



HAL
open science

A Proximal Approach for Nonnegative Tensor Decomposition

Xuan Vu, Caroline Chaux, Nadège Thirion-Moreau, Sylvain Maire

► **To cite this version:**

Xuan Vu, Caroline Chaux, Nadège Thirion-Moreau, Sylvain Maire. A Proximal Approach for Nonnegative Tensor Decomposition. 13th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2017), Feb 2017, Grenoble, France. hal-01485331

HAL Id: hal-01485331

<https://hal.science/hal-01485331>

Submitted on 8 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A proximal approach for nonnegative tensor decomposition

Xuan Vu^{1,2}, Caroline Chaux^{2*}, Nadège Thirion-Moreau¹, and Sylvain Maire¹

¹ Aix-Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, F-13397 Marseille
Université de Toulon, CNRS, LSIS, UMR 7296, F-83957 La Garde, France
{thirion, maire}@univ-tln.fr thi-thanh-xuan.vu@lsis.org

² Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France
caroline.chaux@univ-amu.fr

Abstract. This communication deals with N -th order tensor decompositions. More precisely, we are interested in the (Canonical) Polyadic Decomposition. In our case, this problem is formulated under a variational approach where the considered criterion to be minimized is composed of several terms: one accounting for the fidelity to data and others that can represent not only regularization (such as sparsity prior) but also hard constraints (such as nonnegativity). The resulting optimization problem is solved by using the Block-Coordinate Variable Metric Forward-Backward (BC-VMFB) algorithm. The robustness and efficiency of the suggested approach is illustrated on realistic synthetic data such as those encountered in the context of environmental data analysis and fluorescence spectroscopy. Our simulations are performed on 4-th order tensors.

Keywords: Constrained optimization - Proximal algorithm - Block alternating minimization - Nonnegative tensor factorization (NTF)

1 Introduction

In numerous applications, the data sets that are collected can be organized into multi-way (or N -way with $N \geq 3$) arrays of numerical values. Consequently, a growing interest has been dedicated to the development of efficient methods and derived algorithms, capable of both processing such multi-way arrays and extracting as much relevant information as possible. The most famous tensor decomposition certainly remains the (Canonical) Polyadic Decomposition (CPD) since it has been proven effective in many application fields (see [13] and [6] for an overview of applications). Another of its main advantages is its uniqueness under mild conditions [14][20]. In some leading applications of CPD particularly those linked to the image processing field (examples include 3D fluorescence spectroscopy and functional magnetic resonance imaging (fMRI) for brain mapping), some specific properties are generally known about the latent variables due to their physical meaning. Standing for concentrations, percentage, fractional abundance, spectra, and so on, these latent variables should be nonnegative and/or

* Caroline Chaux is the corresponding author.

smooth and/or sparse quantities and imposing these physical constraints can “help” the algorithms to recover more “relevant” pure constituent compounds. As a consequence, our main aim, here, is to properly tackle the relatively general problem of the CPD of N -way tensors subject to a certain number of constraints linked to *a priori knowledge* we may have about the involved latent variables. To that purpose, this general problem is formulated under the general framework of variational approaches where the cost function to be minimized is composed of several terms: the classical one accounting for the fidelity to data and additional ones that can either stand for regularization (such as sparsity prior) or represent hard constraints such as nonnegativity. The resulting optimization problem can become numerically difficult; the adopted algorithm is based upon a “Block Coordinate Variable Metric Forward-Backward” (BC-VMFB) approach [5] that gathers four main stages: 1) a gradient step involved in the forward stage, 2) a proximal step involved in the backward stage, 3) a preconditioning step (“variable metric”) and finally 4) a block arrangement (“Block Coordinate”) of the unknown (latent) variables that will be swept according to a random (or cyclic or other [21]) rule. Such an approach but without preconditioning has been used to deal with third order tensor decompositions (CPD and Tucker) [23]. Alternatively, an alternating optimization approach based on an alternating direction method of multipliers has been recently proposed in [12]. Finally, this algorithm recently proved its effectiveness in third order tensor decomposition for 3D fluorescence spectroscopy [22].

