* of Jacobian matrices is then a  
55 matrix composed of gradient vectors. On the other hand, the approach of [24] can be  
56 applied to non-polynomial functions, which may in some cases be of interest, e.g., in [9]  
57 a neural network was decoupled.

58 We aim at obtaining a deeper understanding of the connections between the solution  
59 approaches. This is profitable when extending the applicability range of the methods,  
60 e.g., when moving from polynomials to any differentiable functions. Furthermore, such  
61 connections may provide a way to transfer theoretical properties from one formulation to  
62 another. For example, as we argue in Section 6, exploring the previously ignored structure  
63 in the tensor decomposition in one of the settings enlarges the range of decomposable  
64 functions. This knowledge may lead to improved algorithms in another setting as well.

65 The remainder of this article is organized as follows: Section 2 formalizes the problem  
66 of decoupling multivariate polynomials. Section 3 explains the link between the decou-  
67 pling problem and the symmetric tensor decomposition problem. Section 4 discusses  
68 the construction of the tensor of unfoldings [26] and the Jacobian tensor [24]. Section 5  
69 presents our first contribution, namely the relation between the two tensorizations. The  
70 second main contribution of the paper is Section 6, which clarifies the need of dealing  
71 with structure in the decompositions and proposes a coupled CP decomposition approach

72 for solving the structured problem. Section 7 draws the conclusions and points out open  
73 problems for future work.

#### 74 Notation

75 Scalars are denoted by lowercase or uppercase letters. Vectors are denoted by low-  
76 ercase boldface letters, e.g.,  $\mathbf{u}$ . Elements of a vector are denoted by lowercase letters  
77 with an index as subscript, e.g.,  $\mathbf{x} = [x_1 \cdots x_m]^\top$ . Matrices are denoted by up-  
78 percase boldface letters, e.g.,  $\mathbf{V}$ . The entry in the  $i$ -th row and  $j$ -th column of a matrix  
79  $\mathbf{V}$  is denoted by  $v_{ij}$ , and the matrix  $\mathbf{V} \in \mathbb{R}^{m \times r}$  may be represented by its columns  
80  $\mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_r]$ . The Kronecker product of matrices is denoted by “ $\otimes$ ”.

81 Tensors of order  $d$  are denoted by uppercase calligraphical letters, e.g.,  $\mathcal{J} \in \mathbb{R}^{n \times m \times N}$ .  
82 The outer product is denoted by “ $\circ$ ” and is defined as follows: For  $\mathcal{T} = \mathbf{u} \circ \mathbf{v} \circ \mathbf{w}$ , the  
83 entry in position  $(i, j, k)$  is equal to  $u_i v_j w_k$ . The canonical polyadic (CP) decomposition  
84 expresses a tensor  $\mathcal{T}$  as a (minimal) sum of rank-one tensor terms [30, 31, 2] as  $\mathcal{T} =$   
85  $\sum_{i=1}^R \mathbf{u}_i \circ \mathbf{v}_i \circ \mathbf{w}_i$ , and is sometimes denoted in a short-hand notation as  $\mathcal{T} = \llbracket \mathbf{U}, \mathbf{V}, \mathbf{W} \rrbracket$ .  
86 The CP rank  $r$  is defined as the (minimal) number of terms that is required to represent  
87  $\mathcal{T}$  as a sum of  $r$  rank-one terms. To refer to elements of matrices or tensors, or subsets  
88 thereof, we may use MATLAB-like index notation (including MATLAB’s colon wildcard):  
89 for instance,  $\mathcal{T}_{i,j,k,\ell}$  is the element at position  $(i, j, k, \ell)$  of a fourth-order tensor  $\mathcal{T}$ , and  
90  $\mathcal{T}_{:, :, 2}$  is the second frontal slice of a third-order tensor  $\mathcal{T}$ . The mode- $n$  product is denoted  
91 by “ $\bullet_n$ ” and is defined as follows. Let  $\mathcal{X}$  be an  $I_1 \times I_2 \times \cdots \times I_N$  tensor, and let  $\mathbf{u}$  be a  
92 vector of length  $I_n$ , then we have  $(\mathcal{X} \bullet_n \mathbf{u}^\top)_{i_1 \cdots i_{n-1} i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \cdots i_N} u_{i_n}$ . Notice  
93 that the result is a tensor of order  $N - 1$ , as mode  $n$  is summed out. Similarly, for an  
94  $I_1 \times I_2 \times \cdots \times I_N$  tensor  $\mathcal{X}$  and a matrix  $\mathbf{M} \in J \times I_n$ , the mode- $n$  product is defined  
95 as  $(\mathcal{X} \bullet_n \mathbf{M}^\top)_{i_1 \cdots i_{n-1} i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \cdots i_N} m_{j, i_n}$ . Let  $\text{vec}(\mathcal{T})$  denote the column-  
96 major vectorization of a tensor  $\mathcal{T}$ . The first-mode unfolding of an  $I_1 \times I_2 \times \cdots \times I_N$   
97 tensor  $\mathcal{X}$  is the matrix  $\mathcal{X}_{(1)}$  of size  $I_1 \times I_2 \cdots I_N$ , where each row is the vectorized slice  
98 of the tensor  $\mathcal{X}$ , i.e.  $(\mathcal{X}_{(1)})_{i,:} = (\text{vec}(\mathcal{X}_{i, :, \dots, :}))^\top$  (see, for example, [2] for more details).

## 99 2. The polynomial decoupling model

100 First, we describe the model, following the notation of [24] as illustrated in Fig. 1.  
101 Consider a multivariate polynomial map  $\mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ , i.e., a vector

$$\mathbf{f}(\mathbf{u}) = [f_1(\mathbf{u}) \cdots f_n(\mathbf{u})]^\top$$

102 of multivariate polynomials (of total degree at most  $d$ ) in variables  $\mathbf{u} = [u_1 \cdots u_m]^\top$ .  
103 We say that  $\mathbf{f}$  has a *decoupled representation*, if it can be expressed as

$$\mathbf{f}(\mathbf{u}) = \mathbf{W} \mathbf{g}(\mathbf{V}^\top \mathbf{u}), \quad (1)$$

104 where  $\mathbf{V} \in \mathbb{R}^{m \times r}$ ,  $\mathbf{W} \in \mathbb{R}^{n \times r}$  are transformation matrices, and  $\mathbf{g} : \mathbb{R}^r \rightarrow \mathbb{R}^r$  is defined as

$$\mathbf{g}(x_1, \dots, x_r) = [g_1(x_1) \cdots g_r(x_r)]^\top,$$

105 where  $g_k : \mathbb{R} \rightarrow \mathbb{R}$  are univariate polynomials of degree at most  $d$ , i.e.,

$$g_k(t) = c_1 t + \cdots + c_d t^d. \quad (2)$$

106 Note that we omitted the constant terms of the polynomials, since they are not uniquely  
 107 identifiable [24]. In this paper we limit ourselves to the model (1).

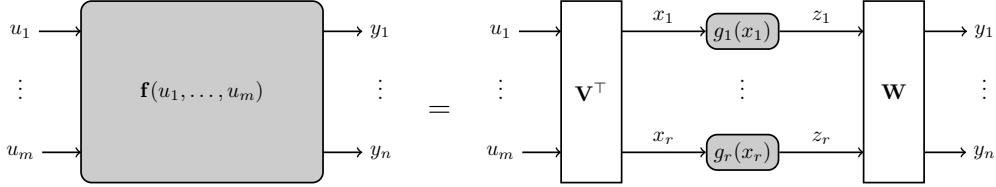


Figure 1: Every multivariate polynomial vector function  $\mathbf{f}(\mathbf{u})$  can be represented by a linear transformation of a set of univariate functions (in linear combinations of the original variables).

107  
 108

The decoupled representation (1) can be also equivalently rewritten as

$$\mathbf{f}(\mathbf{u}) = \mathbf{w}_1 g_1(\mathbf{v}_1^\top \mathbf{u}) + \cdots + \mathbf{w}_r g_r(\mathbf{v}_r^\top \mathbf{u}), \quad (3)$$

109 where  $\mathbf{v}_k$  and  $\mathbf{w}_k$  are the columns of  $\mathbf{V}$  and  $\mathbf{W}$ , respectively. As shown in [32, 33], the  
 110 decomposition (3) is a special case of the  $X$ -rank decomposition [34, §5.2.1], where the  
 111 set of “rank-one” terms is the set of polynomial maps of the form  $\mathbf{w}g(\mathbf{v}^\top \mathbf{u})$ . The  $X$ -rank  
 112 framework is useful [33] for studying the identifiability of the model (3).

113 The following example shows a decoupled representation for a simple case. This  
 114 example will be used throughout the paper to illustrate the main ideas of the various  
 115 aspects that we will explore.

