



HAL
open science

Breaking the curse of dimensionality for coupled matrix-tensor factorization

Abdelhak Boudehane, Yassine Zniyed, Arthur Tenenhaus, Laurent Le Brusquet, Remy Boyer

► **To cite this version:**

Abdelhak Boudehane, Yassine Zniyed, Arthur Tenenhaus, Laurent Le Brusquet, Remy Boyer. Breaking the curse of dimensionality for coupled matrix-tensor factorization. 8th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2019), Dec 2019, Le Gosier, Guadeloupe. 10.1109/camsap45676.2019.9022462 . hal-02417126

HAL Id: hal-02417126

<https://hal.science/hal-02417126v1>

Submitted on 18 Dec 2019

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.

BREAKING THE CURSE OF DIMENSIONALITY FOR COUPLED MATRIX-TENSOR FACTORIZATION

Abdelhak Boudehane* Yassine Zniyed* Arthur Tenenhaus* Laurent Le Brusquet* Remy Boyer‡

*Laboratoire des Signaux et Systèmes, CentraleSupélec, CNRS, Univ. Paris-Sud, Univ. Paris-Saclay, Gif-sur-Yvette, France

‡ Centre de Recherche en Informatique, Signal et Automatique de Lille, Univ. Lille 1, Villeneuve-d’Ascq, France

ABSTRACT

In different application fields, heterogeneous data sets are structured into either matrices or higher-order tensors. In some cases, these structures present the property of having common underlying factors, which is used to improve the efficiency of factor-matrices estimation in the process of the so-called coupled matrix-tensor factorization (CMTF). Many methods target the CMTF problem relying on alternating algorithms or gradient approaches. However, computational complexity remains a challenge when the data sets are tensors of high-order, which is linked to the well-known “curse of dimensionality”. In this paper, we present a methodological approach, using the Joint dDimensionality Reduction And Factors rEtrieval (JIRAFE) algorithm for joint factorization of high-order tensor and matrix. This approach reduces the high-order CMTF problem into a set of 3-order CMTF and canonical polyadic decomposition (CPD) problems. The proposed algorithm is evaluated on simulation and compared with a gradient-based method.

Index Terms— Coupled matrix tensor decomposition, Tensor train, Heterogeneous data analysis, Joint estimation, Fast algorithms.

1. INTRODUCTION

The last decade has witnessed a considerable evolution in data processing and a major improvement in data analysis techniques [1]. Moreover, the performance of analysis algorithms is subject to great advancement, setting as objective to reduce the processing time. The different types of data are exposed to noise that requires sophisticated robust algorithms to have an acceptable level of performance. The heterogeneous data sets are present, nowadays, in multiple fields such as the bio-medical [2], or recommendation systems and social networks [3], to mention a few. These data sets may be structured in matrices or tensors of low and high-order. According to the structure of the data sets, specific tools has to be used for a proper processing. Many of these data structures present the property of sharing common underlying factors, which may be used to improve the factor estimation. Therefore, joint analysis of coupled data set is promising

framework for common factor estimation. This joint analysis has been applied in the context of clustering application [4], where the coupled estimation of the common factor is more efficient than its estimation using the tensor or the matrix only, or when only the common factor is required, as in [5], where the data sets are multiple images recording the same object/phenomenon, under different lightening/angle conditions. On the other hand, high-order tensors are widely used in data modeling [6], where the goal is to extract the factor matrices. However, high-order tensors factorization remains a challenge as the number of elements and the computational complexity increase exponentially with respect to the order, which requires more memory and computational power [6]. Nevertheless, as the need of big data sets is emerging in more and more fields, the development of new techniques for high-order tensors factorization seems to be inevitable.

Various solutions have been proposed in literature to the CMTF problem, such as the constrained alternating least squares (ALS) algorithm presented in [7], treating the particular case of pattern and topic detection. In this framework, the physical interpretability required for factors is behind the use of the constraints, which need to be suited for each case with respect to the physical requirements. Furthermore, an all-at-once optimization algorithm, based on gradient approach, was presented in [4], for different CMTF scenarios. However, the well known all-at-once optimization methods’ performance is still a difficult issue, especially in high-order tensors case.

