

# Decoupling multivariate functions using second-order information and tensors

Philippe Dreesen, Jeroen De Geeter, and Mariya Ishteva

Vrije Universiteit Brussel (VUB), Dept. VUB-ELEC, Brussels, Belgium

## Abstract

The power of multivariate functions is their ability to model a wide variety of phenomena, but have the disadvantages that they lack an intuitive or interpretable representation, and often require a (very) large number of parameters. We study decoupled representations of multivariate vector functions, which are linear combinations of univariate functions in linear combinations of the input variables. This model structure provides a description with fewer parameters, and reveals the internal workings in a simpler way, as the nonlinearities are one-to-one functions. In earlier work, a tensor-based method was developed for performing this decomposition by using first-order derivative information. In this article, we generalize this method and study how the use of second-order derivative information can be incorporated. By doing this, we are able to push the method towards more involved configurations, while preserving uniqueness of the underlying tensor decompositions. Furthermore, even for some non-identifiable structures, the method seems to return a valid decoupled representation. These results are a step towards more general data-driven and noise-robust tensor-based framework for computing decoupled function representations.

## 1 Introduction

### 1.1 Towards interpretability of nonlinear models

Nonlinear models are used in a wide variety of science and engineering fields, such as data analytics, signal processing, system identification, and control engineering. While nonlinear models are able to capture wild nonlinear effects, this often comes at the cost of high parametric complexity, and a lack of ‘model interpretability’.

This paper studies the question how a given nonlinear multivariate vector function  $\mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^n$  can be decomposed into a simpler structure, as in [5, 9, 18, 19, 21]. In particular, we investigate a structure of the form

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

where  $\mathbf{W}$  and  $\mathbf{V}$  are transformation matrices, and the vector function  $\mathbf{g}(\mathbf{z}) = [g_1(z_1) \ \cdots \ g_r(z_r)]^T$  is composed of univariate functions  $g_i(z_i)$  in its  $r$  components. The decoupled representation is visualized in Figure 1.

The decomposition (1) provides a representation that is easier to comprehend, as the nonlinearity is contained in a set of univariate components. Moreover, it typically has a lower parametric complexity, and could hence be viewed as a form of nonlinear model order reduction.

### 1.2 Tensorization methods for decoupling polynomials

When  $\mathbf{f}$  is polynomial, the decomposition (1) has connections to the canonical polyadic decomposition (CPD) of a partially symmetric tensor (possibly a joint decomposition), see [20]. Indeed, typical tensorization methods make use of the connection between homogeneous polynomials and their coefficient tensors [20, 4]. The (homogeneous) scalar polynomial case, *i.e.*,  $n = 1$ , is known as the Waring decomposition [2, 13], and is a fundamental problem in algebraic geometry. The problem (1) we study

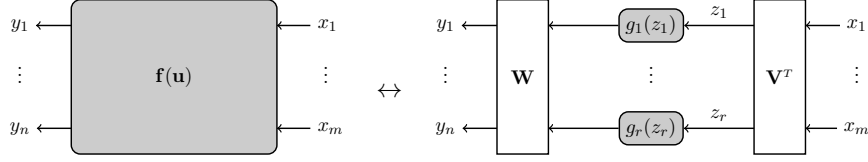


Figure 1: A multivariate nonlinear vector function  $\mathbf{f}(\mathbf{x})$  can be represented in a decoupled representation  $\mathbf{f}(\mathbf{x}) = \mathbf{W}\mathbf{g}(\mathbf{V}^T\mathbf{x})$ . A decoupled representation typically has fewer parameters, and reveals in an intuitive way, the internal nonlinearities of  $\mathbf{f}(\mathbf{x})$ .

is hence very reminiscent of the classical Waring problem, but we consider the *non-homogeneous case of several polynomials*. The non-homogeneous Waring problem is studied in [1, 14]. The simultaneous Waring problem for several homogeneous polynomials is studied in [2, 18].

