Low rank canonical polyadic decomposition of tensors based on group sparsity

Xu Han[1,2,4], Laurent Albera[1,2,4], Amar Kachenoura[1,2,4], Lotfi Senhadji[1,2,4], and Huazhong Shu[3,4]
[1]LTSI, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France
[2]INSERM, U 1099, 35402 Rennes Cedex, France
[3]LIST, Southeast University, 2 Sipailou, 210096, Nanjing, China
[4]Centre de Recherche en Information Biomédicale Sino-Français (CRIBs), Rennes, France
Email: hanxu.list@gmail.com, {laurent.albera, amar.kachenoura, lotfi.senhadji}@univ-rennes1.fr, shu.list@seu.edu.cn

***Abstract*—A new and robust method for low rank Canonical Polyadic (CP) decomposition of tensors is introduced in this paper. The proposed method imposes the Group Sparsity of the coefficients of each Loading (GSL) matrix under orthonormal subspace. By this way, the low rank CP decomposition problem is solved without any knowledge of the true rank and without using any nuclear norm regularization term, which generally leads to computationally prohibitive iterative optimization for large-scale data. Our GSL-CP technique can be then implemented using only an upper bound of the rank. It is compared in terms of performance with classical methods, which require to know exactly the rank of the tensor. Numerical simulated experiments with noisy tensors and results on fluorescence data show the advantages of the proposed GSL-CP method in comparison with classical algorithms.**

I. Introduction

The development of multi-way array decomposition methods always attracts the attention in numerous domains. The low rank Canonical Polyadic model is the most famous one [1]. It is widely used in signal processing [2] [3], image processing [4] and brain source imaging [5] [6]. Since a few decades and Harshman's pioneer work [7], many algorithms have been proposed to compute the CP decomposition. Among them, the Alternating Least Squares (ALS) method, simultaneously introduced by Carroll and Chang [8] and Harshman [7], is the most famous one. Several modifications have been proposed in order to improve its behavior, especially in the presence of bottlenecks [9]. For instance, Enhanced Line Search (ELS) procedures, based on a sophisticated extrapolation scheme, using information on nonlinear trends in the parameters, was designed [10] [11]. Despite the practical good results of the ELS-ALS technique, no global minimization of the used data-fit objective function is guaranteed. More recently, a semi-algebraic method, namely the DIrect AlGorithm for CP decomposition (DIAG), has been proposed [12]. Instead of minimizing the ALS data-fit objective function, it formulates the CP decomposition as a Joint EigenValue Decomposition (JEVD) problem [13], followed by rank-1 tensor approximations.

The aforementioned methods require that the rank has been estimated in a preprocessing step. Unfortunately, learning the exact rank of the tensor can be difficult in some particular cases, such as for low Signal-to-Noise Ratio (SNR) values. Thus, the rank of the tensor could be added to the objective function to be minimized in order to fit the CP model with the lowest rank. Nevertheless, since the rank of a multi-way array is discrete, rank minimization problems are usually hard to solve and sometimes NP hard. Then researchers usually replace rank in the objective function with nuclear norm [14]. It corresponds to the sum of singular values, which can be used as convex envelope of the rank function [15]. However, iterative nuclear norm minimization can not obtain the group sparse result of objective matrix.

In this paper, we propose a new method to achieve efficient low rank CP decomposition. We use a rank measure proposed by Shu et al. [16] for low rank matrices. This rank measure is lower bounded by nuclear norm and it has the same global minimum as the latter [16]. Based on it, we reformulate the low rank CP decomposition problem by imposing Group Sparsity on the coefficients of each Loading matrix (GSL) under orthonormal subspace. This allows us to solve the low rank CP decomposition problem without any knowledge of the true rank. Numerical simulated experiments with noisy tensors and results on fluorescence data show the advantages of the proposed GSL-CP method in comparison with classical algorithms such as ALS [7], ELS-ALS [10] and DIAG [12]. In particular, a series of experiments including amino acids fluorescence data [1] show that GSL-CP gives good results when only an upper bound of the rank is used for initialization. Although the concept of the GSL-CP approach is presented with third order tensors, the proposed method can be easily generalized to higher orders.

The paper is organized as follows. Section II gives some notations and recalls some definitions. In Section III, we present the GSL-CP algorithm and give its numerical complexity. Numerical simulated experiments are reported in Section IV. Conclusion is drawn in the fifth section.