116 **Example 1.** Consider a function  $\mathbf{f}(\mathbf{u}) = [f_1(u_1, u_2) \quad f_2(u_1, u_2)]^\top$  given as

$$\begin{aligned} f_1(u_1, u_2) &= -3u_1^3 - 9u_1^2 u_2 - 27u_1 u_2^2 - 15u_2^3 - 8u_1^2 - 8u_1 u_2 - 20u_2^2 + 3u_1 + 9u_2, \\ f_2(u_1, u_2) &= -7u_1^3 - 6u_1^2 u_2 + 6u_1 u_2^2 + 7u_2^3 + 10u_1^2 + 16u_1 u_2 + 10u_2^2 - 3u_2. \end{aligned}$$

117 It can be verified that  $\mathbf{f}$  has a decomposition (3) with  $m = n = 2$  and  $r = 3$  as

$$\mathbf{V} = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \quad \text{and} \quad \mathbf{W} = \begin{bmatrix} 0 & 1 & -2 \\ -1 & 0 & 1 \end{bmatrix},$$

118 and  $g_1(x_1) = x_1^3 - 2x_1^2 - x_1$ ,  $g_2(x_2) = x_2^3 - 4x_2^2 + x_2$ ,  $g_3(x_3) = x_3^3 + 2x_3^2 - 2x_3$  (see  
 119 Figure 2).

120 **Remark 1.** In general, the coupled representation  $\mathbf{f}(\mathbf{u})$  has  $n \binom{m+d}{d} - 1$  coefficients,  
 121 while the decoupled representation  $\mathbf{W}\mathbf{g}(\mathbf{V}^\top \mathbf{u})$  has  $r(m+n+d)$  coefficients. Due to the  
 122 combinatorial increase of the number of coefficients in the coupled representation, the  
 123 decoupled representation is especially beneficial for large values of  $m$ ,  $n$ , and  $d$ . But even  
 124 for small values of  $m$ ,  $n$ , and  $d$ , the parametric reduction can be significant, for example,  
 125 if  $m = n = 3$ ,  $d = 5$ , and  $r = 3$ , the coupled representation has 168 coefficients, while  
 126 the decoupled one has only 36 coefficients.

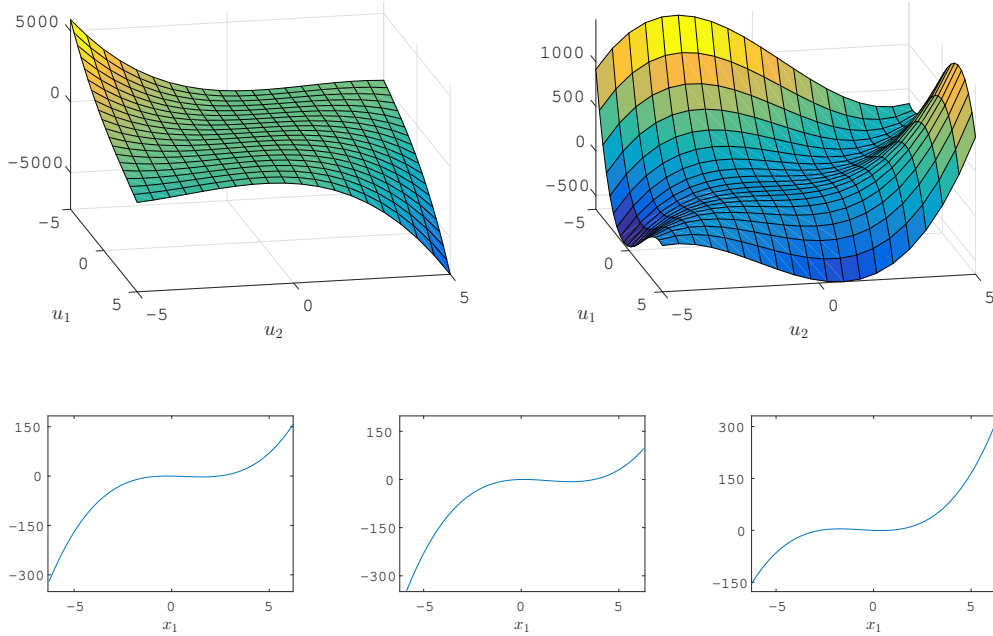


Figure 2: The decoupling problem in Example 1 consists of decomposing the multivariate functions  $f_1(u_1, u_2)$  and  $f_2(u_1, u_2)$  (top row) to the univariate functions  $g_1(x_1)$ ,  $g_2(x_2)$ , and  $g_3(x_3)$  (bottom row), using suitable transformation matrices as in (1).

### 127 3. Decoupling polynomials and symmetric tensor decompositions

128 Let us review some well-known facts that connect polynomials with symmetric tensors  
 129 [35, 36], and that connect some special cases of the representation (1) with symmetric  
 130 tensor decompositions.

#### 131 3.1. Homogeneous polynomials, symmetric tensors and Waring decomposition

132 It is well-known that there is a one-to-one correspondence between homogeneous  
 133 polynomials and symmetric tensors [35]. For instance, the polynomial  $-8u_1^2 - 8u_1u_2 -$   
 134  $20u_2^2$  can be written as

$$-8u_1^2 - 8u_1u_2 - 20u_2^2 = \mathbf{u}^\top \Psi^{(2)} \mathbf{u}, \quad \text{where } \Psi^{(2)} = \begin{bmatrix} -8 & -4 \\ -4 & -20 \end{bmatrix}. \quad (4)$$

135 In general, let  $p^{(d)}(u_1, \dots, u_m)$  be a homogeneous polynomial (also called a  $d$ -ary form)  
 136 of degree  $d$  in  $m$  variables. Then there is a unique symmetric tensor  $\Psi^{(d)}$  of order  $d$  and  
 137 dimension  $m$  such that

$$p(\mathbf{u}) = \Psi^{(d)} \bullet_1 \mathbf{u} \cdots \bullet_d \mathbf{u}. \quad (5)$$

138 Next, it is easy to see that the decoupling problem for the polynomial (5) takes the form

$$p(u_1, \dots, u_m) = \sum_{i=1}^r w_i (v_{1i}u_1 + \cdots + v_{mi}u_m)^d, \quad (6)$$

139 which is known as the Waring decomposition [19, 20] of  $p(u_1, \dots, u_m)$ . The Waring  
 140 decomposition, in its turn, is equivalent to the symmetric CP decomposition of  $\Psi^{(d)}$ :

$$\Psi^{(d)} = \sum_{i=1}^r w_i (\mathbf{v}_i \circ \dots \circ \mathbf{v}_i).$$

141 The symmetric CP decomposition of  $\Psi^{(d)}$  reveals possible values for the unknowns  $v_{ij}$   
 142 and  $w_i$ .

143 **Example 2.** Consider the polynomial given in (4). Then the corresponding symmetric  
 144 matrix  $\Psi^{(2)}$  admits the decomposition

$$\begin{bmatrix} -8 & -4 \\ -4 & -20 \end{bmatrix} = \begin{bmatrix} -2 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ 2 & 4 \end{bmatrix}, \quad (7)$$

145 such that  $p(u_1, u_2) = \mathbf{u}^\top \Psi^{(2)} \mathbf{u}$  has the Waring decomposition

$$p(u_1, u_2) = -(-2u_1 + 2u_2)^2 - (2u_1 + 4u_2)^2.$$

146 Notice that the symmetric decomposition of  $\Psi^{(2)}$  from Example 2 is not unique (nor  
 147 ‘essentially unique’ [2]). Indeed, the eigenvalue decomposition

$$\begin{bmatrix} -8 & -4 \\ -4 & -20 \end{bmatrix} \approx \begin{bmatrix} 0.2898 & -0.9571 \\ 0.9571 & 0.2898 \end{bmatrix} \begin{bmatrix} -21.2111 & 0 \\ 0 & -6.7889 \end{bmatrix} \begin{bmatrix} 0.2898 & 0.9571 \\ -0.9571 & 0.2898 \end{bmatrix}$$

148 provides another valid factorization. For  $d > 2$ , however, the Waring decomposition (6)  
 149 possesses uniqueness properties even in the case of quite large ranks [37, 38].

150 Along the same lines, it is possible to decouple jointly several homogeneous polyno-  
 151 mials. Consider the case of  $n$  homogeneous polynomials of degree  $d$ , denoted by

$$\begin{aligned} p_1(u_1, \dots, u_m) &= \Psi_1^{(d)} \bullet_1 \mathbf{u} \cdots \bullet_d \mathbf{u}, \\ &\vdots \\ p_n(u_1, \dots, u_m) &= \Psi_n^{(d)} \bullet_1 \mathbf{u} \cdots \bullet_d \mathbf{u}. \end{aligned} \quad (8)$$

152 Then the decoupling problem (1) corresponds to the simultaneous Waring decomposition  
 153 of several forms or, equivalently, the coupled CP decomposition of several symmetric  
 154 tensors. The rank and identifiability properties of simultaneous Waring decompositions  
 155 were also studied in the literature, see [23, 37, 39] and references therein.

### 156 3.2. The case of non-homogeneous polynomials

157 Next, consider the case of non-homogeneous polynomials. Any non-homogeneous  
 158 polynomial of degree  $d$  can hence be written as

$$p(\mathbf{u}) = \mathbf{u}^\top \Psi^{(1)} + \mathbf{u}^\top \Psi^{(2)} \mathbf{u} + \Psi^{(3)} \bullet_1 \mathbf{u} \bullet_2 \mathbf{u} \bullet_3 \mathbf{u} + \dots + \Psi^{(d)} \bullet_1 \mathbf{u} \cdots \bullet_d \mathbf{u}, \quad (9)$$

159 where  $\Psi^{(1)} \in \mathbb{R}^m$ ,  $\Psi^{(2)} \in \mathbb{R}^{m \times m}$  is a symmetric matrix, and each  $\Psi^{(s)} \in \mathbb{R}^{m \times \dots \times m}$ ,  
 160  $3 \leq s \leq d$ , is a symmetric tensor of order  $s$ .