To the best of our knowledge, the CMTF problem for the case of high-order tensors has not been treated in literature. However, in [8, 9], the “Joint dDimensionality Reduction And Factors rEtrieval” (JIRAFE) scheme, based on the tensor train (TT) model [7], has been introduced for high-order canonical polyadic decomposition (CPD) [10]. The latter work uses the equivalence CPD-TT to break the curse of dimensionality for high-order CPD tensors problem, *i.e.* reducing a high-order CPD into 3-order CPDs. As a result, it yields to slower growth of the computational complexity with respect to the order of the tensor.

In this work, we propose a methodological approach to solve a coupled matrix and high-order tensor factorization problem. The goal of this approach is to reduce this prob-

lem into 3-order CMTF and a set of 3-order CPD problems. For this, we propose a JIRAFE-based scheme, adapted to the case of coupled matrix and high-order tensor factorization, to obtain a semi-closed-form solution.

Notations: Vectors, matrices and tensors are represented by \mathbf{x} , \mathbf{X} and \mathcal{X} , respectively. The symbols $(\cdot)^T$ and \odot denote, respectively, the transpose and Khatri-Rao product. The Frobenius norm is defined by $\|\cdot\|_F$. The matrix $\mathbf{X}^{(k)}$ of size $N_k \times N_1 \cdots N_{k-1} N_{k+1} \cdots N_Q$ refers to the k -mode unfolding of \mathcal{X} of size $N_1 \times \cdots \times N_Q$. The n -mode product is denoted by \times_n and the tensor contraction by \times_q^p [11].

2. COUPLED MATRIX-TENSOR MODELIZATION

2.1. Generalized Coupled Matrix and High-Order Tensor Model

In this section, we will address the generalized CMTF problem of a Q -order tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_Q}$ coupled with matrices $\mathbf{Y}_k \in \mathbb{R}^{N_k \times N_m}$ on the k -th modes, i.e., each k -th factor matrix is common in both the tensor \mathcal{X} and the corresponding matrix \mathbf{Y}_k . The tensor \mathcal{X} is assumed to follow a CP model, i.e., $\mathcal{X} = \llbracket \mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_Q \rrbracket$ where $\mathbf{P}_q \in \mathbb{R}^{N_q \times R}$ is the q -th mode factor matrix and R is the canonical rank of the tensor, while the matrices \mathbf{Y}_k are subject to a rank factorization, given by $\mathbf{Y}_k = \mathbf{P}_k \mathbf{V}_k^T$ where \mathbf{P}_k is the common factor matrix and $\mathbf{V}_k \in \mathbb{R}^{N_m \times R}$. In order to jointly estimate the factor matrices of the tensors and the matrix, the objective function to be minimized is given by [4, eq. (3)] as:

$$f(\mathbf{P}_1, \dots, \mathbf{P}_Q, \mathbf{V}_1, \dots, \mathbf{V}_K) = \|\mathcal{X} - \llbracket \mathbf{P}_1, \dots, \mathbf{P}_Q \rrbracket\|_F^2 + \sum_{k=1}^K \|\mathbf{Y}_k - \mathbf{P}_k \mathbf{V}_k^T\|_F^2$$

For the sake of simplicity, and without a loss of generality, we solve the above problem with a single matrix coupled with the tensor in eq. (1), i.e.,

$$f(\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_Q, \mathbf{V}_k) = \|\mathcal{X} - \llbracket \mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_Q \rrbracket\|_F^2 + \|\mathbf{Y}_k - \mathbf{P}_k \mathbf{V}_k^T\|_F^2. \quad (1)$$

The optimization problem (1) is depicted in Fig. 1 using factor graphs, where the tensor \mathcal{X} is represented by an identity tensor node surrounded by the factors represented by the nodes $\mathbf{P}_1, \dots, \mathbf{P}_Q$, while the common factor \mathbf{P}_k with the matrix \mathbf{V}_k constitute the matrix \mathbf{Y}_k . Note that the edges represent the matrices dimensions.