II. Notations and Preliminaries

Throughout this paper, we denote multi-way arrays by a bold calligraphic letter, e.g., $\boldsymbol{\mathcal{A}}$. A matrix is denoted by a bold capital letter, e.g., $\mathbf{A}$, vectors are denoted by bold lowercase letters, e.g., $\mathbf{a}$ and scalars are denoted by lowercase letters, e.g., a. $\odot$ and $\otimes$ denote Khatri-Rao product and Kronecker product,

ISBN 978-0-9928626-7-1

respectively. The mode-i unfolding matrix of the tensor $\boldsymbol{\mathcal{A}}$ is denoted by $\mathbf{A}^{(i)}$. The inner product of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, assuming that both tensors have the same dimensions, is defined as:

$$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=2}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N} y_{i_1 i_2 \cdots i_N}$$

with $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}) \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N} \times \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. The Frobenius norm of $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined by:

$$\|\boldsymbol{\mathcal{X}}\|_F = \sqrt{\sum_{i_1 i_2 \ldots i_N} (\mathcal{X}_{i_1 i_2 \ldots i_N})^2}$$

Given a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, the symbol $\mathbf{X}_{i,:}$ denotes the ith row of $\mathbf{X}$. The vectorization of $\mathbf{X}$ is denoted by $\mathrm{vec}(\mathbf{X}) \in \mathbb{R}^{mn \times 1}$. The nuclear norm and and the mixed-norm of $\mathbf{X}$ are denoted by $\|\mathbf{X}\|_*$ and $\|\mathbf{X}\|_{2,1}$, respectively, with:

$$\|\mathbf{X}\|_{2,1} = \sum_{i=1}^{m} \sqrt{\sum_{j=1}^{n} (\mathbf{X}_{i,j})^2} = Tr[\mathbf{X}^\intercal \boldsymbol{\Phi} \mathbf{X}]$$

where $Tr[.]$ is the trace operator and $\boldsymbol{\Phi}$ is a diagonal matrix with $\Phi_{i,i} = 1/\sqrt{\sum_{j=1}^{n} (\mathbf{X}_{i,j})^2}$ standing for the (i,i)-th component of $\boldsymbol{\Phi}$. Note that since the denominator may be equal to or close to zero, it is necessary to add a small value ε, leading to $\Phi_{i,i} = 1/\big(\sqrt{\sum_{j=1}^{n} (\mathbf{X}_{i,j})^2} + \varepsilon\big)$.

III. The GSL-CP Method

The rank-R CP decomposition of a three-way array $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ is given by:

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{B}} = \sum_{r=1}^{R} \mathbf{f}_r^{(1)} \circ \mathbf{f}_r^{(2)} \circ \mathbf{f}_r^{(3)} + \boldsymbol{\mathcal{B}} \qquad (1)$$

where $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ is a noise three-way array. Let $\mathbf{F}^{(1)} = [\mathbf{f}_1^{(1)}, \ldots, \mathbf{f}_R^{(1)}]$, $\mathbf{F}^{(2)} = [\mathbf{f}_1^{(2)}, \ldots, \mathbf{f}_R^{(2)}]$, and $\mathbf{F}^{(3)} = [\mathbf{f}_1^{(3)}, \ldots, \mathbf{f}_R^{(3)}]$ be the loading matrices of the three-way array $\boldsymbol{\mathcal{A}}$. Then $\mathbf{A}^{(1)} = \mathbf{F}^{(1)} (\mathbf{F}^{(3)} \odot \mathbf{F}^{(2)})^\intercal$, $\mathbf{A}^{(2)} = \mathbf{F}^{(2)} (\mathbf{F}^{(3)} \odot \mathbf{F}^{(1)})^\intercal$ and $\mathbf{A}^{(3)} = \mathbf{F}^{(3)} (\mathbf{F}^{(2)} \odot \mathbf{F}^{(1)})^\intercal$ will denote the three unfolding matrices of $\boldsymbol{\mathcal{A}}$. Without any loss of generality, we will assume in the sequel that $N_1 \leq N_2 \leq N_3$. Since we deal with low rank tensors, we can assume that $R \ll N_1$.