161 **Example 3.** We continue Example 1. We can write  $f_1(u_1, u_2)$  and  $f_2(u_1, u_2)$  as

$$f_1(u_1, u_2) = \mathbf{u}^\top \begin{bmatrix} 3 \\ 9 \end{bmatrix} + \mathbf{u}^\top \begin{bmatrix} -8 & -4 \\ -4 & -20 \end{bmatrix} \mathbf{u} + \Psi^{(3)} \bullet_1 \mathbf{u} \bullet_2 \mathbf{u} \bullet_3 \mathbf{u},$$

162 with

$$\Psi_{::,1}^{(3)} = \begin{bmatrix} -3 & -3 \\ -3 & -9 \end{bmatrix}, \quad \Psi_{::,2}^{(3)} = \begin{bmatrix} -3 & -9 \\ -9 & -15 \end{bmatrix},$$

163 and

$$f_2(u_1, u_2) = \mathbf{u}^\top \begin{bmatrix} 0 \\ -3 \end{bmatrix} + \mathbf{u}^\top \begin{bmatrix} 10 & 8 \\ 8 & 10 \end{bmatrix} \mathbf{u} + \Psi^{(3)} \bullet_1 \mathbf{u} \bullet_2 \mathbf{u} \bullet_3 \mathbf{u},$$

164 with

$$\Psi_{::,1}^{(3)} = \begin{bmatrix} -7 & -2 \\ -2 & 2 \end{bmatrix}, \quad \Psi_{::,2}^{(3)} = \begin{bmatrix} -2 & 2 \\ 2 & 7 \end{bmatrix}.$$

165 The decomposition of a single non-homogeneous polynomial as in (3) is hence equiv-  
166 alent to joint decomposition of several symmetric tensors but of different orders [32].

167 Finally, several non-homogeneous polynomials can be jointly decomposed in a similar  
168 way. Consider  $n$  non-homogeneous polynomials of maximal degree  $d$ , denoted as

$$\begin{aligned} p_1(u_1, \dots, u_m) &= \mathbf{u}^\top \Psi_1^{(1)} + \dots + \Psi_1^{(d)} \bullet_1 \mathbf{u} \cdots \bullet_d \mathbf{u}, \\ &\vdots \\ p_n(u_1, \dots, u_m) &= \mathbf{u}^\top \Psi_n^{(1)} + \dots + \Psi_n^{(d)} \bullet_1 \mathbf{u} \cdots \bullet_d \mathbf{u}, \end{aligned} \tag{10}$$

169 The full decomposition in (1) can be also viewed as a coupled tensor decomposition,  
170 which will be presented in Section 6.2.

#### 171 4. Tensorizations and their decompositions

172 In this section, we recall tensorizations proposed in the literature to find the decom-  
173 position (1) by a CP decomposition of a single tensor constructed from  $\mathbf{f}$ , namely the  
174 tensorizations of [26] and [24]. We recall basic properties and give short proofs for com-  
175 pleteness, although these proofs are already present in [26, 24]. We also use a slightly  
176 different notation to simplify the exposition.

##### 177 4.1. Tensor of unfoldings [26]

178 The above link between polynomials, (partially) symmetric tensors and their CP  
179 decompositions gives rise to the tensorization approach of [26], in which a tensor is  
180 constructed from the coefficients of the polynomials  $f_1(u_1, \dots, u_m)$  up to  $f_n(u_1, \dots, u_m)$ .  
181 This tensorization offers the advantage that several polynomials can be represented as a  
182 single tensor, and the decoupling task can be solved using a single (but structured) CP  
183 decomposition. In this approach, the tensor (shown in Figure 3) is constructed from the  
184 coefficients of the polynomial map of degree  $d$ , as follows:

- 185 • The tensor has size  $n \times m \times \delta$ , where  $\delta = \sum_{k=1}^d m^{k-1}$ .



- 186 • The tensor is constructed by slices

$$\mathcal{Q}_{i,\dots} := \Psi(f_i),$$

187 where  $\Psi$  is a structured  $m \times \delta$  matrix built from the coefficients of  $f_i(\mathbf{u})$ .

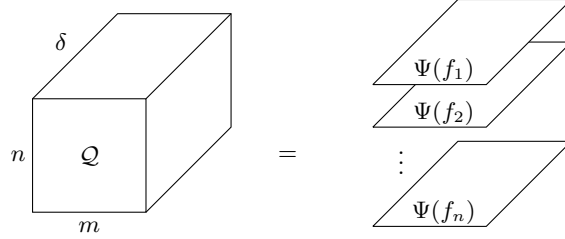


Figure 3: The coefficients of a polynomial map  $\mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^n$  of degree  $d$  can be arranged into an  $n \times m \times \delta$  tensor  $\mathcal{Q}$ , where  $\delta = \sum_{k=1}^d m^{k-1}$ .

188 Now let us describe the construction of the structured coefficient matrix  $\Psi(p)$  for a  
 189 given polynomial of degree  $d$ . Recall that each such polynomial can be written as in  
 190 (9), where  $\Psi^{(1)} \in \mathbb{R}^m$ ,  $\Psi^{(2)} \in \mathbb{R}^{m \times m}$  is a symmetric matrix and  $\Psi^{(s)} \in \mathbb{R}^{m \times \dots \times m}$  are  
 191 symmetric tensors of order  $s$ . Then the matrix  $\Psi(p) \in \mathbb{R}^{m \times \delta}$  is constructed<sup>1</sup> as

$$\Psi(p) = \left[ \Psi^{(1)} \mid \Psi^{(2)} \mid \Psi^{(3)} \mid \dots \mid \Psi^{(d)} \right], \quad (11)$$

192 where  $\mathcal{G}_{(1)}$  denotes the first-mode unfolding of a tensor  $\mathcal{G}$ .

193 **Example 4.** A third-degree polynomial in two variables

$$p(u_1, u_2) = a_1 u_1 + a_2 u_2 + b_1 u_1^2 + 2b_2 u_1 u_2 + b_3 u_2^2 + d_1 u_1^3 + 3d_2 u_1^2 u_2 + 3d_3 u_1 u_2^2 + d_4 u_2^3$$

194 has the representation

$$p(u_1, u_2) = \mathbf{u}^\top \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \mathbf{u}^\top \begin{bmatrix} b_1 & b_2 \\ b_2 & b_3 \end{bmatrix} \mathbf{u} + \Psi^{(3)} \bullet_1 \mathbf{u} \bullet_2 \mathbf{u} \bullet_3 \mathbf{u}, \quad (12)$$

195 where

$$\Psi^{(3)}_{\dots,1} = \begin{bmatrix} d_1 & d_2 \\ d_2 & d_3 \end{bmatrix}, \quad \Psi^{(3)}_{\dots,2} = \begin{bmatrix} d_2 & d_3 \\ d_3 & d_4 \end{bmatrix}.$$

196 By putting all the unfoldings together, we get

$$\Psi(p) = \left[ \begin{array}{c|cc|cccc} a_1 & b_1 & b_2 & d_1 & d_2 & d_2 & d_3 \\ a_2 & b_2 & b_3 & d_2 & d_3 & d_3 & d_4 \end{array} \right]. \quad (13)$$

197 Hence, for  $f_1$  and  $f_2$  in Example 1, the slices of the tensor  $\mathcal{Q}$  are given by

$$\mathcal{Q}_{1,\dots} = \Psi(f_1) = \left[ \begin{array}{c|cc|cccc} 3 & -8 & -4 & -3 & -3 & -3 & -9 \\ 9 & -4 & -20 & -3 & -9 & -9 & -15 \end{array} \right],$$

<sup>1</sup>In [26] the linear term is skipped, and  $\delta = \sum_{k=2}^d m^{d-1}$ . In [40] the matrix  $\Psi$  is denoted as  $\Gamma$ .

198 and

$$\mathcal{Q}_{2,,:,} = \Psi(f_2) = \left[ \begin{array}{c|cc|ccc} 0 & 10 & 8 & -7 & -2 & -2 & 2 \\ -3 & 8 & 10 & -2 & 2 & 2 & 7 \end{array} \right].$$

199 As proved in [26], the tensor  $\mathcal{Q}$  has a CP decomposition, which reveals the decompo-  
200 sition (1). We repeat here a simplified version of the proof for completeness.