2.2. Tensor-Train representation of CMTF

Tensor \mathcal{X} of rank R , under a CP model assumption and all the factor matrices ($\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_Q$) in Section 2 have a rank equal to the tensor's canonical rank, can be decomposed into

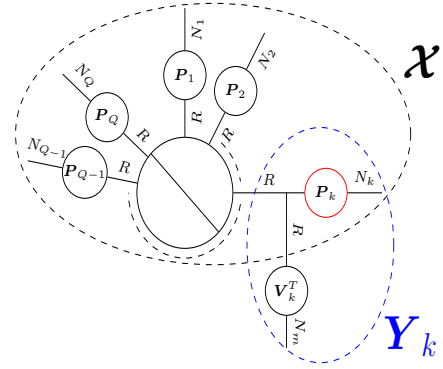


Fig. 1. Factor graph modelization of CMTF

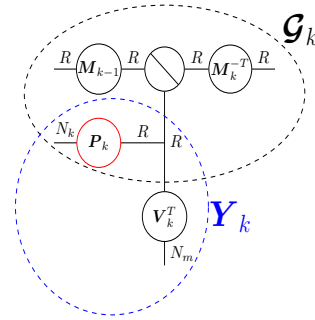


Fig. 2. Factor graph modelization of 3-order CMTF

a train of 3-order tensors and two matrices, of rank R , using the tensor train singular value decomposition (TT-SVD) algorithm [12]. we can write :

$$\mathcal{X} = \mathbf{G}_1 \times_2^1 \mathbf{G}_2 \times_3^1 \dots \times_{Q-1}^1 \mathbf{G}_{Q-1} \times_Q^1 \mathbf{G}_Q$$

where $\mathbf{G}_1 = \mathbf{P}_1 \mathbf{M}_1^{-1}$, $\mathbf{G}_Q = \mathbf{M}_{Q-1} \mathbf{P}_Q^T$ are the TT matrices, subject to matrix rank factorization, and $\mathbf{G}_q = \llbracket \mathbf{M}_{q-1}, \mathbf{P}_q, \mathbf{M}_q^{-T} \rrbracket$, for $q = \{2, 3, \dots, Q-1\}$, are the core tensors following a CP model, while \mathbf{M}_q of size $R \times R$, are the change-of-basis matrices for $q = \{1, 2, \dots, Q-1\}$ [8]. Note that the TT-core tensors and matrices are of a rank equal to the canonical rank of \mathcal{X} . However, the k^{th} TT-core tensor \mathbf{G}_k shares the common mode \mathbf{P}_k with the matrix \mathbf{Y}_k as represented in Fig. 2 using factor graphs.

3. PROPOSED SCHEME: C-JIRAFE

The structure of the core tensor \mathbf{G}_k resulting from the TT decomposition and sharing the common factor with the matrix $\mathbf{Y}_k = \mathbf{P}_k \mathbf{V}_k^T$ is given by:

$$\mathbf{G}_k = \llbracket \mathbf{M}_{k-1}, \mathbf{P}_k, \mathbf{M}_k^{-T} \rrbracket$$

The equivalence between the CMTF problem presented in Section 2.1 and the tensor-train representation described in

Section 2.2 allow us to rewrite the objective function in eq. (1) in two steps. The first step targets the estimation of the core tensors \mathcal{G}_q .

$$f_1(\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_{Q-1}, \mathbf{G}_Q) = \|\mathcal{X} - \mathbf{G}_1 \times_2^1 \mathbf{G}_2 \times_3^1 \dots \times_{Q-1}^1 \mathbf{G}_{Q-1} \times_Q^1 \mathbf{G}_Q\|_F^2 \quad (2)$$

This step can be handled by estimating the core tensors \mathcal{G}_q in eq. (2) using the TT-SVD algorithm.

The second step is the estimation of the common factor thanks to the prior estimation of the correspondding TT-core \mathcal{G}_k sharing the common \mathbf{P}_k mode with the matrix \mathbf{Y}_k . The estimation of the common factor boils down to minimize the following objective function:

$$f_2(\mathbf{M}_{k-1}, \mathbf{P}_k, \mathbf{M}_k, \mathbf{V}_k) = \lambda_1 \|\mathcal{G}_k - \llbracket \mathbf{M}_{k-1}, \mathbf{P}_k, \mathbf{M}_k \rrbracket\|_F^2 + \lambda_2 \|\mathbf{Y}_k - \mathbf{P}_k \mathbf{V}_k^T\|_F^2 \quad (3)$$

where λ_1 and λ_2 are introduced to balance the weight of the two terms in eq. (3). In the literature, a usual choice is to set $\lambda_1 = \frac{1}{R^2 N_k}$ and $\lambda_2 = \frac{1}{N_k N_m}$, which are the inverse of number of elements of the tensor and the matrix, respectively [13].

