Some rank conditions for the identifiability of the sparse Paralind model

Sebastian Miron, David Brie

Centre de Recherche en Automatique de Nancy (CRAN), UMR 7039, Université de Lorraine, Vandœuvre, F-54506, France.

Abstract. In this paper we study the identifiability of the PARALIND model with sparse interaction matrices (*i.e.* S-PARALIND). We provide some theoretical results on how to obtain the sparsest interaction matrices in some particular configurations and when these matrices are unique. These results could be use for the design and analysis of ℓ_0-based decomposition algorithms.

1 Introduction

The PARAFAC [8, 14] decomposition of an $\boldsymbol{\mathcal{X}}$ $(I \times J \times K)$ 3-way array (or tensor) into sum of R rank-1 tensors is given by $\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$, where $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$ are vectors of dimensions I, J and K, respectively, and "$\circ$" denotes the outer vector product. For simplicity, the noise/error term in the model is ignored at this point of the presentation. By regrouping the vectors of the three dimensions (or "modes") of $\boldsymbol{\mathcal{X}}$ into three component matrices $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1 \dots \mathbf{b}_R]$ and $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_R]$, an alternative notation for the PARAFAC decomposition of $\boldsymbol{\mathcal{X}}$ is obtained:

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]. \tag{1}$$

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 $\boldsymbol{\Psi}(R_1 \times R)$, $\boldsymbol{\Phi}(R_2 \times R)$, $\boldsymbol{\Omega}(R_3 \times R)$, containing the linear dependency patterns between the columns of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, respectively. Thus, instead of $[\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ the decomposition is given by

$$\boldsymbol{\mathcal{X}} = [\![\tilde{\mathbf{A}}\boldsymbol{\Psi}, \tilde{\mathbf{B}}\boldsymbol{\Phi}, \tilde{\mathbf{C}}\boldsymbol{\Omega}]\!], \tag{2}$$

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 [6] and previous versions, and named PARALIND[1]. A slightly different version, CONFAC[2], with the constraint matrices having canonical vectors as columns, was proposed in [3, 2]. A less general framework (involving structured types of linear dependencies), but often highly interpretable, called Block Component Model (BCM) was introduced in [10]. These

[1] PARAllel profiles with LINear Dependencies

[2] CONstrained FACtor decomposition

decompositions proved their usefulness in various domains such as telecommunications [3, 17, 20, 19], spectroscopy [4, 9, 5] or direction finding [16, 21].

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 squares (ALS) estimator for the PARALIND model, referred to as ALS-PARALIND, was proposed in [6]. 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. To regularize this ill-posed inverse problem, we proposed in [7] to impose sparsity constraints on the interaction matrices, leading to sparse PARALIND (S-PARALIND). These contraints are physically meaningful as they aim at explaining the interactions between the mechanisms generating the data in the simplest way possible. However, no results were given in [7] regarding the identifiability of the S-PARALIND model; the objective of this paper is to shed some light on this aspect.

2 Identifiability of S-Paralind model

2.1 Preliminaries

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 (1) is identifiable if the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ can be uniquely estimated from $\boldsymbol{\mathcal{X}}$ up to simultaneous column permutation and column-wise rescaling. The most well-known PARAFAC identifiability condition is due to Kruskal [15] and is based on the Kruskal-rank[3] of the component matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$. Following [6], identifiability of the PARALIND model is essentially the same as that of the PARAFAC model. If the interaction matrices are fixed and known, identifiability conditions specific to PARALIND can be found in [18]. 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 [13] for details.

Let us now assume that the uniqueness of matrix $\mathbf{A}$ is fulfilled and that we aim at estimating the constraint matrix $\boldsymbol{\Psi}$ together with the full column rank matrix $\tilde{\mathbf{A}}$. The identifiability of $\boldsymbol{\Psi}$ and $\tilde{\mathbf{A}}$ comes down to the uniqueness of the bilinear decomposition $\mathbf{A} = \tilde{\mathbf{A}}\boldsymbol{\Psi}$. Without any further constraints, such a decomposition is not unique since an alternative decomposition can be obtained as $\mathbf{A} = \tilde{\mathbf{A}}\boldsymbol{\Psi} = (\tilde{\mathbf{A}}\mathbf{T}^{-1})(\mathbf{T}\boldsymbol{\Psi}) = \tilde{\mathbf{A}}'\boldsymbol{\Psi}'$, for any non-singular matrix $\mathbf{T}$. By imposing

[3] The Kruskal-rank of a matrix $\mathbf{A}$ (denoted $k_{\mathbf{A}}$) is the maximum number ℓ such that every ℓ columns of $\mathbf{A}$ are linearly independent.

sparsity on the constraint matrix $\boldsymbol{\Psi}$ (which should have a minimum number of non-zero entries), we try to explain the rank deficiency of matrix $\mathbf{A}$ by considering the the simplest dependency pattern between its columns. This problem has close connection with the problem of dictionary identification using sparse matrix factorization and sparse component analysis, which has been studied in different papers such as [11, 1, 12]. Basically, in [11, 1], the problem is addressed as a $\ell_2 - \ell_0$ optimization problem. Using a geometrical point of view, identifiability conditions are obtained requiring that the size of the training set grows exponentially with the number of atoms. In contrast, the work of [12] addresses the problem as a $\ell_2 - \ell_1$ (non combinatorial) optimization problem. It is shown that the size of training set only needs to grow quadratically with the number of atoms .

All these works consider the problem of overcomplete dictionary recovery. We to stress up the fact that this is one of the main differences with the problem addressed in this paper where only full column rank dictionaries are considered.

The rest of the section aims at giving some answers to the following questions:

- when is the matrix $\tilde{\mathbf{A}}$, yielding the sparsest $\boldsymbol{\Psi}$, a submatrix of $\mathbf{A}$?
- when is the decomposition $\mathbf{A} = \tilde{\mathbf{A}}\boldsymbol{\Psi}$ unique ?

Before addressing analytically these problems, let us consider some examples to illustrate the purpose. Let $\mathbf{A}$ be given by

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{a}_1 + \mathbf{a}_2] = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3] \begin{bmatrix} 1\,0\,0\,1 \\ 0\,1\,0\,1 \\ 0\,0\,1\,0 \end{bmatrix} \tag{3}$$

$$= [\mathbf{a}_1 + \mathbf{a}_2 \quad \mathbf{a}_1 + \mathbf{a}_3 \quad \mathbf{a}_2 + \mathbf{a}_3] \begin{bmatrix} 1/2 & 1/2 & -1/2 & 1 \\ 1/2 & -1/2 & 1/2 & 0 \\ -1/2 & 1/2 & 1/2 & 0 \end{bmatrix} \tag{4}$$

As illustrated by (3) and (4), it appears that the sparsest matrix $\boldsymbol{\Psi}$ is obtained by selecting R_1 independent columns of $\mathbf{A}$ to form $\tilde{\mathbf{A}}$. It is worth noting that imposing sparsity of $\boldsymbol{\Psi}$ does not ensure the uniqueness of the bilinear decomposition. For example, another possible decomposition of $\mathbf{A}$ is

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_3 \quad \mathbf{a}_1 + \mathbf{a}_2] \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}. \tag{5}$$

One can see that $\boldsymbol{\Psi}$ matrix in (5) has the same sparsity degree as the one in (3).

2.2 Choosing a basis of the column space of A that yields the sparsest $\boldsymbol{\Psi}$

Now we are ready to provide some results on what is the "best" basis, that is the "best" matrix $\tilde{\mathbf{A}}$, for having the sparsest $\boldsymbol{\Psi}$ matrix. However, it is first necessary to introduce some notations. Let $\mathbf{A}$ be a matrix of dimension $(M \times N), M \geq N$

and let $r_{\mathbf{A}} = \text{rank}(\mathbf{A}) \leq N$. We aim at finding a factorization of the matrix $\mathbf{A} = \tilde{\mathbf{A}}\boldsymbol{\Psi}$ where $\tilde{\mathbf{A}}$ is a $(M \times r_{\mathbf{A}})$ (tall) matrix and $\boldsymbol{\Psi}$ is $(r_{\mathbf{A}} \times N)$ (fat) matrix. The considered factorization problem is known to be subject to permutation and scale ambiguities. To remove scale ambiguities we impose to have the maximum value of each column of $\boldsymbol{\Psi}$ equal to 1.

Let us denote the set of admissible bases of the column space of $\mathbf{A}$ by $\mathcal{A} = \{\tilde{\mathbf{A}} \text{ of size } (M \times r_{\mathbf{A}}) \ / \ \text{span}(\tilde{\mathbf{A}}) = \text{span}(\mathbf{A})\}$. The problem of finding the factorization of $\mathbf{A}$ having the sparsest $\boldsymbol{\Psi}$ can then be formulated as follows:

$$\min_{(\tilde{\mathbf{A}},\boldsymbol{\Psi}),\ \tilde{\mathbf{A}}\in\mathcal{A},\ \mathbf{A}=\tilde{\mathbf{A}}\boldsymbol{\Psi}} \|\boldsymbol{\Psi}\|_0 \tag{6}$$

where $\|\boldsymbol{\Psi}\|_0$ stands for the ℓ_0 pseudo-norm of matrix $\boldsymbol{\Psi}$, that is the number of non-zero entries of $\boldsymbol{\Psi}$.

Proposition 1. *Let $\mathbf{A} = \tilde{\mathbf{A}}_1\boldsymbol{\Psi}_1 = \tilde{\mathbf{A}}_2\boldsymbol{\Psi}_2$ where both $\tilde{\mathbf{A}}_1$ and $\tilde{\mathbf{A}}_2$ are full column-rank matrices in $\mathcal{A}$ such that $\tilde{\mathbf{A}}_1$ is composed of $r_{\mathbf{A}} \leq N$ independent columns of $\mathbf{A}$ and $\tilde{\mathbf{A}}_2$ is composed of $r_{\mathbf{A}} \leq N$ independent linear combination of the columns of $\mathbf{A}_1$ which are not proportional to the columns of $\mathbf{A}$. If $r_{\mathbf{A}}$ satisfies $r_{\mathbf{A}}^2 - (N+1)r_{\mathbf{A}} + 2N \geq 0$, then $\|\boldsymbol{\Psi}_1\|_0 \leq \|\boldsymbol{\Psi}_2\|_0$.*

Proof. Let $\mathbf{A} = \tilde{\mathbf{A}}_1\boldsymbol{\Psi}_1 = \tilde{\mathbf{A}}_2\boldsymbol{\Psi}_2$ where both $\tilde{\mathbf{A}}_1$ and $\tilde{\mathbf{A}}_2$ are full column-rank matrices in $\mathcal{A}$. The matrix $\boldsymbol{\Psi}_1$ can be written as $\boldsymbol{\Psi}_1 = [\boldsymbol{\psi}_1(1)\cdots\boldsymbol{\psi}_1(N)]$ and $\boldsymbol{\Psi}_2 = [\boldsymbol{\psi}_2(1)\cdots\boldsymbol{\psi}_2(N)]$. Now we assume that $\tilde{\mathbf{A}}_1$ is composed of $r_{\mathbf{A}} \leq N$ independent columns of $\mathbf{A}$, which, without loss of generality, are assumed to be the first $r_{\mathbf{A}}$ columns: $\boldsymbol{\psi}_1(1), \cdots, \boldsymbol{\psi}_1(r_{\mathbf{A}})$. Thus, the number of non-zero elements of $\boldsymbol{\Psi}_1$ is given by :

$$\|\boldsymbol{\Psi}_1\|_0 = \sum_{i=1}^{N} \|\boldsymbol{\psi}_1(i)\|_0 = \sum_{i=1}^{r_{\mathbf{A}}} \|\boldsymbol{\psi}_1(i)\|_0 + \sum_{i=r_{\mathbf{A}}+1}^{N} \|\boldsymbol{\psi}_1(i)\|_0 = r_{\mathbf{A}} + \sum_{i=r_{\mathbf{A}}+1}^{N} \|\boldsymbol{\psi}_1(i)\|_0$$

As $\forall i = r_{\mathbf{A}}+1, \cdots, N,\ k_{\mathbf{A}} \leq \|\boldsymbol{\psi}_1(i)\|_0 \leq r_{\mathbf{A}}$, $\|\boldsymbol{\Psi}_1\|_0$ is bounded by : $r_{\mathbf{A}} + (N - r_{\mathbf{A}})k_{\mathbf{A}} \leq \|\boldsymbol{\Psi}_1\|_0 \leq r_{\mathbf{A}} + (N - r_{\mathbf{A}})r_{\mathbf{A}}$. Let us now consider the matrix $\mathbf{A}_2$. Since $\tilde{\mathbf{A}}_2$ is composed of $r_{\mathbf{A}} \leq N$ independent linear combination of the columns of $\mathbf{A}_1$ which are not proportional to the columns of $\mathbf{A}$, we have $\forall i = 1, \cdots, N, \|\boldsymbol{\psi}_2(i)\|_0 \geq 2$ and $\|\boldsymbol{\Psi}_2\|_0 \geq 2N$. Thus, $\|\boldsymbol{\Psi}_1\|_0 \leq \|\boldsymbol{\Psi}_2\|_0$ if $r_{\mathbf{A}} + (N - r_{\mathbf{A}})r_{\mathbf{A}} \leq 2N$, that is : $r_{\mathbf{A}}^2 - (N+1)r_{\mathbf{A}} + 2N \geq 0$.

Remark 1. It can be noticed that $r_{\mathbf{A}}^2 - (N+1)r_{\mathbf{A}} + 2N \geq 0$ is satisfied for all $r_{\mathbf{A}} \leq N \leq 6$. This is no longer true when $N > 6$.

Remark 2. The result of proposition 1 is based on the worst case scenario since it corresponds to the least favorable case $\|\boldsymbol{\psi}_1(i)\|_0 \leq r_{\mathbf{A}}$. This condition can be relaxed by imposing a more favorable situation such as : $\|\boldsymbol{\psi}_1(i)\|_0 \leq r_{\mathbf{A}} - k$ which results in $r_{\mathbf{A}} + (N - r_{\mathbf{A}})(r_{\mathbf{A}} - k) \leq 2N$, that is $r_{\mathbf{A}}^2 - (N + k + 1)r_{\mathbf{A}} + (2 + k)N \geq 0$.

Remark 3. In proposition 1, it is assumed that $\tilde{\mathbf{A}}_2$ does not include any column of $\mathbf{A}$. Let us now examine what happens if $\tilde{\mathbf{A}}_2$ does include a number of $l \leq r_{\mathbf{A}}$ columns of $\mathbf{A}$. In such a case, the sufficient condition for having $\|\boldsymbol{\Psi}_1\|_0 \leq \|\boldsymbol{\Psi}_2\|_0$ is $r_{\mathbf{A}}^2 - (N + k + 1)r_{\mathbf{A}} + (2 + k)N \geq l$.

The results corresponding to proposition 1 and remark 2 are shown in figure 1 for different values of N. The bottom curve corresponds to the case of proposition 1 ($k = 0$). The other curves are obtained for increasing values of k. For all the cases corresponding to the values of the plotted curves greater that the threshold (set to 0), the solutions obtained by considering independent columns of $\mathbf{A}$ are sparser than those obtained by considering linear combinations of columns of $\mathbf{A}$ which are not proportional to the columns of $\mathbf{A}$. As mentioned in remark 1, proposition 1 is always true for $N \leq 6$. The case $N = 6$ is depicted on the left-hand side of figure 1. The case of remark 3, which corresponds to having l of columns of $\mathbf{A}$ in $\tilde{\mathbf{A}}_2$, is simply obtained by shifting the threshold to a value equal to l.

To further illustrate these results, we provide next an example in which the condition of proposition 1 is not fulfilled and for which it is possible to find a sparsest decomposition with a basis not consisting of columns of $\mathbf{A}$. Let $\mathbf{A}$ be the following matrix :

$$\begin{aligned}\mathbf{A} &= [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{a}_4 \quad \mathbf{a}_1 - \mathbf{a}_2 - \mathbf{a}_4 \quad \mathbf{a}_1 + 2\mathbf{a}_2 - \mathbf{a}_3 + \mathbf{a}_4 \quad \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_3 \quad \mathbf{a}_2 - \mathbf{a}_3 - \mathbf{a}_4] \\ &= [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{a}_4] \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 2 & -1 & 1 \\ 0 & 0 & 1 & 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 1 & 0 & -1 \end{bmatrix} \qquad (7) \\ &= [(\mathbf{a}_1 + \mathbf{a}_4)/2 \quad (\mathbf{a}_1 - \mathbf{a}_4)/2 \quad \mathbf{a}_2 - (\mathbf{a}_1 - \mathbf{a}_4)/2 \quad \mathbf{a}_3 - \mathbf{a}_2 + (\mathbf{a}_1 - \mathbf{a}_4)/2] \\ &\quad \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}. \qquad (8)\end{aligned}$$

The number of non-zero elements of the matrix $\boldsymbol{\Psi}$ corresponding to the first decomposition equals 17 while that of the second is 16. Indeed, in that case $N = 8$ and $r_{\mathbf{A}} = 4$. Thus $r_{\mathbf{A}}^2 - (N+1)r_{\mathbf{A}} + 2N < 0$, and the sparsest decomposition cannot be guaranteed to correspond to a matrix $\tilde{\mathbf{A}}$ including only columns of $\mathbf{A}$. For example, the second decomposition (8) is sparser than the decomposition (7) .

Finally, to conclude this part, we can consider a special case where the dependencies take only the form of collinear loadings. This corresponds to having $k = r_{\mathbf{A}} - 1$ in $r_{\mathbf{A}}^2 - (N + k + 1)r_{\mathbf{A}} + (2 + k)N \geq 0$, yielding $r_{\mathbf{A}} \geq 0$ which is always satisfied. In other words, in the case of collinear loading only, the matrix $\tilde{\mathbf{A}}$ yielding the sparsest solution consists (obviously) in a selection of $r_{\mathbf{A}}$ independent columns of $\mathbf{A}$.

2.3 Uniqueness of the sparsest decomposition

In this part we aim at studying the uniqueness of the sparsest decomposition $\mathbf{A} = \tilde{\mathbf{A}}\boldsymbol{\Psi}$. For all the uniqueness results presented in this part we assume that the full column rank matrix $\tilde{\mathbf{A}}$ yielding the sparsest solution is a submatrix of $\mathbf{A}$. The uniqueness properties of the case where $\tilde{\mathbf{A}}$ is not a sub matrix of $\mathbf{A}$ are much more difficult to analyze. Thus, the matrix $\mathbf{A}$ can be