201 **Lemma 1.** *For the polynomial map (1), the tensor  $\mathcal{Q}$  has the following CP decomposi-*  
202 *tion:*

$$\mathcal{Q} = \sum_{k=1}^r \mathbf{w}_k \circ \mathbf{v}_k \circ \mathbf{z}_k, \quad (14)$$

203 where

$$\mathbf{z}_k = \left[ \begin{array}{c|c|c|c|c} c_{k,1} & c_{k,2}\mathbf{v}_k^\top & c_{k,3}(\mathbf{v}_k \otimes \mathbf{v}_k)^\top & \cdots & c_{k,d}(\mathbf{v}_k \otimes \cdots \otimes \mathbf{v}_k)^\top \end{array} \right]^\top. \quad (15)$$

204 *Proof.* Consider  $q_k(\mathbf{u}) := g_k(\mathbf{v}_k^\top \mathbf{u})$ , where  $g_k$  is as in (2). Easy calculations show that

$$\Psi(q_k) = \mathbf{v}_k \mathbf{z}_k^\top,$$

205 see also [40, eqn. (A.7)]. Since, from (3)  $f_i(\mathbf{u}) = \sum_{k=1}^r (\mathbf{w}_k)_i q_k(\mathbf{u})$ , we have that

$$\Psi(f_i) = \sum_{k=1}^r (\mathbf{w}_k)_i \mathbf{v}_k \mathbf{z}_k^\top$$

206 which implies (14). □

207 **Example 5.** *We continue Examples 1, 4. The Kronecker products of columns of  $\mathbf{V}$  are:*

$$(\mathbf{v}_1 \otimes \mathbf{v}_1)^\top = [4 \ 2 \ 2 \ 1], (\mathbf{v}_2 \otimes \mathbf{v}_2)^\top = [1 \ -1 \ -1 \ 1], (\mathbf{v}_3 \otimes \mathbf{v}_3)^\top = [1 \ 2 \ 2 \ 4].$$

208 Hence, the matrix  $\mathbf{Z} = [\mathbf{z}_1 \ \mathbf{z}_2 \ \mathbf{z}_3]$  is given by

$$\mathbf{Z}^\top = \left[ \begin{array}{c|cc|ccc} -1 & -4 & -2 & 4 & 2 & 2 & 1 \\ 1 & 4 & -4 & 1 & -1 & -1 & 1 \\ -2 & 2 & 4 & 1 & 2 & 2 & 4 \end{array} \right].$$

209 *4.2. The tensor of Jacobian matrices of [24]*

210 The tensorization method of [24] does not use the coefficients of  $\mathbf{f}(\mathbf{u})$  directly, but  
211 proceeds by collecting the first-order information of  $\mathbf{f}(\mathbf{u})$  (i.e., the partial derivatives)  
212 in a set of sampling points. The thusly obtained Jacobian matrices are arranged into a  
213 third-order tensor, of which the CP decomposition reveals the decomposition (1).

214 As in [24], we consider the Jacobian of  $\mathbf{f}$ :

$$\mathbf{J}_f(\mathbf{u}) := \begin{bmatrix} \frac{\partial f_1}{\partial u_1}(\mathbf{u}) & \cdots & \frac{\partial f_1}{\partial u_m}(\mathbf{u}) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial u_1}(\mathbf{u}) & \cdots & \frac{\partial f_n}{\partial u_m}(\mathbf{u}) \end{bmatrix}. \quad (16)$$

215 Using Lemma 2, the tensorization is constructed as follows (see Figure 4):

- 216 •  $N$  points  $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(N)} \in \mathbb{R}^m$  are chosen (so-called *sampling points*).
- 217 • An  $n \times m \times N$  tensor  $\mathcal{J}$  is constructed by stacking the Jacobian evaluations at  $\mathbf{u}^{(k)}$

$$\mathcal{J}_{::,k} := \mathbf{J}_{\mathbf{f}}(\mathbf{u}^{(k)}).$$

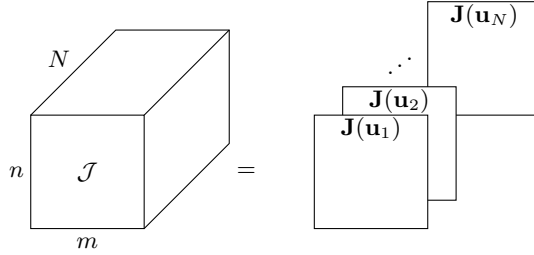


Figure 4: The third-order tensor  $\mathcal{J}$  is constructed by stacking behind each other a set of Jacobian matrices  $\mathbf{J}$  evaluated at the sampling points  $\mathbf{u}^{(k)}$ . Its CP decomposition is equivalent to joint matrix diagonalization of the Jacobian matrix slices.

218 **Example 6.** We continue Example 1. As a set of sampling points, we choose

$$\mathbf{u} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{u}^{(2)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{u}^{(3)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

219 By evaluating  $\mathbf{J}_{\mathbf{f}}(\mathbf{u})$  at these points, we get the tensor  $\mathcal{J}$  given by

$$\mathcal{J}_{::,1} = \begin{bmatrix} 3 & 9 \\ 0 & -3 \end{bmatrix}, \quad \mathcal{J}_{::,2} = \begin{bmatrix} -22 & -8 \\ -1 & 7 \end{bmatrix}, \quad \mathcal{J}_{::,3} = \begin{bmatrix} -32 & -76 \\ 22 & 38 \end{bmatrix}. \quad (17)$$

220 If  $\mathbf{f}(\mathbf{u})$  has a decoupled representation (1), the following lemma holds true.

221 **Lemma 2** ([24, Lemma 2.1]). The first order derivatives of (1) are given by

$$\mathbf{J}_{\mathbf{f}}(\mathbf{u}) = \mathbf{W} \text{diag}(g'_1(\mathbf{v}_1^\top \mathbf{u}), \dots, g'_r(\mathbf{v}_r^\top \mathbf{u})) \mathbf{V}^\top, \quad (18)$$

222 where  $g'_i(t) := \frac{dg_i}{dt}(t)$ .

223 The proof, given in [24], follows by chain rule:

$$\mathbf{J}_{\mathbf{f}}(\mathbf{u}) = \mathbf{W} \mathbf{J}_{\mathbf{g}}(\mathbf{V}^\top \mathbf{u}) \mathbf{V}^\top.$$

224 By Lemma 2, the evaluations of the Jacobians can be jointly factorized:

$$\begin{aligned} \mathbf{J}(\mathbf{u}^{(1)}) &= \mathbf{W} \mathbf{D}^{(1)} \mathbf{V}^\top, \\ &\vdots \\ \mathbf{J}(\mathbf{u}^{(N)}) &= \mathbf{W} \mathbf{D}^{(N)} \mathbf{V}^\top, \end{aligned} \quad (19)$$

225 where  $\mathbf{D}^{(k)} = \text{diag}(g'_1(\mathbf{v}_1^\top \mathbf{u}^{(k)}), \dots, g'_r(\mathbf{v}_r^\top \mathbf{u}^{(k)}))$ . Therefore,  $\mathcal{J}$  admits a CP decomposi-  
 226 tion

$$\mathcal{J} = \llbracket \mathbf{W}, \mathbf{V}, \mathbf{H} \rrbracket = \sum_{k=1}^r \mathbf{w}_k \circ \mathbf{v}_k \circ \mathbf{h}_k, \quad (20)$$

227 where  $\mathbf{w}_k, \mathbf{v}_k$  are as in (3), and  $\mathbf{h}_k$  contains the evaluations of  $g'_k(\mathbf{v}_k^\top \mathbf{u})$  in  $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(N)}$ :

$$\mathbf{h}_k = [g'_k(\mathbf{v}_k^\top \mathbf{u}^{(1)}) \quad \dots \quad g'_k(\mathbf{v}_k^\top \mathbf{u}^{(N)})]^\top. \quad (21)$$

228 **Example 7.** We continue Examples 1 and 6. By differentiation, we get

$$g'_1(t) = 3t^2 - 4t - 1, \quad g'_2(t) = 3t^2 - 8t + 1, \quad g'_3(t) = 3t^2 + 4t - 2,$$

229 and hence, by substitution,

$$\mathbf{H} = \begin{bmatrix} -1 & 1 & -2 \\ 3 & 12 & 5 \\ -2 & -4 & 18 \end{bmatrix}. \quad (22)$$

230 Straightforward calculations show indeed that  $\mathcal{J}$  given in (17) admits a decomposition  
 231 (20) with  $\mathbf{H}$  as in (22).

## 232 5. Relation between tensorizations $\mathcal{J}$ and $\mathcal{Q}$

233 In this section, we show how CP decompositions of (14) and (20) are related. More-  
 234 over, we establish the relation between the ranks of the tensors and uniqueness of CP  
 235 decompositions.

236 First, we show the relation between the vectors  $\mathbf{z}_k$  and  $\mathbf{h}_k$ , defined in (15) and (21),  
 237 respectively. We give the proof of this basic fact for completeness.