### 1.3 Contributions and organization of this paper

We start from the tensorization method of [9]. In this framework, the function  $\mathbf{f}$  and its first-order derivative information are evaluated in a number of sampling points. A tensor is constructed from the set of corresponding Jacobian matrices, which admits a CPD that allows for the reconstruction of the decomposition (1). This approach has the following advantages: (i) the order and size of the constructed tensor do not increase with the degree of  $\mathbf{f}$ , and, (ii) the approach is not limited to the use of polynomials.

In the current paper, we generalize the method [9] to incorporate second-order derivative information. Since second-order derivative information leads to partially symmetric tensors, we are ultimately able to formulate the decomposition as a partially symmetric joint tensor decomposition. By involving the second-order derivatives, we impose additional constraints on the (joint) tensor decompositions, hence it is expected to enjoy more relaxed uniqueness conditions. In the article, we assume that an exact and uniquely identifiable [5] representation of  $\mathbf{f}(\mathbf{x})$  exists. Nevertheless, the resulting joint tensor decomposition will ultimately be phrased as an optimization problem, and provides a natural starting point for studying both the noisy decoupling problem, as well as a model reduction interpretation, but this is beyond the scope of the current paper.

The current article is organized as follows. Section 2 outlines the tensor-based decoupling method, leading to a joint tensor decomposition formulation. We illustrate how the uniqueness properties improve by including second-order derivatives. Section 3 validates the method on three simulation examples. Section 4 summarizes the results and points out a few future research directions.

### 1.4 Notation

Scalars are denoted by lowercase or uppercase letters and vectors are denoted by lowercase bold-face letters. Elements of a vector are denoted by lowercase letters with an index as subscript, *e.g.*,  $\mathbf{x} = [x_1 \dots x_m]^T$ . Matrices are denoted by uppercase bold-face letters, *e.g.*,  $\mathbf{V} \in \mathbb{R}^{m \times r}$ . The entry in the  $i$ -th row and  $j$ -th column of the matrix  $\mathbf{V}$  is denoted by  $v_{ij}$ . A matrix  $\mathbf{V} \in \mathbb{R}^{m \times r}$  has columns  $\mathbf{v}_i$  as in  $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_r]$ . The transpose of a matrix  $\mathbf{V}$  is denoted by  $\mathbf{V}^T$ . A diagonal matrix with diagonal elements  $a_1, a_2, a_3$  is denoted by  $\text{diag}(a_1, a_2, a_3)$  or  $\text{diag}(a_i)$ . Higher-order tensors are denoted by bold-face uppercase calligraphical letters, *e.g.*,  $\mathcal{J} \in \mathbb{R}^{n \times m \times N}$ . For scalar, vector, matrix and higher-order tensor *functions*, we employ the same conventions. The outer product is denoted by  $\circ$  and defined as follows: For  $\mathcal{X} = \mathbf{u} \circ \mathbf{v} \circ \mathbf{w}$ , the entry in position  $(i, j, k)$  is  $u_i v_j w_k$ . The Frobenius norm of a tensor  $\mathcal{X}$  is denoted as  $\|\mathcal{X}\|_F$ . The Euclidean norm of a vector  $\mathbf{x}$  is denoted as  $\|\mathbf{x}\|$ . The first-order and second-order derivatives of a univariate function  $g(z)$  are denoted by  $g'(z)$  and  $g''(z)$ , respectively.

## 2 Decoupling multivariate functions using tensors

### 2.1 The Canonical Polyadic Decomposition

The canonical polyadic decomposition (CPD) [3, 10, 11] is the decomposition of a tensor into a minimal sum of rank-one components. For instance, a third-order tensor  $\mathcal{T}$  has a CPD of the form

$$\mathcal{T} = \sum_{i=1}^R \mathbf{a}_i \circ \mathbf{b}_i \circ \mathbf{c}, \quad (2)$$

or in a short-hand notation  $\mathcal{T} = \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket$ , where  $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_R]$  (similar for  $\mathbf{B}$  and  $\mathbf{C}$ ). The CPD is a celebrated tensor decomposition, which has found a variety of applications in signal processing and data sciences. One of the attractive properties of the CPD is its more relaxed uniqueness conditions. In contrast with matrix factorization, where uniqueness is only possible by imposing additional constraints (*e.g.*, orthogonality in the singular value decomposition (SVD)), the CPD has milder uniqueness properties [12, 6, 11, 7, 8].