A. Problem formulation

One distinctive characteristic of our low rank CP decomposition method is to impose the low rank constraint on the loading matrices $\mathbf{F}^{(i)}$ directly. So the objective function that we propose to minimize in this paper is given by:

$$\min_{\mathbf{F}^{(1)}, \mathbf{F}^{(2)}, \mathbf{F}^{(3)}} \sum_{i=1}^{3} \mathrm{rank}(\mathbf{F}^{(i)}) \quad \text{s.t.} \quad \boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{B}} \qquad (2)$$

According to [17, theorem 1], the nuclear norm $\|\mathbf{X}\|_*$ can be treated as convex envelope of $\mathrm{rank}(\mathbf{X})$. However, it cannot guarantee simultaneously the low rank and the sparsity of $\mathbf{X}$ if the latter matrix enjoys both properties. So, let's see how to avoid the use of the nuclear norm using the following results [16]:

***Lemma** 1: Consider a thin matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ (i.e. $m \geq n$), its SVD and orthonormal subspace decomposition are denoted by $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^\intercal$ and $\mathbf{A} = \mathbf{D}\boldsymbol{\alpha}$, respectively, where $\mathbf{D} \in \mathbb{R}^{m \times n}$, $\boldsymbol{\alpha} \in \mathbb{R}^{n \times n}$ and $\mathbf{D}^\intercal \mathbf{D} = \mathbf{I}_n$ without any loss of generality. The minima of the row-0 and row-1 group sparsity measures of $\mathbf{A}$ with respect to $\boldsymbol{\alpha}$ such that $\mathbf{A} = \mathbf{D}\boldsymbol{\alpha}$ and $\mathbf{D}^\intercal \mathbf{D} = \mathbf{I}_n$ are given by $\mathrm{rank}(\mathbf{A})$ and $\|\mathbf{A}\|_*$, respectively.*

Note that the row-0 norm of $\boldsymbol{\alpha}$ is the number of non-zero rows of $\boldsymbol{\alpha}$ while the row-1 norm of $\boldsymbol{\alpha}$ is defined as the mixed-norm of $\boldsymbol{\alpha}$, i.e. $\|\boldsymbol{\alpha}\|_{\mathrm{row}-1} = \|\boldsymbol{\alpha}\|_{2,1}$ and the relationship between $\mathbf{D}$ and $\mathbf{U}$ is constructed by a rotation matrix $\boldsymbol{\Omega}$, i.e. $\mathbf{D} = \mathbf{U}\boldsymbol{\Omega}$.

***Proposition** 1: Let $\mathbf{F} \in \mathbb{R}^{m \times n}$ with $m \geq n$ be a rank-R matrix, then we have the following inequality:*

$$\|\mathbf{D}\mathbf{D}^\intercal \mathbf{F}\|_* \leq \|\mathbf{D}^\intercal \mathbf{F}\|_{2,1} \qquad (3)$$

where $\mathbf{D} \in \mathbb{R}^{m \times n}$ is a random orthonormal basis.

Proof. The proof of proposition 1 is straightforward. Indeed, from lemma 1, we can easily conclude that $\|\boldsymbol{\alpha}\|_{2,1}$ is greater than $\|\mathbf{A}\|_*$. If we replace $\boldsymbol{\alpha}$ and $\mathbf{A}$ by $\mathbf{D}^T\mathbf{F}$ and $\mathbf{D}\mathbf{D}^T\mathbf{F}$, respectively, then we obtain equation (3).

Note that if $n = m$, then we have $\mathbf{D}\mathbf{D}^\intercal = \mathbf{I}$ and proposition 1 means $\|\mathbf{F}\|_* \leq \|\mathbf{D}^\intercal \mathbf{F}\|_{2,1}$. Furthermore, when $\mathbf{D}$ is the identity matrix, inequality (3) becomes $\|\mathbf{F}\|_* \leq \|\mathbf{F}\|_{2,1}$. Based on lemma 1 and proposition 1, it is then possible to minimize the nuclear norm by minimizing the mixed-norm, which will ensure the low rank property. This will also ensure group sparsity along the row direction due to the definition of the mixed-norm. The minimization problem defined by equation (2) is then be reformulated as follows:

$$\min_{\mathbf{F}^{(1)}, \mathbf{F}^{(2)}, \mathbf{F}^{(3)}} \sum_{i=1}^{3} \|\mathbf{D}^{(i)^\intercal} \mathbf{F}^{(i)}\|_{2,1} \quad \text{s.t.} \quad \boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{B}} \qquad (4)$$

where the $\mathbf{D}^{(i)}$ matrices are square random orthonormal bases.