The remaining of this communication is organized as follows. Section 2 is devoted to the presentation of the considered multilinear model and the objective to be reached. Section 3 describes the proposed approach which consists of two steps: after formulating the problem under a variational approach and introducing the resulting criterion to be minimized, the proximal algorithm based on the Block coordinate Variable Metric Forward-Backward algorithm is presented. The efficiency of the proposed approach is emphasized through numerical experiments conducted in Section 4. A complicated ill-posed scenario (noisy overestimated model) is considered for the decomposition of a synthetic, yet, realistic 4-th order tensor. Finally, a conclusion is drawn and perspectives are delineated.

2 Canonical Polyadic Decomposition of N -th order tensors

2.1 Model

The Canonical Polyadic Decomposition (CPD) of tensors, also known as Parafac (PARAllel FACtor analysis [10]), CanDecomp (Canonical Decomposition [2]) and CP (for CanDecomp/Parafac [8]), constitutes a compact and informative model. It consists of decomposing an original tensor $\bar{\mathcal{T}}$ into a minimal sum of rank-1 terms:

$$\bar{\mathcal{T}} = \sum_{r=1}^{\bar{R}} \bar{\mathbf{a}}_r^{(1)} \circ \bar{\mathbf{a}}_r^{(2)} \circ \dots \circ \bar{\mathbf{a}}_r^{(N)} = \llbracket \bar{\mathbf{A}}^{(1)}, \bar{\mathbf{A}}^{(2)}, \dots, \bar{\mathbf{A}}^{(N)} \rrbracket, \quad (1)$$

where $N \in \mathbb{N}$ is the tensor order and \circ is the outer product of vectors. The minimal $\bar{R} \in \mathbb{N}$ such that Eq. (1) holds is called the tensor rank. For every $n \in \{1, \dots, N\}$, $r \in \{1, \dots, \bar{R}\}$, the real column vector $\bar{\mathbf{a}}_r^{(n)} = (\bar{a}_{1r}^{(n)}, \bar{a}_{2r}^{(n)}, \dots, \bar{a}_{I_n r}^{(n)})^\top \in \mathbb{R}^{I_n}$ is called a loading factor (where $(\cdot)^\top$ stands for the transpose operator) and the unknown latent matrices $\bar{\mathbf{A}}^{(n)} = [\bar{\mathbf{a}}_1^{(n)}, \bar{\mathbf{a}}_2^{(n)}, \dots, \bar{\mathbf{a}}_{\bar{R}}^{(n)}] = (\bar{a}_{i_n r}^{(n)})_{i_n, r} \in \mathbb{R}^{I_n \times \bar{R}}$ are called the loading matrices.

2.2 Objective

Given a tensor \mathcal{T} (which can be an observation, possibly noisy, of an original tensor $\bar{\mathcal{T}}$), we aim at approximating it using the CP model *i.e.* we intend to determine for all $n \in \{1, \dots, N\}$ an estimation of the loading matrices $\bar{\mathbf{A}}^{(n)}$. To estimate the loading matrices $\bar{\mathbf{A}}^{(n)}$ for all $n \in \{1, \dots, N\}$, it can be more convenient to rewrite Eq. (1) under a matrix form by using flattening. Indeed, let $\bar{\mathbf{T}}_{I_n, I_{-n}}^{(n)} \in \mathbb{R}^{I_n \times I_{-n}}$ be the matrix obtained by unfolding tensor $\bar{\mathcal{T}}$ along mode n , where $n \in \{1, \dots, N\}$ and $I_{-n} = I_1 \dots I_{n-1} I_{n+1} \dots I_N$, then the model given in Eq. (1) can be written in a compact matrix form [6, p. 352] as follows

$$\bar{\mathbf{T}}_{I_n, I_{-n}}^{(n)} = \bar{\mathbf{A}}^{(n)} (\bar{\mathbf{Z}}^{(-n)})^\top, \quad \forall n \in \{1, \dots, N\} \quad (2)$$

where

$$\bar{\mathbf{Z}}^{(-n)} = \bar{\mathbf{A}}^{(N)} \odot \dots \odot \bar{\mathbf{A}}^{(n+1)} \odot \bar{\mathbf{A}}^{(n-1)} \odot \dots \odot \bar{\mathbf{A}}^{(1)} \in \mathbb{R}^{I_{-n} \times \bar{R}}, \quad (3)$$

and \odot denotes the Khatri-Rao product.

3 Optimization problem and proximal algorithm

We choose, here, to express the problem of estimating the loading matrices under a variational framework *i.e.* to solve an optimization problem whose solution constitutes an estimation of the initial loading matrices.