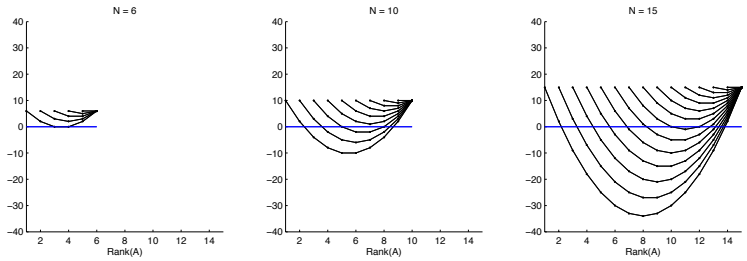

Fig. 1: Illustration of the sparsity properties : the bottom curve corresponds to the case of proposition 1. The others correspond to the case of remark 2 for different values of k. As k is increasing, there are much more situations in which the condition $\|\boldsymbol{\Psi}_1\|_0 \leq \|\boldsymbol{\Psi}_2\|_0$ is fulfilled.

written as: $\mathbf{A} = [\tilde{\mathbf{A}} \quad \breve{\mathbf{A}}] = [\tilde{\mathbf{a}}_1, \cdots, \tilde{\mathbf{a}}_{r_\mathbf{A}}, \breve{\mathbf{a}}_{r_\mathbf{A}+1}, \cdots, \breve{\mathbf{a}}_N]$. Defining the matrix $\mathbf{A}^i = [\tilde{\mathbf{A}} \quad \breve{\mathbf{a}}_i]$, with $i = r_\mathbf{A}+1, \ldots, N$, its Kruskal-rank satisfies $k_{\mathbf{A}^i} \leq k_\mathbf{A} \leq r_\mathbf{A}$ and each $\breve{\mathbf{a}}_i$ can be expressed as a linear combination of exactly $k_{\mathbf{A}^i}$ columns of $\tilde{\mathbf{A}}$. In other words : $\breve{\mathbf{a}}_i = \tilde{\mathbf{A}}\boldsymbol{\psi}(i)$ and $\|\boldsymbol{\psi}(i)\|_0 = k_{\mathbf{A}^i}$. We define the set $\tilde{\mathcal{A}} = \{\tilde{\mathbf{A}} \ / \ \tilde{\mathbf{A}} \text{ is a submatrix of } \mathbf{A} \text{ of size } (M \times r_\mathbf{A}), \text{span}(\tilde{\mathbf{A}}) = \text{span}(\mathbf{A})\} \subset \mathcal{A}$.

Proposition 2. *Let $\tilde{\mathbf{A}}_1 \neq \tilde{\mathbf{A}}_2$ two matrices of $\tilde{\mathcal{A}}$ and $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ the two matrices satisfying $\mathbf{A} = \tilde{\mathbf{A}}_1\boldsymbol{\Psi}_1 = \tilde{\mathbf{A}}_2\boldsymbol{\Psi}_2$. Then $\|\boldsymbol{\Psi}_1\|_0 \leq \|\boldsymbol{\Psi}_2\|_0$ if and only if $\sum_{r_\mathbf{A}+1}^N k_{\mathbf{A}_1^i} \leq \sum_{r_\mathbf{A}+1}^N k_{\mathbf{A}_2^i}$.*

Proof. After a proper column permutation, the matrix $\mathbf{A}$ can be written as $\mathbf{A} = [\tilde{\mathbf{A}}_1 \quad \breve{\mathbf{A}}_1]$, thus we have : $\mathbf{A} = \tilde{\mathbf{A}}_1\boldsymbol{\Psi}_1 = \tilde{\mathbf{A}}_1 [\mathbf{I}_{r_\mathbf{A}}, \boldsymbol{\psi}(r_\mathbf{A}+1) \cdots \boldsymbol{\psi}(N)]$ where $\mathbf{I}_{r_\mathbf{A}}$ is the identity matrix of dimension $r_\mathbf{A}$. The number of non-zero elements of $\boldsymbol{\Psi}_1$ is given by $\|\boldsymbol{\Psi}_1\|_0 = r_\mathbf{A} + \sum_{r_\mathbf{A}+1}^N \|\boldsymbol{\psi}_1(i)\|_0$, which, as shown earlier, is equivalent to $\|\boldsymbol{\Psi}_1\|_0 = r_\mathbf{A} + \sum_{r_\mathbf{A}+1}^N k_{\mathbf{A}_1^i}$. Similarly we can write $\|\boldsymbol{\Psi}_2\|_0 = r_\mathbf{A} + \sum_{r_\mathbf{A}+1}^N k_{\mathbf{A}_2^i}$. It follows immediately that $\|\boldsymbol{\Psi}_1\|_0 \leq \|\boldsymbol{\Psi}_2\|_0$ if and only if $\sum_{r_\mathbf{A}+1}^N k_{\mathbf{A}_1^i} \leq \sum_{r_\mathbf{A}+1}^N k_{\mathbf{A}_2^i}$.

