

Joint rank and factor estimation for canonical polyadic decomposition using group sparsity: Application to fluorescence spectroscopy

Nasrin TAHERI^{1,2}, Amar KACHENOURA¹, Xavier LUCIANI³, Laurent ALBERA¹, Lotfi SENHADJI¹

¹Univ Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France,

²Department of Electrical Engineering, Faculty of Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran,

³Aix Marseille Université, Université de Toulon, CNRS, LIS, F-13000, Marseille, France

nasrin.taheri.ir@gmail.com, amar.kachenoura@univ-rennes1.fr, lucianix@gmail.com

Résumé – L’analyse tensorielle suscite de l’intérêt depuis plusieurs décennies, tant d’un point de vue théorique qu’applicatif. Différentes décompositions tensorielles permettent d’étendre la SVD matricielle aux tableaux d’ordre supérieur. Parmi elles, la décomposition polyadique canonique (CPD en anglais) permet sous certaines conditions d’écrire de manière (essentiellement) unique un tenseur comme un nombre minimal de tenseurs de rang 1. En pratique, ce modèle est bruité et requière la mise en œuvre d’algorithmes d’approximation sous contraintes de rang faible. Malheureusement le rang est rarement connu en pratique. Nous proposons une nouvelle méthode (AGS-CPD), capable d’estimer simultanément le rang et les matrices de facteurs. Elle est basée sur l’exploitation de la parcimonie de groupe et montre une plus grande robustesse en présence de bruit que les approches classiques. En effet, les résultats obtenus sur des données simulées et des données réelles non-négatives de spectroscopie de fluorescence montrent que, sans tenir compte d’aucune contrainte de non-négativité, la méthode AGS-CPD est très compétitive à la fois pour l’estimation du rang et des matrices de facteurs.

Abstract – This study addresses the problem of simultaneously estimating the Canonical Polyadic (CP) rank and factors of noisy third order tensors. The presented algorithm, namely Alternating Group Sparse CP Decomposition (AGS-CPD), promotes group sparsity of the row-wise concatenated loading matrices under the low-rank CP model, without any knowledge of the true rank. The good performance of this approach is assessed on both simulated data and real fluorescence spectroscopy data which are non-negative. The simulated and real experiments show the advantage of AGS-CPD both in rank and factor estimation without considering any non-negativity constraint.

1 Introduction

During the last decades, Canonical Polyadic Decomposition (CPD) has been widely used in many applications such as chemometrics [1], neural imaging [2] and brain source imaging [3, 4]. Among classical approaches [1, 5], even if some of them are more robust to the overfactoring problem such as the semi-algebraic DIAG technique [6], knowing the true rank of a tensor is essential to decompose it efficiently. Therefore, these algorithms require a preprocessing step in order to estimate the rank. In practice, the considered tensor is noisy. In order to address this challenging issue, several approaches have been presented in literature. For instance, in the n -mode Minimum Description Length approach (ND-MDL) [7], MDL [8] is applied to the n -mode unfolding matrices. In the Generalized ND-MDL (GND-MDL) method [9], different sets of eigenvalues are calculated and combined to estimate the rank accurately. Even if the results of the aforementioned methods are promising, finding methods robust with respect to low Signal-to-Noise Ratio (SNR) values, which estimates jointly the rank and the factors, is still needed. To tackle this problem, our previous work [10] considered the low-rank assumption by promoting the group sparsity of each over-estimated loading matrix. The resulting methods, namely R-CPD₁ and R-CPD₂, allow us to solve the low-rank CPD problem without any knowledge of the true rank.

In chemometrics applications, the interest of CP decomposition coupled with 3D fluorescence spectroscopy to identify and track chemical components has been well established. A common issue in fluorescence spectroscopy is the number of fluorophores (i.e. the rank), which is usually unknown and must be

estimated correctly in order to obtain a good estimation of the non-negative loading matrices. In BC-VMFB [11], authors successfully combined sparsity and non-negativity constraints upon the factors in order to deal with the problem of non-negative CPD with a low unknown rank. Another constrained CPD approach, namely AO-ADMM [12], accommodates the non-negative constraint on the loading matrices, but it requires to know the true rank. In the case of non-negative data, AO-ADMM can be used as a reference method to assess the performance of other methods.

In this paper, we propose the Alternating Group Sparse CPD (AGS-CPD), which promotes the group sparsity of the row-wise concatenated loading matrices. It is robust to noise and allows us to estimate simultaneously the rank and the factors efficiently without considering any non-negativity constraint. In order to assess the good performance of AGS-CPD, we simulate two sets of data using the normal and folded normal distribution for different SNR values. Note that the latter distribution generates random non-negative data. Eventually, the proposed algorithm is applied to a set of real non-negative fluorescence data.

2 Methods

2.1 Low rank- R CPD

The rank- R CPD of any tensor $\mathcal{T} \in \mathbb{R}^{I_1, I_2, I_3}$ is given by :

$$\mathcal{T} = \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket_R = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \quad (1)$$

where \mathbf{a}_r , \mathbf{b}_r and \mathbf{c}_r are the r -th columns of the loading matrices \mathbf{A} , \mathbf{B} and \mathbf{C} , respectively. In order to estimate the rank, we assume $R \ll \min(I_1, I_2, I_3)$. According to [13, Theorem 4.a, p. 123] this assumption not only helps to estimate the rank but also guarantees the generic and deterministic uniqueness of the rank- R CPD. In practice, the recorded tensor data is corrupted by some noise :

$$\mathcal{X} = \mathcal{T} + \mathcal{N} \quad (2)$$

where \mathcal{N} is a $(I_1 \times I_2 \times I_3)$ noise tensor.

2.2 Existing (group) sparse CPD methods

To deal with different applications, we can impose other constraints to the CPD problem such as non-negativity and low-rank constraints. Thus, equation (2) can be written as :

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \lambda \mathcal{R}(\mathbf{A}, \mathbf{B}, \mathbf{C}) \quad \text{s.t.} \quad \mathcal{X} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]_R \quad (3)$$

where \mathcal{R} is a function of the three loading matrices which accommodates the considered constraints and where λ is a penalty parameter. The R-CPD approach [10], which led to the R-CPD₁ and R-CPD₂ methods, imposes the low-rank constraint by promoting group sparsity on each loading matrix. This allows us to solve the low-rank CPD problem without any knowledge of the true rank. For R-CPD₁, we have :

$$\mathcal{R}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \|\mathbf{A}\|_{1,2} + \|\mathbf{B}\|_{1,2} + \|\mathbf{C}\|_{1,2} \quad (4)$$

while for R-CPD₂, we get :

$$\mathcal{R}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \|\mathbf{A}\|_{2,1} + \|\mathbf{B}\|_{2,1} + \|\mathbf{C}\|_{2,1} \quad (5)$$

where $\|\cdot\|_{2,1}$ and $\|\cdot\|_{1,2}$ are the mixed-norms $\ell_{2,1}$ and $\ell_{1,2}$. The mixed-norm is a good convex envelope of the matrix rank. For a given matrix \mathbf{X} ($I \times J$), the mixed norms $\|\cdot\|_{2,1}$ and $\|\cdot\|_{1,2}$ are computed as the traces, $Tr[\mathbf{X}^T \Phi \mathbf{X}]$ and $Tr[\mathbf{X} \Psi \mathbf{X}^T]$, respectively, where Φ and Ψ are diagonal matrices given by $\Phi_{ii} = 1/\sqrt{\sum_{j=1}^J X_{ij}^2}$ and $\Psi_{jj} = 1/\sqrt{\sum_{i=1}^I X_{ij}^2}$. Problem (3) was solved by R-CPD₁ and R-CPD₂ using Alternating Direction Method of Multipliers (ADMM) [14]. The Block-coordinate Metric Variable Forward-Backward (BC-VMFB) algorithm [11], a block alternating proximal method which can be seen as a special case of majorize-minimize approaches, was designed to solve the following minimization problem :

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \frac{1}{2} \|\mathcal{X} - [\mathbf{A}, \mathbf{B}, \mathbf{C}]_R\|_F^2 + \mathcal{R}(\mathbf{A}, \mathbf{B}, \mathbf{C}) \quad (6)$$

where the penalization term $\mathcal{R}(\mathbf{A}, \mathbf{B}, \mathbf{C})$ promotes sparsity and non-negativity constraint of each overestimated loading matrix. For instance, sparsity can be enforced for \mathbf{A} and \mathbf{B} while non-negativity of \mathbf{C} is promoted.

2.3 The proposed method : AGS-CPD

In order to estimate the rank and the functions simultaneously, we propose an Alternating Group Sparse CPD (AGS-CPD) procedure. First, we define the matrix \mathbf{G} which is the row-wise concatenation of the three loading matrices. To estimate the rank of a tensor using the loading matrices, we impose the mixed-norm $\|\cdot\|_{1,2}$ on \mathbf{G} . Moreover, we add the CPD constraint to accurately estimate each loading matrix at the same time. Here, the objective function is similar to problem (6), where $\mathcal{R}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \lambda \|\mathbf{G}\|_{1,2}$

and $\mathbf{G} = [\mathbf{A}^T, \mathbf{B}^T, \mathbf{C}^T]$. The minimization problem can be solved by minimizing the following function :

$$\mathcal{K}(\mathbf{G}) = \lambda Tr[\mathbf{G} \Psi \mathbf{G}^T] + \|\mathcal{X} - \mathcal{T}\|_F^2 = \lambda \left\{ Tr[\mathbf{A} \Psi_{\mathbf{G}} \mathbf{A}^T] + Tr[\mathbf{B} \Psi_{\mathbf{G}} \mathbf{B}^T] + Tr[\mathbf{C} \Psi_{\mathbf{G}} \mathbf{C}^T] \right\} + \|\mathcal{X} - \mathcal{T}\|_F^2 \quad (7)$$

where λ is a penalty parameter that should be tuned. Here, it is noted that the matrix Ψ is computed using \mathbf{G} and it is common in the traces of equation (7). To compute the update functions of the loading matrices, the derivatives of \mathcal{K} with respect to each loading matrix are vanished. The loading matrices are then updated as follows :

$$\mathbf{A} = \mathbf{X}^{(1)} (\mathbf{C} \odot \mathbf{B}) \{ \lambda \Psi_{\mathbf{G}} + (\mathbf{C} \odot \mathbf{B})^T (\mathbf{C} \odot \mathbf{B}) \}^{-1} \quad (8)$$

$$\mathbf{B} = \mathbf{X}^{(2)} (\mathbf{C} \odot \mathbf{A}) \{ \lambda \Psi_{\mathbf{G}} + (\mathbf{C} \odot \mathbf{A})^T (\mathbf{C} \odot \mathbf{A}) \}^{-1} \quad (9)$$

$$\mathbf{C} = \mathbf{X}^{(3)} (\mathbf{B} \odot \mathbf{A}) \{ \lambda \Psi_{\mathbf{G}} + (\mathbf{B} \odot \mathbf{A})^T (\mathbf{B} \odot \mathbf{A}) \}^{-1} \quad (10)$$

where $\mathbf{X}^{(i)}$ denotes the i -th unfolding matrix of the tensor \mathcal{X} .

2.4 Rank estimation

In order to estimate the rank of a tensor in R-CPD and AGS-CPD methods, we initialize the loading matrices by an over-estimated value using HOSVD [15]. Thus, by applying MDL on $\hat{\mathbf{C}}$ ($I_3 \times \hat{R}$) in R-CPD _{i} and on $\hat{\mathbf{G}}$ ($(I_1 + I_2 + I_3) \times \hat{R}$) in AGS-CPD, the rank of the tensor \mathcal{X} can be estimated. First, the singular values σ_i of the matrix ($I \times \hat{R}$) are sorted in descending order, i.e $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\hat{R}}$. Then, the R_{est} can be calculated as follows :

$$R_{est} = \arg \min_n -2 \log \left(\frac{\prod_{i=1+n}^I \sigma_i^{1/(I-n)}}{\frac{1}{I-n} \sum_{i=1+n}^I \sigma_i} \right)^{\hat{R}(I-n)} + r(2I - n) \log(\hat{R}) \quad (11)$$

where σ_i represents the i -th highest singular value and \hat{R} denotes the over-estimated rank. Using this method, R_{est} is computed by finding the breakpoint in the singular value curve.

3 Simulations and Results

To evaluate the efficiency of the methods both in rank estimation and decomposition, we consider two scenarios : constructing the tensor using the loading matrices generated randomly by a normal distribution and folded normal ones to have a non-negative tensor and loading matrices. The simulated tensor \mathcal{T} with rank R is built from loading matrices of size $(50 \times R)$. Here, the true rank is three. The noise tensor \mathcal{N} is also sampled from the same distribution. Hence, the noisy tensor \mathcal{X} is obtained as follows :

$$\mathcal{X} = \mathcal{T} + \sigma \frac{\|\mathcal{T}\|_F}{\|\mathcal{N}\|_F} \mathcal{N} \quad (12)$$

where the parameter σ controls the SNR (Signal to Noise Ratio) value defined by $\text{SNR} = -20 \log_{10}(\sigma)$. We consider different SNR values (-15 dB to 15 dB with a 5 dB step size) and run 100 independent Monte Carlo (MC) trials for each SNR value.

Performance criteria We propose two criteria to evaluate the rank estimation and decomposition results. The first criterion calculates the percentage of good estimation, which is called Accuracy Rate (AR) and defined by : AR : Times of $R = R_{est}$. The other one is the Relative Factor Error (RFE) which shows the accuracy of each loading matrix estimation. It measures the sum of

errors between the set of the true loading matrices $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$, and the set of the estimated ones $\{\mathbf{A}_{est}, \mathbf{B}_{est}, \mathbf{C}_{est}\}$ (see [16] for details).

Rank estimation performance For both scenarios the behavior of AGS-CPD method is evaluated and compared to the recent classical methods. In terms of rank estimation, AGS-CPD is compared to RCPD₁ and RCPD₂. Figure 1 shows the AR criterion on 100 MC runs as a function of SNR in the case of normal distribution. Clearly the performance of the three methods are quasi-equivalent, whatever the SNR level. In addition, the AR results demonstrate that all the methods succeed in the estimation of the rank. Indeed, even for the very bad SNR=-15dB, the AR values are greater than 91%. The AR results of the non-negative case are depicted in figure 2. For SNR values ranging from 5dB to 15dB the three methods always succeeded in the estimation of the rank. In the cases of SNR equal to -5dB and 0dB, AGS-CPD seems to be very efficient (AR about 85% for SNR=-5dB and 95% for SNR = 0dB, respectively). For RCPD₁ and RCPD₂, even if their behavior remain very satisfying (AR about 71% for SNR=-5dB, and 90% and 87% for SNR=0dB), they are less efficient than AGS-CPD. Finally, all methods cannot guarantee a favorable performance when the SNR=-10dB and -15dB.

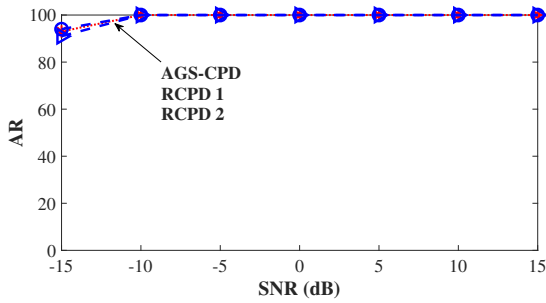


FIGURE 1 – AR values in different SNR values of the tensor constructed using normal distribution

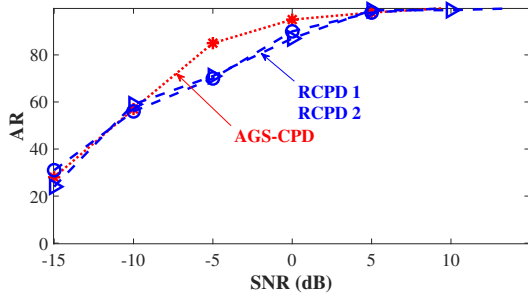


FIGURE 2 – AR values in different SNR values of the tensor constructed using Folded normal distribution (non-negative).

Decomposition Performance Since the methods DIAG [6] and AO-ADMM are not able to estimate the rank, the true rank is given to them directly. Besides, the original versions of AO-ADMM and BC-VMFB are not used and we modified the initialization part as for AGS-CPD and R-CPD. After these modifications, we discovered a significant improvement in their performances. In this study, RFE is just computed in the cases that the estimated rank equal to three (true rank) is given to the decomposition algorithms. Figure 3 shows the RFE values as a function of SNR in the case of normal distribution. As can be seen, RCPD₁ and RCPD₂ do not show a stable behavior as a function of SNR. The AGS-CPD and DIAG algorithms provide decreasing performances as SNR increases. Moreover, for SNR values

equal and higher than -10 dB, the RFE values are very small, which means that the decomposed factors are really similar to the ground truth. Regarding very low SNR of -15 dB, AGS-CPD approach outperforms DIAG with very well decomposition quality (RFE value is 0.041). Figure 4 depicts the obtained RFE in the case of non-negative data. As for the first scenario, RCPD₁ and RCPD₂ are not able to estimate the loading matrices efficiently. For SNR values greater than 0 dB, the four other methods, namely AGS-CPD, AO-ADMM, BC-VMFB and DIAG work very well, with RFE values close to zero. Interestingly, AGS-CPD and AO-ADMM algorithms give the best results in comparison to DIAG and BC-VMFB, for low SNR values ranging from -15 to 0 dB, with really reasonable decomposition results.