3.1 Criterion formulation, assumptions and properties

In classical variational approaches, the criterion is divided into two main terms: a data fidelity term denoted by \mathcal{F} and a regularization term which is here constituted of the sum of N regularization functions, each linked to one of the loading matrices.

Mathematically, this problem is formulated as

$$\underset{\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times \bar{R}}, n \in \{1, \dots, N\}}{\text{minimize}} \quad \mathcal{F}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) + \sum_{n=1}^N \mathcal{R}_n(\mathbf{A}^{(n)}) \quad (4)$$

where \mathcal{F} and $(\mathcal{R}_n)_{n \in \{1, \dots, N\}}$ are assumed to be proper lower semi-continuous functions such that \mathcal{F} is differentiable with a β -Lipschitz gradient where $\beta \in]0, +\infty[$ and such that for all $n = 1, \dots, N$, $\mathcal{R}_n : \mathbb{R}^{I_n \times \bar{R}} \rightarrow]-\infty, +\infty[$ is bounded from below by an affine function, and its restriction to its domain is continuous. The numerical method used to solve Eq. (4) is described in the next section.

3.2 Proposed algorithm

Here, we suggest to use the Block Coordinate Variable Metric Forward Backward (BC-VMFB) algorithm [5, 4] to solve the problem described by Eq. (4). The general principle of the resulting iterative method is detailed in the following paragraph and is summed up in Algorithm 1. Our different choices (of cost function, preconditioning matrix, etc.) to tackle specifically the CPD problem are discussed in section 3.3. The approach chosen here mainly consists of two steps:

- ❶ a gradient step (linked to \mathcal{F} which is assumed to be differentiable with a β -Lipschitz gradient). It requires to compute the partial gradient matrices of \mathcal{F} with respect to $\mathbf{A}^{(n)}$ for all $n = 1, \dots, N$. In the following, they are denoted by $\nabla_n \mathcal{F}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)})$.
- ❷ a proximal step (linked to $(\mathcal{R}_n)_{n \in \{1, \dots, N\}}$): for all $n = 1, \dots, N$ it requires to compute the proximity operator of \mathcal{R}_n associated to the metric $\mathbf{P}^{(n)}$. The definition of the proximity operator is recalled hereafter.

The proximity operator of a proper, lower semicontinuous function from \mathbb{R}^I to $] -\infty, +\infty[$ associated with a Symmetric Positive Definite (SPD) matrix \mathbf{P} is defined as [11]

$$\text{prox}_{\mathbf{P}, \varphi}: \mathbb{R}^I \rightarrow \mathbb{R}^I: v \mapsto \arg \min_{u \in \mathbb{R}^I} \frac{1}{2} \|u - v\|_{\mathbf{P}}^2 + \varphi(u). \quad (5)$$

where $\forall x \in \mathbb{R}^I$, $\|x\|_{\mathbf{P}}^2 = \langle x, \mathbf{P}x \rangle$, and $\langle \cdot, \cdot \rangle$ stands for the inner product. The original definition of the proximity operator [17] is recovered when \mathbf{P} reduces to the identity matrix.

To simplify the notations, the partial gradient matrices $\nabla_n \mathcal{F}(\mathbf{A}^{(1)}[k], \dots, \mathbf{A}^{(N)}[k])$ associated to k -th iteration are simply denoted by $\nabla_n[k]$. The Hadamard division between two matrices is denoted by \oslash . Finally, we recall that under some technical assumptions [5, Sec. 2.2] (concerning the preconditioning matrices $\mathbf{P}^{(n)}$, the step-size γ , the block scanning rule, and the fact that $\mathcal{F} + \mathcal{R}$ satisfies the Kurdyka-Łojasiewicz inequality) the convergence of the algorithm to a critical point is guaranteed [5, Theorem 3.1].

3.3 Criterion choice: related gradient and proximity operators

The algorithm described hereabove was presented in a very general way. We now introduce the objective that we have chosen to minimize, explain some of our choices and provide the resulting involved quantities (partial gradient matrices, preconditioning matrix, and so on).