B. Optimization scheme

Let's begin by writing the augmented Lagrangian objective function derived from (4):

$$\begin{aligned} \mathcal{L}(\{\mathbf{F}^{(i)}\}, \boldsymbol{\mathcal{Y}}) \;=\; & \sum_{i=1}^{3} \lambda_i Tr\big[(\mathbf{D}^{(i)^\intercal} \mathbf{F}^{(i)})^\intercal \boldsymbol{\Phi}^{(i)} (\mathbf{D}^{(i)^\intercal} \mathbf{F}^{(i)})\big] \\ & + \big\langle \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{A}} \big\rangle + \frac{u}{2} \|\boldsymbol{\mathcal{T}} - \boldsymbol{\mathcal{A}}\|_F^2 \end{aligned} \qquad (5)$$

where λ_i are penalty parameters and $\boldsymbol{\mathcal{Y}}$ is the multiplier tensor. Then the function $\mathcal{L}$ is minimized using the method described in [18], derived from the alternating direction method of multipliers [19]. More particularly, at the $(k+1)$-th iteration of the method, $\mathbf{F}_{(k+1)}^{(1)}$ is computed by vanishing the gradient of $\mathcal{L}$ with respect to $\mathbf{F}^{(1)}$:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{F}^{(1)}} = 2\lambda_1 \mathbf{D}^{(1)} \boldsymbol{\Phi}^{(1)} \mathbf{D}^{(1)^\intercal} \mathbf{F}^{(1)} + u \mathbf{F}^{(1)} (\mathbf{F}^{(3)} \odot \mathbf{F}^{(2)})^\intercal (\mathbf{F}^{(3)} \\ \odot \mathbf{F}^{(2)}) - \big(\mathbf{Y}^{(1)} + u \mathbf{T}^{(1)}\big)(\mathbf{F}^{(3)} \odot \mathbf{F}^{(2)}) = \mathbf{0} \end{aligned} \qquad (6)$$

ISBN 978-0-9928626-7-1 © EURASIP 2017

Now, equation $\mathbf{AX} + \mathbf{XB} = \mathbf{C}$ implies (see Lyapunov equation):

$$\mathrm{vec}(\mathbf{X}) = (\mathbf{I} \otimes \mathbf{A} + \mathbf{B}^{\intercal} \otimes \mathbf{I})^{-1}\mathrm{vec}(\mathbf{C}) \tag{7}$$

Consequently, we obtain:

$$\begin{aligned}\mathrm{vec}(\mathbf{F}^{(1)}_{(k+1)}) = \Big(\mathbf{I} \otimes \big(2\lambda_1 \mathbf{D}^{(1)}\mathbf{\Phi}^{(1)}_k\mathbf{D}^{(1)^{\intercal}}\big) \\ + \big[u_k\big(\mathbf{F}^{(3)}_k \odot \mathbf{F}^{(2)}_k\big)^{\intercal}\big(\mathbf{F}^{(3)}_k \odot \mathbf{F}^{(2)}_k\big)\big] \otimes \mathbf{I}\Big)^{-1} \\ \mathrm{vec}\Big(\big(\mathbf{Y}^{(1)}_k + u_k\mathbf{T}^{(1)}\big)\big(\mathbf{F}^{(3)}_k \odot \mathbf{F}^{(2)}_k\big)\Big)\end{aligned} \tag{8}$$

The vectors $\mathrm{vec}(\mathbf{F}^{(2)}_{(k+1)})$ and $\mathrm{vec}(\mathbf{F}^{(3)}_{(k+1)})$ can be computed similarly:

$$\begin{aligned}\mathrm{vec}(\mathbf{F}^{(2)}_{(k+1)}) = \Big(\mathbf{I} \otimes \big(2\lambda_2 \mathbf{D}^{(2)}\mathbf{\Phi}^{(2)}_k\mathbf{D}^{(2)^{\intercal}}\big) \\ + \big[u_k\big(\mathbf{F}^{(3)}_k \odot \mathbf{F}^{(1)}_{(k+1)}\big)^{\intercal}\big(\mathbf{F}^{(3)}_k \odot \mathbf{F}^{(1)}_{(k+1)}\big)\big] \otimes \mathbf{I}\Big)^{-1} \\ \mathrm{vec}\Big(\big(\mathbf{Y}^{(2)}_k + u_k\mathbf{T}^{(2)}\big)\big(\mathbf{F}^{(3)}_k \odot \mathbf{F}^{(1)}_{(k+1)}\big)\Big)\end{aligned} \tag{9}$$

and:

$$\begin{aligned}\mathrm{vec}(\mathbf{F}^{(3)}_{(k+1)}) = \Big(\mathbf{I} \otimes \big(2\lambda_3 \mathbf{D}^{(3)}\mathbf{\Phi}^{(3)}_k\mathbf{D}^{(3)^{\intercal}}\big) \\ + \big[u_k\big(\mathbf{F}^{(2)}_{(k+1)} \odot \mathbf{F}^{(1)}_{(k+1)}\big)^{\intercal}\big(\mathbf{F}^{(2)}_{(k+1)} \odot \mathbf{F}^{(1)}_{(k+1)}\big)\big] \otimes \mathbf{I}\Big)^{-1} \\ \mathrm{vec}\Big(\big(\mathbf{Y}^{(3)}_k + u_k\mathbf{T}^{(3)}\big)\big(\mathbf{F}^{(2)}_{(k+1)} \odot \mathbf{F}^{(1)}_{(k+1)}\big)\Big)\end{aligned} \tag{10}$$