A straightforward extension of proposition 2 is given by the following proposition which gives the condition for having the sparsest and possibly unique decomposition of $\mathbf{A}$:

Proposition 3. *If $\exists\, \tilde{\mathbf{A}}_0 \in \tilde{\mathcal{A}}$ such as $\forall \tilde{\mathbf{A}} \in \tilde{\mathcal{A}}, \tilde{\mathbf{A}} \neq \tilde{\mathbf{A}}_0$, $\sum_{r_\mathbf{A}+1}^N k_{\mathbf{A}_0^i} \leq \sum_{r_\mathbf{A}+1}^N k_{\mathbf{A}^i}$ then the decomposition $\mathbf{A} = \tilde{\mathbf{A}}_0\boldsymbol{\Psi}_0$ is the sparsest decomposition. If the inequality is strict ($<$), the sparsest decomposition is unique.*

It should be noted that proposition 3 does not provide any effective means to find the sparsest decomposition of $\mathbf{A}$. Indeed, finding it would require to find all the possible basis consisting in $r_\mathbf{A}$ columns of $\mathbf{A}$ which is actually an NP-complete combinatorial problem.

The result of proposition 3 can be specialized into the following cases:

– the linear dependencies between the columns of $\mathbf{A}$ are only colinearities. In that case $\tilde{\mathcal{A}}$ only includes a single element since any selection of $r_{\mathbf{A}}$ independent columns of $\mathbf{A}$ will result in the same basis up to scale and permutation. Thus the decomposition is unique.
– $\mathbf{A}$ has a Kruskal-rank equal to its rank *i.e.* $k_{\mathbf{A}} = r_{\mathbf{A}}$. In that case, any selection of $r_{\mathbf{A}}$ columns of $\mathbf{A}$ is a basis and $\forall \tilde{\mathbf{A}} \in \tilde{\mathcal{A}}$, $k_{\mathbf{A}^i} = k_{\mathbf{A}}, \forall i = r_{\mathbf{A}} + 1, \cdots, N$. Thus $\forall \tilde{\mathbf{A}} \in \tilde{\mathcal{A}}$, $\|\boldsymbol{\Psi}\|_0 = (N - r_{\mathbf{A}} + 1)r_{\mathbf{A}}$.

In practice, having a unique sparsest matrix $\boldsymbol{\Psi}$ is not crucial. Indeed, from a PARALIND point of view, having a number of decompositions yielding the same degree of sparsity simply means that all these decompositions are equivalent.

3 Conclusion

In this paper we provided some rank-based results for the identifiability of the PARALIND model with sparse interaction matrices (S-PARALIND). More precisely, we prove a condition that indicates in which cases, choosing the PARALIND loadings between the loadings of the associated PARAFAC decomposition yields the sparsest interaction matrix. These results could be helpful for the design and the analysis of ℓ_0-based algorithms for the decomposition of bilinear/multilinear arrays.