In our framework, the uniqueness conditions of the CPD will allow us to ensure that the proposed decoupling method retrieves the uniquely identifiable model structure: for identifiable models, uniqueness of the CPD is a sufficient condition for uniqueness of the model.

### 2.2 Decoupling functions using first-order information

Consider the function  $\mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^n$  that admits a decoupled representation (1). The Jacobian  $\mathbf{J}(\mathbf{x})$  is represented with an  $n \times m$  matrix function defined as

$$\mathbf{J}_{ij}(x) = \frac{\partial f_i(\mathbf{x})}{\partial x_j}. \quad (3)$$

The chain rule for derivation shows that the Jacobian matrix  $\mathbf{J}(\mathbf{x})$  can be factorized as  $\mathbf{J}(\mathbf{x}) = \mathbf{W} \text{diag}(g'_i(\mathbf{v}_i^T \mathbf{x})) \mathbf{V}^T$ , or alternatively, in a CPD formulation  $\mathcal{J} = \llbracket \mathbf{W}, \mathbf{V}, (\mathbf{g}'(\mathbf{V}^T \mathbf{x}))^T \rrbracket$ . Then an  $n \times m \times N$  tensor  $\mathcal{J}$ , built from evaluating the Jacobian matrix  $\mathbf{J}(\mathbf{x}^{(k)})$  in a set of  $N$  sampling points  $\mathbf{x}^{(k)}, k = 1, \dots, N$ , admits a CPD of the form

$$\mathcal{J} = \llbracket \mathbf{W}, \mathbf{V}, \mathbf{G}' \rrbracket, \quad (4)$$

where  $\mathbf{G}'$  contains the first-order derivatives of the functions  $g_i$  in the  $N$  points, *i.e.*,  $\mathbf{G}'_{ki} = g'_i(\mathbf{v}_i^T \mathbf{x}^{(k)})$ . In this way, the decoupled representation can be reconstructed from a simultaneous matrix diagonalization, or a CPD (Figure 2).

### 2.3 Decoupling functions using second-order information

Along the same lines, we may represent the Hessian  $\mathcal{H}(\mathbf{x})$  by an  $n \times m \times m$  tensor function defined as

$$\mathcal{H}_{ijk}(\mathbf{x}) = \frac{\partial^2 f_i(\mathbf{x})}{\partial x_j \partial x_k}, \quad (5)$$

which is symmetric in the second and third mode since  $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ . The Hessian tensor function has a CPD representation of the form

$$\mathcal{H}(\mathbf{x}) = \llbracket \mathbf{W}, \mathbf{V}, \mathbf{V}, (\mathbf{g}''(\mathbf{V}^T \mathbf{x}))^T \rrbracket. \quad (6)$$

By evaluating the Hessian in a set of  $N$  sampling points  $\mathbf{x}^{(k)}, k = 1, \dots, N$ , we find the CPD of the  $n \times m \times m \times N$  tensor  $\mathcal{H}$  as

$$\mathcal{H} = \llbracket \mathbf{W}, \mathbf{V}, \mathbf{V}, \mathbf{G}'' \rrbracket, \quad (7)$$

where the columns of  $\mathbf{G}''$  contain the second-order derivatives of the functions  $g_i$  in the  $N$  points, *i.e.*,  $\mathbf{G}''_{ki} = g''_i(\mathbf{v}_i^T \mathbf{x}^{(k)})$ .

