



**HAL**  
open science

## Correcting inner filter effects, a non multilinear tensor decomposition method

Jérémy E Cohen, Pierre Comon, Xavier Luciani

► **To cite this version:**

Jérémy E Cohen, Pierre Comon, Xavier Luciani. Correcting inner filter effects, a non multilinear tensor decomposition method. *Chemometrics and Intelligent Laboratory Systems*, 2016, 150, pp.29-40. 10.1016/j.chemolab.2015.11.002 . hal-01158152

**HAL Id: hal-01158152**

**<https://hal.science/hal-01158152>**

Submitted on 29 May 2015

**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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Correcting inner filter effects, a non multilinear tensor decomposition method

Jeremy Emile Cohen<sup>a,\*</sup>, Pierre Comon<sup>a</sup>, Xavier Luciani<sup>b,c</sup>

<sup>a</sup>*Department of Image and Signal-processing, Gipsa-lab, CNRS, Grenoble*

<sup>b</sup>*Aix Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, 13397 Marseille, France.*

<sup>c</sup>*Université de Toulon, CNRS, LSIS, UMR 7296, 83957 La Garde, France., France.*

---

## Abstract

Among measurement used in analytical chemistry, fluorescence spectroscopy is widely spread and its applications are numerous. To recover various information on unknown components in chemical mixtures, multilinear tensor decomposition of multiway fluorescence spectra have proven extremely powerful. However, inner filter effects induce a systematic error on measurements, disturbing the decomposition. In this paper, we fully describe a non multilinear approach to include inner filter effects in the model instead of neglecting them or correcting them by linearization methods. A theoretical framework on non multilinear tensor decomposition is developed, an algorithm to recover the factors in the decomposition is detailed, and real data computer results are reported.

*Keywords:* Fluorescence spectroscopy, Tensor decomposition, Inner filter effects, Identification

---

## Acknowledgement

This work has been funded by the European Research Council under the European Community's Seventh Framework Programme FP7/2007-2013 Grant Agreement no.320594, "Decoda" project.

## 1. Introduction

Fluorescence spectroscopy is a non invasive method for identifying components and determining their relative contribution in samples. Measurements consist of the response of the studied solution to a monochromatic source, repeated for multiple excitation wavelengths and recorded at multiple emission wavelengths [1]. The data is then stored in a Fluorescence Emission Excitation Matrix (FEEM). Because fluorescent chemical component (fluorofore) can be identified given its excitation and emission spectra, analyzing fluorescence data is typically a blind source separation problem, where the sources are the unknown fluorophores, and the observations are the mixing of their spectrum stored in the FEEM.

---

\*Corresponding author. Address:

*Email addresses:* jeremy.cohen@gipsa-lab.grenoble-inp.fr (Jeremy Emile Cohen), pierre.comon@gipsa-lab.grenoble-inp.fr (Pierre Comon), luciani@univ-tln.fr (Xavier Luciani)

*Preprint submitted to Elsevier*

*May 29, 2015*

During the last decade, the Canonical Polyadic decomposition (CPD) [2] of tensors (seen as multiway arrays), also known in the community as CANonical DECOMPosition (CANDECOMP) [3] or PARAllel FACtor analysis (PARAFAC) [4], has proven very efficient at solving this blind source separation problem [5]. The core idea behind tensor decomposition is that using only one FEEM and without any other *a priori* knowledge, it is theoretically and practically impossible to recover the components linearly because usually their contribution to the FEEM overlap. However, using multiple FEEM from different samples provides a third diversity: the fluorophore concentration profile through the sample set and hence a 3-way data tensor. This ensures that information can be obtained on each component individually. A sampling campaign is then an easy way to get a tensor whose decomposition reveals the different components in the mixture. This approach is now currently used in analytical chemistry [6] or environmental sciences, in particular for Dissolved Organic Matter (DOM) tracing and characterization purpose [7, 8]. However there is a fundamental problem using the CPD to separate sources in fluorescence spectroscopy. Indeed it is well known that the suitability of the linear fluorescence model for describing a FEEM decreases with the solution absorbency [1]. Hence, in many practical situations, the gradual absorption by the solution of both exciting and fluorescent lights cannot be neglected. These effects are known as Inner Filter Effects (IFE) [9, 10]. IFE affect both FEEM magnitude and patterns and are still noticeable at quite low concentrations, since the absorbency can still be very high. For instance, this is the case for protein at low excitation wavelengths.

Most IFE correction methods consist of deducing the linear contribution from the measured FEEM, which is then called the linearized FEEM. The non linear contribution directly depends on the solution absorbance spectrum. Therefore a first linearization may be achieved by strongly diluting the solution until reaching a maximal absorbance threshold [11]. However the procedure can be very tedious and can lead to contamination or physico-chemical changes, thus modifying the fluorescence properties of the sample. The most common alternative is to measure the absorbance spectrum of the solution and then deduce the non linear contribution and finally linearize the FEEM [12, 13]. However absorbance measurement is much less sensitive than fluorescence measurement and can lead to poor results [14]. In addition, it often requires another experimental device. In order to avoid these complications, more sophisticated correction methods that require neither a strong dilution of the solution nor the absorbance spectrum knowledge have been recently proposed in [14] and [15].

In other words, to perform the three-way decomposition of a set of FEEM suspected to be affected by IFE, FEEM are linearized independently, one after another. Then a tensor gathering the linearized FEEM is built and the CPD can be computed to identify individual spectra and concentration profiles. Thereby these approaches do not directly exploit the tensorial structure of the initial data set. In addition these methods require additional measurements for each solution in order to linearize the corresponding FEEM, which is time consuming and not always possible.

We show in this contribution that such a fluorescent tensor can be directly decomposed without any linearization step, and we give a complete analysis of all aspects of the resolution of the inverse problem. Following the non-linear model including inner filter effects [14], some mathematical aspects are explored, focusing on local identifiability in Section 2.3. The optimization problem and an efficient algorithm are then described in Section 3. Finally, in Sections 4 and 5 we study its assets on two real data sets.

## 2. A Non Linear Fluorescence Tensor Decomposition

### 2.1. Modeling inner filter effect

We consider here a set of  $K$  FEEM measured from  $K$  mixtures of  $R$  fluorophores.

Each fluorophore  $r$  can be characterized by the evolution of its concentration throughout the mixture set (its concentration profile), the evolution of its molar extinction coefficient wrt (the excitation) wavelength and the evolution of its light emission probability as a function of (the emission) wavelength. In the following these values are denoted by vector  $\mathbf{a}_r$  and functions  $\tilde{b}_r(\lambda^{ex})$  and  $\tilde{c}_r(\lambda^{em})$  respectively. Note that any function of the form  $\alpha\tilde{b}_r(\lambda^{ex})$  defines the excitation spectrum of  $r$  and any function of the form  $\beta\tilde{c}_r(\lambda^{em})$  defines its emission spectrum in arbitrary units. Let  $x_k(\lambda^{ex}, \lambda^{em})$  be the fluorescence intensity measured from a given mixture (sample)  $k$  at a given couple  $(\lambda^{ex}, \lambda^{em})$  of excitation and emission wavelengths. A classical continuous fluorescence model, taking into account IFE is then given by the following non linear relationship [14]:

$$x_k(\lambda^{ex}, \lambda^{em}) = \sum_{r=1}^R (a_r)_k \tilde{b}_r(\lambda^{ex}) \tilde{c}_r(\lambda^{em}) \nu_r \prod_{r=1}^R e^{-\mu[(a_r)_k(\tilde{b}_r(\lambda^{ex}) + \tilde{b}_r(\lambda^{em}))]} + e_k(\lambda^{ex}, \lambda^{em}), \quad (1)$$

where  $(a_r)_k$  denote the  $k^{th}$  entry of  $\mathbf{a}_r$ ,  $\nu_r$  and  $\mu$  are unknown values modeling non observable or experimental parameters (such as fluorophore quantum yields, optical path length...) and  $e$  is an error term. Indeed, we have to recall here that although it takes into account IFE, this model is still an approximation of the (noisy) fluorescence measurement [16, 14]. Since excitation and emission spectra are unnormalized, it is then interesting to define functions  $b_r(\lambda^{ex})$  and  $c_r(\lambda^{em})$  as

$$b_r(\lambda) = \mu \tilde{b}_r(\lambda), \quad (2)$$

$$c_r(\lambda^{em}) = \frac{\nu_r}{\mu} \tilde{c}_r(\lambda^{em}), \quad (3)$$

yielding a simpler model:

$$x_k(\lambda^{ex}, \lambda^{em}) = \sum_{r=1}^R (a_r)_k b_r(\lambda^{ex}) c_r(\lambda^{em}) \prod_{r=1}^R e^{-(a_r)_k(b_r(\lambda^{ex}) + c_r(\lambda^{em}))} + e_k(\lambda^{ex}, \lambda^{em}). \quad (4)$$

Now turning back to our set of FEEM, we respectively denote  $[\lambda_{ex}^{min}; \lambda_{ex}^{max}]$ ,  $\Delta_{ex}$ ,  $[\lambda_{em}^{min}; \lambda_{em}^{max}]$  and  $\Delta_{em}$  the excitation range, the excitation sampling step, the emission range and the emission sampling step used to measure each FEEM. Excitation and emission range sizes are denoted  $L$  and  $M$  respectively. We assume for the moment that excitation and emission sampling steps are equal and that in the wavelength range  $[\lambda_{em}^{min}; \lambda_{ex}^{max}]$  all excitation wavelengths and emission wavelengths coincide *i.e.*  $\Delta_{ex} = \Delta_{em}$  and  $\lambda_{em}^{min}$  coincides with a value of  $\lambda_{ex}$ . At this point this assumption is fundamental but we will see in the algorithm description how it can be relaxed.

In practice, for numerical computations, excitation and emission wavelength ranges are substituted by two ranges of integer index :  $[1; L]$  and  $[1; M]$  respectively so that if we call  $\mathcal{X}$  the fluorescence tensor of size  $(K \times L \times M)$  that gathers these  $K$  FEEM we have:

$$\mathcal{X}_{klm} = x_k(\lambda_{ex}^{min} + (l-1)\Delta_{ex}, \lambda_{em}^{min} + (m-1)\Delta_{em}). \quad (5)$$

In the same way, for each fluorophore  $r$  we can define discrete version of excitation and emission spectra truncated in the considered excitation and emission ranges as vectors of size  $L$  and  $M$  and denoted respectively  $\mathbf{b}_r$  and  $\mathbf{c}_r$ , whose entries are defined by:

$$(b_r)_l = b_r(\lambda_{ex}^{min} + (l-1)\Delta_{ex}), \quad (6)$$

$$(c_r)_m = c_r(\lambda_{em}^{min} + (m-1)\Delta_{em}). \quad (7)$$

We also assume that  $\lambda_{ex}^{min}$  and  $\lambda_{em}^{min}$  can be different and we define the wavelength index shift  $s$  as:

$$s = \frac{\lambda_{em}^{min} - \lambda_{ex}^{min}}{\Delta_{ex}} + 1. \quad (8)$$

Note that according to the previous hypothesis, it appears clearly that  $s$  is a strictly positive integer.

As a consequence, in order to develop a rigorous discrete version of the continuous non-linear fluorescence model (4) we need to define a "shifted" excitation  $\mathbf{b}'_r$  spectra of size  $M$  as:

$$(b'_r)_m = (b_r)_{m+s-1} \quad \text{if } m \leq L - s + 1, \quad (9)$$

$$(b'_r)_m = 0 \quad \text{if } m > L - s + 1. \quad (10)$$

$$(11)$$

Finally  $\mathcal{X}$  can be decomposed as :

$$\mathcal{X}_{klm} = \sum_{r=1}^R A_{k,r} B_{l,r} C_{m,r} \prod_{r=1}^R e^{-A_{k,r}(B_{l,r} + B'_{m,r})} + \mathcal{E}_{klm}, \quad (12)$$

where  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{B}'$  and  $\mathbf{C}$  are matrices of size  $(K \times R)$ ,  $(L \times R)$ ,  $(M \times R)$  and  $(M \times R)$  respectively so that column  $r$  of  $\mathbf{A}$  (respectively  $\mathbf{B}$ ,  $\mathbf{B}'$  and  $\mathbf{C}$ ) contains vector  $\mathbf{a}_r$  (respectively  $\mathbf{b}_r$ ,  $\mathbf{b}'_r$  and  $\mathbf{c}_r$ ). This decomposition is called the Non Linear Fluorescence Decomposition (NLFD) and matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  are the factor matrices of the decomposition ( $\mathbf{B}'$  being directly deduced from  $\mathbf{B}$ ).

From an algorithmic point of view, it will be useful to stack all the unknown parameters of the decomposition, *i.e.* all entries of  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  in an unique parameter vector  $\boldsymbol{\theta}$ :

$$\boldsymbol{\theta} = \begin{bmatrix} \mathbf{vec}(\mathbf{A}) \\ \mathbf{vec}(\mathbf{B}) \\ \mathbf{vec}(\mathbf{C}) \end{bmatrix}, \quad (13)$$

where  $\mathbf{vec}()$  is the operator that maps a matrix or a tensor to a column vector by stacking its columns one below the other in a prescribed order<sup>1</sup>. In the following we define tensors  $\mathcal{G}_{CPD}(\boldsymbol{\theta})$  and  $\mathcal{G}_{NLFD}(\boldsymbol{\theta})$  as

$$[\mathcal{G}_{CPD}(\boldsymbol{\theta})]_{klm} = \sum_{r=1}^R A_{k,r} B_{l,r} C_{m,r}, \quad (14)$$

$$[\mathcal{G}_{NLFD}(\boldsymbol{\theta})]_{klm} = [\mathcal{G}_{CPD}(\boldsymbol{\theta})]_{klm} \prod_{r=1}^R e^{-A_{k,r}(B_{l,r} + B'_{m,r})}, \quad (15)$$

<sup>1</sup>The exact definition of this reorganization is not important as long as it is known, bijective, and consistent with subsequent operations.

so that:

$$\mathcal{X}_{klm} = [\mathcal{G}_{NLFD}(\boldsymbol{\theta})]_{klm} + \mathcal{E}_{klm} \quad (16)$$

$$= [\mathcal{G}_{CPD}(\boldsymbol{\theta})]_{klm} \prod_{r=1}^R e^{-A_{k,r}(B_{l,r}+B'_{m,r})} + \mathcal{E}_{klm}. \quad (17)$$

Therefore, if IFE can be neglected, then the argument of the exponential term of  $\mathcal{G}_{NLFD}$  tends to 0 and the fluorescence model becomes a classical truncated CPD of the fluorescence tensor. As it has been briefly recalled in the introduction, standard approaches resort to additional measurements to estimate  $\mathcal{G}_{CPD}(\boldsymbol{\theta})$  (FEEM linearization step) and then deduce the factor matrices from its CPD. The goal of the remainder is to give an insight on how to work directly on the NLFD instead of the CPD of linearized FEEM.

## 2.2. Optimisation problem

Finding factors that best explain the fluorescence data is equivalent to solving a minimization problem. Because equation (12) is parametric and the function  $\mathcal{G}_{NLFD}$  is smooth, we are able to use descent algorithms. However, CP decomposition is still known to be a difficult non-convex optimization problem, and as shown further, NLFD suffers from similar issues.

$\mathcal{X}$  is the  $\mathbb{R}^{K \times L \times M}$  tensor containing the fluorescence data. The objective function  $S(\boldsymbol{\theta})$  is defined (up to a multiplicative constant) as the squared Euclidian norm between identical vectorizations of  $\mathcal{X}$  and  $\mathcal{G}_{NLFD}$ . It also corresponds (up to another multiplicative constant) to the likelihood in the presence of Gaussian noise. Because of the physical meaning factors bear, they cannot be negative and this constraint can be included in the problem statement. Thus the constrained optimization problem is to find  $\boldsymbol{\theta}_0$  such that :

$$\begin{cases} \boldsymbol{\theta}_0 = \operatorname{argmin} S(\boldsymbol{\theta}) = \operatorname{argmin} \frac{1}{2} \|\operatorname{vec}(\mathcal{G}_{NLFD}(\boldsymbol{\theta}) - \mathcal{X})\|_2^2 \\ \boldsymbol{\theta} \in \mathbb{R}_+^{KLM} \end{cases} \quad (18)$$

Solving (18) could seem to be an easy question at first glance because we have all the information we need to build a simple descent algorithm. However,  $S$  is firstly not a convex function. It contains numerous local minima, as a polynomial in  $\boldsymbol{\theta}$ . Moreover, the set of all tensors admitting an exact CPD of rank  $R$  is not closed if  $\boldsymbol{\theta} \in \mathbb{R}$ , so it may not be closed also in the non multilinear case. Actually, this means that we do not know whether a solution to the approximation problem (18) actually exists nor if it is unique.

The problem of existence of the best low-rank approximation has been recently addressed in [17]. It has been shown therein that the set of nonnegative tensors of fixed rank is closed. This means that the nonnegativity constraint ensures the existence of a best approximation of the measurement tensor by low-rank nonnegative tensors. In other words, a global minimum to the objective function always exists for the CPD. However, this does not hold true for NLFD, because the multiplicative exponential term tends to zero when its arguments tend to infinity. Moreover, if measurements are too noisy or if the model appears to be too far from the reality of the situation, we might eventually end up with a local minimum instead of the best approximation, even if it exists. Because we will use a descent algorithm, much attention will thus be paid to the initialization of the algorithm.

Another issue is how to include positivity constraints in a suitable descent algorithm. Many authors choose to add a barrier function to the objective function to ensure positivity [18], others

solve the dual problem. Parameterization is also an option [19]. Actually, a pertinent comparison between all those solutions for tensor decomposition is yet to be done. Because initialization is nonnegative and carefully computed in our setting, no positiveness constraints will be added in the descent algorithm.

### 2.3. Identifiability of the non linear model

When building a parametric model for measurements, the most fundamental question to ask is whether the parameters of the model can be inferred from the data or not. When they can be uniquely recovered, the model is said to be identifiable. Identifiability can be understood here as global uniqueness of a best fit of the noiseless data by the model for some parameter set. In this section, local identifiability will be studied, that is, we wonder if around a set of given parameters, the NLFD is well-defined and unique. Note that local identifiability is a weaker property than the global identifiability, but the latter is much harder to check numerically. Most of the time, in concrete applications, this identifiability question is eluded, and optimization algorithms are run without any attention being paid to the credibility of their output. We will see that local identifiability is hopefully almost always ensured in most practical cases, with a strong restriction however in the matrix case where we have only one FEEM.

The local identifiability is a necessary condition for the model to be well-defined. In other words, if the NLFD is identifiable, this does not mean that the approximation problem (18), or even the exact decomposition problem, will admit a solution. Actually, in the approximation problem, i.e. in the presence of noise, very little is known on the global uniqueness of the best approximation [20]. However, results subsequently presented on local identifiability of the NLFD hold for both the approximate and exact decompositions, as we simply try to fit a NLFD on the data in both cases.

Concerning the existence of a solution of (18), finding the best low-rank approximation in the NLFD model (15) is guaranteed to be well-posed only in the attraction basin of any (possibly local) minimum and under conditions conjectured in subsection 2.4. In fact, in that case the search is limited to a bounded set. Again, global uniqueness of the best low rank approximation is a difficult problem for the (multilinear) CPD [20], and it is by no means an easier problem in the non multilinear case. Yet, local uniqueness can be studied by simple derivative tools without taking into account the non-negativity of the tensor.

#### 2.3.1. Matrix bilinear decomposition

Regarding tensor decomposition, the most prominent example of non uniqueness of the decomposition is the matrix case. Take for example  $K = 1$  in (14). The measurement tensor is then simply a  $L \times M$  matrix that we want to decompose in the following manner :

$$\mathcal{G}(\theta) = \sum_{r=1}^R \mathbf{b}_r \otimes \mathbf{c}_r = \mathbf{B}\mathbf{C}^T \quad (19)$$

It is clear that for the same tensor (which is here a matrix)  $\mathcal{G}$ , there exist an infinite number of matrices  $\mathbf{B}$  and  $\mathbf{C}$  containing candidates factors  $\mathbf{b}_r$  and  $\mathbf{c}_r$ . Indeed, let  $\mathbf{P}$  be any orthogonal matrix, so that  $\mathbf{P}\mathbf{P}^T = \mathbf{I}$ . Then  $\mathcal{G}(\theta) = \mathbf{B}\mathbf{C}^T = (\mathbf{B}\mathbf{P})(\mathbf{C}\mathbf{P})^T$ . Thus it is impossible to recover the true factors  $\theta$ . To restore identifiability, an arbitrary constraint of orthogonality between factors can be imposed, and leads to the well known singular value decomposition (SVD). But this constraint may have no real physical meaning. CP decomposition for tensors of order higher

than 2 is known to generally restore this identifiability without adding any constraint on the factors. This is the main reason why it has been used extensively in fluorescence spectroscopy [5].

### 2.3.2. Jacobian of the NLF D

The local identifiability of the NLF D can be checked at any fixed point by using the local inversion theorem. Take any  $\theta$  in the parameter space. If the Jacobian matrix of the model at this point is full column rank, then the model is locally invertible. If the set of  $\theta$  for which the NLF D is not locally unique has zero volume in the parameters space, then we say that the NLF D model is generically identifiable.

The link between the column rank of the Jacobian and the local uniqueness can be explicated in many ways, and we give now a understandable rationale to address it. Let us use the gradient  $\mathbf{g}$  of the objective function  $S$  defined in (18) in the previous section:

$$\mathbf{g}(\theta) = \mathbf{J}(\theta)^T \mathbf{e}(\theta) \quad (20)$$

where  $\mathbf{J}$  denotes the Jacobian matrix of the NLF D defined as  $J_{ij} = \frac{\partial \text{vec}(\mathcal{G}_{NLF D})_i}{\partial \theta_j}(\theta)$ , and

$$\mathbf{e}(\theta) = \text{vec}(\mathcal{G}_{NLF D}(\theta) - \mathcal{X}) \quad (21)$$

is the error vector. Now suppose that  $\mathbf{J}(\theta)$  is not full column rank, which means that there exist a non null tensor  $\mathcal{T}$  so that  $\mathbf{J}(\theta)^T \text{vec}(\mathcal{T}) = 0$  for some  $\theta$  minimizing  $S$ . Then this means equivalently that we have several tensors cancelling the gradient because:

$$\mathbf{g}(\theta) = \mathbf{J}(\theta)^T \text{vec}(\mathcal{G}_{NLF D}(\theta) - \mathcal{X}) = \mathbf{J}(\theta)^T \text{vec}([\mathcal{G}_{NLF D}(\theta) + \mathcal{T}] - \mathcal{X}).$$

In other words, for a non identifiable model, there are subspaces of the parameter space mapped to the same point in the tensor space by the model, seen as a mapping from  $\mathbb{R}^{R(K+L+M)}$  to  $\mathbb{R}^K \otimes \mathbb{R}^L \otimes \mathbb{R}^M \sim \mathbb{R}^{K \times L \times M}$ . Further explanations can be found in [21, 22]. Notice that working in  $\mathbb{R}_+$  does not change this property as long as the zero border is not reached.

We have seen that the gradient of the objective function is related to the Jacobian (this holds true in any least squares problem). Moreover, identifiability of the model is completely described by it. Calculating the Jacobian is not a difficult task, but it is tedious. We will thus refer to our previous paper [21] for the full calculus and give the following compact expression using the Kronecker product  $\boxtimes$ :

$$\mathbf{J}(\theta) = \mathbf{H}(\mathbf{J}_{CP} - \mu \mathbf{G} \mathbf{J}_{Exp})(\theta), \quad (22)$$

where

$$\left\{ \begin{array}{l} \mathbf{H} = \text{Diag} \left( \exp \left[ - \sum_{r=1}^R \mathbf{a}_r \boxtimes (\mathbf{b}_r \boxtimes \mathbf{1}_M + \mathbf{1}_L \boxtimes \mathbf{b}'_r) \right] \right), \\ \mathbf{J}_{CP} = [\mathbf{a}_1 \boxtimes \mathbf{b}_1 \boxtimes \mathbf{I}_M \mid \mathbf{a}_1 \boxtimes \mathbf{I}_L \boxtimes \mathbf{c}_1 \mid \mathbf{I}_K \boxtimes \mathbf{b}_1 \boxtimes \mathbf{c}_1 \mid \mathbf{a}_2 \boxtimes \mathbf{b}_2 \boxtimes \mathbf{I}_M \mid \dots \mid \mathbf{I}_K \boxtimes \mathbf{b}_R \boxtimes \mathbf{c}_R], \\ \mathbf{G} = \text{Diag} \left( \sum_{r=1}^R \mathbf{a}_r \boxtimes \mathbf{b}_r \boxtimes \mathbf{c}_r \right), \\ \mathbf{J}_{Exp} = [0 \mid \mathbf{a}_1 \boxtimes (\mathbf{I}_L \boxtimes \mathbf{1}_M + \mathbf{1}_L \boxtimes \mathbf{I}_M') \mid \mathbf{I}_K \boxtimes (\mathbf{b}_1 \boxtimes \mathbf{1}_M + \mathbf{1}_L \boxtimes \mathbf{b}'_1) \mid \dots \mid \mathbf{I}_K \boxtimes (\mathbf{b}_R \boxtimes \mathbf{1}_M + \mathbf{1}_L \boxtimes \mathbf{b}'_R)], \end{array} \right.$$



where  $\exp[\cdot]$  is taken entry-wise. Moreover, the form taken by  $\mathbf{I}_M'$  in  $\mathbf{J}_{Exp}$  depends on the shift as follows :

$$\begin{cases} \mathbf{I}_M' = [0_{LM \times s} \mid \mathbf{I}_M \mid 0_{LM \times (L-M-s)}] & \text{if } L - s > M, \\ \mathbf{I}_M' = [0_{LM \times s} \mid \mathbf{I}_M(\cdot, 1 : (L-s))] & \text{if } L - s \leq M. \end{cases}$$

Using linear algebra and numerical simulations, it is then possible to study the column rank of the Jacobian at a general point  $\theta$ , or at least give strong conjectures on the identifiability of the NLFD. Remember that all the following results deal with exact decompositions, the noiseless identifiability being necessary but not sufficient to design a converging algorithm to solve (18).

#### 2.4. Generic results on identifiability

All results above are true with overwhelming probability, i.e. for any tensor drawn randomly in all possible tensors of given dimensions, which we call a generic tensor. This is needed both to run simulations, and to build constructive proofs on the Jacobian column rank.

Firstly, it should be clear that the model suffers from a scaling indetermination between the first two factors. This can be seen also on the  $KLM \times R(K + L + M)$  Jacobian as pointed out in [23], and leads to the following proposition :

*Proposition 1.* For a generic tensor  $\mathcal{T} = \mathcal{G}_{NLFD}(\theta)$  in  $\mathbb{R}_+^{*K \times L \times M}$ ,

$$\text{rank}(\mathbf{J}(\theta)) \leq (K + L + M - 1)R.$$

We call  $\bar{R}$  the critical value of NLFD rank defined by

$$\bar{R} = \left\lfloor \frac{KLM}{K + L + M - 1} \right\rfloor \quad (23)$$

where  $\lfloor x \rfloor$  defines the integer part of  $x$ .

This proposition means that the scaling ambiguity reduces the column rank of  $\mathbf{J}$  by  $R$ . The difference with the CPD is noticeable ( $K + L + M - 1$  instead of  $K + L + M - 2$ ) since the norm of the third factor in the NLFD is indeed fixed.

Even though the Jacobian cannot have full column rank, the scaling indeterminacy is in most cases the only cause of column rank deficiency. The remainder of this section deals with cases where the column rank of  $\mathbf{J}$  is strictly lower than the critical value for which the column rank of the Jacobian is full, up to the scaling problem. In other words, if  $R = \bar{R}$ , and if  $R$  well-chosen columns are deleted in  $\mathbf{J}$ , the latter becomes full column rank.

We shall not take into account the shift in the remaining theoretical study for the sake of simplicity, without loss of generality. Yet it will be included in the formula of the Jacobian for the descent algorithm.

At this point, two main questions arise when studying identifiability of the NLFD. **Q1.** First, does the non linearity restore the identifiability of the matrix case ? In other words, knowing that the fluorescence data follows a non multilinear model due to the inner filter effects, could we use only one FEEM to recover uniquely the concentrations and spectra of the  $R$  components ? **Q2.** Another issue is what are the dimensions of the measurement set and the corresponding ranks of the exact CPD for which identifiability will not be met ? Also, can we find a boundary on such dimensions, and is it restrictive for real life data ?

*Question Q1.* As exposed earlier, CPD for matrices is not a well posed problem if orthogonality constraints are not imposed on the factors. One could however hope that the non-linearity due to inner filter effects could restore the identifiability of the decomposition. The next proposition proved in [21] shows that this is not the case.

*Proposition 2.* For the matrix case, *i.e.* with  $\mathbf{a}_r$  playing no role in the NLFD, the non linear decomposition with rank  $R > 1$  does not restore generic identifiability, *i.e.*, the Jacobian is full column rank only when  $R = 1$ .

Once again, this means that the NLFD cannot be applied to only one FEEM and that the tensorial structure is needed, except for a rank one approximation (that is, with only one component in the mixture), as the Jacobian column rank is full only when  $R = 1$ .

*Question Q2.* The other case of non-identifiability for the NLFD, more difficult to apprehend, happens when the rank of the measurement tensor is equal to the critical rank  $\bar{R}$ , but the number of unknown parameters is still larger than the number of free equations. This happens only when one dimension is much larger than the two others, as precised in the conjecture below. In such cases, the rank  $R$  of model (15) should be chosen such that  $R < \bar{R}$  to ensure local uniqueness. There exist no theoretical result for the non multilinear case, but we give an empirical bound on the dimensions of the measurement tensor to ensure identifiability, obtained by simulations of the Jacobian rank.

*Conjecture 1.* For a generic tensor  $\mathcal{T} = \mathcal{G}_{NLFD}(\theta)$  in  $\mathbb{R}_+^{*K \times L \times M}$ , if  $R = \bar{R}$ , and  $M \geq (K - 1)(L - 1) + 3$  or any permutation of this inequality is true, then the NLFD is generically not locally identifiable and we say that the model is defective unbalanced.<sup>2</sup> This does not hold for  $R < \bar{R}$ .

Note that the value of the critical rank  $\bar{R}$  grows very fast with the dimension of the measurement tensor, so that the case  $R = \bar{R}$  is highly unlikely to be encountered if the dimensions of the tensor are not very small. If anyway  $R$  happens to be strictly larger than the critical rank, then the model is still not identifiable because the Jacobian has more columns than rows (taking the scaling indeterminacy into account), and therefore the column rank cannot be full.

To summarize the identifiability issue, necessary conditions for a data set to be decomposable efficiently using the NLFD model are two-fold: the measurement tensor must be at least of order 3, and the number of components to be extracted must not be very large compared to the dimensions of the tensor, which is absolutely not a problem with typical fluorescence inverse problems. These conditions are necessary, because the optimization problem addressed in this paragraph is noiseless, but we conjecture they are also sufficient in practice. In the next section, an algorithm is described to solve the optimization problem at hand.

### 3. NLFD algorithm

#### 3.1. Description of the optimization algorithm

The optimization problem (18) is constrained (cf. Section 2.2) and highly non-convex. It differs from the usual nonnegative CP decomposition problem because of the non-linearity induced

---

<sup>2</sup>The terminology *defective unbalanced* is already used in the literature for multilinear models [24].

by the exponential term. However it is still smooth, and standard tensor decomposition routines can be easily adapted here.

Many different algorithms exist to decompose tensors following a CPD model, but the two main trends are alternating algorithms and all-at-once descent algorithms. Alternating algorithms are the easiest to implement and give satisfying results. However, all-at-once algorithms are known to prevent some of the problems encountered with alternating algorithms with similar convergence speed and computation time. Because the gradient and the Jacobian can be obtained easily from our previous study on identifiability, we adapted the Levenberg-Marquardt descent method described in [25] to solve (18). This NLFD-LM algorithm is described in Algorithm 1.

---

**Algorithm 1** The NLFD-LM algorithm

---

```

1:  $\lambda=2; \eta=2;$ 
2: Initialize  $\mathbf{A}, \mathbf{B}$  and  $\mathbf{C}$  from a fast CPD of  $\mathcal{X}$ ;
3: build  $\theta^0$  from this initial factor matrices
4: Compute  $\mathbf{e}(\theta^0) = \text{vec}(\mathcal{G}_{NLFD}(\theta^0) - \mathcal{X})$ ;
5: Compute  $\mathbf{J}$  at  $\theta^0$ ;
6: Compute  $\mathbf{g} = \mathbf{J}^t \mathbf{e}(\theta^0)$ 
7: flag = false;
8:  $i = 0$ ;
9: while flag = false and  $i < \text{imax}$  do
10:   Compute  $\mathbf{h} = -(\mathbf{J}^t \mathbf{J} + \lambda \mathbf{I})^{-1} \mathbf{g}$ ;
11:   Compute  $\theta = \theta^i + \mathbf{h}$ ;
12:   Compute  $\mathbf{e}(\theta)$ ;
13:   Compute  $S(\theta)$  and  $S(\theta^i)$ ;
14:   Compute  $S_{app} = S(\theta^i) + \mathbf{h}^t \mathbf{g} + \mathbf{h}^t \mathbf{J}^t \mathbf{J} \mathbf{h} / 2$ ;
15:   Compute  $Q = \frac{S(\theta^i) - S(\theta)}{S(\theta^i) - S_{app}}$ ;
16:   if  $Q > 0$  ( $\theta$  is accepted and we decrease  $\lambda$ ) then
17:      $\theta^{i+1} = \theta$ ;
18:     Compute  $\mathbf{J}$  at  $\theta^{i+1}$ ;
19:     Compute  $\mathbf{g} = \mathbf{J}^t \mathbf{e}(\theta)$ 
20:      $\lambda = \lambda * \max(1/3, (1 - (2Q - 1)^3))$ ;
21:      $\eta = 2$ ;
22:     Compute a stopping criterion and according to its value, set flag to true;
23:   else
24:      $\theta^{i+1} = \theta^i$ ; ( $\theta$  is rejected and we increase  $\lambda$ )
25:      $\lambda = \lambda \eta$ ;
26:      $\eta = 2\eta$ ;
27:   end if
28:    $i = i + 1$ ;
29: end while

```

---

This second order gradient-based descent algorithm is known to perform well for the multi-linear CPD problem, but it is rather demanding computationally when problem dimensions are large. Yet, an alternating algorithm is unadapted here since each iteration would require to solve a non-linear constrained problem. Also, recent compression methods [26] will not apply because of the non-linearity of the NLFD. However, for very large data set, conjugate gradient algorithms

may lead to faster computation.

Another issue is how to include nonnegativity constraints in a descent algorithm. This issue has already been tackled in Section 2.2, and we will not dig into this problem here. In our practical simulations on real data, the unconstrained version of the algorithm described here worked well and no negative values were observed after convergence. Should this be the case, a very simple approach is to use the absolute value of the factors at each iteration, or to set negative values to zero.

The main difficulty with the Levenberg-Marquardt (LM) descent method is the choice of the initial value for the estimated factors. In our preliminary tests we found that the NLFD-LM algorithm was indeed very sensitive to the initial guess. However, the NLFD is close to CPD numerically. For this reason, the initialization of the LM algorithm can be done efficiently by computing the CPD of the measurement tensor.

### 3.2. Handling non-uniform spectral sampling

The NLFD-LM algorithm described in the previous section assumes as a working hypothesis that FEEM excitation and emission sampling steps and extremal values are chosen so that in the wavelength range  $[\lambda_{em}^{min}, \lambda_{ex}^{max}]$  all excitation and emission wavelengths coincide *i.e.*  $\Delta_{ex} = \Delta_{em}$  and  $\lambda_{em}^{min}$  coincides with a value of  $\lambda_{ex}$ . However, this is not always the case in practice. In particular, it is quite common to have a greater sampling step for the excitation. As a consequence we now introduce a generalized version of the algorithm that allows to relax this assumption. Our working hypothesis now states that (i)  $\lambda_{em}^{min}$  coincides with a value of  $\lambda_{ex}$  and that (ii)  $\Delta_{ex} = \beta\Delta_{em}$  with  $\beta$  an integer greater than or equal to 1. For instance, we can have the following discrete excitation range: [280 : 20 : 500] and discrete emission range: [340 : 5 : 600] since here  $\beta = 4$  (20/5) and  $\lambda_{em}^{min}$  coincides with the fourth wavelength of the excitation range (340 nm). The two latter properties are not mandatory for the algorithm to work, but they will be also included in our working hypothesis since they are generally available and significantly simplify the implementation.

In order to merge the two sampling grids, we resort to a linear interpolation of the excitation spectra, which has shown to be broadly sufficient because they are usually very smooth. Let  $b(\lambda_{ex})$  be an excitation spectrum and  $\lambda_{em}$  an emission wavelength belonging to  $[\lambda_{em}^{min}, \lambda_{em}^{max}]$  for which we have to compute the value of  $b$ . Now, let  $\lambda_{ex}^1$  and  $\lambda_{ex}^2$  be the excitation wavelengths included in the discrete excitation range and closest to  $\lambda_{em}$ , so that  $\lambda_{ex}^1 < \lambda_{em} < \lambda_{ex}^2$ . For instance, in the previous numerical example, for  $\lambda_{em} = 355\text{nm}$  we will have  $\lambda_{ex}^1 = 340\text{nm}$  and  $\lambda_{ex}^2 = 360\text{nm}$ . The idea is to approximate  $b(\lambda_{em})$  as a function of  $b(\lambda_{ex}^1)$  and  $b(\lambda_{ex}^2)$ . One can always express  $\lambda_{em}$  as:  $\lambda_{em} = \alpha_{\lambda_{em}}\lambda_{ex}^1 + (1 - \alpha_{\lambda_{em}})\lambda_{ex}^2$  with  $\alpha_{\lambda_{em}} = \frac{\lambda_{ex}^2 - \lambda_{em}}{\lambda_{ex}^2 - \lambda_{ex}^1}$ . Therefore a linear approximation of  $b(\lambda_{em})$  in  $[\lambda_{ex}^1; \lambda_{ex}^2]$  leads to:

$$b(\lambda_{em}) = \alpha_{\lambda_{em}} b(\lambda_{ex}^1) + (1 - \alpha_{\lambda_{em}}) b(\lambda_{ex}^2). \quad (24)$$

Now, denoting  $\mathbf{b}$  the sampled version of  $b$  with a step  $\Delta_{ex} = \beta\Delta_{em}$  and  $m$  the position of  $\lambda_{em}$  in the emission sampling interval, according to our working hypothesis we have:

- $\lambda_{ex}^2 = \lambda_{ex}^1 + \Delta_{ex}$ ,
- $\lambda_{em} = (m - 1)\Delta_{em} + \lambda_{em}^{min}$ ,
- $\lambda_{em}^{min} = (s - 1)\Delta_{ex} + \lambda_{ex}^{min}$ ,

- $b(\lambda_{ex}^1) = b_{l_1}$ ,
- $b(\lambda_{ex}^2) = b_{l_2}$

with:  $l_1 = \lfloor \frac{m-1}{\beta} \rfloor + s$  and  $l_2 = \min(l_1 + 1, L)$ . We then immediately deduce the value of  $\alpha_m$ :

$$\alpha_m = 1 + \lfloor \frac{m-1}{\beta} \rfloor - \frac{m-1}{\beta}. \quad (25)$$

Finally the NLFD (12) is rewritten as:

- if  $m > (L - s)\beta + 1$  then:

$$[\mathcal{G}_{NLFD}(\theta)]_{k,l,m} = [\mathcal{G}_{CPD}(\theta)]_{k,l,m} \prod_{r=1}^R e^{A_{k,r} B_{l,r}}. \quad (26)$$

- Otherwise:

$$[\mathcal{G}_{NLFD}(\theta)]_{k,l,m} = [\mathcal{G}_{CPD}(\theta)]_{klm} \prod_{r=1}^R e^{A_{k,r}(B_{l,r} + \alpha_{m,r} B_{l_1 r + (1 - \alpha_{m,r}) B_{l_2 r}})}. \quad (27)$$

One can easily verify that for  $\beta = 1$  (26) and (27) are equivalent to (15). Of course, the decomposition can still be performed using the NLFD-LM algorithm, only the expression of the error vector  $\mathbf{e}$  and of the Jacobian matrix have to be modified. The error vector is computed easily by reporting (26) and (27) (instead of (15) in (21) and we show in Appendix A how to compute the Jacobian matrix in this more general case (matlab codes are also provided).

#### 4. Materials and methods

The proposed NLFD approach has been tested on two real data sets gathering mixtures of two and three fluorophores respectively. Each data set was constituted and treated independently. All 2D spectra and FEEM were measured with a fluorometer Hitachi F4500 at the PROTEE laboratory (university of Toulon). Fluorescence intensity was corrected from PM response using manufacturer setting. Data for FEEM treatment were extracted by FLWinLab software. Reference excitation and emission spectra of the fluorophores were measured from diluted solutions of each fluorophore. These 2D spectra were recorded at 240 nm/min scan speed by step of 5 nm with 5 nm bandwidth in excitation and 2.5 nm bandwidth in emission.

##### 4.1. First data set: highly concentrated mixtures of fluorescein and quinine sulphate

Five mixtures of fluorescein (Aldrich) and quinine sulphate (Merck) were prepared in 0.1M  $H_2SO_4$  (Aldrich). All chemical are analytical grade. Fluorophore concentrations are reported in table 1 and were chosen so that each mixture is affected by strong IFE effects. FEEM of the five mixtures were recorded at 30000 nm/min scan speed from 350 to 600 nm in emission, and from 280 to 500 nm for excitation wavelength with excitation and emission steps of 5 nm ( $\Delta_{ex} = \Delta_{em} = 5nm$ ). Excitation and emission bandwidth were 5 nm. We thus obtain a first fluorescence tensor denoted  $\mathcal{X}^1$  of size  $(5 \times 45 \times 51)$  with  $s^1 = 15$  and  $\beta^1 = 1$ .

Table 1: Concentrations in parts per million (ppm), maximal absorbances and mean absorbances for data set 1.

Solution number	1	2	3	4	5
$c_{QS}$ (ppm)	11.02	32.6	54.38	76.15	97.73
$c_F$ (ppm)	74.72	58.23	41.58	24.92	8.43
Absorbance max.	2.18	1.83	1.32	1.17	1.38
Mean Absorbance	0.40	0.37	0.33	0.29	0.27

Then, in order to validate the non-uniform spectral sampling version of our algorithm a second tensor  $\mathcal{X}^2$  was built by downsampling the excitation mode by a factor 2 (*i.e.* we have now  $\Delta_{ex} = 10nm$ ). This tensor is thus of size  $(5 \times 23 \times 51)$  with  $s^2 = 8$  and  $\beta^2 = 2$ .

Eventually, we have also tested our algorithm in a more critical situation by downsampling  $\mathcal{X}^1$  by a factor 4 (*i.e.*  $\Delta_{ex} = 20nm$ ). This third tensor is denoted  $\mathcal{X}^3$ , it is of size  $(5 \times 11 \times 51)$  and we have  $s^3 = 4$  and  $\beta^3 = 4$ . (Note that for this last case we had first suppressed tensor entries corresponding to  $\lambda_{ex} = 280$  and  $\lambda_{ex} = 285$  so that the working conditions of our non-uniform spectral sampling NLFD algorithm were respected).

#### 4.2. Second data set: highly concentrated mixtures of fluorescein, quinine sulphate and tryptophane

Table 2: Concentrations in  $mg.L^{-1}$  and maximal absorbances for data set 2.

Solution number	1	2	3	4	5
$c_{Trip}$ ( $mg.L^{-1}$ )	53.5	53.5	53.5	53.5	53.5
$c_{QS}$ ( $mg.L^{-1}$ )	0	14.66	22	7.3	14.6
$c_F$ ( $mg.L^{-1}$ )	0	14.4	14.4	21.6	21.6
Absorbance max.	2.4	2.4	2.5	2.8	2.8

We now consider mixtures of 3 fluorophores. Our second data set is then constituted of 5 highly concentrated mixtures of fluorescein, quinine sulphate and tryptophane. Fluorophore concentrations along with absorbance values are given in table 2. 3D spectra were recorded at 2400 nm/min scan speed, with excitation and emission slit width of 2.5 nm and from 270 to 550 nm by step of 5 nm both in excitation and emission. Consequently, we obtain a fluorescing tensor  $\mathcal{X}^4$  of size  $(5 \times 57 \times 57)$  with  $s^4 = 1$  and  $\beta^4 = 1$ . We then build two other tensors by downsampling the excitation mode in the same way we built the first data set. Thereby  $\mathcal{X}^5$  is of size  $(5 \times 29 \times 57)$  with  $s^5 = 1$  and  $\beta^5 = 2$  ( $\Delta_{ex} = 10nm$ ) and  $\mathcal{X}^6$  is of size  $(5 \times 15 \times 57)$  with  $s^6 = 1$  and  $\beta^6 = 4$  ( $\Delta_{ex} = 20nm$ ).

#### 4.3. Data treatment

In order to help the reader to navigate through our different experiments corresponding to the six fluorescing tensors, all their characteristics are summed up by Table 3.

All data treatments were made using MATLAB software. Rayleigh and Raman scatters were first removed numerically from each measured FEEM using the method proposed by Zepp [27]. Then each of the six tensors were fitted by both a non negative truncated CPD and the NLFD.

Table 3: Characteristics of the six fluorescing tensors used for NLFD-LM validation

Tensor	Data set	Fluorophore number ( $R$ )	tensor dimensions ( $K \times L \times M$ )	Shift ( $s$ )	Excitation downsampling factor ( $\beta$ )
$\mathcal{X}^1$	1	2	$(5 \times 45 \times 51)$	15	1
$\mathcal{X}^2$	1	2	$(5 \times 23 \times 51)$	15	2
$\mathcal{X}^3$	1	2	$(5 \times 11 \times 51)$	15	4
$\mathcal{X}^4$	2	3	$(5 \times 57 \times 57)$	1	1
$\mathcal{X}^5$	2	3	$(5 \times 29 \times 57)$	1	2
$\mathcal{X}^6$	2	3	$(5 \times 15 \times 57)$	1	4

CPD was computed using the non negative ALS algorithm of the NWAY package provided by R.Bro [28, 29]. The NLFD was computed by the NLFD-LM algorithm proposed in section 3.

In our simulations we thus use the absolute value of the the factor matrices estimated by a recent fast CPD algorithm called DIAG [30] as initial guess for NLFD-LM. Indeed, this strategy were first tested to synthetic data and improved convergence properties significantly with respect to random initialization. Note that a direct CPD algorithm such as DIAG provides a very quick estimation of the factor matrices so that this initialization step does not slow up the whole process. Then the iterative part of the NLFD-LM algorithm is stopped only when a maximal number of iteration is reached. We empirically found that a maximal number of 50 iterations was enough to achieve convergence. After convergence, permutation and scaling indeterminacies are removed in order to compare NLFD-LM and ALS estimated factors with reference ones.

## 5. Results and discussions

### 5.1. Separation of 2 components mixtures

Results from the NLFD decomposition of tensors  $\mathcal{X}^1$ ,  $\mathcal{X}^2$  and  $\mathcal{X}^3$  are plotted on Figures 1, 2 and 3 respectively. They are compared to CPD results and reference loadings corresponding to the spectra of the pure fluorophores measured at low concentration (i.e. without any IFE) and to the actual concentration profiles used to make the mixtures.

Estimation quality of the factor matrices can also be assessed by means of normalized mean squared errors (NMSE) that quantify the deviation between a reference matrix and its estimate. These values are denoted here  $r_A$ ,  $r_B$  and  $r_C$  for matrices **A**, **B** and **C** respectively.  $r_A$  is given by:

$$r_A = \sum_{n=1}^R \frac{\|\mathbf{A}_n - \hat{\mathbf{A}}_n\{CPD\}\|_2^2}{N\|\mathbf{A}_n\|_2^2}, \quad (28)$$

where  $\mathbf{A}_n$  is the  $n$ -th column of **A**,  $R$  the number of fluorophores, and  $r_B$  and  $r_C$  are constructed in the same way. For each tensor, CPD and NLFD NMSE values are reported in table 4.

This first data set furnishes a classical example of IFE which is mainly due to the conjugation of quite high absorbance values and a strong overlap between quinine sulfate emission spectrum and fluorescein excitation spectrum.

Even in this simple case, impact of IFE appears clearly when looking at CPD results obtained from the original tensor  $\mathcal{X}^1$  in figure 1. Indeed, the fluorescein excitation spectrum produced by the CPD is totally distorted and the evolution of its concentration throughout the mixture set is also badly estimated. Besides the estimated emission spectrum of quinine sulphate suffers

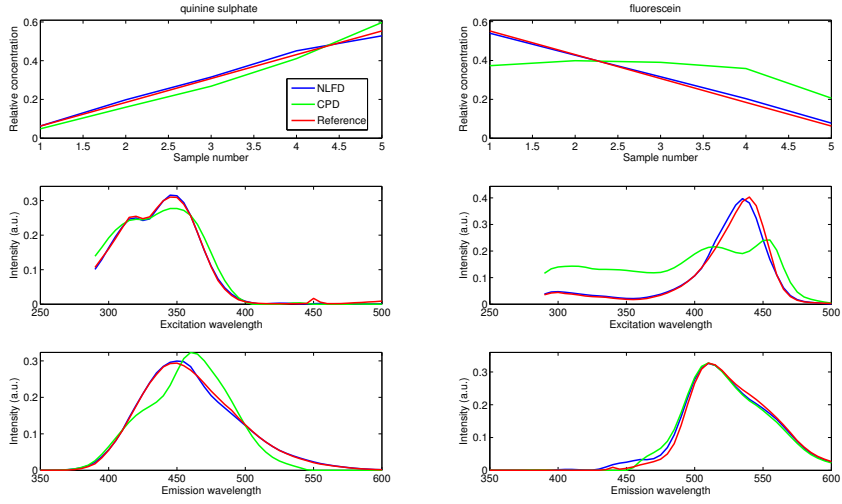


Figure 1: Comparison between CPD (green lines) and NLFD (blue lines) of fluorescence tensor  $\mathcal{X}^1$  (2 components,  $\Delta_{ex} = \Delta_{em}$ ). Reference spectra and concentration profiles are plotted in red.

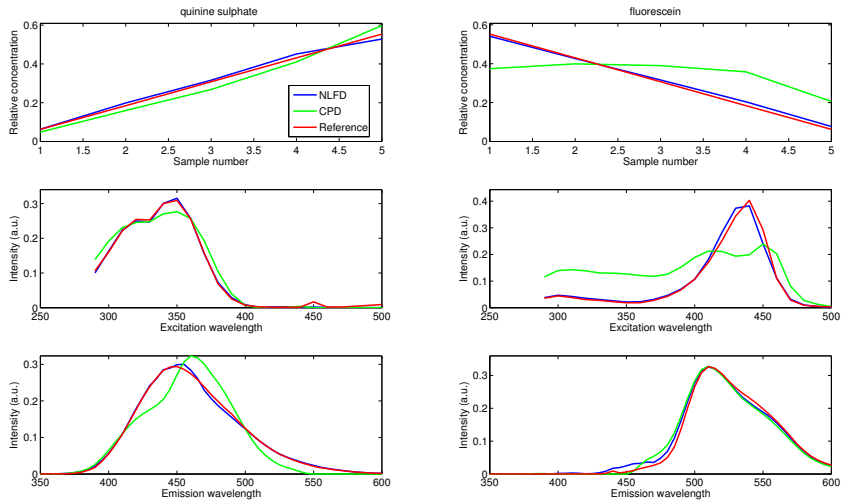


Figure 2: Comparison between CPD (green lines) and NLFD (blue lines) of fluorescence tensor  $\mathcal{X}^2$  (2 components,  $\Delta_{ex} = 2\Delta_{em}$ ). Reference spectra and concentration profiles are plotted in red.



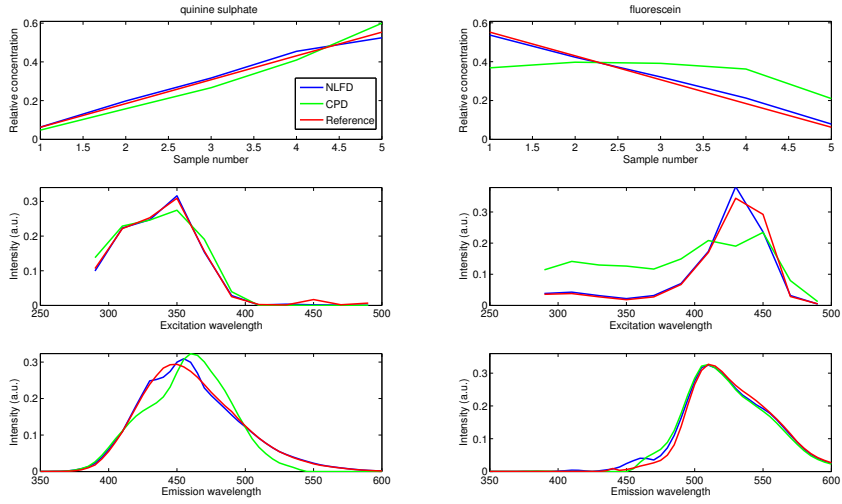


Figure 3: Comparison between CPD (green lines) and NLFD (blue lines) of fluorescence tensor  $\mathcal{X}^3$  (2 components,  $\Delta_{ex} = 4\Delta_{em}$ ). Reference spectra and concentration profiles are plotted in red.

Table 4: Characteristic of the six fluorescing tensors used for NLFD-LM validation

Tensor	CPD NMSE values			NLFD NMSE values		
	$r_A$	$r_B$	$r_C$	$r_A$	$r_B$	$r_C$
$\mathcal{X}^1$	0.0764	0.2039	0.0299	0.0014	0.0057	0.0030
$\mathcal{X}^2$	0.0760	0.2055	0.0298	0.0015	0.0058	0.0030
$\mathcal{X}^3$	0.0809	0.1828	0.0303	0.0026	0.0105	0.0042
$\mathcal{X}^4$	0.1148	0.0771	0.0945	0.0097	0.0108	0.0057
$\mathcal{X}^5$	0.1127	0.0766	0.0937	0.0111	0.0107	0.0059
$\mathcal{X}^6$	0.0924	0.0775	0.0965	0.0174	0.0177	0.0113

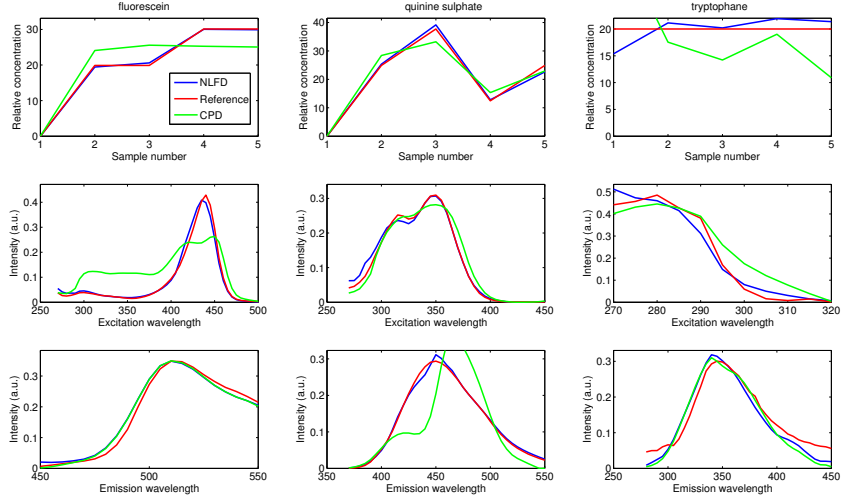


Figure 4: Comparison between CPD (green lines) and NLFD (blue lines) of fluorescence tensor  $\mathcal{X}^4$  (3 components,  $\Delta_{ex} = \Delta_{em}$ ). Reference spectra and concentration profiles are plotted in red.

from a 10 nm shift and its estimated excitation spectrum is smoothed. NMSE values confirm that the worst results are obtained for the excitation mode ( $r_B = 0.2$ ) while emission spectra are almost correctly estimated ( $r_C = 0.03$ ). The disappointing concentration value ( $r_A = 0.076$ ) is an average between the good estimation of quinine sulphate concentration profile and the bad estimation of that of fluorescein.

Conversely, NLFD estimates almost perfectly fit the reference spectra and concentrations profile for both components :  $r_A = 0.001$ ,  $r_B = 0.006$ ,  $r_C = 0.003$ . In fact, only a slight shift of the fluorescein excitation maximum can be observed. This explains the higher value of  $r_B$ . These observations and conclusions still hold for downsampled tensor  $\mathcal{X}^2$  ( $\beta = 2$ , Figure 2). Indeed NMSE value obtained from CPD and NLFD of  $\mathcal{X}^2$  are identical to those obtained from CPD and NLFD of  $\mathcal{X}^1$ . It is then worth mentioning that NLFD results are still very good for  $\beta = 4$  (all spectra and concentration profiles are well recovered) even though the size of the excitation mode is now very small (only 11 points). This points out that for regular functions such as excitation spectra, our linear interpolation is suitable. In fact further studies (data not shown) showed that NLFD results are still good for  $\beta = 8$  (i.e.  $\Delta_{ex} = 40nm$ ) and acceptable for  $\beta = 10$  (i.e.  $\Delta_{ex} = 50nm$ ). A quick look to the NMSE values confirms these conclusion. Indeed, values computed from the CPD decomposition are comprised between 0.0298 and 0.2055 while they are comprised between 0.0014 and 0.0105 for the NLFD.

### 5.2. Separation of 3-component mixtures

This second data set constitutes a difficult case. First because solution absorbances are very high. Second, the excitation spectra of the tryptophan is very narrow, in addition its overall fluorescence signal is quite weak with respect to the two others fluorophore. Finally, its concentration profile is constant. If we first look at the results of the CPD of the original tensor

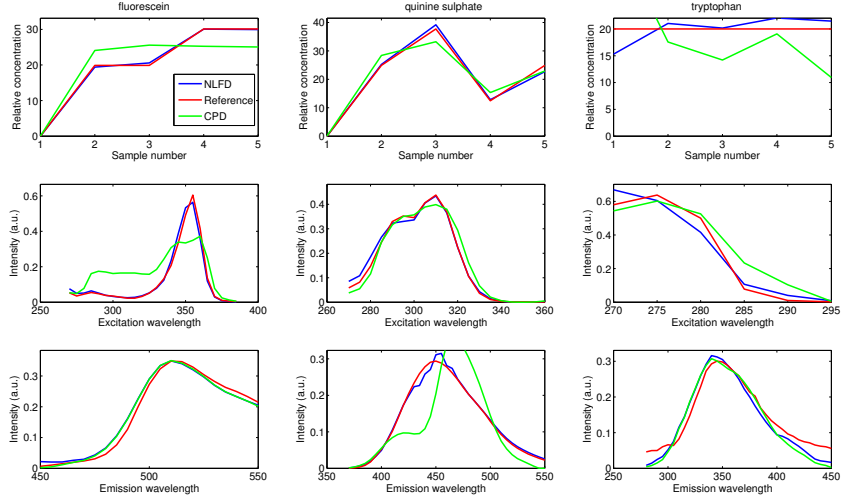


Figure 5: Comparison between CPD (green lines) and NLFD (blue lines) of fluorescence tensor  $\mathcal{X}^5$  (3 components,  $\Delta_{ex} = 2\Delta_{em}$ ). Reference spectra and concentration profiles are plotted in red.

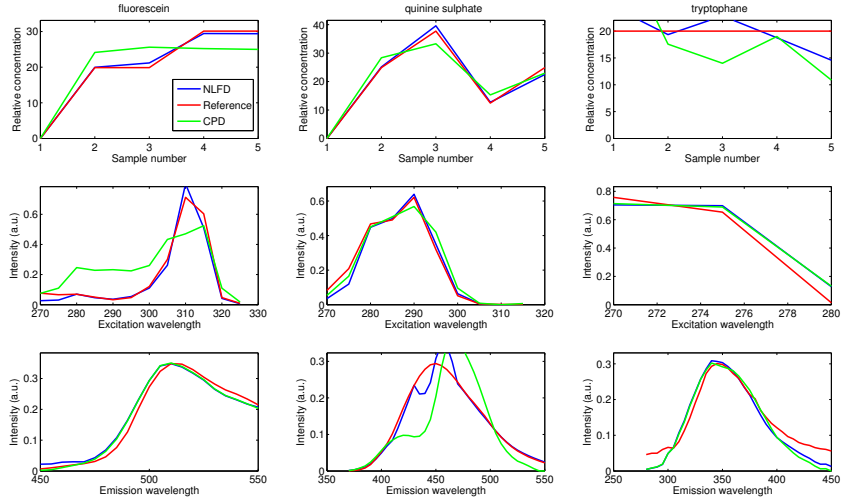


Figure 6: Comparison between CPD (green lines) and NLFD (blue lines) of fluorescence tensor  $\mathcal{X}^6$  (3 components,  $\Delta_{ex} = 4\Delta_{em}$ ). Reference spectra and concentration profiles are plotted in red.

$\mathcal{X}^4$  in figure 4 we observe the same behaviour than for data set 1: some factors are correctly estimated (fluorescein and tryptophan emission spectra, quinine sulfate concentration profile and excitation spectra) whereas some others are severely distorted : fluorescein excitation spectrum, quinine sulfate emission spectrum and tryptophan concentration profile. However these large distortions now affect the three modes (and the three components):  $r_A = 0.11$ ,  $r_B = 0.077$  and  $r_C = 0.094$ . This is mainly due to higher absorbance values. In spite of those difficulties, NLFD still provides very good results, notably regarding the emission mode ( $r_C = 0.006$ ). Concerning the two other modes ( $r_A = 0.01$  and  $r_B = 0.011$ ), fluorescein and quinine sulphate factors are perfectly estimated whereas some slight distortions appears in the estimated concentration profile and excitation spectrum of the tryptophan. This is mainly due to previously stated difficulties. The same kind of results are observed for downsampled tensor  $\mathcal{X}^5$  ( $\beta = 2$ ) in Figure 5. This is confirmed by the NMSE values since all of them remain unchanged with respect to the case  $\beta = 1$ . Finally, increasing the downsampling factor (figure 6) does not change CPD conclusions but slightly affects NLFD results and notably quinine sulphate emission spectrum and tryptophan concentration profile.

As a general conclusion, whatever the considered tensor (*i.e.* whatever the considered number of factors or the sampling step) it appears clearly that applying the CPD to a set of FEEM affected by IFE may lead to incorrect identification of the component of the mixtures. Conversely, in both studied data sets the NLFD approach has consistently provided a very satisfying estimation of all the factors.

## 6. Conclusion and Perspectives

Tensor decomposition methods in fluorescence spectroscopy is becoming the workhorse method for blindly retrieving unknown fluorescent components in sample mixtures. The well known CANDECOMP/PARAFAC decomposition (CPD) is a powerful tool, but the linearity assumption of the fluorescence response is made with respect to functionals of concentrations and spectra of the components, which is not acceptable in some circumstances, e.g. at higher concentrations.

In this paper we have worked directly on the nonlinear response model, called NLFD, which appears naturally when taking into account the absorbance by each component in the mixture. We have shown numerically that the NLFD enjoys properties similar to CPD with respect to identifiability of the unknown concentrations and spectra. We have designed a Newton algorithm to compute the NLFD, and proved its superiority in various contexts. Other nonlinearities could be handled the same way as long as they can be modeled properly. Also, other kinds of algorithms may be designed to compute the NLFD, especially if dealing with large datasets.

## Appendix A. Jacobian matrix computation: formulas and matlab code

The Jacobian matrix is very sparse. Its non-zero values are given by differentiating (26) and (27):

1. if  $m > (L - s)\beta + 1$  then,

- $\frac{\partial[\mathcal{G}_{NLFD}(\theta)]_{k,l,m}}{\partial A_{k,r}} = B_{l,r}(C_{m,r} - 1)S_{k,l,m}^1 e^{-S_{k,l}^3}$
- $\frac{\partial[\mathcal{G}_{NLFD}(\theta)]_{k,l,m}}{\partial B_{l,r}} = A_{k,r}(C_{m,r} - 1)S_{klm}^1 e^{-S_{k,l}^3}$

- $\frac{\partial[\mathcal{G}_{NLFD}(\theta)]_{k,l,m}}{\partial C_{k,m}} = A_{k,r} B_{l,r} e^{-S_{k,l}^3}$

2. Otherwise,

- $\frac{\partial[\mathcal{G}_{NLFD}(\theta)]_{k,l,m}}{\partial A_{k,r}} = (B_{l,r} C_{m,r} - (B_{l,r} + \alpha_{m,r} B_{l_1,r} + (1 - \alpha_{m,r}) B_{l_2,r}) S_{k,l,m}^1) e^{-S_{k,l,m}^2}$
- $\frac{\partial[\mathcal{G}_{NLFD}(\theta)]_{k,l,m}}{\partial B_{l,r}} = (A_{k,r} C_{m,r} - A_{k,r} (1 + \delta_{l,l_1} \alpha_{m,r} + \delta_{l,l_2} (1 - \alpha_{m,r})) S_{k,l,m}^1) e^{-S_{k,l,m}^2}$
- $\frac{\partial[\mathcal{G}_{NLFD}(\theta)]_{k,l,m}}{\partial B_{l_1,r}} = (\delta_{l,l_1} - 1) A_{k,r} \alpha_{m,r} S_{k,l,m}^1 e^{-S_{k,l,m}^2}$
- $\frac{\partial[\mathcal{G}_{NLFD}(\theta)]_{k,l,m}}{\partial B_{l_2,r}} = (\delta_{l,l_2} - 1) A_{k,r} (1 - \alpha_{m,r}) S_{k,l,m}^1 e^{-S_{k,l,m}^2}$
- $\frac{\partial[\mathcal{G}_{NLFD}(\theta)]_{klm}}{\partial C_{m,r}} = A_{k,r} B_{l,r} e^{-S_{k,l,m}^2}$

where:

$$S_{k,l,m}^1 = \sum_{r=1}^R A_{k,r} B_{l,r} C_{m,r}$$

$$S_{k,l,m}^2 = \sum_{r=1}^R A_{k,r} (B_{l,r} + \alpha_{m,r} B_{l_1,r} + (1 - \alpha_{m,r}) B_{l_2,r})$$

$$S_{k,l}^3 = \sum_{r=1}^R A_{k,r} B_{l,r}$$

Then these values have to be correctly placed in the matrix. The following matlab function computes these values and forms the Jacobian matrix at the point

$$\theta = \begin{bmatrix} \text{vec}(\mathbf{A}) \\ \text{vec}(\mathbf{B}) \\ \text{vec}(\mathbf{C}) \end{bmatrix}. \quad (\text{A.1})$$

```

function Jac=construct_jacobian(A,B,C,s,beta)
% INPUTS: (A,B,C) = factor matrices, s = shift, beta = downsampling factor
% OUTPUT: Jac = Jacobian matrix
[K,R]=size(A);
L=size(B,1);
M=size(C,1);
I=K*L*M;
J=(K+L+M)*R;
Jac= zeros(I,J);
for i=1:I
    m=floor((i-1)/(K*L))+1;
    x=i-(m-1)*K*L;
    l=floor((x-1)/K)+1;
    k=x-(l-1)*K;
    l1=floor((m-1)/beta)+s;
    l2=min(mp+1,L);
    alpha=(beta+beta*floor((m-1)/beta)+1-m)/beta;
    S1=sum(A(k,:).*B(l,:).*C(m,:));
    if m<=(L-s)*beta+1
        S2=sum(A(k,:).* (B(l,:)+alpha*B(l1,:)+(1-alpha)*B(l2,:)));
    for r=1:R

```

```

        Jac(i, (k-1)*R+r) = (B(l, r) * C(m, r) - (B(l, r) + alpha * B(l1, r) ...
        + (1-alpha) * B(l2, r)) * S1) * exp(-S2);
        Jac(i, K*R + (l-1) * R + r) = (A(k, r) * C(m, r) - A(k, r) * (1 + (l == (l1)) * alpha ...
        + (l == (l2)) * (1-alpha)) * S1) * exp(-S2);
        Jac(i, K*R + (l1-1) * R + r) = -(l == (l1)) * A(k, r) * S1 * alpha * exp(-S2);
        Jac(i, K*R + (l2-1) * R + r) = -(l == (l2)) * A(k, r) * S1 * (1-alpha) * exp(-S2);
        Jac(i, (K+L) * R + (m-1) * R + r) = (A(k, r) * B(l, r)) * exp(-S2);
    end
else
    S3 = sum(A(k, :) .* B(l, :));
    for r = 1 : R
        Jac(i, (k-1) * R + r) = (B(l, r) * C(m, r) - (B(l, r)) * S1) * exp(-S3);
        Jac(i, K*R + (l-1) * R + r) = (A(k, r) * C(m, r) - A(k, r) * S1) * exp(-S3);
        Jac(i, (K+L) * R + (m-1) * R + r) = (A(k, r) * B(l, r)) * exp(-S3);
    end
end
end
end

```

## References

- [1] J. Lakowicz, Principles of Fluorescence Spectroscopy, New York, Plenum Press, 1983.
- [2] F. L. Hitchcock, Multiple invariants and generalized rank of a p-way matrix or tensor, J. Math. and Phys. 7 (1927) 3979.
- [3] J. Carroll, J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition, Psychometrika (1970) 267–283.
- [4] R. Harshman, Foundations of the parafac procedure: Model and conditions for an "explanatory" multi-mode factor analysis", UCLA Working Papers in phonetics 16 (1970) 1–84.
- [5] R. Bro, PARAFAC. tutorial and applications, Chemometrics and intelligent laboratory systems 38 (1997) 149–171.
- [6] E. Sikorska, T. Gorecki, I. Khmelinskii, M. Sikorski, D. de Keukeleire, Monitoring beer during storage by fluorescence spectroscopy, Food Chemistry 96 (2006) 632–639.
- [7] C. Stedmon, R. Bro, Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial, Limnology and Oceanography 6 (2008) 572–579.
- [8] X. Luciani, S. Mounier, H. Paraquetti, R. Redon, Y. Lucas, A. Bois, L. Lacerda, M. Raynaud, M. Ripert, Tracing of dissolved organic matter from the sepetiba bay (brazil) by parafac analysis of total luminescence matrice, Marine Environmental Research 65 (2008) 148–157.
- [9] C. Parker, W. Barnes, Some experiments with spectrofluorimeters and filter fluorimeters, The analyst 82 (1957) 606–618.
- [10] J. F. Holland, R. E. Teets, P. M. Kelly, A. Timnick, Correction of right-angle fluorescence measurements for the absorption of excitation radiation, Analytical chemistry 49 (1977) 706–710.
- [11] B. Valeur, Molecular Fluorescence. Principles and Applications, Wiley-VCH, Weinheim, 2002.
- [12] F. Guimet, J. Ferr, R. Boqu, F. X. Rius, Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy, Analytica Chimica Acta 515 (2004) 75–85.
- [13] A. Rinnan, C. M. Andersen, Handling of first-order Rayleigh scatter in parafac modelling of fluorescence excitation-emission data, Chemometrics and intelligent laboratory systems 76 (2005) 91–99.
- [14] X. Luciani, S. Mounier, R. Redon, A. Bois, A simple correction method of inner filter effects affecting feem and its application to the parafac decomposition, Chemometrics and Intelligent Laboratory Systems 96 (2009) 227–238.
- [15] X. Luciani, R. Redon, S. Mounier, How to correct inner filter effects altering 3d fluorescence spectra by using a mirrored cell, Chemometrics and Intelligent Laboratory Systems 126 (2013) 91 – 99.
- [16] B. C. MacDonald, S. J. Lvin, H. Patterson, Correction of fluorescence inner filter effects and the partitioning of pyrene to dissolved organic carbon, Analytica chimica acta 338 (1997) 155–162.
- [17] L.-H. Lim, P. Comon, Nonnegative approximations of nonnegative tensors, Journal of chemometrics (2009) 432–441.
- [18] P. Paatero, A weighted non-negative least squares algorithm for three-way 'parafac' factor analysis, Chemometrics and Intelligent Laboratory Systems 38 (1997) 223–242.
- [19] N. T. J.P. Royer, P. Comon, Computing the polyadic decomposition of nonnegative third order tensors, Elsevier, Signal Processing 91 (2011) 2159–2171.

- [20] Y. Qi, P. Comon, L. H. Lim, Uniqueness of non-negative tensor approximations, submitted (2014). ArXiv:1410.8129.
- [21] J. Cohen, P. Comon, On almost sure identifiability of non multilinear tensor decomposition, in: Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 22nd European, IEEE, pp. 2245–2249.
- [22] P. Comon, J. M. F. T. Berge, L. DeLathauwer, J. Castaing, Generic and typical ranks of multi-way arrays, *Linear Algebra Appl.* 430 (2009) 2997–3007.
- [23] P. Comon, X. Luciani, A. L. F. de Almeida, Tensor decompositions, alternating least squares and other tales, *Journal of Chemometrics* 23 (2009) 393–405.
- [24] L. Chiantini, G. Ottaviani, On generic identifiability of 3-tensors of small rank, *SIAM J. Matrix Ana. Appl.* 33 (2012) 1018–1037.
- [25] K. Madsen, H. B. Nielsen, O. Tingleff, *Methods for non-linear least squares problems* (2nd ed.), 2004.
- [26] J. E. