

A BLIND SPARSE APPROACH FOR ESTIMATING CONSTRAINT MATRICES IN PARALIND DATA MODELS

F. Caland^{1,2}, S. Miron²

D. Brie², C. Mustin¹

¹ LIMOS, Nancy-Université, CNRS
Boulevard des Aiguillettes BP 70239
54506 Vandoeuvre-lès-Nancy, France

² CRAN, Nancy-Université, CNRS
Boulevard des Aiguillettes BP 70239
54506 Vandoeuvre-lès-Nancy, France

ABSTRACT

In this paper we address the problem of estimating the interaction matrices of PARALIND decomposition. In general, this is an ill-posed problem admitting an infinite number of solutions. First we study the gain of imposing sparsity constraints on the interaction matrices, in terms of model identifiability. Then, we propose a new algorithm (S-PARALIND) for fitting the PARALIND model, using a $\ell_2 - \ell_1$ optimization step for estimating the interaction matrix. This new approach provides more accurate and robust estimates of the constraint matrices than ALS-PARALIND, thus improving the interpretability of the PARALIND decomposition.

Index Terms— PARAFAC, linear constraints, PARALIND/CONFAC, sparse, ALS-PARALIND, S-PARALIND

1. INTRODUCTION

PARAFAC-based methods [1, 2] are presently standard tools for factor or component modeling in various domains such as psychometry, spectroscopy, signal processing or telecommunications systems. A general overview of PARAFAC applications can be found in [3, 4]. The PARAFAC decomposition of a \mathcal{X} ($I \times J \times K$) 3-way array (or tensor) into R rank-1 terms is given by

$$\mathcal{X} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad (1)$$

where $\mathbf{a}_r(I \times 1)$, $\mathbf{b}_r(J \times 1)$ and $\mathbf{c}_r(K \times 1)$ are vectors and “ \circ ” denotes the outer vector product. For simplicity, the noise/error term in (1) is ignored at this point of the presentation. The three dimensions of \mathcal{X} are referred to as “modes”. An alternative notation for (1) is

$$\mathcal{X} = \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket, \quad (2)$$

where $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1 \dots \mathbf{b}_R]$ and $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_R]$ denote the component matrices. The mode-1 matrix unfold-

ing of the PARAFAC model (1) is given by

$$\mathbf{X}_1 = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T, \quad (3)$$

with \odot the Khatri-Rao product of two matrices. The mode-2 and the mode-3 unfolding matrices, \mathbf{X}_2 and \mathbf{X}_3 , can be obtained by switching $\mathbf{A}, \mathbf{B}, \mathbf{C}$ in (3).

In some applications, prior knowledge on the existence of linear dependencies between the columns of the component matrices is available. This information can be explicitly taken into account by introducing some constraint (or interaction) matrices $\Psi(R_1 \times R)$, $\Phi(R_2 \times R)$, $\Omega(R_3 \times R)$, containing the linear dependency patterns between the columns of $\mathbf{A}, \mathbf{B}, \mathbf{C}$, respectively. Thus, instead of $\llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket$ the decomposition is given by

$$\mathcal{X} = \llbracket \tilde{\mathbf{A}}\Psi, \tilde{\mathbf{B}}\Phi, \tilde{\mathbf{C}}\Omega \rrbracket, \quad (4)$$

with $\tilde{\mathbf{A}}(I \times R_1)$, $\tilde{\mathbf{B}}(J \times R_2)$ and $\tilde{\mathbf{C}}(K \times R_3)$ full column rank matrices. This type of decomposition was introduced in [5] and previous versions, and named PARALIND¹. A slightly different version, CONFAC², with the constraint matrices having canonical vectors as columns, was proposed in [6, 7]. In order to illustrate PARALIND model we consider a simple example, similar to those given in [5]. Suppose that the columns of the first component matrix \mathbf{A} are $[\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3 \ \mathbf{a}_2 \ \mathbf{a}_3 \ \mathbf{a}_1 + \mathbf{a}_2]$. Then, PARALIND expresses \mathbf{A} as the matrix product of $\tilde{\mathbf{A}} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3]$ and the interaction matrix

$$\Psi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}. \quad (5)$$

In general, the algorithms for fitting the PARALIND model assumes that the constraint matrices are *a priori* known. However, this is not always the case in practice. Moreover, in some real life applications it may be of practical interest to estimate these constraint matrices, as they provide important information on the interactions between the physical mechanisms generating the data. A blind alternating least

¹This work has been supported by the French ANR program through grant ANR-09-BLAN-0336-04.

¹PARALLEL profiles with LINear Dependencies
²CONSTRAINED FACTOR decomposition

squares (ALS) estimator for the PARALIND model, referred to as ALS-PARALIND, was proposed in [5]. However, for identifiability reasons (as explained in the next section), the interaction matrices estimated by this approach are highly dependent on the algorithm initialization, which limits their practical utility.

In this paper we propose a blind approach for estimating the interaction matrices of the PARALIND model that imposes sparsity of the elements of these matrices. This new approach, called S-PARALIND, when compared to the ALS-PARALIND algorithm, yields more robust estimates of the constraints matrices. Moreover, it also improves the interpretability of the decomposition since the linear dependencies are expressed using a small number of interacting components. The remainder of this paper is organized as follows: in section 2 identifiability issues of PARAFAC and PARALIND are addressed; section 3 introduces the S-PARALIND algorithm and some results on synthetic data are given in section 4. Finally, conclusions are drawn in section 5.

2. IDENTIFIABILITY OF PARAFAC AND PARALIND MODELS

A model is said *identifiable* if all its parameters can be *uniquely* estimated from the data, up to some trivial indeterminacies. Thus, in this paper, identifiability can be understood as a uniqueness problem. For example, the PARAFAC model given by (2) is identifiable if the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ can be uniquely estimated from \mathcal{X} up to simultaneous column permutation and column-wise rescaling. An attractive feature of this decomposition is its identifiability under mild conditions. The most well-known PARAFAC identifiability condition is due to Kruskal [8] and states that the decomposition in (3) is unique if

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2R + 2, \quad (6)$$

where $k_{(\cdot)}$ denotes the Kruskal-rank³ of a matrix.

Following [5], identifiability of the PARALIND model is essentially the same as that of the PARAFAC model. Meanwhile, if the interaction matrices are fixed and known, identifiability conditions specific to PARALIND can be found in [9]. If these interaction matrices are not known, the identifiability problem can be much more complicated. In particular, it may happen that only some components of the three matrices or only one matrix (among the three) are identifiable, resulting in the so-called *partial uniqueness* or *uni-mode uniqueness* results. The interested reader is referred to [10] for details.

Let us now assume that the uniqueness of matrix \mathbf{A} is fulfilled and that we aim at estimating the constraint matrix Ψ together with the full column rank matrix $\tilde{\mathbf{A}}$. The identifiability of Ψ and $\tilde{\mathbf{A}}$ comes down to the uniqueness of the bilinear

decomposition $\mathbf{A} = \tilde{\mathbf{A}}\Psi$. Without any further constraints, such a decomposition is not unique since an alternative decomposition can be obtained as

$$\mathbf{A} = \tilde{\mathbf{A}}\Psi = (\tilde{\mathbf{A}}\mathbf{T}^{-1})(\mathbf{T}\Psi) = \tilde{\mathbf{A}}'\Psi',$$

for any non-singular matrix \mathbf{T} . In this paper we propose to impose sparsity on the constraint matrix Ψ which should have a minimum number of non-zero entries. This comes down to explaining the rank deficiency of matrix \mathbf{A} by considering the simplest dependency pattern between its columns. This problem is somewhat connected with the problem of dictionary identification using sparse matrix factorization, which has been intensely studied lately in different papers such as [12]. However, the main difference with these approaches is the fact that the problem addressed in this paper considers only full column rank dictionaries. Therefore, the proposed approach is somewhat closer to sparse singular value decomposition methods [13]. As we do not dispose of any definitive result on the uniqueness of the decomposition using this kind of sparsity constraint, we consider next some examples to illustrate the purpose. Let \mathbf{A} be given by

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{a}_1 + \mathbf{a}_2] \quad (7)$$

$$= [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3] \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (8)$$

$$= [\mathbf{a}_1 + \mathbf{a}_2 \quad \mathbf{a}_1 + \mathbf{a}_3 \quad \mathbf{a}_2 + \mathbf{a}_3] \begin{bmatrix} 1/2 & 1/2 & -1/2 & 1 \\ 1/2 & -1/2 & 1/2 & 0 \\ -1/2 & 1/2 & 1/2 & 0 \end{bmatrix} \quad (9)$$

As illustrated by (8) and (9), it appears that the sparsest matrix Ψ is obtained by selecting R_1 independent columns of \mathbf{A} to form $\tilde{\mathbf{A}}$. It is worth noting that imposing sparsity of Ψ does not ensure the uniqueness of the bilinear decomposition. For example, another possible decomposition of \mathbf{A} is

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_3 \quad \mathbf{a}_1 + \mathbf{a}_2] \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}. \quad (10)$$

One can see that Ψ matrix in (10) has the same sparsity degree as the one in (8). From an interpretation point of view, there is no reason to favor either decomposition (10) or (8) since the number of non-zero entries in the constraint matrix is the same in both cases. However, if some additional physical arguments are available, then the number of possible solutions can be reduced. For example it is possible to impose both sparsity and positivity of the entries of the Ψ matrix to ensure uniqueness of the bilinear decomposition. The reader may consult [11] for a discussion on this topic.

³The Kruskal-rank of a matrix \mathbf{A} is the maximum number ℓ such that every ℓ columns of \mathbf{A} are linearly independent.

3. A SPARSE PARALIND ALGORITHM (S-PARALIND)

In this section, we present an algorithm (S-PARALIND) for estimating the PARALIND model with sparse constraints on the interaction matrix. This algorithm is currently implemented for estimating linear dependencies in one mode only, *i.e.* mode-1, but the extension to two or three modes simultaneously is straightforward.

In the presence of noise or model errors, evaluating Ψ from a PARAFAC estimate of \mathbf{A} (as presented in the previous section) is not judicious because of the error accumulation effect. Therefore, in [5] the constraint matrix is estimated directly within the main ALS loop for updating the PARAFAC matrix components. The approach presented in this section is based on the algorithm proposed in [5] with the difference that the LS estimation step of Ψ is replaced by the following $\ell_2 - \ell_1$ optimization problem [14]

$$\min_{\Psi} \left[\|\mathbf{X}_1 - \tilde{\mathbf{A}}\Psi(\mathbf{C} \odot \mathbf{B})^T\|_2^2 + \lambda \|\Psi\|_1 \right], \quad (11)$$

where the hyperparameter λ controls the sparsity degree of the constraint matrix. The minimization in (11) can be formulated as a LASSO problem for which a number of efficient algorithms have been developed lately (see [15] and references therein). Table 1 illustrates the main steps of the proposed approach, where $\text{vec}(\cdot)$, \otimes and “ $*$ ” denotes the matrix vectorization operator, the Kronecker product and the Hadamard (element wise) product, respectively. For simultaneous linear dependencies in all the three modes, each of the steps 4 and 5 in table 1 should be replaced by two other steps (analogous to steps 2 and 3).

Input : $\mathcal{X}, \lambda, R, R_1$
1: Initialize $\tilde{\mathbf{A}}, \mathbf{B}, \mathbf{C}$
2: $\text{vec}\Psi = \arg \min_{\Psi} \{ \ \text{vec}\mathbf{X}_1 - [(\mathbf{C} \odot \mathbf{B}) \otimes \tilde{\mathbf{A}}] \text{vec}\Psi\ _2^2 + \lambda \ \text{vec}\Psi\ _1 \}$
3: $\tilde{\mathbf{A}} = \mathbf{X}_1(\mathbf{C} \odot \mathbf{B})\Psi^T \{ \Psi[(\mathbf{B}^T\mathbf{B}) * (\mathbf{C}^T\mathbf{C})]\Psi^T \}^{-1}$
4: $\mathbf{B} = \mathbf{X}_2[\mathbf{C} \odot (\tilde{\mathbf{A}}\Psi)] \left[(\Psi^T \tilde{\mathbf{A}}^T \tilde{\mathbf{A}}\Psi) * (\mathbf{C}^T\mathbf{C}) \right]^{-1}$
5: $\mathbf{C} = \mathbf{X}_3[\mathbf{B} \odot (\tilde{\mathbf{A}}\Psi)] \left[(\Psi^T \tilde{\mathbf{A}}^T \tilde{\mathbf{A}}\Psi) * (\mathbf{B}^T\mathbf{B}) \right]^{-1}$
6: If stop condition not satisfied, go to Step 2
Output : Estimated $\tilde{\mathbf{A}}, \Psi, \mathbf{B}, \mathbf{C}$

Table 1. S-PARALIND algorithm

Steps 3 – 5 in Table 1 are similar to ALS-PARALIND algorithm presented in [5]. In the next section we compare S-PARALIND with ALS-PARALIND on synthetic examples.

4. RESULTS

In this section, we aim at illustrating the benefit of including a sparsity constraint in the estimation of the Ψ matrix.

To that end, we simulated a positive data array \mathcal{X} of size $700 \times 30 \times 30$ that mimics spectroscopy data. In order to improve the uniqueness properties of the PARALIND decomposition as well as result interpretability, we also imposed non-negativity constraints on the estimated modes. In S-PARALIND, to solve the $\ell_2 - \ell_1$ optimization problem of step 2 in table 1, the LASSO implementation⁴ of [14] was used. The results provided by S-PARALIND are compared with those of ALS-PARALIND. In all the experiments the number of sources was set to $R = 4$ and the number of columns of matrix $\tilde{\mathbf{A}}$ to $R_1 = 3$. The same initializations, non-negativity constraints and number of iterations were used for both algorithms. The columns of the simulated matrices $\tilde{\mathbf{A}}, \mathbf{B}$ and \mathbf{C} are depicted in Fig. 1.

In a first example the constraint matrix is given by

$$\Psi = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (12)$$

In this case, the k-rank of \mathbf{A} equals 1, implying the non-uniqueness of the PARAFAC decomposition. However, the uni-mode uniqueness conditions in [10] state that the first mode matrix is still identifiable. Meanwhile \mathbf{B} and \mathbf{C} are subject to rotational ambiguities which, due to the positive offset (background), are not resolved by the non-negativity constraints. Figure 2 and 3 show the three modes estimates for

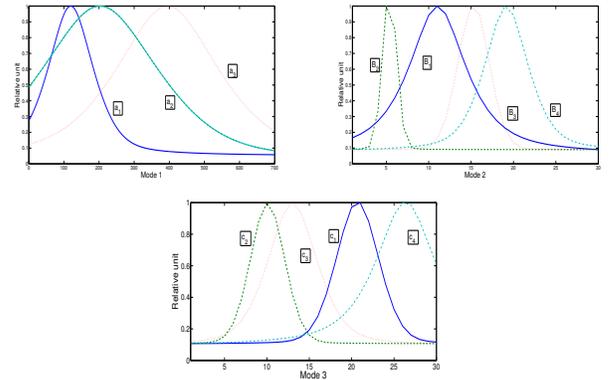


Fig. 1. Simulated data

ALS-PARALIND and S-PARALIND, respectively. One can observe that the S-PARALIND estimates are more accurate than ALS-PARALIND for the first mode. However, for both algorithms, the estimated components of \mathbf{B} and \mathbf{C} present artifacts that are typical for rotational indeterminacies. The effect of the sparse constraints can be clearly seen on the estimated constraint matrices

⁴Available at <http://www.di.ens.fr/~mschmidt/Software/lasso.html>

$$\hat{\Psi}_{S\text{-PARALIND}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0.008 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad (13)$$

$$\hat{\Psi}_{ALS\text{-PARALIND}} = \begin{bmatrix} 1 & 0.591 & 0.155 & 0.59 \\ 0.003 & 1 & 0.103 & 1 \\ 0.004 & 0.157 & 1 & 0.156 \end{bmatrix}. \quad (14)$$

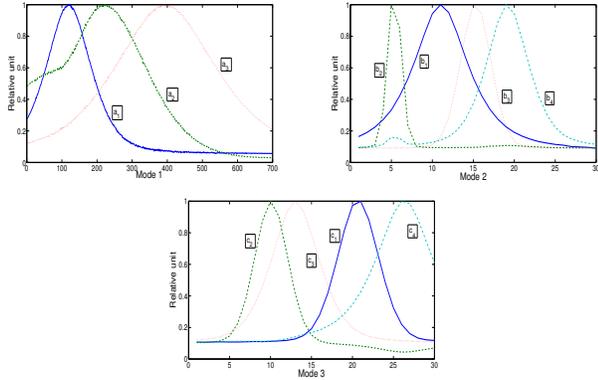


Fig. 2. Results of ALS-PARALIND for the first example

$$\hat{\Psi}_{S\text{-PARALIND}} = \begin{bmatrix} 1 & 0 & 0 & 0.749 \\ 0.171 & 1 & 0.024 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad (16)$$

$$\hat{\Psi}_{ALS\text{-PARALIND}} = \begin{bmatrix} 1 & 0.468 & 0.006 & 1 \\ 0.014 & 1 & 0.007 & 0.569 \\ 0.071 & 0.401 & 1 & 0.276 \end{bmatrix}. \quad (17)$$