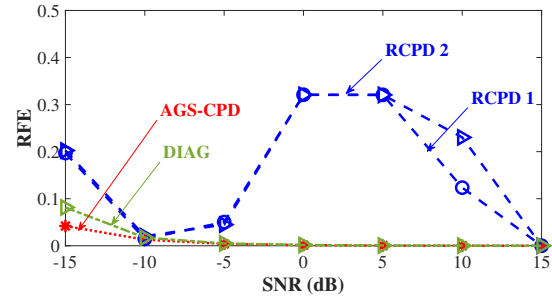


FIGURE 3 – RFE values in different SNR values of the tensor constructed using normal distribution.

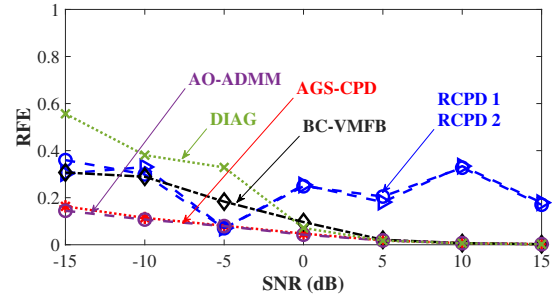


FIGURE 4 – RFE values in different SNR values of the tensor constructed using Folded normal distribution (non-negative).

4 Fluorescence spectroscopy data

CPD is commonly used to decompose sets of 2D discrete fluorescence signals in order to identify and track chemical components [1]. Considering a set of liquid samples, we can measure in laboratory the fluorescence intensity of each sample at several discrete couples of excitation and emission wavelengths. Thus, we are able to build a fluorescence tensor by gathering the measured signals. Each sample can be seen as a mixture of R fluorescing components (fluorophores). Under some assumptions [17], it can be shown that the three estimated loading matrices \mathbf{A} , \mathbf{B} and \mathbf{C} , of the R -rank fluorescence tensor, are related to the emission spectra, the excitation spectra of the R fluorophores present in the sample set, and the concentration profiles of the fluorophores throughout the sample set, respectively. A common issue with this approach is that the number of fluorophores is usually unknown and must be estimated correctly in order to obtain a good estimation of the factors. Moreover, the factors are often correlated in one or several modes and can have very weak contributions, making the decomposition difficult from an algorithmic point of view. Here, all the factors are non-negative.

The obtained results using real data show that AGS-CPD achieves the same performances as the non-negative based me-

thods without any non-negativity constraint and without giving the true rank (see table 1 and figure 5). Indeed, we compare the performances of AGS-CPD with those of BC-VMFB and AO-ADMM for the decomposition of a fluorescence tensor obtained from 11 real laboratory mixtures of three fluorophores (tryptophan, quinine sulfate and fluorescein).¹ The tensor size is (57, 57, 11). Note that for AGS-CPD and BC-VMFB, we overestimated the rank by taking $R = 10$. Regarding the AO-ADMM algorithm, we give the true rank $R = 3$. After convergence, the three algorithms AGS-CPD, R-CPD₁ and R-CPD₂ have successfully estimated the correct rank value (*i.e.* 3). Regarding the estimation quality of the factors, RFE values of each method and loading matrix are reported in table 1. As can be seen, the RFE values show that AGS-CPD gives quasi-similar results to those of BC-VMFB and AO-ADMM algorithms with non-negativity constraint. These results are confirmed in figure 5, which display the estimated factors compared to those of the ground truth. We can see that all methods clearly identify the excitation and emission spectra. Two concentration profiles out of three are perfectly estimated. The same small error is observed on the last concentration profile (for samples 6 to 11) for all algorithms. The real experiment illustrates that our approach can reach the same level of performance than the reference non-negative algorithms, without using this *apriori* information.