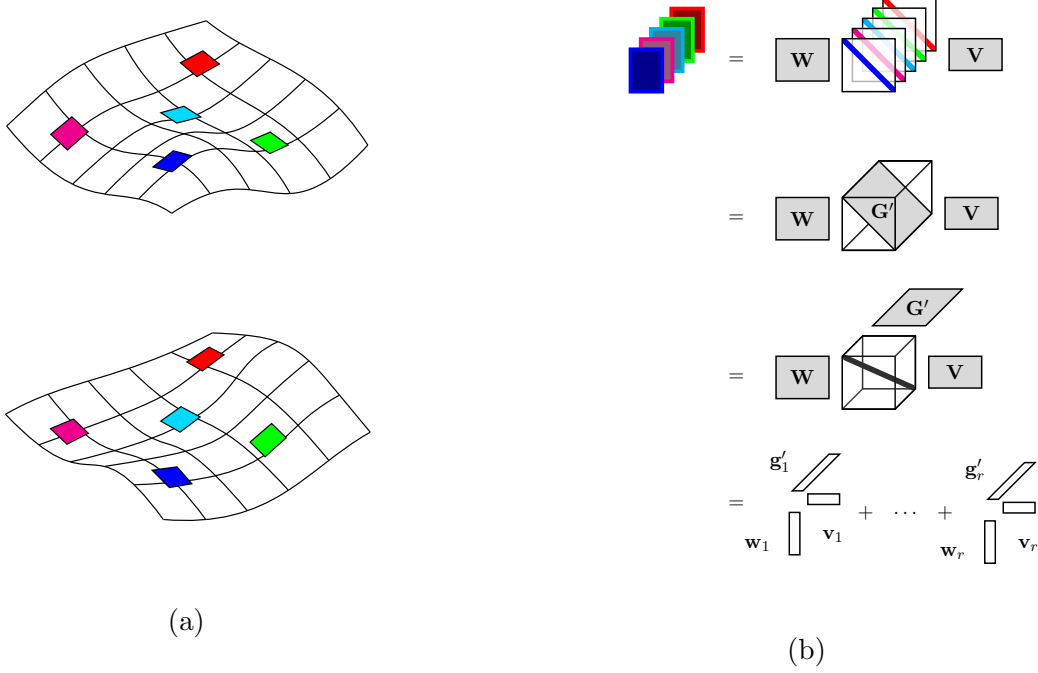


Figure 2: The first-order information of  $\mathbf{f}$  is collected in a set of sampling points  $\mathbf{x}^{(k)}$ , with  $k = 1, \dots, N$  (indicated by the colored patches on the surfaces in (a)). The corresponding Jacobian matrices  $\mathbf{J}(\mathbf{x}^{(k)})$  are arranged into a three-way tensor (b). Each Jacobian matrix can be written as  $\mathbf{J}(\mathbf{x}^{(k)}) = \mathbf{W} \text{diag}(\mathbf{g}'_i(\mathbf{v}_i^T \mathbf{x}^{(k)})) \mathbf{V}^T$ . This results in a simultaneous matrix diagonalization problem, which is computed by the CPD.

## 2.4 A joint tensor decomposition with partial symmetry

The first-order and second-order derivative information can be combined into a joint tensor decomposition with partial symmetry. This can be phrased into the Structured Data Fusion framework [15] and is implemented in tensorlab [22] for MATLAB. The underlying optimization problem is

$$\underset{\mathbf{W}, \mathbf{V}, \mathbf{G}', \mathbf{G}''}{\text{minimize}} \alpha_1 \|\mathcal{J} - [[\mathbf{W}, \mathbf{V}, \mathbf{G}']]\|_F^2 + \alpha_2 \|\mathcal{H} - [[\mathbf{W}, \mathbf{V}, \mathbf{V}, \mathbf{G}'']]\|_F^2, \quad (8)$$

where the two terms in the cost function can be given different weights  $\alpha_1$  and  $\alpha_2$ . The factor matrices  $\mathbf{W}$  and  $\mathbf{V}$  are shared among both decompositions. The partial symmetry in the Hessian tensor can be recognized in the fact that the factor  $\mathbf{V}$  occurs twice. The joint decomposition approach is visualized in Figure 3.