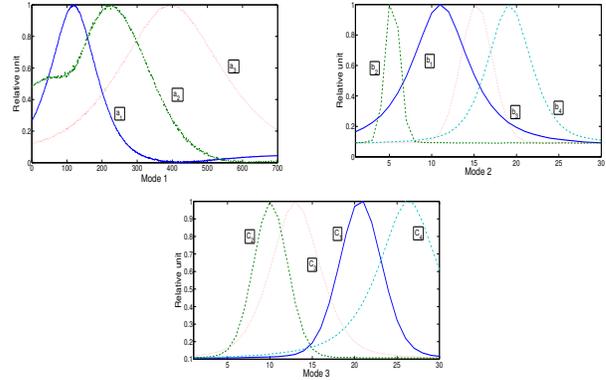


Fig. 4. Results of ALS-PARALIND for the second example

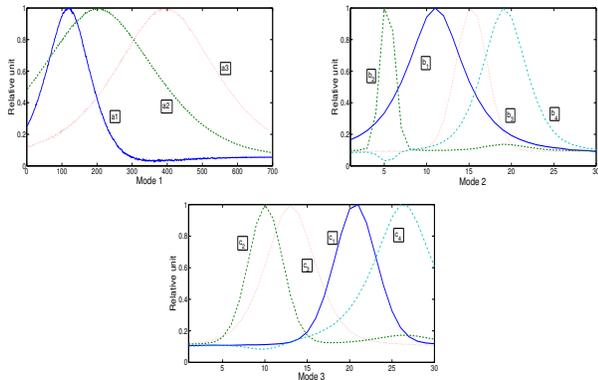


Fig. 3. Results of S-PARALIND for the first example

In a second example matrix Ψ is fixed to

$$\Psi = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (15)$$

In this case, $k_A = 2$ and the PARAFAC uniqueness is achieved, resulting in a unique estimation of all the three matrices \mathbf{A} , \mathbf{B} and \mathbf{C} . As one can see in fig. 4 and fig. 5, there are no more ambiguities in the estimation of modes 2 and 3. However, in this case also, imposing sparse constraints on Ψ yields better estimation of the first mode component matrix and constraint matrix

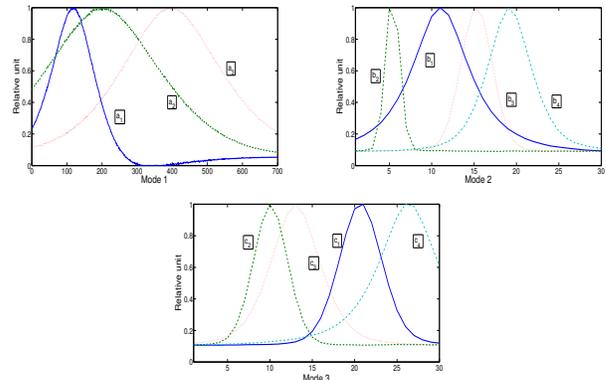


Fig. 5. Results of S-PARALIND for the second example

Next we present the results of the S-PARALIND study of the response of a bacterial biosensor to a varying concentration of IPTG (isopropyl β -D-1-thiogalactopyranoside). The employed biosensor is a bacteria genetically modified to produce two fluorescent proteins when exposed to IPTG. A same gene ($lacZ$) of the bacteria is instrumented by two other genes encoding the synthesis of different fluorescent proteins. As the production of these fluorescent proteins is controlled by the same gene, they have the same response to IPTG concentration variation, resulting in collinearities in the associated PARAFAC model (mode 1). The second diversity, necessary to source separation, is provided by the spectral mode, each

source having a different fluorescence spectrum. The third diversity in the data is created by considering the time evolution of the fluorescence, as each source has a different maturation time. In other words, the time between the beginning of the synthesis of the protein and the beginning of the fluorescence light emission is different for each protein. The analyzed data contains spectral measurements (between 450 and 600 nm with a step of 3 nm), performed every 45 minutes after an initial time lapse of 1 hour, for 6 different IPTG concentrations. Figure 6 shows the decomposition results of the data using the S-PARALIND algorithm. The red and the green sources correspond to the two fluorescent proteins, while the blue source corresponds to the autofluorescence of the bacteria which is theoretically independent of time and IPTG concentration. The first plot corresponds to the estimated sources as a function of the concentration of IPTG. The other two modes correspond to the estimated biosensor spectra and temporal evolution of fluorescence, respectively. Because the collinearity in the first mode, the red and the green sources are theoretically subject to rotational indeterminacies in modes 2 and 3 (see [10]). This explains the partial overlapping of the source spectra on the second plot, despite the non-negativity constraints imposed in the estimation process. Further constraints, such as unimodality, can be imposed to tackle this indeterminacy problem. Nevertheless, the S-PARALIND decomposition results comply with theory.