For the computer simulations provided in this communication, the data fidelity term \mathcal{F} takes a quadratic form. It thus leads to the following definition

$$\mathcal{F}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) = \frac{1}{2} \|\mathcal{T} - \llbracket \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket\|_F^2 = \frac{1}{2} \|\mathbf{T}_{I_n, I_{-n}}^{(n)} - \mathbf{A}^{(n)} \mathbf{Z}^{(-n)\top}\|_F^2, \quad (6)$$

where $\|\cdot\|_F$ stands for the Frobenius norm. As a consequence

Algorithm 1 BC-VMFB algorithm to minimize Eq. (4).

```

1: Let  $\mathbf{A}^{(n)} \in \text{dom } \mathcal{R}_n$ ,  $n \in \{1, \dots, N\}$ ,  $k \in \mathbb{N}$  and  $\gamma[k] \in ]0, +\infty[$  // Initialization
2: for  $k = 0, 1, \dots, K$  do //  $k$ -th iteration
3:   Choose a block  $n \in \{1, \dots, N\}$  // Quasi cyclic rule
4:   Compute  $\mathbf{P}^{(n)}[k] = \mathbf{P}^{(n)}(\mathbf{A}^{(1)}[k], \dots, \mathbf{A}^{(N)}[k])$  // Preconditioner construction
5:   Compute the Gradient Matrix  $\nabla_n[k]$  // Calculation of Gradient
6:    $\mathbf{A}^{(n)}[k + \frac{1}{2}] = \mathbf{A}^{(n)}[k] - \gamma[k]\nabla_n[k] \oslash \mathbf{P}^{(n)}[k]$  // Gradient step
7:    $\mathbf{A}^{(n)}[k + 1] \in \text{prox}_{\gamma[k]-1\tilde{\mathbf{P}}^{(n)}[k], \mathcal{R}_n}(\mathbf{A}^{(n)}[k + \frac{1}{2}])$  // Proximal step
8:    $\mathbf{A}^{(\bar{n})}[k + 1] = \mathbf{A}^{(\bar{n})}[k]$  where  $\bar{n} = \{1, \dots, N\} \setminus \{n\}$  // Other blocks kept unchanged
9: end for
10:  $\forall n \in \{1, \dots, N\}$ ,  $\hat{\mathbf{A}}^{(n)} = \mathbf{A}^{(n)}[K]$  // Convergence reached at  $K$ -th iteration

```

❶ the associated partial gradient matrices are given by [9]

$$\nabla_n \mathcal{F}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) = -(\mathbf{T}_{I_n, I_{-n}}^{(n)} - \mathbf{A}^{(n)} \mathbf{Z}^{(-n)\top}) \mathbf{Z}^{(-n)}. \quad (7)$$

❷ in the same spirit as in [16, 19], the preconditioning matrix $\mathbf{P}^{(n)}$, $\forall n \in \{1, \dots, N\}$, can be defined as follows

$$\mathbf{P}^{(n)}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) = \left(\mathbf{A}^{(n)} (\mathbf{Z}^{(-n)\top} \mathbf{Z}^{(-n)}) \right) \oslash \mathbf{A}^{(n)}. \quad (8)$$

It is based on the n -th mode unfolding of the tensor (see (2)) and on the definition of a majorant function of the restriction of \mathcal{F} to the n -th loading matrix on the domain of \mathcal{R}_n . Additional details about preconditioning matrix construction can be found in [22].

Concerning the regularization terms, they may account at the same time for the nonnegativity constraint we want to impose on the solution and to the sparsity of the data (possible overfactoring). For all $\mathbf{A}^{(n)} = (a_{i_n r}^{(n)})_{(i_n, r) \in \{1, \dots, I_n\} \times \{1, \dots, R\}}$, we thus choose [5, p. 18-20],

$$\mathcal{R}_n(\mathbf{A}^{(n)}) = \sum_{i_n=1}^{I_n} \sum_{r=1}^R \rho_n(a_{i_n r}^{(n)}) \quad (9)$$

where $\forall n \in \{1, \dots, N\}$

$$\rho_n(\omega) = \begin{cases} \alpha^{(n)} |\omega|^{\pi^{(n)}} & \text{if } \eta_{\min}^{(n)} \leq \omega \leq \eta_{\max}^{(n)} \\ +\infty & \text{otherwise} \end{cases} \quad (10)$$

and $\alpha^{(n)} \in]0, +\infty[$, $\pi^{(n)} \in \mathbb{N}^*$, $\eta_{\min}^{(n)} \in [-\infty, +\infty[$, and $\eta_{\max}^{(n)} \in [\eta_{\min}^{(n)}, +\infty[$ (block dependent regularization parameters³). This choice enables to ensure nonnegativity by taking for example $\eta_{\min}^{(n)} = 10^{-10}$ and $\eta_{\max}^{(n)} = +\infty$ and to promote

³ In our case, the easiest way to proceed is to consider that each block matches a loading matrix, but other choices could have been made.

sparsity by choosing the exponent $\pi^{(n)} = 1$ (hence performing an ℓ_1 -norm regularization).

To be properly computed, the associated proximity operator requires first to define⁴ i) $\mathbf{a}^{(n)} = \text{vec}(\mathbf{A}^{(n)}) \in \mathbb{R}^{RI_n}$ (vectorization of loading matrices and operator $\text{vec}(\cdot)$ stacks the columns of the matrix given in argument into a vector) and ii) $\tilde{\mathbf{P}}^{(n)} = \text{Diag}(\text{vec}(\mathbf{P}^{(n)})) \in \mathbb{R}^{RI_n \times RI_n}$ (vectorization and diagonalization of preconditioning matrices, the $\text{Diag}(\cdot)$ operator builds a diagonal matrix whose diagonal elements are the elements of the vector passed as a parameter).

By using definition in Eq. (5), we can derive the expression of $\text{prox}_{\gamma^{-1}\tilde{\mathbf{P}}^{(n)}, \mathcal{R}_n}(\mathbf{a}^{(n)})$ as

$$\begin{aligned} & \left(\forall \mathbf{A}^{(n)} = (a_{i_n r}^{(n)})_{(i_n, r) \in \{1, \dots, I_n\} \times \{1, \dots, R\}} \right) \\ & \text{prox}_{\gamma^{-1}\tilde{\mathbf{P}}^{(n)}, \mathcal{R}_n}(\mathbf{A}^{(n)}) = \left(\text{prox}_{\gamma^{-1}p_{i_n r}^{(n)}, \rho_n}(a_{i_n r}^{(n)}) \right)_{(i_n, r) \in \{1, \dots, I_n\} \times \{1, \dots, R\}} \end{aligned} \quad (11)$$

where $\forall (i_n, r) \in \{1, \dots, I_n\} \times \{1, \dots, R\}$, we have [7] ($\forall v \in \mathbb{R}$)

$$\text{prox}_{\gamma^{-1}p_{i_n r}^{(n)}, \rho_n}(v) = \min \left\{ \eta_{\max}^{(n)}, \max \left\{ \eta_{\min}^{(n)}, \text{prox}_{\gamma \alpha^{(n)}(p_{i_n r}^{(n)})^{-1} \cdot |\pi^{(n)}(v)|} \right\} \right\}. \quad (12)$$

A closed form expression of the proximity operator presented in Eq. (12) can be found in [3]. Note that in Algorithm 1, at iteration k , the proximity operator is associated with metric $\mathbf{P}^{(n)}[k]$ and is computed at $\mathbf{A}^{(n)}[k + \frac{1}{2}]$ with stepsize $\gamma[k]$.

4 Numerical simulations: application to 4-th order CPD

We consider here a tensor of order $N = 4$. It has been constructed synthetically but following realistic guidelines. Inspired by 3D fluorescence spectroscopy, we build this tensor simulating: (uni or bimodal type) emission and excitation spectra, smooth (either linear or unimodal) concentrations (the 3 classical components of 3D fluorescence spectroscopy) and an additional 4-th dimension modelling the lifetime (exponential decay) of compounds (such as those observed when time resolved spectroscopy is performed [15]).

The tensor rank has been fixed here to $\bar{R} = 5$. The resulting tensor $\bar{\mathcal{T}}$ is of size $I_1 = I_2 = I_3 = I_4 = 100$ and $\bar{\mathcal{T}} \in \mathbb{R}_+^{100 \times 100 \times 100 \times 100}$. Original spectra, concentrations, lifetimes are displayed in Fig. 1 (black curves).