## 2.5 Some remarks on uniqueness

Our framework contains two notions that relate to ‘uniqueness’ or ‘identifiability’, which could be confusing. Therefore, it is useful to elaborate briefly on this matter. There is a notion of uniqueness at the level of the function decomposition (1), as well as at the level of the CPD of a (corresponding) tensor. It is important to realize that in both cases, we consider so-called ‘essential uniqueness’, which makes abstraction of trivial scaling and permutation invariances.

On the one hand, in the polynomial case of (1), the problem at hand has a rich algebraic structure, which has recently been studied in the X-rank framework, leading to novel identifiability results [5]. These results assert that, for certain choices for  $m$ ,  $n$  and  $r$ , the function (1) has a single ‘unique’ representation for generic choices of  $\mathbf{W}$ ,  $\mathbf{V}$  and  $\mathbf{g}(\mathbf{z})$ . In other words, there does not exist an equivalent representation (1) having different  $\mathbf{W}$ ,  $\mathbf{V}$  and  $\mathbf{g}(\mathbf{z})$ .

On the other hand, tensor decompositions have uniqueness properties themselves [12, 6, 11, 7, 8], which ensure that if the tensor decomposition has converged to a (numerical) zero error, then the

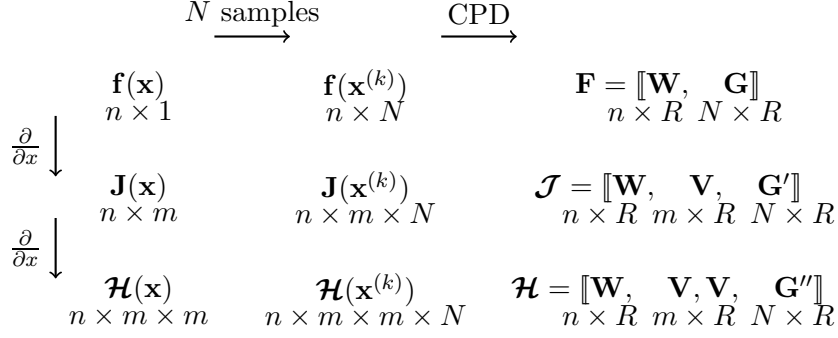


Figure 3: The proposed method takes into account a combination of first-order and second-order derivatives evaluations. These evaluations are organized into two tensors which admit a joint canonical polyadic decomposition with partial symmetry. The corresponding cost function is  $\alpha_1 \|\mathcal{J} - \llbracket \mathbf{W}, \mathbf{V}, \mathbf{G}' \rrbracket\|_F^2 + \alpha_2 \|\mathcal{H} - \llbracket \mathbf{W}, \mathbf{V}, \mathbf{V}, \mathbf{G}'' \rrbracket\|_F^2$ .

(essentially) unique underlying factorization has been retrieved. Observe that uniqueness of the tensor decomposition is a sufficient condition for uniqueness of the function decomposition (1).

Uniqueness properties of joint CPDs have only been studied recently in [16, 17]. Intuitively, it can be expected that, by imposing additional constraints on a decomposition, it is likely that uniqueness conditions are more easily met.

### 3 Numerical examples

In the current section, we will illustrate the proposed method on a number of examples. The joint decomposition (8) is implemented in tensorlab [22].