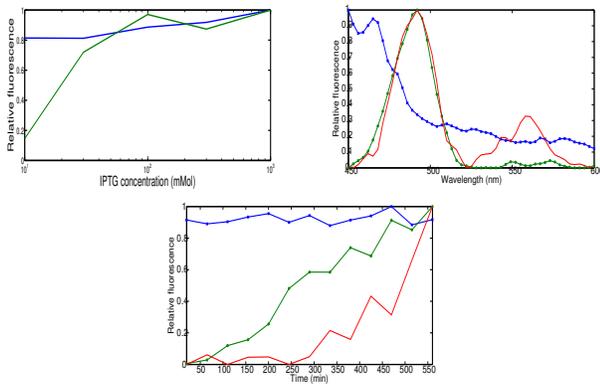


Fig. 6. Results of S-PARALIND decomposition of the dataset.

5. CONCLUSIONS

In this paper, an approach has been proposed for estimating the constraint matrices in PARALIND models. First, the identifiability of the PARALIND model has been investigated and we have provided some evidences showing the interest of imposing sparsity of the constraint matrix. The proposed S-PARALIND approach has been compared to the ALS-PARALIND in two different cases: partially and fully identifiable models. In both cases, the effectiveness of S-PARALIND algorithm has been confirmed. Finally, the proposed algorithm was applied on a real dataset resulted from bacterial biosensors interaction with IPTG.

6. REFERENCES

- [1] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, Sept. 1970.
- [2] 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, Dec. 1970.
- [3] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Sept. 2009.
- [4] E. Acar and B. Yener, "Unsupervised multiway data analysis: A literature survey," *IEEE Trans. Knowledge Data Eng.*, vol. 21, no. 1, pp. 6–20, Jan. 2009.
- [5] R. Bro, R. A. Harshman, N. D. Sidiropoulos, and M. E. Lundy, "Modeling multi-way data with linearly dependent loadings," *J. Chemometr.*, vol. 23, no. 7-8, pp. 324–340, July-Aug. 2009, Special Issue: In Honor of Professor Richard A. Harshman.
- [6] A. L. F. de Almeida, G. Favier, and J. C. M. Mota, "Constrained tensor modeling approach to blind multiple-antenna CDMA schemes," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2417–2428, June 2008.
- [7] A. L. F. de Almeida, G. Favier, and J. C. M. Mota, "A constrained factor decomposition with application to MIMO antenna systems," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2429–2442, June 2008.
- [8] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Applicat.*, vol. 18, no. 2, pp. 95–138, 1977.
- [9] A. Stegeman and A. L. F. de Almeida, "Uniqueness conditions for constrained three-way factor decompositions with linearly dependent loadings," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 3, pp. 1469–1499, 2009.
- [10] X. Guo, S. Miron, D. Brie, and A. Stegeman, "Uni-mode and partial uniqueness conditions for CANDECOMP/PARAFAC of three-way arrays with linearly dependent loadings," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 1, pp. 111–129, 2012.
- [11] S. Miron, M. Dossot, C. Carteret, S. Margueron, and D. Brie, "Joint processing of the parallel and crossed polarized Raman spectra and uniqueness in blind nonnegative source separation," *Chemometr. Intell. Lab.*, vol. 105, no. 1, pp. 7–18, 2011.
- [12] R. Gribonval and K. Schnass, "Dictionary identification - sparse matrix-factorisation via ℓ_1 -minimisation," *IEEE Trans. Inform. Theory*, vol. 56, no. 7, pp. 3523–3539, July 2010.
- [13] M. Li, H. Shen, J.Z. Huang, and J.S. Marron, "Biclustering via sparse singular value decomposition," *Biometrics*, vol. 66, no. 4, pp. 1087–1095, Dec. 2010.
- [14] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996.
- [15] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," *IEEE Signal Proc. Mag.*, vol. 27, no. 3, pp. 76–88, May 2010.