The following scenario has been considered: the observed tensor \mathcal{T} is assumed to be a perturbed version of the original tensor, that is $\mathcal{T} = \bar{\mathcal{T}} + \mathcal{B}$ where \mathcal{B} stands for an additive white Gaussian noise with $\sigma = 0.001$ resulting in an initial SNR of 18.46dB. Furthermore, the tensor rank \bar{R} is assumed to be unknown and the decomposition is performed assuming a tensor rank $\hat{R} = 7$ (which corresponds to an overestimation of a factor 2).

⁴ In practice, elementwise operations are performed instead making it possible to avoid memory issues.

We compare our algorithm (with $\pi^{(n)} \equiv \pi = 1$, $\alpha^{(n)} \equiv \alpha = 0.05$, $\eta_{\min}^{(n)} \equiv \eta_{\min} = 10^{-16}$ and $\eta_{\max}^{(n)} \equiv \eta_{\max} = 10^2$) performances to two state-of-the-art methods: 1) fast HALS algorithm [18] and 2) Bro's N -way algorithm [1] for which, to be fair, we used the non-negativity constrained versions). Algorithm initialization is random.

In addition to visual results, we compute three error measures. Let $\widehat{\mathbf{A}}^{(n)} = [\widehat{\mathbf{a}}_1^{(n)}, \widehat{\mathbf{a}}_2^{(n)}, \dots, \widehat{\mathbf{a}}_{\widehat{R}}^{(n)}]$ denote the normalized permuted estimate of $\bar{\mathbf{A}}^{(n)} = [\bar{\mathbf{a}}_1^{(n)}, \bar{\mathbf{a}}_2^{(n)}, \dots, \bar{\mathbf{a}}_{\bar{R}}^{(n)}]$. The considered error measures are given by

1. Signal to Noise Ratio (SNR) defined as $\text{SNR} = 20 \log_{10} \frac{\|\bar{\mathcal{T}}\|_F}{\|\bar{\mathcal{T}} - \widehat{\mathcal{T}}\|_F}$
2. Estimation error

$$\mathbf{E}_1 = 10 \log_{10} \left(\frac{\sum_{n=1}^4 \|\widehat{\mathbf{A}}^{(n)}(1 : \bar{R}) - \bar{\mathbf{A}}^{(n)}\|_1}{\sum_{n=1}^4 \|\bar{\mathbf{A}}^{(n)}\|_1} \right) \quad (13)$$

3. Over-factoring error \mathbf{E}_2 :

$$\mathbf{E}_2 = 10 \log_{10} \left(\left\| \sum_{r=\bar{R}+1}^{\widehat{R}} \widehat{\mathbf{a}}_r^{(1)} \circ \widehat{\mathbf{a}}_r^{(2)} \circ \widehat{\mathbf{a}}_r^{(3)} \circ \widehat{\mathbf{a}}_r^{(4)} \right\|_1 \right) \quad (14)$$

All the considered approaches being iterative, the following stopping conditions were used: either the maximum number of iterations fixed to $K = 10^5$ has been reached or the relative diminishing rate of the quadratic criterion reads $\frac{\|\mathcal{F}[\cdot+1] - \mathcal{F}[\cdot]\|}{\mathcal{F}[\cdot]} < 10^{-8}$.

The estimated spectra are displayed in Fig. 1. We can see that despite the over-estimation factor and the noise, the proposed algorithm, contrary to fHALS or N -way approaches, allows to accurately recover the original data without creating phantoms in the artificially added compounds. This is confirmed by the numerical results given in Tab. 1 where we can see that the estimation error is equal or higher for fHALS and N -way methods and that the over-factoring error is much more smaller for the proposed method. Concerning the algorithm computation times, we can see that the proposed approach is very competitive but requires more iterations to reach the stopping criterion.

	fHALS	N -way	BC-VMFB
Time (in s) for 50 iterations	15.4	251	9.7
Time (in s) to reach stopping criterion	139	554	1689
Iteration number to reach stopping criterion	506	176	17000
Associated (SNR, \mathbf{E}_1 , \mathbf{E}_2) in dB	(29.61, 0.21, 47.74)	(29.61, 0.11, 46.61)	(34.15, 0.11, -543)

Table 1. Computation times and numerical performances of fHALS, N -way with non negativity constraints and BC-VMFB algorithms. Simulations were performed on a 8 cores Intel i7 @3.40GHz.