238 **Lemma 3.** The vectors  $\mathbf{z}_k$  and  $\mathbf{h}_k$  defined in (15) and (21), respectively, satisfy

$$\mathbf{h}_k = \mathbf{A}^\top \mathbf{z}_k, \quad (23)$$

239 where  $\mathbf{A} \in \mathbb{R}^{\delta \times N}$  is a Vandermonde-like matrix whose columns are

$$\mathbf{A}_{:,j} = [1 \mid 2(\mathbf{u}^{(j)})^\top \mid 3(\mathbf{u}^{(j)} \otimes \mathbf{u}^{(j)})^\top \mid \dots \mid d(\mathbf{u}^{(j)} \otimes \dots \otimes \mathbf{u}^{(j)})^\top]^\top. \quad (24)$$

240 *Proof.* Recall that by the properties of the Kronecker product

$$\underbrace{(\mathbf{u} \otimes \dots \otimes \mathbf{u})}_{d \text{ times}}^\top (\mathbf{v} \otimes \dots \otimes \mathbf{v}) = (\mathbf{u}^\top \mathbf{v})^d.$$

241 Then from (15) have that

$$(\mathbf{A}^\top \mathbf{z}_k)_j = \mathbf{A}_{:,j}^\top \mathbf{z}_k = c_{k,1} + c_{k,2}(\mathbf{v}_k^\top \mathbf{u}^{(j)}) + \dots + c_{k,d}(\mathbf{v}_k^\top \mathbf{u}^{(j)})^{d-1} = (\mathbf{h}_k)_j,$$

242 where the last equality follows from (21) and the fact that

$$g'_k(t) = c_{k,1} + c_{k,2}t + \dots + c_{k,d}t^{d-1}.$$

243

□

244 **Example 8.** In Example 6, the matrix  $\mathbf{A}$  can be found as

$$\mathbf{A}^\top = \left[ \begin{array}{c|cc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 3 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 & 3 \end{array} \right].$$

245 It is easy to see that  $\mathbf{H} = \mathbf{A}^\top \mathbf{Z}$ .

246 As a consequence, we get that the two tensors and their ranks are also related.

247 **Theorem 1.** 1. For any polynomial map  $\mathbf{f}$ ,  $\mathcal{J}$  and  $\mathcal{Q}$  are related as

$$\mathcal{J} = \mathcal{Q} \bullet_3 \mathbf{A}^\top. \quad (25)$$

248 2. The rank of  $\mathbf{A}$  is bounded as

$$\text{rank } \mathbf{A} \leq M := \binom{m+d-1}{d-1}.$$

249 In addition, if  $M \leq N$ , and  $M$  points in  $\{\mathbf{u}^{(j)}\}$  are in general position, then  
 250  $\text{rank } \mathbf{A} = M$ . For example, if points  $\{\mathbf{u}^{(j)}\}$  are independent and sampled from a  
 251 continuous probability distribution, then  $\text{rank } \mathbf{A} = M$  with probability 1.

252 3. If  $\mathbf{A}$  has maximal possible rank (i.e.  $\text{rank } \mathbf{A} = M$ ), then

$$\text{rank } \mathcal{J} = \text{rank } \mathcal{Q},$$

253 and all the minimal CP decompositions differ only by the third factors, which are  
 254 linked as in (23). Moreover, if the CP decomposition of  $\mathcal{Q}$  is unique, then the CP  
 255 decomposition of  $\mathcal{J}$  is also unique.

256 *Proof of Lemma 3.* Let us express  $g'_k(\mathbf{v}_k^\top \mathbf{u})$  in an explicit form. First,  $g'_k(t) = c_{k,1} +$   
 257  $2c_{k,2}t + 3c_{k,3}t^2 + \dots + dc_{k,d}t^{d-1}$ , from which it follows that

$$g'_k(\mathbf{v}_k^\top \mathbf{u}) = c_{k,1} + 2c_{k,2}\mathbf{v}_k^\top \mathbf{u} + 3c_{k,3}(\mathbf{v}_k^\top \mathbf{u})^2 + \dots + dc_{k,d}(\mathbf{v}_k^\top \mathbf{u})^{d-1}.$$

258 Since  $(\mathbf{v}^\top \mathbf{u})^s = (\mathbf{v} \otimes \dots \otimes \mathbf{v})^\top (\mathbf{u} \otimes \dots \otimes \mathbf{u})$ , the  $j$ -th element of  $\mathbf{h}_k$  is equal to

$$(\mathbf{h}_k)_j = h_{j,k} = g'_k(\mathbf{v}_k^\top \mathbf{u}^{(j)}) = \mathbf{A}_{:,k}^\top \mathbf{z}_k,$$

259 which completes the proof.  $\square$

260 *Proof of Theorem 1.* 1. First, any polynomial map  $\mathbf{f}$  can be decomposed as (1) with  $r$   
 261 sufficiently large. Let us take such a decomposition; then it holds that

$$(\mathcal{Q}) \bullet_3 \mathbf{A}^\top = \left( \sum_{k=1}^r \mathbf{w}_k \circ \mathbf{v}_k \circ \mathbf{z}_k \right) \bullet_3 \mathbf{A}^\top = \sum_{k=1}^r \mathbf{w}_k \circ \mathbf{v}_k \circ \mathbf{A}^\top \mathbf{z}_k = \mathcal{J},$$

262 where the last equality follows from (23).

263 2. By construction, each element in the image of  $\mathbf{A}$  lies in the following subspace:

$$\begin{aligned} \mathcal{A} := \{ [a_0 \quad \mathbf{a}_1^\top \quad \mathbf{a}_2^\top \quad \dots \quad \mathbf{a}_{d-1}^\top]^\top \in \mathbb{R}^\delta \mid \\ \mathbf{a}_k \in \mathbb{R}^{n^k} \text{ is a vectorization of a symmetric } m \times \dots \times m \text{ tensor.} \} \end{aligned} \quad (26)$$

264 Taking into account that the dimension of the space of  $m \times \cdots \times m$  symmetric tensors  
 265 of order  $s$  is  $\binom{m+s-1}{s}$ , we get that the maximal possible rank of  $\mathbf{A}$  is

$$1 + m + \binom{m+1}{2} + \cdots + \binom{m+d-2}{d-1} = M.$$

266 Next, from (24), we have that the  $k$ -th column contains evaluations of all  $M$  monomials  
 267  $\{u_1^{j_1} \cdots u_m^{j_m} \leq d-1\}_{j_1, \dots, j_m=0}^{j_1+\dots+j_m \leq d}$  at a point  $\mathbf{u}^{(k)}$  (scaled by a constant). If, without loss of  
 268 generality, the first  $M$  points  $\{\mathbf{u}^{(k)}\}_{k=1}^M$  are in general position, then the columns of  $\mathbf{A}$   
 269 corresponding to different monomials are linearly independent by [41, Multiplicity One  
 270 Theorem], hence  $\text{rank } \mathbf{A} = M$ .

271 3. Note that each tube  $\mathcal{Q}_{i,j,:}$  of the tensor  $\mathcal{Q}$ , by construction, lies in  $\mathcal{A}$ . If  $\text{rank } \mathbf{A} = M$ ,  
 272 then its row span coincides with  $\mathcal{A}$ . Hence the following identity holds true:

$$\mathcal{J} \bullet_3 (\mathbf{A}^\dagger)^\top = \mathcal{Q} \bullet_3 (\mathbf{A}\mathbf{A}^\dagger)^\top = \mathcal{Q}. \quad (27)$$

273 The remaining properties follow from (27) and (25).  $\square$

## 274 6. Structured tensor decompositions

### 275 6.1. From CPD to a decomposition with structured rank-one terms

276 The CP decomposition of  $\mathcal{J}$  and  $\mathcal{Q}$ , although related, are not always equivalent to  
 277 the original decomposition (3). This happens because there are still nontrivial linear de-  
 278 pendencies between the elements of  $\mathcal{Q}$  and  $\mathcal{J}$ . In what follows, we establish relationships  
 279 between the CP decompositions and the original decomposition (3).

280 First, we prove that for the rank-one case, these decompositions coincide.

281 **Proposition 1.** *Consider a polynomial map  $\mathbf{f}(\mathbf{u})$  of degree  $d$ , and the tensor  $\mathcal{Q}$  built*  
 282 *from it. Then the following holds*

$$\text{rank}(\mathcal{Q}) \leq 1 \iff \mathbf{f}(\mathbf{u}) = \mathbf{w}g(\mathbf{v}^\top \mathbf{u}),$$

283 where  $\mathbf{w} \in \mathbb{R}^n$ ,  $\mathbf{v} \in \mathbb{R}^m$  and  $g(t)$  is a polynomial of degree  $d$ .

284 *Proof.* The  $\boxed{\Leftarrow}$  follows from Lemma 1. Let us prove the  $\boxed{\Rightarrow}$  part. Assume that

$$\mathcal{Q} = \mathbf{w} \circ \mathbf{v} \circ \mathbf{y}.$$

285 First, since the tensor  $\mathcal{Q}$  contains all the coefficients of the derivatives, we have that there  
 286 exists a polynomial  $\tilde{f}(\mathbf{u})$  such that  $\nabla f_k(\mathbf{u}) = (\mathbf{w})_k \nabla \tilde{f}(\mathbf{u})$ . Since the polynomials  $f_k(\mathbf{u})$   
 287 do not have constant terms, we have that

$$\mathbf{f}(\mathbf{u}) = \mathbf{w} \tilde{f}(\mathbf{u}),$$

288 where  $\Psi(\tilde{f}) = \mathbf{v}\mathbf{y}^\top$ .