#### 3.1 Second-order derivatives for tackling the Waring decomposition

In the single-output case, the Jacobian of the multivariate scalar function  $f(\mathbf{x})$  is a vector function rather than a matrix function. Its representation simplifies to  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{g}(\mathbf{V}^T \mathbf{x})$ . Notice that  $\mathbf{w}$  could be absorbed into the function  $\mathbf{g}$ , but we have chosen to keep it explicitly in the formula to show the resemblance to the vector function case. For a scalar function  $f(\mathbf{x})$ , the Jacobian reduces to a vector function, rather than a matrix function, *i.e.*,  $\mathbf{j}^T(\mathbf{x}) = \llbracket \mathbf{w}^T, \mathbf{V}, (\mathbf{g}'(\mathbf{V}^T \mathbf{x}))^T \rrbracket$ . Evaluating the Jacobian in a set of sampling points then gives rise to a matrix, rather than a tensor. Summarizing, the  $n \times m \times N$  Jacobian tensor  $\mathcal{J}$  reduces in this situation to an  $m \times N$  matrix  $\mathbf{J}$ , and we obtain a matrix factorization question, rather than the third-order tensor CPD (4). It is easy to understand that there is no unique solution, since one can insert  $\mathbf{M}\mathbf{M}^{-1}$ , with  $\mathbf{M}$  an invertible  $R \times R$  matrix, to obtain an equivalent factorization  $\mathbf{J} = (\mathbf{V}\mathbf{M})(\mathbf{G}'\mathbf{M}^{-T}) = \tilde{\mathbf{V}}\tilde{\mathbf{G}}'^T$ .

A possible solution to resolve this lack of uniqueness is to consider second-order derivative information of  $f(\mathbf{x})$ . We evaluate the  $m \times m$  Hessian function

$$\mathbf{H}_{ij}(\mathbf{x}) = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}, \quad (9)$$

in a set of  $N$  sampling points  $\mathbf{x}^{(k)}$ . This gives an  $m \times m \times N$  tensor  $\mathcal{H}$  with

$$\mathcal{H} = \llbracket \mathbf{w}^T, \mathbf{V}, \mathbf{V}, \mathbf{G}'' \rrbracket, \quad (10)$$

as in (7). Now, we have again a case in which uniqueness of the CPD is attainable.

For instance, we consider the function

$$f(x_1, x_2) = -37x_1^3 - 213x_1^2x_2 - 399x_1x_2^2 + 5x_1 - 239x_2^3 + 9x_2 - 2, \quad (11)$$

which can be decomposed with

$$\mathbf{w}^T = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad (12)$$

and

$$\begin{aligned} g_1(z_1) &= 3z_1^3 - z_1 + 5, \\ g_2(z_2) &= -5z_2^3 + 3z_2 - 7. \end{aligned} \quad (13)$$

We draw an i.i.d. set of sampling points  $\mathbf{x}^{(k)}, k = 1, \dots, 200$  from a uniform distribution between  $-10$  and  $10$  in both components. The Jacobian matrix is hence a  $2 \times 200$  matrix  $\mathbf{J}$ , which admits a non-unique rank-two factorization. If we take into account the Hessian information, the problem again becomes a uniquely defined tensor question. The CPD of  $\mathcal{H}$  is computed using tensorlab [22], and retrieves up to a scaling and permutation invariance, the true factors  $\mathbf{V}$  and  $\mathbf{G}''$ .

### 3.2 Second-order derivatives improve uniqueness of the CPD

For a general vector function  $\mathbf{f}(\mathbf{x})$ , including the Hessian information can improve the uniqueness properties beyond the Jacobian tensor method. For instance, in the case  $m = n = 2$ , the bound by [9] ensures uniqueness up to  $r \leq 2$ . It can be verified that the Jacobian-based CPD method is not able to retrieve the underlying model. However, [5] asserts that  $r = 3$  is still identifiable for polynomial models (with degree  $d \geq 3$ ). Considering the second-order information then leads to a CPD that is generically unique.

We consider the function  $\mathbf{f}(\mathbf{x})$  which is defined as

$$\begin{aligned} f_1(x_1, x_2) &= 24x_1^3 + 36x_1^2x_2 - 4x_1^2 + 18x_1x_2^2 - 4x_1x_2 + 84x_2^3 - x_2^2 - 6x_2 + 7, \\ f_2(x_1, x_2) &= -43x_1^3 - 72x_1^2x_2 + 8x_1^2 - 36x_1x_2^2 + 8x_1x_2 - 3x_1 \\ &\quad + 75x_2^3 + 2x_2^2 - 6x_2 - 1, \end{aligned} \quad (14)$$

and admits a representation with

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 1 \\ -2 & -1 & 1 \end{bmatrix} \text{ and } \mathbf{V} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix}, \quad (15)$$

and

$$\begin{aligned} g_1(z_1) &= 3z_1^3 - z_1^2 + 5, \\ g_2(z_2) &= -5z_2^3 + 3z_2 - 7, \\ g_3(z_3) &= 3z_3^3 - 2z_3 + 2. \end{aligned} \quad (16)$$