The objective function in eq. (3) can be minimized using a joint order-three ALS, where the common factor matrix is estimated jointly. For the joint estimation, we use the ALS-based solution given in [4]. We define the matrix $\tilde{\mathbf{G}}_k^{(2)}$ resulting from the second mode unfolding of tensor \mathcal{G}_k concatenated with the matrix \mathbf{Y}_k :

$$\tilde{\mathbf{G}}_k^{(2)} = \begin{bmatrix} \lambda_1 \mathbf{G}_k^{(2)} \\ \lambda_2 \mathbf{Y}_k \end{bmatrix} = \mathbf{P}_k \begin{bmatrix} \lambda_1 \mathbf{\Pi}^T \\ \lambda_2 \mathbf{V}_k^T \end{bmatrix}$$

where $\mathbf{\Pi} = \mathbf{M}_k^{-T} \odot \mathbf{M}_{k-1}$

As a consequence, we define the joint criterion as :

$$\tilde{f} = \left\| \begin{bmatrix} \lambda_1 \mathbf{G}_k^{(2)} \\ \lambda_2 \mathbf{Y}_k \end{bmatrix} - \mathbf{P}_k \begin{bmatrix} \lambda_1 \mathbf{\Pi}^T \\ \lambda_2 \mathbf{V}_k^T \end{bmatrix} \right\|_F^2$$

for which, the least-squares solution for the common factor matrix is given by:

$$\mathbf{P}_k = \begin{bmatrix} \lambda_1 \mathbf{G}_k^{(2)} & \lambda_2 \mathbf{Y}_k \end{bmatrix} \begin{bmatrix} \lambda_1 \mathbf{\Pi}^T & \lambda_2 \mathbf{V}_k^T \end{bmatrix}^\dagger \quad (4)$$

where the matrix \mathbf{V}_k is estimated as:

$$\mathbf{V}_k = (\mathbf{P}_k^T)^\dagger \mathbf{Y}_k \quad (5)$$

A closed-form expression is obtained for the common factor matrix. The estimation of the factors of the tensor is obtained sequentially, using the JIRAFE algorithm [8]. The coupled JIRAFE (C-JIRAFE) algorithm is fully detailed in Algorithm 1.

Note that **C-TriALS** is a 3-order ALS where the estimation of the common mode is done using eq. (4) and an extra

Algorithm 1 Coupled JIRAFE

Input: Tensor \mathcal{X} , matrix \mathbf{Y}_k , order Q , rank R and common mode k

Output: Estimated factor matrices $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_Q, \mathbf{V}_k$

1: TT-cores estimation

$[\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_{Q-1}, \mathbf{G}_Q] = \text{TT-SVD}(\mathcal{X}, R)$

2: Joint estimation of \mathbf{P}_k and \mathbf{V}_k :

$[\hat{\mathbf{M}}_{k-1}, \hat{\mathbf{P}}_k, \hat{\mathbf{M}}_k^{-T}, \hat{\mathbf{V}}_k] = \text{C-TriALS}(\mathcal{G}_k, \mathbf{Y}_k, R)$

3: Estimation of the rest of factor matrices

4: **for** $q = k - 1 \dots 2$

5: $[\hat{\mathbf{M}}_{q-1}, \hat{\mathbf{P}}_q] = \text{BiALS}(\mathcal{G}_q, \hat{\mathbf{M}}_q^{-T}, R)$

6: **end for**

7: **for** $q = k + 1 \dots Q - 1$

8: $[\hat{\mathbf{P}}_q, \hat{\mathbf{M}}_q^{-T}] = \text{BiALS}(\mathcal{G}_q, \hat{\mathbf{M}}_{q-1}, R)$

9: **end for**

10: $\hat{\mathbf{P}}_1 = \mathbf{G}_1 \hat{\mathbf{M}}_1$

11: $\hat{\mathbf{P}}_Q = \mathbf{G}_Q^T \hat{\mathbf{M}}_{Q-1}^{-T}$

step is added to estimate the factor matrix \mathbf{V}_k according to eq. (5). Moreover, it is worth mentioning that **BiALS** is a **TriALS** with a pre-estimated factor.

At last, steps 4-6 and 7-9 are independent and can be done in parallel (see fig.3). Furthermore, in applications, such as clustering, where only the common factor is needed [5], only the steps 1 and 2 of algorithm 1 are to be executed since the estimation of the common factor is done independently, which means that there is no need to estimate the rest of the factor matrices.