289 Next, let us show that the polynomial  $\tilde{f}$  should necessarily the form  $\tilde{f}(\mathbf{u}) = g(\mathbf{v}^\top \mathbf{u})$ .  
 290 Since  $\Psi(\tilde{f}) = \mathbf{v}\mathbf{y}^\top$ , then it follows from (11) that all the unfoldings  $\Psi^{(1)}, \Psi^{(2)}, \Psi^{(3)}, \dots, \Psi^{(d)}$

291 have rank at most one and their column space is spanned by the vector  $\mathbf{v}$ . Therefore, we  
 292 have that

$$\begin{aligned}\Psi^{(1)} &= c_1 \mathbf{v}, \\ \Psi^{(2)} &= c_2 \mathbf{v} \mathbf{v}^\top, \\ \Psi^{(3)} &= c_3 \mathbf{v} \circ \mathbf{v} \circ \mathbf{v}, \\ &\vdots \\ \Psi^{(d)} &= c_d \mathbf{v} \circ \cdots \circ \mathbf{v},\end{aligned}$$

293 and hence  $\tilde{f}(\mathbf{u}) = g(\mathbf{v}^\top \mathbf{u})$  where

$$g(t) = c_1 t + c_2 t^2 + \cdots + c_d t^d,$$

294 which completes the proof.  $\square$

295 **Remark 2.** *The fact that  $\text{rank } \Psi(\tilde{f}) \leq 1$  implies  $\tilde{f}(\mathbf{u}) = g(\mathbf{v}^\top \mathbf{u})$  also can be proved  
 296 alternatively, by noting that the matrix  $\tilde{\Psi} \tilde{f}$ , after removing duplicate columns, can be  
 297 reduced to the form  $S(f)$  in [33, Proposition 22]. Hence, by [33, Proposition 4.1], the  
 298 polynomial  $\tilde{f}$  has necessarily the form  $\tilde{f}(\mathbf{u}) = g(\mathbf{v}^\top \mathbf{u})$ . However, this alternative proof  
 299 requires introducing extra notation, which would be much longer than the proof presented  
 300 in this paper.*

301 **Corollary 1.** *If the  $N$  sampling points are chosen such that  $\text{rank}(\mathbf{A}) = M$ , then*

$$\text{rank}(\mathcal{J}) \leq 1 \iff \mathbf{f}(\mathbf{u}) = \mathbf{w}g(\mathbf{v}^\top \mathbf{u}).$$

302 As a corollary of Proposition 1, we get that the original polynomial decomposition (3)  
 303 is equivalent to a structured CP decomposition.

304 **Corollary 2.** *Let  $\mathcal{L}_{\mathcal{Q}} \subset \mathbb{R}^{n \times m \times \delta}$  be the linear subspace of tensors with the structure of  
 305  $\mathcal{Q}$ . Let the sampling points be chosen such that  $\text{rank } \mathbf{A} = M$ , and  $\mathcal{L}_{\mathcal{J}} \subset \mathbb{R}^{n \times m \times N}$  be the  
 306 linear subspace of tensors with the structure of  $\mathcal{J}$ .*

307 *Then the following three statements are equivalent:*

- 308 1. *the polynomial map  $\mathbf{f}(\mathbf{u})$  admits a decomposition (3);*
- 309 2. *the tensor  $\mathcal{Q}(\mathbf{f})$  admits the structured CP decomposition*

$$\mathcal{Q} = \mathcal{Q}_1 + \cdots + \mathcal{Q}_r, \quad \text{rank}(\mathcal{Q}_k) = 1, \quad \mathcal{Q}_k \in \mathcal{L}_{\mathcal{Q}}; \quad (28)$$

- 310 3. *the tensor  $\mathcal{J}(\mathbf{f})$  admits the structured CP decomposition*

$$\mathcal{J} = \mathcal{J}_1 + \cdots + \mathcal{J}_r, \quad \text{rank}(\mathcal{J}_k) = 1, \quad \mathcal{J}_k \in \mathcal{L}_{\mathcal{J}}. \quad (29)$$

311 The structure constraint is important: indeed, the CP decomposition of the tensor  
 312  $\mathcal{Q}$  or  $\mathcal{J}$  is not necessarily structured. In general, we do not know even if the CP rank  
 313 is equal to the structured CP rank (minimal number of terms in (28) or (29)). This  
 314 is similar to the Comon's conjecture [35, §5] about symmetric tensors: it is not known  
 315 whether the symmetric rank of a symmetric tensor equals its non-symmetric rank.

316 However, if the CP decomposition of a tensor is unique (for example, if it satisfies  
 317 Kruskal's uniqueness conditions), then it should necessarily be a structured CP decom-  
 318 position.

319 *6.2. Computing coupled/structured CP decomposition*

320 Earlier attempts to tackle the structured case were made by [25, 13] and [42, §8,  
 321 pp. 133–136]. The attempts of [25, 13] have the disadvantage that a tensor is built  
 322 that has missing values, which increase in number as the polynomial degree grows. The  
 323 attempt of [42] consisted of parameterizing the internal nonlinear functions  $g_k$  using their  
 324 coefficients. Although this seems a promising approach, it turned out to be problematic  
 325 in practice to build a working algorithm, as the decoupling method led to strongly  
 326 nonlinear/nonconvex optimization problems.

327 We propose to tackle the problem by solving a coupled and structured CP decomposi-  
 328 tion instead. First, let us consider simultaneous decoupling of homogeneous polynomials  
 329 (8). Let us arrange the  $\Psi_i^{(d)}$ , for  $i = 1, \dots, n$  into a tensor  $\mathcal{T}^d$ , such that  $\mathcal{T}_{i_1, \dots, i_d}^d = \Psi_i^{(d)}$ ,  
 330 for all  $i = 1, \dots, n$ . Then it is easy to verify that  $\mathcal{T}^d$  admits a partially symmetric CP  
 331 decomposition

$$\mathcal{T}^d = \llbracket \mathbf{W}, \underbrace{\mathbf{V}, \dots, \mathbf{V}}_{d \text{ times}} \rrbracket,$$

332 which, in our decoupled representation (1), takes the form  $\mathbf{W}\mathbf{g}(\mathbf{V}^\top \mathbf{u})$ , where  $\mathbf{g}(\mathbf{x}) =$   
 333  $\left[ x_1^d \ \dots \ x_r^d \right]^\top$ .

334 Decoupling non-homogeneous polynomials can be achieved by means of a coupled  
 335 structured CP decomposition of the  $\mathcal{T}^d$  tensors. Let us arrange all  $\Psi_i^{(d)}$ , for  $i = 1, \dots, n$   
 336 into the tensors  $\mathcal{T}^d$  (like in the previous paragraph), such that  $\mathcal{T}_{i_1, \dots, i_d}^d = \Psi_i^{(d)}$ , for all  $i =$   
 337  $1, \dots, n$ . We now have for each degree a coupled partially symmetric CP decomposition  
 338 as

$$\begin{aligned} \mathcal{T}^1 &= \llbracket \mathbf{W}, \mathbf{V}, \mathbf{c}_1^\top \rrbracket, \\ \mathcal{T}^2 &= \llbracket \mathbf{W}, \mathbf{V}, \mathbf{V}, \mathbf{c}_2^\top \rrbracket, \\ &\vdots \\ \mathcal{T}^d &= \llbracket \mathbf{W}, \underbrace{\mathbf{V}, \dots, \mathbf{V}}_{d \text{ times}}, \mathbf{c}_d^\top \rrbracket, \end{aligned} \tag{30}$$

339 where the  $\mathbf{c}_i$ , for  $i = 1, \dots, d$ , are the  $i$ -th degree coefficients for each of the  $r$  nonlinear  
 340 functions  $g_k$ .

341 Remark that these coefficients were not required in the previous paragraphs when  
 342 homogeneous polynomials were considered: in such cases the nonlinear functions  $g_k$  are  
 343 of the form  $c_k t^d$ , i.e., they differ only by a scaling factor, which can be assumed to be fully  
 344 absorbed by  $\mathbf{W}$ . Also remark that there are redundancies in the representation (30):  
 345 for example, an equivalent problem can be obtained if one rescales a coefficient vector  $\mathbf{c}_\delta$   
 346 to a vector containing ones, in which case a rescaling has to take place on the remaining  
 347 coefficients as well as on  $\mathbf{W}$ . Finally we want to mention that the framework of structured  
 348 data fusion [43, 27] allows for computing tensor decompositions as in (30), where several  
 349 tensors (and possibly matrices) are jointly decomposed while sharing factors, possibly  
 350 while imposing structure on the factors.