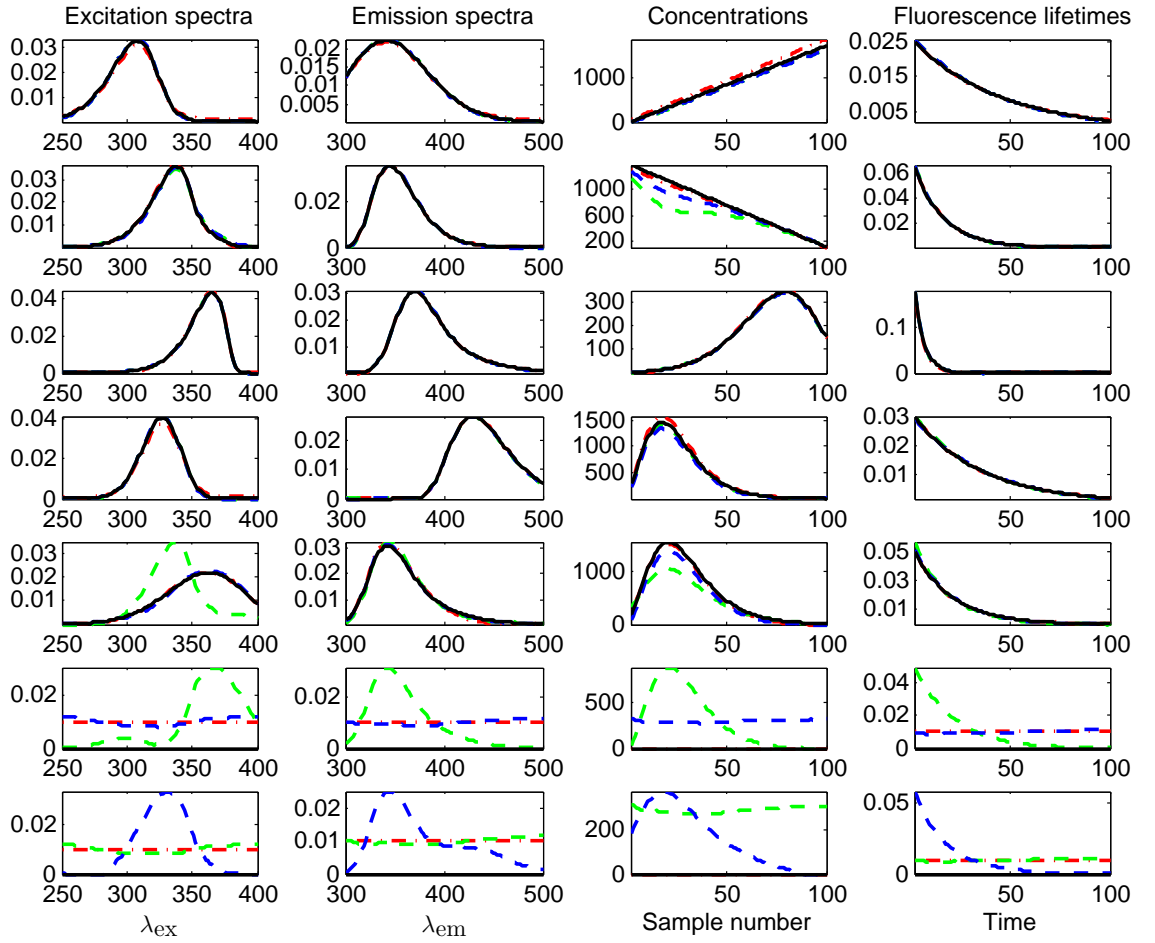


Fig. 1. Estimated scaled spectra using N -way (blue) with non negativity constraints, fHALS (green) and BC-VMFB (red). Black curves: ground truth.

5 Conclusion

In this communication, we addressed the problem of the CP decomposition of N -th order tensors subject to given constraints such as nonnegativity, sparsity, regularity, etc. We tackled this problem within the very general framework of Block Coordinate Variable Metric Forward-Backward (BC-VMFB) approaches. An algorithm was provided and its robustness and efficiency was demonstrated on synthetic yet realistic 4-th order data inspired by those encountered in the context of environmental data analysis and more precisely fluorescence spectroscopy. The obtained results are encouraging, and future developments could be to apply this algorithm on raw data sets, to test other cost functions, other preconditioning matrices, etc. It could be also interesting to test other kind of regularization functions and to better understand their impact on the performance of the algorithm.