4. SIMULATION RESULTS

In this section we present the numerical simulations of our approach, that we compare with the gradient based all-at-once optimization presented in [4]. In order to calculate the gradient, we use the derivatives given in [4], and add α_1 and α_2 to the expression for equal contribution between the tensor and the matrix [13]. The resulting objective function becomes:

$$\tilde{f}(\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_Q, \mathbf{V}_k) = \alpha_1 \|\mathcal{X} - \llbracket \mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_Q \rrbracket\|_F^2 + \alpha_2 \|\mathbf{Y}_k - \mathbf{P}_k \mathbf{V}_k^T\|_F^2 \quad (6)$$

where $\alpha_1 = \frac{1}{N_1 N_2 \dots N_Q}$ and $\alpha_2 = \frac{1}{N_k N_m}$.

The initial factor matrices are generated randomly following the standard normal distribution. A random noise is then added to the tensor and the matrix with the same noise-variance (same signal to noise ratio). Furthermore, we allow the same maximum number of iterations for both Algorithm 1 and the gradient based approach. While the number of iterations for Algorithm 1 is equal to the number of iterations within the ALS steps (10^5 iteration per ALS), we allow a number of evaluation functions equal to the iteration number in the gradient based approach. Both algorithms are

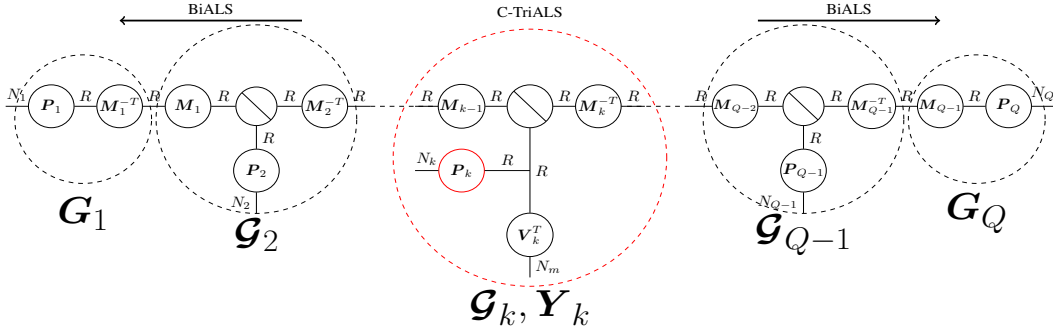


Fig. 3. Coupled JIRAFE

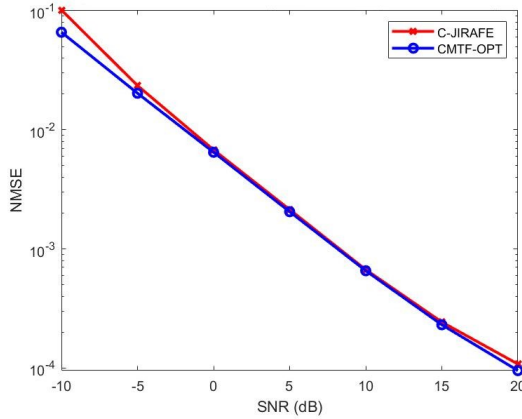


Fig. 4. NMSE (log-scale) in function of signal to noise ratio (dB) for a 4-order tensor coupled with a matrix ($Q = 4, R = 2, N = 10$)

Tensor order	C-JIRAFE	CMTF-OPT	Gain
4	0.0201 (s)	0.1124 (s)	5.5920
5	0.0310 (s)	1.3369 (s)	43.1258
6	0.1057 (s)	20.5572 (s)	194.4863

Table 1. Execution time in function of the order for $SNR = 10dB$

stopped if $\frac{\|f_{old} - f_{new}\|}{f_{old}} \leq 10^{-8}$, or when $f \leq 10^{-8}$, where f is given by eq. (3) for the **C-TriALS**, $\|\hat{\mathcal{G}}_q - \mathcal{G}_q\|$ for the **BiALS** and given by eq. (6) for CMTF-OPT. For the evaluation, we calculate the reconstruction normalized mean square error (NMSE) given by $\frac{\|\hat{\mathcal{X}} - \mathcal{X}\|^2}{\|\mathcal{X}\|^2}$. We evaluate the estimation performance by calculating the calculation time for both algorithms. In Fig. 4, C-JIRAFE stands for sequential approach that we present in this paper, while CMTF-OPT (CMTF Optimization) is the gradient-based approach presented in [4]. The NMSE of both algorithm presents no significant differ-

ence. For a fixed signal to noise ratio, and for the same NMSE value, we notice in Table 1 an important difference in execution time, in the favor of the sequential approach (C-JIRAFE). It is also clear that the gain in terms of execution time increases with respect to the order of the tensor, due to the slow increase of computational complexity in C-JIRAFE case, versus the exponential increase for CMTF-OPT.

We mention that, in our simulations, C-JIRAFE is run on a moncore architecture, and a more interesting gain in execution time on multicore architecture is expected.

5. CONCLUSION

The coupled factorization takes advantages of the particular structures present in heterogeneous data sets in order to improve the factors estimation. Yet, the high-order tensors factorization remains a challenge due to the exponential increase of the number of elements, and so the algorithmic complexity with respect to the order. In this paper, we presented a coupled matrix tensor factorization approach named C-JIRAFE. This method targets the joint factorization of high-order tensors with matrices, taking advantage of the dimensionality reduction provided by the tensor train model. Moreover, we have shown that, for equivalent error level, the execution time for C-JIRAFE was significantly lower than the state of art method based on the gradient approach. Also, as we increased the tensor order, the gap of execution time between our approach and the reference kept increasing, thanks to the slower increase of complexity provided by the tensor train model. Finally, this approach allows us to extract the common factor independently, which means that, for applications where only the common factor is targeted, our approach helps estimating this factor without the obligation to estimate all the factors.

References

- [1] J. Pokorný, P. Škoda, I. Zelinka, D. Bednárek, F. Zavoral, M. Kruliš, and P. Šaloun, *Big Data Movement: A Challenge in Data Processing*, pp. 29–69, Springer International Publishing, Cham, 2015.
- [2] F. Cong, Q. H. Lin, L. D. Kuang, X. F. Gong, P. Astikainen, and T. Ristaniemi, “Tensor decomposition of eeg signals: A brief review,” *Journal of Neuroscience Methods*, vol. 248, pp. 59 – 69, 2015.
- [3] V. N. Ioannidis, A. S. Zamzam, G. B. Giannakis, and N. D. Sidiropoulos, “Coupled Graphs and Tensor Factorization for Recommender Systems and Community Detection,” *IEEE Transactions on Knowledge and Data Engineering*, 2018.
- [4] E. Acar, T. G. Kolda, and D. M. Dunlavy, “All-at-once Optimization for Coupled Matrix and Tensor Factorizations,” in *9th Workshop on Mining and Learning with Graphs, San Diego, CA*, 2011.
- [5] I. Kisil, G. G. Calvi, and D. P. Mandic, “Tensor Valued Common and Individual Feature Extraction: Multi-dimensional Perspective,” *International Conference on Acoustics, Speech and Signal Processing*, 2017.
- [6] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, “Tensor decomposition for signal processing and machine learning,” *Trans. Sig. Proc.*, vol. 65, no. 13, pp. 3551–3582, 2017.
- [7] S. Bahargam and E. E. Papalexakis, “A constrained coupled matrix-tensor factorization for learning time-evolving and emerging topics,” *CoRR*, vol. abs/1807.00122, 2018.
- [8] Y. Zniyed, R. Boyer, A. L. F. de Almeida, and G. Favier, “High-order cpd estimation with dimensionality reduction using a tensor train model,” pp. 2613–2617, 2018.
- [9] Y. Zniyed, R. Boyer, A. L.F. de Almeida, and G. Favier, “Multidimensional harmonic retrieval based on vandermonde tensor train,” *Signal Processing*, vol. 163, pp. 75 – 86, 2019.
- [10] R. A. Harshman, “Foundations of the PARAFAC procedure: Models and conditions for an ”explanatory” multimodal factor analysis,” *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.
- [11] A. Cichocki, “Era of big data processing: A new approach via tensor networks and tensor decompositions,” *CoRR*, 2014.
- [12] I. Oseledets, “Tensor-train decomposition,” *SIAM Journal on Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, 2011.
- [13] Q. Wu, J. Wang, J. Fan, G. Xu, J. Wu, B. Johnson, X. Li, Q. Do, and R. Ge, “Improved coupled tensor factorization with its applications in health data analysis,” *Complexity*, p. 16, 2019.