A set of sampling points  $\mathbf{x}^{(k)}, k = 1, \dots, 200$  is sampled again uniformly on  $[-10, 10]^2$ . Applying the Jacobian method results in a  $2 \times 2 \times 200$  tensor which does not satisfy the uniqueness conditions. Indeed, although the CPD has an error of the order  $10^{-11}$ , it does not return the correct factors. However, if we compute the CPD of the Hessian tensor, the underlying representation is found.

### 3.3 Can we go beyond identifiable structures?

In our experiments, we have observed a number of cases where adding Hessian info ensures interpretability, while the underlying model structure does not seem to be identifiable. For instance consider the  $m = n = 2$  and  $r = 4$  case of polynomials of degree  $d = 3$ . We consider a function  $\mathbf{f}(\mathbf{x})$  of the form (1) with

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 1 & 2 \\ -2 & -1 & 1 & 3 \end{bmatrix}, \text{ and } \mathbf{V} = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 0 & 3 & -1 \end{bmatrix}, \quad (17)$$

and

$$\begin{aligned} g_1(z_1) &= 3z_1^3 - z_1^2 + 5, \\ g_2(z_2) &= -5z_2^3 + 3z_2 - 7, \\ g_3(z_3) &= 3z_3^3 - 2z_3 + 2, \\ g_4(z_4) &= z_4^3 - 2z_4^2 + 1. \end{aligned} \quad (18)$$

We observe that by considering the second-order derivatives information, we are able to retrieve a decomposition having (numerical) zero error. However, the linear transformations  $\mathbf{W}$  and  $\mathbf{V}$  are *not equal* (up to scaling and permutation) to the underlying factors. Nevertheless, when investigating the factor  $\mathbf{G}''$ , we see that the retrieved factor does have in its components a set of linear relations as expected from  $\mathbf{G}''_{ki} = g''_i(\mathbf{v}_i^T \mathbf{x}^{(k)})$ . The fact that an interpretable model is obtained is a surprising result: it seems to suggest that, considering the second-order information enforces that only ‘interpretable’ models are retrieved. However, we should mention that we have observed that this effect does not always hold for other cases.

## 4 Conclusions and perspectives

In this article, we generalized a tensor-based method for finding a decoupled representation of a given nonlinear multivariate vector function. The method works by evaluating second-order derivatives in a set of sampling points. First-order and second-order information can be combined in this way into a simultaneous higher-order tensor decomposition task with partial symmetry. We illustrated the promising abilities of this approach on a number of simulation examples. The method was shown to outperform the existing approach: uniquely identifiable structures are recovered in a greater number of configurations, such as the single-output case, and cases in which the Jacobian tensor method was not able to ensure uniqueness. We also observed that the method seems to extract meaningful representations even in some cases when the model structure is not uniquely identifiable, which is a property that merits further investigation.

In future work, we want to investigate how function evaluations can be taken into account: this seems to make sense when the corresponding (matrix) factorization is a low-rank approximation, which occurs when  $r < n$ . Also the use of higher-order derivatives is a possible extension, which might improve the uniqueness conditions even further. In this sense, the results in this article are a step towards a more general data-driven and noise-robust tensor-based framework for decoupling function representations.

## Acknowledgments

This work was supported in part by the Flemish Government (Methusalem), and by the Fonds Wetenschappelijk Onderzoek – Vlaanderen under EOS Project no 30468160 and research projects G.0280.15N and G.0901.17N.