References

1. Bro, R.: Parafac: tutorial and applications. *Chemom. Intell. Lab. Syst.* 38(2), 149–171 (Oct 1997)
2. Carroll, P., Chang, J.J.: Analysis of individual differences in multi-dimensional scaling via n -way generalization of Eckart-Young decomposition. *Psychometrika* 35(3), 283–319 (Jan 1970)
3. Chaux, C., Combettes, P.L., Pesquet, J.C., Wajs, V.R.: A variational formulation for frame based inverse problems. *Inverse Probl.* 23(4), 1495–1518 (Aug 2007)
4. Chouzenoux, E., Pesquet, J.C., Repetti, A.: Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. *Journal of Optimization Theory and Applications* 162(1), 107–132 (Jul 2014)
5. Chouzenoux, E., Pesquet, J.C., Repetti, A.: A block coordinate variable metric forward-backward algorithm. *J. Global Optim.* pp. 1–29 (2016)
6. Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.I.: Non negative matrix and tensor factorizations: Application to exploratory multi-way data analysis and blind separation. Wiley (2009)
7. Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: Bauschke, H.H., Burachik, R., Combettes, P.L., Elser, V., Luke, D.R., Wolkowicz, H. (eds.) *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer Verlag (2010)
8. Comon, P., Jutten, C.: *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. ISBN: 978-0-12-374726-6, Academic Press, Oxford UK, Burlington USA (2010)
9. Franc, A.: *Etude algébrique des multi-tableaux : apport de l’algèbre tensorielle*. Phd thesis, University of Montpellier II, Montpellier, France (1992)
10. Harshman, R.A.: Foundation of the Parafac procedure: models and conditions for an explanatory multimodal factor analysis. *UCLA Working papers in phonetics* 16, 1–84 (1970)
11. Hiriart-Urruty, J.B., Lemarechal, C.: *Convex Analysis and Minimization Algorithms*. Springer-Verlag Berlin Heidelberg (1993)
12. Huang, K., Sidiropoulos, N.D., Liavas, A.P.: A flexible and efficient algorithmic framework for constrained matrix and tensor factorization. *IEEE Transactions on Signal Processing* 64(19), 5052–5065 (Oct 2016)

13. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. *SIAM Rev.* 51(3), 455–500 (Sep 2009)
14. Kruskal, J.B.: Rank, decomposition and uniqueness for 3-way and n-way arrays, pp. 7–18. North-Holland Publishing Co. (Apr 1989)
15. Lakowicz, J.R., Szmajda, H., K. Nowaczyk, K., Berndt, K.W., Johnson, M.: Fluorescence lifetime imaging. *Analytical Biochemistry* 202(2), 316–330 (1992)
16. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: Leen, T.K., Dietterich, T.G., Tresp, V. (eds.) *Advances in Neural Information Processing Systems 13*, pp. 556–562. MIT Press (2001), <http://papers.nips.cc/paper/1861-algorithms-for-non-negative-matrix-factorization.pdf>
17. Moreau, J.J.: Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France* 93, 273–299 (1965)
18. Phan, A.H., Tichavský, P., Cichocki, A.: Fast alternating ls algorithms for high order CANDECOMP/PARAFAC tensor factorizations. *IEEE Trans. on Sig. Proc.* 61(19), 4834–4846 (Jun 2013)
19. Repetti, A., Chouzenoux, E., Pesquet, J.C.: A preconditioned Forward-Backward approach with application to large-scale nonconvex spectral unmixing problems. In: 39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014). Florence, Italie (May 2014)
20. Sidiropoulos, N., Bro, R.: On the uniqueness of multilinear decomposition of N -way arrays. *J. Chemom.* 14(3), 229–239 (May/June 2000)
21. Vervliet, N., Lathauwer, L.D.: A randomized block sampling approach to canonical polyadic decomposition of large-scale tensors. *IEEE Journal on Selected Topics in Signal Processing* 10(2), 284–295 (2016)
22. Vu, X.T., Chau, C., Thirion-Moreau, N., Maire, S.: A new penalized nonnegative third order tensor decomposition using a block coordinate proximal gradient approach: application to 3D fluorescence spectroscopy. *J. Chemometrics* (2017)
23. Xu, Y., Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.* 6(3), 1758–1789 (2013)