351 **Example 9.** *Let us continue with Examples 1, 6 and 7. We have that  $m = n = 2$  and*  
 352  *$r = 3$ , which does not guarantee a unique CP decomposition of  $\mathcal{J}$  (under assumptions*  
 353 *of genericity, see [24]). Indeed, if we compute a numerical CP decomposition of ten-*  
 354 *sor  $\mathcal{J}$ , we find that, up to a relative norm-wise error  $2.3546 \times 10^{-16}$ ,  $\mathcal{J}$  admits a CP*



355 *decomposition with factors*

$$\begin{aligned} \tilde{\mathbf{W}} &= \begin{bmatrix} 1.1628 & -3.2951 & 3.0252 \\ 0.5705 & 1.1349 & -2.1791 \end{bmatrix}, \quad \tilde{\mathbf{V}} = \begin{bmatrix} 3.5822 & -0.7705 & -2.2959 \\ -0.0226 & -3.4455 & -2.9785 \end{bmatrix}, \\ \tilde{\mathbf{H}} &= \begin{bmatrix} 0.2736 & 0.8181 & 0.0312 \\ -3.3900 & 0.2647 & 1.2313 \\ 0.5334 & -3.9194 & 3.4945 \end{bmatrix}, \end{aligned}$$

356 *the columns of which are not scaled and permuted versions of the columns of  $\mathbf{W}$ ,  $\mathbf{V}$ ,  $\mathbf{H}$ .*  
 357 *It can be shown that the ‘structured CP’ approaches are able to correctly return the*  
 358 *underlying factors  $\mathbf{W}$ ,  $\mathbf{V}$  and  $\mathbf{H}$  (up to scaling and permutation invariances). For in-*  
 359 *stance, the structured data fusion framework [43, 27] is able to compute the coupled and*  
 360 *partially symmetric decomposition (30). This returns*

$$\tilde{\mathbf{W}} = \begin{bmatrix} 1.2767 & 1.7112 & 0 \\ 0 & -0.8556 & -1.9980 \end{bmatrix}, \quad \tilde{\mathbf{V}} = \begin{bmatrix} -0.9218 & -1.0534 & 1.5879 \\ 0.9218 & -2.1067 & 0.7940 \end{bmatrix},$$

361 *as well as computed values for the coefficient vectors of  $g_i(x_i)$ , which are omitted here.*  
 362 *It can be verified that  $\tilde{\mathbf{W}}$  and  $\tilde{\mathbf{V}}$  are scaled and permuted versions of  $\mathbf{W}$  and  $\mathbf{V}$ .*

363 *Remark that if one uses  $m = n = r = 2$ , both the structured and non-structured*  
 364 *CP decomposition return the same decomposition (up to scaling and permutation of the*  
 365 *columns of the factors). Indeed, in this case, uniqueness is guaranteed (generically),*  
 366 *ensuring that the underlying factors are identifiable. This could be checked easily by*  
 367 *generating a variation of the equations that we are decoupling where the third columns of*  
 368  *$\mathbf{V}$  and  $\mathbf{W}$  are removed, so that  $g_3(x_3)$  is not considered.*

### 369 6.3. Linking $\mathcal{Q}$ and $\mathcal{T}^1, \dots, \mathcal{T}^d$ tensors

370 In this section, we show how  $\mathcal{Q}$  and its CPD is connected with the tensors  $\mathcal{T}^s$  and  
 371 their joint decomposition (30). Let  $(1, 2)$ -reshapings of the tensors  $\mathcal{T}^s$  to be the third  
 372 order tensors  $\mathcal{T}_{(1,2)}^s \in \mathbb{R}^{n \times m \times m^{s-1}}$  defined as

$$(\mathcal{T}_{(1,2)}^s)_{i,j,:} = \text{vec}(\mathcal{T}_{i,j,:\dots,:}^s).$$

373 Then it is easy to see that the tensor  $\mathcal{T}_{(1,2)}^s$  can be split into slices as shown in Fig. 5,  
 374 where the  $\Psi^{(s,j)}$  is the symmetric tensor corresponding to the  $s$ -th degree homogeneous  
 375 part of the polynomial  $f_k$ .

376 By taking into account the definition (11) of the slices of the tensor  $\mathcal{Q}$ , we can easily  
 377 see that  $\mathcal{Q}$  can be constructed by stacking the tensors is equivalent to reshaping the  
 378 tensors  $\mathcal{T}_{(1,2)}^s$  along the third mode together, as shown in Fig. 6.

379 Remark that in Lemma 1 we see that the structure appearing in the CP decomposition  
 380 of  $\mathcal{Q}$  is closely connected to the simultaneous decomposition described in Section 6.2.  
 381 Indeed, Lemma 1 can be alternatively deduced from (30), because the outer products of  
 382 vectors become Kronecker products after reshaping.

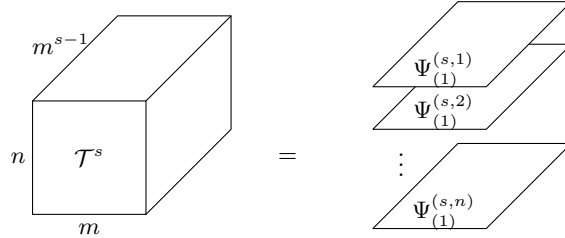


Figure 5: The slices of the tensor  $\mathcal{T}^s$  are the unfoldings of the symmetric tensors corresponding to  $s$ -degree homogeneous parts of polynomials  $f_1, \dots, f_n$ .

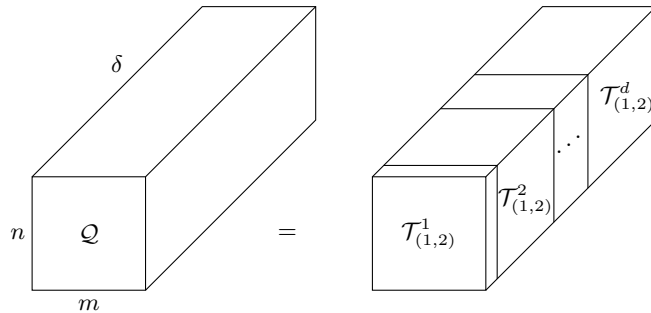


Figure 6: Stacking the reshapings of  $\mathcal{T}^s$  together.

## 383 7. Conclusions and perspectives

384 We have established a link between two tensorization approaches for decoupling  
 385 multivariate polynomials [26, 24]: the tensor of Jacobian matrices [24] can be obtained by  
 386 multiplying the coefficient-based tensor [26] by a Vandermonde-like matrix. As revealed  
 387 by this connection, the two approaches have similar fundamental properties, such as  
 388 equal tensor rank and uniqueness of the CP decomposition under conditions on the  
 389 number and location of the sampling points.

390 The decoupling problem, however, is not equivalent to the CP decomposition of one  
 391 of the tensors. This may lead to loss of uniqueness and identifiability of the CP decom-  
 392 position, in the cases when the original decomposition is still unique. We have shown that  
 393 by adding structure to the CP decomposition we can obtain equivalence between tensor  
 394 decomposition and decoupling problems for polynomials. The structure can be imposed  
 395 either as a joint decomposition of partially symmetric tensors, or can be imposed on  
 396 rank-one factors. Numerical experiments confirm that using structured decompositions  
 397 can restore uniqueness of the polynomial decoupling.

398 Although our results show that different tensor-based approaches are very closely  
 399 related, let us make some remarks on applicability of the approaches and some future  
 400 directions. For (differentiable) non-polynomial functions, the approach based on Jacobian  
 401 matrices would be more appropriate, as it only uses evaluations of the derivatives of the  
 402 functions. Coefficient-based approach seems more relevant in the case when the region  
 403 of interest is unclear, or when some of the coefficients are missing or unreliable. In  
 404 both cases, an interesting open question remains how to impose the structure directly on

405 the rank-one components, without resorting to coupled tensor factorizations. Another  
406 important question is how to address the approximate decoupling problem, i.e., when we  
407 are dealing with noise (see [44] for results on the unstructured case).

## 408 8. Acknowledgements

409 This work was supported in part by the ERC AdG-2013-320594 grant “DECODA”, by  
410 the Flemish Government (Methusalem), by the ERC Advanced Grant SNLSID under  
411 contract 320378, and by the Fonds Wetenschappelijk Onderzoek – Vlaanderen under  
412 EOS Project no 30468160 and FWO projects G.0280.15N and G.0901.17N.