## References

- [1] A. Białynicki-Birula and A. Schinzel. Representations of multivariate polynomials as sums of polynomials in linear forms. *Colloq. Math.*, 112(2):201–233, 2008.
- [2] E. Carlini and J. Chipalkatti. On Waring’s problem for several algebraic forms. *Comment. Math. Helv.*, 78:494–517, 2003.
- [3] J. Carroll and J. Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3):283–319, 1970.
- [4] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM J. Matrix Anal. Appl.*, 30(3):1254–1279, 2008.
- [5] P. Comon, Y. Qi, and K. Usevich. Identifiability of an X-rank decomposition of polynomial maps. *SIAM J. Appl. Algebra and Geometry*, 1(1):388–414, 2017.
- [6] L. De Lathauwer. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM J. Matrix Anal. Appl.*, 28(3):642–666, 2006.

- [7] I. Domanov and L. De Lathauwer. On the uniqueness of the canonical polyadic decomposition of third-order tensors — part I: Basic results and uniqueness of one factor matrix. *SIAM J. Matrix Anal. Appl.*, 34(3):855–875, 2013.
- [8] I. Domanov and L. De Lathauwer. On the uniqueness of the canonical polyadic decomposition of third-order tensors — part II: Uniqueness of the overall decomposition. *SIAM J. Matrix Anal. Appl.*, 34(3):876–903, 2013.
- [9] P. Dreesen, M. Ishteva, and J. Schoukens. Decoupling multivariate polynomials using first-order information and tensor decompositions. *SIAM J. Matrix Anal. Appl.*, 36(2):864–879, 2015.
- [10] R. A. Harshman. Foundations of the PARAFAC procedure: Model and conditions for an “explanatory” multi-mode factor analysis. *UCLA Working Papers in Phonetics*, 16(1):1–84, 1970.
- [11] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, September 2009.
- [12] J. B. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Lin. Algebra Appl.*, 18:95–138, 1977.
- [13] L. Oeding and G. Ottaviani. Eigenvectors of tensors and algorithms for Waring decomposition. *J. Symb. Comp.*, 54:9–35, July 2013.
- [14] A. Schinzel. On a decomposition of polynomials in several variables. *J. de Théorie des Nombres de Bordeaux*, 14(2):647–666, 2002.
- [15] L. Sorber, M. Van Barel, and L. De Lathauwer. Structured data fusion. *IEEE J. Sel. Top. Signal Process.*, 9(4):586–600, 2015.
- [16] M. Sørensen and L. De Lathauwer. Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear Rank- $(L_{r,n}, L_{r,n}, 1)$  terms—part I: Uniqueness. *SIAM. J. Matrix Anal. Appl.*, 36(2):496–522, 2015.
- [17] M. Sørensen and L. De Lathauwer. Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear Rank- $(L_{r,n}, L_{r,n}, 1)$  terms—part II: Algorithms. *SIAM. J. Matrix Anal. Appl.*, 36(3):1015–1045, 2015.
- [18] K. Tiels and J. Schoukens. From coupled to decoupled polynomial representations in parallel Wiener-Hammerstein models. In *Proc. 52nd IEEE Conf. Decis. Control (CDC)*, pages 4937–4942, Florence, Italy, 2013.
- [19] K. Usevich. Decomposing multivariate polynomials with structured low-rank matrix completion. *Proc. 21st Int. Symp. on Mathematical Theory of Networks and Systems (MTNS 2014)*, pages 1826–1833, Groningen, The Netherlands, 2014.
- [20] K. Usevich, P. Dreesen, and M. Ishteva. Decoupling multivariate polynomials: interconnections between tensorizations. 2017. (preprint arXiv:1703.02493).
- [21] A. Van Mulders, L. Vanbeylen, and K. Usevich. Identification of a block-structured model with several sources of nonlinearity. 2014. Proceedings of the 13th European Control Conference.
- [22] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer. Tensorlab 3.0, 2016. Available online, Mar. 2016. URL: <http://www.tensorlab.net/>.