Regarding the update of the multiplier tensor, the rule is given by:

$$\mathcal{Y}_{(k+1)} = \mathcal{Y}_k + u_k(\mathcal{T} - \mathcal{A}_{(k+1)}) \tag{11}$$

In this paper, the orthonormal bases $\mathbf{D}^{(i)}$ for $i = 1, 2, 3$, are fixed and choosen with identity matrix and we expect group sparsity takes more works when the initialized rank R_{ini} is more over-estimated such that $\lambda_i = \sqrt{R_{ini}}, i = 1, 2, 3$. The whole implemented procedure is summarized in **Algorithm** 1.

C. Numerical complexity

The numerical complexity of GSL-CP is analyzed in terms of number of floating point operations (flops). A flop includes a multiplication and an addition. But in practice, multiplication operation is more time-consuming than addition computing, so we only consider the number of multiplications: this does not affect the order of magnitude of the numerical complexity. Note that in order to reduce the cost, we used the $\mathbf{LDL}^{\intercal}$ decomposition to compute the matrix inverse. The computational complexity of GSL-CP, denoted by Γ_{GSL}, mainly depends on the initialized rank R_{ini} and the size $(N_1 \times N_2 \times N_3)$ of the tensor. It is given by:

$$\begin{aligned}\Gamma_{\mathrm{GSL}} = \big[(N_1^3 + N_2^3 + N_3^3)/6\big]R_{ini}^3 + \big[2(N_1^2 + N_2^2 + N_3^2) \\ +(N_1N_2 + N_1N_3 + N_2N_3) + (N_1 + N_2 + N_3) + 3\big]R_{ini}^2 \\ +\big[4N_1N_2N_3 + N_1^2 + N_2^2 + N_3^2 + N_1N_2 + N_1N_3 \\ +N_2N_3 - \frac{7}{6}(N_1 + N_2 + N_3)\big]R_{ini} + (N_1^3 + N_2^3 \\ +N_3^3 + N_1^2 + N_2^2 + N_2^2 + 2N_1N_2N_3)\end{aligned}$$

Algorithm 1: Implementation of the GSL-CP method

Input: tensor data $\mathcal{T}$, the maximum number k_{max} of iterations.

Initialization: tensor $\mathcal{Y}_0$, loading matrices $\mathbf{F}^{(i)}_0$, $\mathbf{\Phi}^{(i)}_0, i = 1, 2, 3$, tolerance value tol, parameter $\rho > 1$, u_0, u_{max}, λ_i for $i = 1, 2, 3$.

Do

Update $\mathbf{F}^{(1)}_{(k+1)}, \mathbf{F}^{(2)}_{(k+1)}, \mathbf{F}^{(3)}_{(k+1)}, \mathcal{Y}_{(k+1)}$ using (8)-(11).

Update $\mathbf{\Phi}^{(i)}_{(k+1)}$ from the $\mathbf{F}^{(i)}_{(k+1)}$ matrices.

Update $u_{(k+1)}$ as $u_{(k+1)} = \min(\rho u_k, u_{max})$.

Update $crit_{(k+1)}$ as $crit_{(k+1)} = (\|\mathcal{T} - \mathcal{A}_{(k+1)}\|_F - \|\mathcal{T} - \mathcal{A}_{(k)}\|_F)/\|\mathcal{T} - \mathcal{A}_{(k)}\|_F$.

Update k as $k := k + 1$.

Until $crit_k < tol$ or $k < k_{max}$.

Output: $\mathbf{F}^{(1)}_k$, $\mathbf{F}^{(2)}_k$ and $\mathbf{F}^{(3)}_k$.

IV. Computer results

In order to study the robustness and effectiveness of GSL-CP, we perform two kinds of experiments: the first one with simulated noisy data and the second one with amino acids fluorescence data [1]. Regarding the first experiment, the rank-R tensor $\mathcal{A} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ is generated with random loading matrices $\mathbf{F}^{(i)} \in \mathbb{R}^{N_i \times R}, i = 1, 2, 3$, whose elements follow a Gaussian distribution. The noise tensor $\mathcal{B}$ is also sampled from a Gaussian distribution. The tensor $\mathcal{T}$ is then obtained as follows:

$$\mathcal{T} = \frac{\mathcal{A}}{\|\mathcal{A}\|_F} + \sigma\frac{\mathcal{B}}{\|\mathcal{B}\|_F} \tag{12}$$

where the parameter σ controls the SNR defined by $\mathrm{SNR} = -20\log_{10}(\sigma)$.

The estimation accuracy is evaluated by means of the following Δ measure:

$$\Delta = \frac{1}{3}\sum_{i=1}^{3}\sum_{n=1}^{R}\min_{(n,n_k)\in I_n^2}\left(d\left(\mathbf{f}_n^{(i)}, \mathbf{f}_{n_{est}}^{(i)}\right)\right) \tag{13}$$

where $\mathbf{F}^{(i)}$ and $\mathbf{F}^{(i)}_k$ denote the i-th true loading matrix and the one estimated from k iterations of the GLS-CP method, respectively, and where $\mathbf{f}_n^{(i)}$ and $\mathbf{f}_{n_k}^{(i)}$ are the n-th column of $\mathbf{F}^{(i)}$ and the n_k-th column of $\mathbf{F}^{(i)}_k$. I_n^2 is defined recursively from $I_1^2 = \{1, \ldots, R\} \times \{1, \ldots, R_k\}$, where R_k is the estimated rank, and from $I_{n+1}^2 = I_n^2 - J_n^2$ with:

$$J_n^2 = \arg\min_{[(n,n_k)\in I_n^2]} d\left(\mathbf{f}_n^{(i)}, \mathbf{f}_{n_k}^{(i)}\right) \tag{14}$$

The definition of the pseudo-distance d between two vectors is given by:

$$d\left(\mathbf{f}_n^{(i)}, \mathbf{f}_{n_k}^{(i)}\right) = 1 - \frac{\left\|\mathbf{f}_n^{(i)^{\intercal}}\mathbf{f}_{n_k}^{(i)}\right\|^2}{\left\|\mathbf{f}_n^{(i)^{\intercal}}\right\|^2\left\|\mathbf{f}_{n_k}^{(i)}\right\|^2} \tag{15}$$

ISBN 978-0-9928626-7-1 © EURASIP 2017

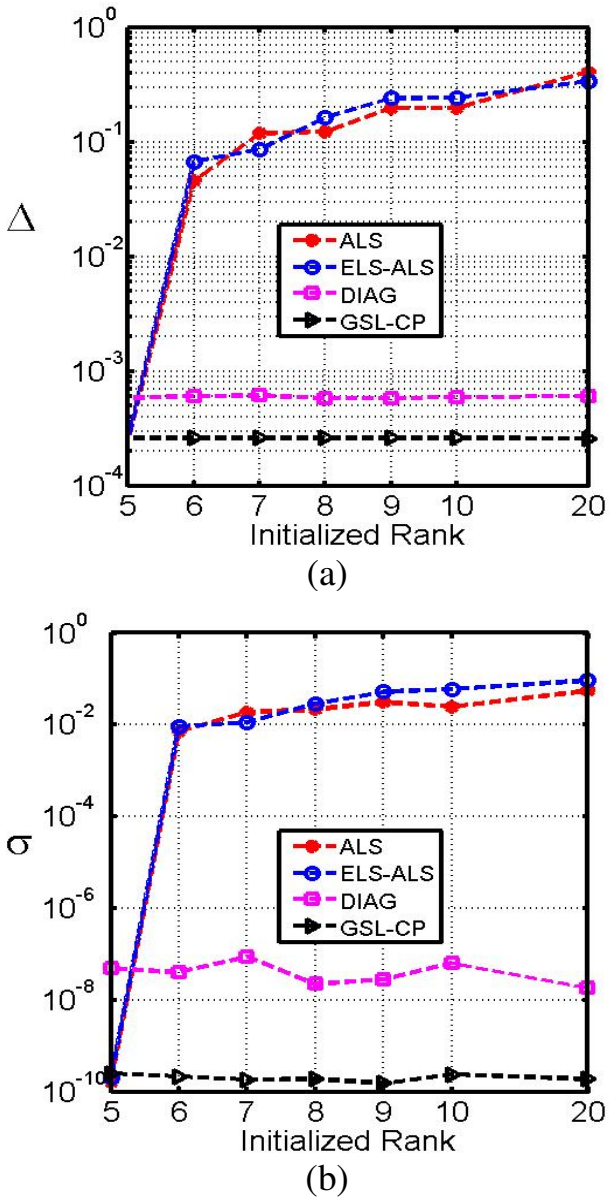

Fig. 1: Low rank CP decomposition with overfactoring for a true rank equal to 5 and an SNR value of 10 dB.

A. Experiment on simulated noisy tensors

We first investigate the behavior of GSL-CP applied to simulated low rank noisy tensors of size $(100 \times 100 \times 100)$ in comparison with ALS [7], ELS-ALS [10] and DIAG [12]. More particularly, we analyse the influence of a overestimated rank used for initialization. All the 50 independent Monte Carlo (MC) trials are stopped when they satisfy the convergence criterion ($tol = 10^{-6}$ and $k_{max} = 1000$). Loading matrices following a normal distribution are used to initialize ALS, ELS-ALS and GSL-CP.

Figure 1 shows the Δ measure averaged over the 50 MC runs at the output of the four methods as a function of the estimated rank used to initialize them for an SNR value of 10 dB and a true rank equal to 5. It appears from figure 1 that all the methods give good results if they use the true rank value as initialization. While ALS and ELS-ALS give poor results as the overestimated rank increases, DIAG and GSL-CP show their robustness with respect to the overfactoring problem. And the σ figure gives the variance of 50 trials on Δ under each initialized case. It is not hard to see that DIAG and GSL-CP methods are more robust than others. As far the computation complexity is concerned, table I gives the averaged numerical complexities for an initialized rank of 10 and SNR=10dB. It is not hard to see that when the rank is not so large, the numerical complexities of the four methods are very similar.

TABLE I: Average numerical complexities with an initialized rank of 10 and SNR=10dB.

ALS	ELS-ALS	DIAG	GSL-CP
6.0690×10^9	1.3125×10^{10}	1.5451×10^9	1.8315×10^{10}

B. Experiment on amino acids fluorescence data

In this experiment, we apply the GSL-CP method to Amino acids fluorescence data, which consists of five simple laboratory-made samples [1]. Each sample contains different amounts of tyrosine, tryptophan and phenylalanine dissolved in phosphate buffered water. The samples were measured by fluorescence (excitation 240-300 nm, emission 250-450 nm, 1 nm intervals) on a PE LS50B spectrofluorometer with excitation slit-width of 2.5 nm, an emission slit-width of 10 nm and a scan-speed of 1500 nm/s. The three-way array to be decomposed is hence of size $5 \times 61 \times 201$.

Figure 2 shows the emission-mode factors by using three different initialized rank values (3, 4 and 5). The first line shows the results of all methods for an accurate rank initialization: all the methods give good results. The results of the second line are obtained by using an initialized rank equal to 4: ALS, ELS-ALS and DIAG still have ability to find three curves (red, pink and blue), but their cyan curve is non-zero showing their limitation contrarily to GSL-CP. Regarding the last line, an initialized rank equal to 5 is used: only DIAG and GSL-CP still guarantee an accurate estimation of the three expected emission-mode factors. But GSL-CP is the only one to not estimate additional emission-mode factors, showing its superiority over the three other methods.

V. Conclusion

In this paper, we provided a novel CP decomposition method called GSL-CP, which estimates both the CP model rank and the loading matrices of the tensor to be decomposed. A less computationally prohibitive norm than the nuclear norm is preferred, promoting group sparsity of the loading matrices under orthonormal subspace. The GSL-CP method can then solve the low rank CP decomposition problem without any knowledge of the expected rank. Numerical experiments verify the superiority of GSL-CP over classical CP decomposition methods, especially when the overfactoring problem occurs.

ISBN 978-0-9928626-7-1 © EURASIP 2017

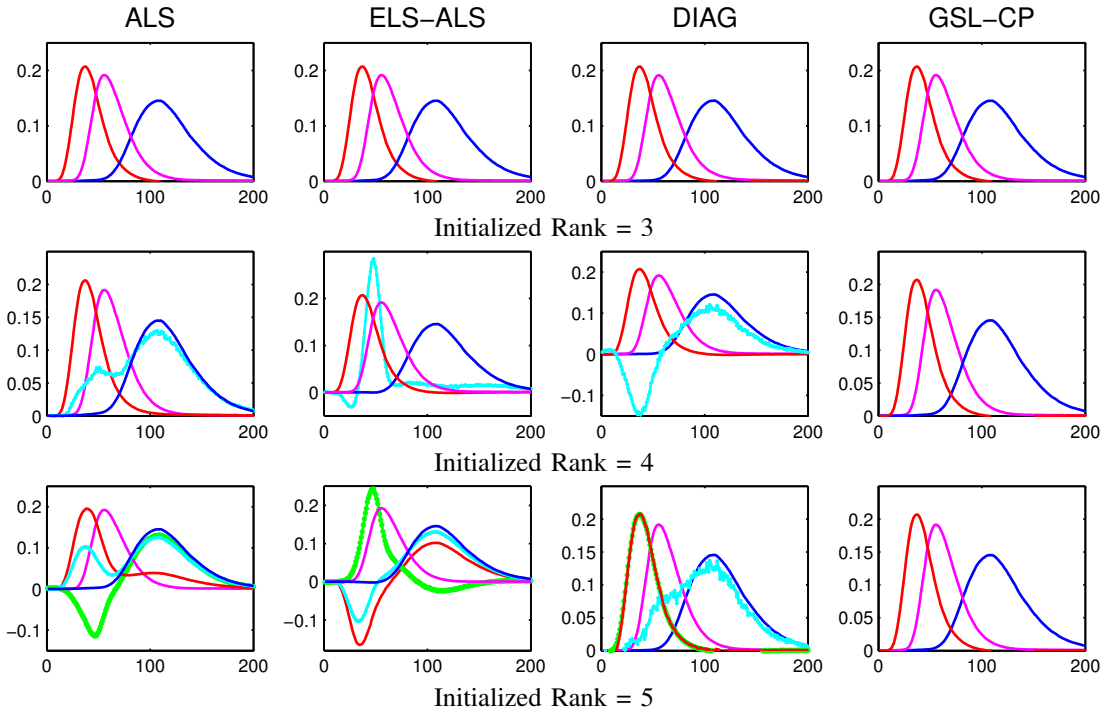

Fig. 2: Estimation of emission factors for different initialized ranks.

REFERENCES

[1] R. Bro, "Parafac. tutorial and applications," *Chemometrics and intelligent laboratory systems*, vol. 38, no. 2, pp. 149–171, 1997.

[2] A. L. de Almeida, G. Favier, and J. C. M. Mota, "Parafac-based unified tensor modeling for wireless communication systems with application to blind multiuser equalization," *Signal Processing*, vol. 87, no. 2, pp. 337–351, 2007.

[3] X. T. Vu, C. Chaux, S. Maire, and N. Thirion-Moreau, "Study of different strategies for the canonical polyadic decomposition of nonnegative third order tensors with application to the separation of spectra in 3d fluorescence spectroscopy," in *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2014.

[4] A. Fargeas, L. Albera, A. Kachenoura, G. Dréan, J.-D. Ospina, J. Cologner, C. Lafond, J.-B. Delobel, R. De Crevoisier, and O. Acosta, "On feature extraction and classification in prostate cancer radiotherapy using tensor decompositions," *Medical engineering & physics*, vol. 37, no. 1, pp. 126–131, 2015.

[5] H. Becker, L. Albera, P. Comon, M. Haardt, G. Birot, F. Wendling, M. Gavaret, C.-G. Bénar, and I. Merlet, "EEG extended source localization: tensor-based vs. conventional methods," *NeuroImage*, vol. 96, pp. 143–157, 2014.

[6] H. Becker, L. Albera, P. Comon, R. Gribonval, F. Wendling, and I. Merlet, "Brain-Source imaging: From sparse to tensor models," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 100–112, 2015.

[7] R. A. Harshman, "Foundations of the parafac procedure: Models and conditions for an" explanatory" multi-modal factor analysis," *UCLA working papers in Phonetics*, no. 16, pp. 1–84, 1970.

[8] 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, 1970.

[9] P. Comon, X. Luciani, and A. L. De Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of chemometrics*, vol. 23, no. 7-8, pp. 393–405, 2009.

[10] M. Rajih, P. Comon, and R. A. Harshman, "Enhanced line search: A novel method to accelerate parafac," *SIAM journal on matrix analysis and applications*, vol. 30, no. 3, pp. 1128–1147, 2008.

[11] Y. Chen, D. Han, and L. Qi, "New ALS methods with extrapolating search directions and optimal step size for complex-valued tensor decompositions," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5888–5898, 2011.

[12] X. Luciani and L. Albera, "Canonical polyadic decomposition based on joint eigenvalue decomposition," *Chemometrics and Intelligent Laboratory Systems*, vol. 132, pp. 152–167, 2014.

[13] X. Luciani and L. Albera, "Joint eigenvalue decomposition of non-defective matrices based on the LU factorization with application to ICA," *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4594–4608, 2015.

[14] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.

[15] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[16] X. Shu, F. Porikli, and N. Ahuja, "Robust orthonormal subspace learning: Efficient recovery of corrupted low-rank matrices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3874–3881.

[17] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, PhD thesis, Stanford University, 2002.

[18] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *Technical report, UILU-ENG-09-2215*, 2009.

[19] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

ISBN 978-0-9928626-7-1 © EURASIP 2017