## 413 References

- 414 [1] P. Comon, Tensors : A brief introduction, *IEEE Signal Processing Magazine* 31 (3) (2014) 44–53.  
415 [doi:10.1109/MSP.2014.2298533](https://doi.org/10.1109/MSP.2014.2298533).
- 416 [2] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, *SIAM Rev.* 51 (3) (2009)  
417 455–500.
- 418 [3] A. Cichocki, D. Mandic, L. D. Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, H. A. PHAN, Tensor  
419 decompositions for signal processing applications: From two-way to multiway component analysis,  
420 *IEEE Signal Processing Magazine* 32 (2) (2015) 145–163.
- 421 [4] M. K. P. Ng, Q. Yuan, L. Yan, J. Sun, An adaptive weighted tensor completion method for the  
422 recovery of remote sensing images with missing data, *IEEE Transactions on Geoscience and Remote*  
423 *Sensing* 55 (6) (2017) 3367–3381. [doi:10.1109/TGRS.2017.2670021](https://doi.org/10.1109/TGRS.2017.2670021).
- 424 [5] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, C. Faloutsos, Tensor  
425 decomposition for signal processing and machine learning, *IEEE Transactions on Signal Processing*  
426 65 (13) (2017) 3551–3582.
- 427 [6] E. Acar, R. Bro, A. K. Smilde, Data fusion in metabolomics using coupled matrix and tensor fac-  
428 torizations, *Proceedings of the IEEE* 103 (9) (2015) 1602–1620. [doi:10.1109/JPROC.2015.2438719](https://doi.org/10.1109/JPROC.2015.2438719).
- 429 [7] C. Chen, X. Li, M. K. Ng, X. Yuan, Total variation based tensor decomposition for multi-  
430 dimensional data with time dimension, *Numerical Linear Algebra with Applications* 22 (6) (2015)  
431 999–1019. [doi:10.1002/nla.1993](https://doi.org/10.1002/nla.1993).
- 432 [8] O. Debals, L. De Lathauwer, Stochastic and deterministic tensorization for blind signal separation,  
433 in: *Proc. 12th International Conference on Latent Variable Analysis and Signal Separation (LVA-*  
434 *ICA 2015)*, Vol. 9237 of *Lecture Notes in Computer Science*, Springer, 2015, pp. 3–13.
- 435 [9] P. Dreesen, M. Schoukens, K. Tiels, J. Schoukens, Decoupling static nonlinearities in a parallel  
436 Wiener-Hammerstein system: A first-order approach, in: *Proc. 2015 IEEE International Instru-*  
437 *mentation and Measurement Technology Conference (I2MTC 2015)*, Pisa, Italy, 2015, pp. 987–992.
- 438 [10] P. Dreesen, A. Fakhrizadeh Esfahani, J. Stoev, K. Tiels, J. Schoukens, Decoupling nonlinear state-  
439 space models: case studies, in: P. Sas, D. Moens, A. van de Walle (Eds.), *International Conference*  
440 *on Noise and Vibration (ISMA2016) and International Conference on Uncertainty in Structural*  
441 *Dynamics (USD2016)*, Leuven, Belgium, 2016, pp. 2639–2646.
- 442 [11] K. Tiels, J. Schoukens, From coupled to decoupled polynomial representations in parallel Wiener-  
443 Hammerstein models, in: *Proc. 52nd IEEE Conf. Decis. Control (CDC)*, Florence, Italy, 2013, pp.  
444 4937–4942.
- 445 [12] M. Schoukens, Y. Rolain, Cross-term elimination in parallel Wiener systems using a linear input  
446 transformation, *IEEE Transactions on Instrumentation and Measurement* 61 (3) (2012) 845–847.
- 447 [13] M. Schoukens, K. Tiels, M. Ishteva, J. Schoukens, Identification of parallel Wiener-Hammerstein  
448 systems with a decoupled static nonlinearity, in: *Proceedings of 19th IFAC World Congress*, Cape  
449 Town (South Africa), August 24–29, 2014, 2014, pp. 505–510.
- 450 [14] A. Fakhrizadeh Esfahani, P. Dreesen, K. Tiels, J.-P. Noël, J. Schoukens, Parameter reduction in  
451 nonlinear state-space identification of hysteresis, *Mechanical Systems and Signal Processing* 104  
452 (2018) 884–895.
- 453 [15] B. F. Logan, L. A. Shepp, Optimal reconstruction of a function from its projections, *Duke Math.*  
454 *J.* 42 (4) (1975) 645–659.

- 455 [16] V. Lin, A. Pinkus, Fundamentality of ridge functions, *Journal of Approximation Theory* 75 (3)  
456 (1993) 295 – 311.
- 457 [17] K. I. Oskolkov, On representations of algebraic polynomials as a sum of plane waves, *Serdica*  
458 *Mathematical Journal* (2002) 379–390.
- 459 [18] Y. Shin, J. Ghosh, Ridge polynomial networks, *IEEE Transactions on Neural Networks* 6 (3) (1995)  
460 610–622.
- 461 [19] A. Iarrobino, V. Kanev, *Power Sums, Gorenstein Algebras, and Determinantal Loci*, Vol. 1721 of  
462 *Lecture Notes in Mathematics*, Springer, 1999.
- 463 [20] J. M. Landsberg, *Tensors: Geometry and Applications*, Vol. 128 of *Grad. Stud. Math.*, American  
464 *Mathematical Society*, Providence, RI, 2012.
- 465 [21] A. Białynicki-Birula, A. Schinzel, Representations of multivariate polynomials as sums of poly-  
466 nomials in linear forms, *Colloquium Mathematicum* 112 (2) (2008) 201–233.
- 467 [22] A. Schinzel, On a decomposition of polynomials in several variables, *J. de Théorie des Nombres de*  
468 *Bordeaux* 14 (2) (2002) 647–666.
- 469 [23] E. Carlini, J. Chipalkatti, On Waring’s problem for several algebraic forms, *Comment. Math. Helv.*  
470 78 (2003) 494–517.
- 471 [24] P. Dreesen, M. Ishteva, J. Schoukens, Decoupling multivariate polynomials using first-order infor-  
472 mation, *SIAM. J. Matrix Anal. Appl.* 36 (2) (2015) 864–879.
- 473 [25] K. Tiels, J. Schoukens, From coupled to decoupled polynomial representations in parallel Wiener-  
474 Hammerstein models, in: *52nd IEEE Conference on Decision and Control*, Florence, Italy, December  
475 10-13, 2013, 2013, pp. 4937–4942.
- 476 [26] A. Van Mulders, L. Vanbeylen, K. Usevich, Identification of a block-structured model with several  
477 sources of nonlinearity, in: *Proceedings of the 14th European Control Conference (ECC 2014)*,  
478 2014, pp. 1717–1722.
- 479 [27] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, L. De Lathauwer, Tensorlab 3.0, available online,  
480 Mar. 2016. URL: <http://www.tensorlab.net/> (2016).
- 481 [28] C. A. Andersson, R. Bro, The N-way toolbox for MATLAB, *Chemometrics & Intelligent Laboratory*  
482 *Systems* 52 (2000) 1–4, <http://www.models.life.ku.dk/source/nwaytoolbox/>.
- 483 [29] B. W. Bader, T. G. Kolda, et al., MATLAB tensor toolbox version 2.5, available online, January  
484 2012. URL: <http://www.sandia.gov/~tgkolda/TensorToolbox/> (2012).
- 485 [30] J. Carroll, J. Chang, Analysis of individual differences in multidimensional scaling via an N-way  
486 generalization of “Eckart-Young” decomposition, *Psychometrika* 35 (3) (1970) 283–319.
- 487 [31] R. A. Harshman, Foundations of the PARAFAC procedure: Model and conditions for an “explana-  
488 tory” multi-mode factor analysis, *UCLA Working Papers in Phonetics* 16 (1) (1970) 1–84.
- 489 [32] P. Comon, Y. Qi, K. Usevich, A polynomial formulation for joint decomposition of symmetric  
490 tensors of different orders, in: E. Vincent, A. Yeredor, Z. Koldovský, P. Tichavský (Eds.), *Latent*  
491 *Variable Analysis and Signal Separation*, Vol. 9237 of *Lecture Notes in Computer Science*, Springer  
492 *International Publishing*, 2015, pp. 22–30.
- 493 [33] P. Comon, Y. Qi, K. Usevich, Identifiability of an x-rank decomposition of polynomial maps, *SIAM*  
494 *Journal on Applied Algebra and Geometry* 1 (1) (2017) 388–414. doi:10.1137/16M1108388.
- 495 [34] J. M. Landsberg, *Tensors: Geometry and applications*, Vol. 128, American Mathematical Soc.,  
496 2012.
- 497 [35] P. Comon, G. H. Golub, L.-H. Lim, B. Mourrain, Symmetric tensors and symmetric tensor rank,  
498 *SIAM. J. Matrix Anal. Appl.* 30 (3) (2008) 1254–1279.
- 499 [36] K. Batselier, N. Wong, Symmetric tensor decomposition by an iterative eigendecomposition algo-  
500 rithm, *J. Comput. Appl. Math.* 308 (2016) 69–82.
- 501 [37] I. Domanov, L. D. Lathauwer, On the uniqueness of the canonical polyadic decomposition of third-  
502 order tensors—part ii: Uniqueness of the overall decomposition, *SIAM Journal on Matrix Analysis*  
503 *and Applications* 34 (3) (2013) 876–903. doi:10.1137/120877258.
- 504 [38] L. Chiantini, G. Ottaviani, N. Vannieuwenhoven, On generic identifiability of symmetric tensors  
505 of subgeneric rank, *Transactions of the American Mathematical Society* 369 (6) (2017) 4021–4042.  
506 doi:10.1090/tran/6762.
- 507 [39] H. Abo, N. Vannieuwenhoven, Most secant varieties of tangential varieties to Veronese varieties are  
508 nondefective, *Transactions of the American Mathematical Society* 370 (1) (2018) 393–420. doi:  
509 10.1090/tran/6955.  
510 URL <https://lirias.kuleuven.be/handle/123456789/538507>
- 511 [40] A. Van Mulders, J. Schoukens, L. Vanbeylen, Identification of systems with localised nonlinearity:  
512 From state-space to block-structured models, *Automatica* 49 (5) (2013) 1392 – 1396.
- 513 [41] R. Miranda, Linear systems of plane curves, *Notices of the AMS* 46 (1999) 192–202.

- 514 [42] G. Hollander, Multivariate polynomial decoupling in nonlinear system identification, Ph.D. thesis,  
515 Vrije Universiteit Brussel (VUB) (2017).
- 516 [43] L. Sorber, M. Van Barel, L. De Lathauwer, Structured data fusion, *IEEE J. Sel. Top. Signal Process.*  
517 9 (4) (2015) 586–600.
- 518 [44] G. Hollander, P. Dreesen, M. Ishteva, J. Schoukens, Approximate decoupling of multivariate poly-  
519 nomials using weighted tensor decomposition, *Numerical Linear Algebra with Applications* 25 (2)  
520 (2018) xxx–xxx. [doi:10.1002/nla.2135](https://doi.org/10.1002/nla.2135).