References

1. Aharon, M., Elad, M., Bruckstein, A.: On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them. Linear Algebra Applicat. 416, 48–67 (Jul 2006)
2. de Almeida, A.L.F., Favier, G., Mota, J.C.M.: A constrained factor decomposition with application to MIMO antenna systems. IEEE Trans. Signal Process. 56(6), 2429–2442 (Jun 2008)
3. de Almeida, A.L.F., Favier, G., Mota, J.C.M.: Constrained tensor modeling approach to blind multiple-antenna CDMA schemes. IEEE Trans. Signal Process. 56(6), 2417–2428 (Jun 2008)
4. Bahram, M., Bro, R.: A novel strategy for solving matrix effect in three-way data using parallel profiles with linear dependencies. Analytica Chimica Acta 584(2), 397–402 (2007)
5. Brie, D., Klotz, R., Miron, S., Moussaoui, S., Mustin, C., Ph., B., Grandemange, S.: Joint analysis of flow cytometry data and fluorescence spectra as a non-negative array factorization problem. Chemometr. Intell. Lab. 137, 21–32 (Oct 2014)
6. Bro, R., Harshman, R.A., Sidiropoulos, N.D., Lundy, M.E.: Modeling multi-way data with linearly dependent loadings. J. Chemometr. 23(7-8), 324–340 (Jul-Aug 2009), special Issue: In Honor of Professor Richard A. Harshman
7. Caland, F., Miron, S., Brie, D., Mustin, C.: A blind sparse approach for estimating constraint matrices in paralind data models. In: Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European. pp. 839–843. Bucharest, Romania (Aug 2012)
8. Carroll, J.D., Chang, J.J.: Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika 35(3), 283–319 (Sep 1970)

9. Chen, H., Zheng, B., Song, Y.: Comparison of PARAFAC and PARALIND in modeling three-way fluorescence data array with special linear dependences in three modes: a case study in 2-naphthol. Chemometr. Intell. Lab. 25(1), 20–27 (2011)
10. De Lathauwer, L.: Decompositions of a higher-order tensor in block terms - Part II: Definitions and uniqueness. SIAM J. Matrix Anal. Appl. 30(3), 1033–1066 (2008)
11. Georgiev, P., Theis, F., Cichocki, A.: Sparse component analysis and blind source separation of underdetermined mixtures. IEEE Trans. Neural Netw. 16(4), 992–996 (Jul 2005)
12. Gribonval, R., Schnass, K.: Dictionary identification - sparse matrix-factorisation via ℓ_1-minimisation. IEEE Trans. Inf. Theory 56(7), 3523–3539 (Jul 2010)
13. Guo, X., Miron, S., Brie, D., Stegeman, A.: Uni-mode and partial uniqueness conditions for CANDECOMP/PARAFAC of three-way arrays with linearly dependent loadings. SIAM J. Matrix Anal. Appl. 33(1), 111–129 (2012)
14. Harshman, R.A.: Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimodal factor analysis. UCLA Working Papers in Phonetics 16, 1–84 (Dec 1970)
15. Kruskal, J.B.: Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra Applicat. 18(2), 95–138 (1977)
16. Liu, X., Guang, L., Yang, L., Zhu, H.: PARALIND-based blind joint angle and delay estimation for multipath signals with uniform linear array. EURASIP J. Appl. Signal Process. 2012(1), 1–13 (2012)
17. Nion, D., De Lathauwer, L.: A block component model-based blind DS-CDMA receiver. IEEE Trans. Signal Process. 56(11), 5567–5579 (2008)
18. Stegeman, A., de Almeida, A.L.F.: Uniqueness conditions for constrained three-way factor decompositions with linearly dependent loadings. SIAM J. Matrix Anal. Appl. 31(3), 1469–1499 (2009)
19. Xiaofei, Z., Fei, W., Dazhuan, X.: Blind signal detection algorithm for mimo-ofdm systems over multipath channel using paralind model. IET communications 5(5), 606–611 (2011)
20. Zhang, X., Gao, X., Wang, Z.: Blind paralind multiuser detection for smart antenna cdma system over multipath fading channel. Progress In Electromagnetics Research 89, 23–38 (2009)
21. Zhang, X., Zhou, M., Li, J.: A PARALIND decomposition-based coherent two-dimensional direction of arrival estimation algorithm for acoustic vector-sensor arrays. Sensors 13(4), 5302–5316 (2013)