TABLE 1 – RFE values of different methods in real data.

	AGS-CPD	BC-VMFB	AO-ADMM
A	0.0023	.0028	.0015
B	0.0095	.0092	.0090
C	0.0157	.0157	.0157
	0.0091	.0092	.0087

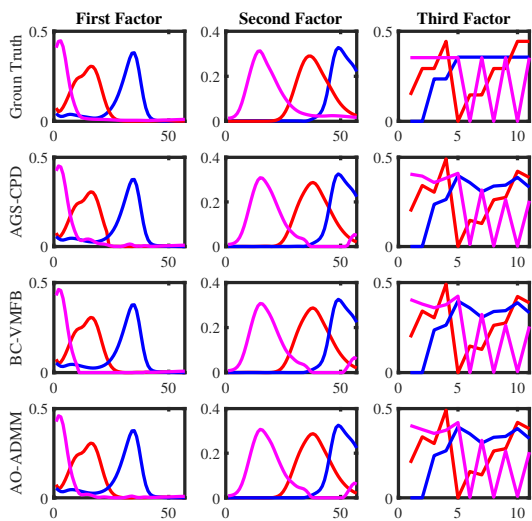


FIGURE 5 – Illustration of the estimated factors using different methods, in real data, compared to the actual ones.

5 Conclusion

In this study, we introduce a novel algorithm called AGS-CPD in order to estimate the rank and decompose the factors in one step. Simulated and real experiments assess the good behavior of the proposed method in comparison with other approaches to overcome the joint problem of rank estimation and tensor factorization even in the case of non-negative data. Moreover, the obtained results show the robustness of the proposed algorithm

with respect to the presence of noise for different data distributions, without imposing any non-negative constraint.

Références

- [1] R. Bro. Parafac. tutorial and applications. *Chemometr. Intell. Lab. Syst.*, 38(2), 1997.
- [2] A. Cichocki, et al. Tensor decompositions for signal processing applications : From two-way to multiway component analysis. *IEEE Signal Process. Mag.*, 32(2), 2015.
- [3] N. Taheri, et al. Brain source localization using constrained low rank canonical polyadic decomposition. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 2018.
- [4] H. Becker, et al. EEG extended source localization : tensor-based vs. conventional methods. *NeuroImage*, 2014.
- [5] J D. Carroll and J.-J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of “eckart-young” decomposition. *Psychometrika*, 35(3), 1970.
- [6] X. Luciani and L.t Albera. Canonical polyadic decomposition based on joint eigenvalue decomposition. *Chemometr. Intell. Lab. Syst.*, 2014.
- [7] JPCL. da Costa, et al. Multi-dimensional model order selection. *Eurasip J. Adv. Signal Process.*, 2011.
- [8] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. on Acoust. Speech. and Signal. Process.*, 33(2), 1985.
- [9] K. Liu, et al. Detection of number of components in CANDECOMP/PARAFAC models via minimum description length. *Digital Signal Process.*, 513, 2016.
- [10] Xu Han, et al. Low rank canonical polyadic decomposition of tensors based on group sparsity. In *2014 25th EUSIPCO*, 2017.
- [11] X. Vu, et al. A new penalized nonnegative third order tensor decomposition using a block coordinate proximal gradient approach : application to 3D fluorescence spectroscopy. *J. of Chemometrics*, 31(4), 2017.
- [12] K. Huang, et al. A flexible and efficient algorithmic framework for constrained matrix and tensor factorization. *IEEE Trans. Signal Process.*, 64(19), 2016.
- [13] J.B. Kruskal. Three-way arrays : rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.*, 18(2), 1977.
- [14] S. Boyd, et al. Distributed optimization and statistical learning via alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1), 2010.
- [15] L. De Lathauwer, et al. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21(4), 2000.
- [16] Lu Wang, et al. Canonical polyadic decomposition of 3rd order semi-nonnegative semi-symmetric tensors using lu and qr matrix factorizations. *Eurasip J. Adv. Signal Process.*, 2014.
- [17] X. Luciani, et al. A simple correction method of inner filter effects affecting FEEM and its application to the PARAFAC decomposition. *Chemometr. Intell. Lab. Syst.*, 96(2), 2009.

1. We thank Dr S. Mounier from the MIO lab. for providing us the data.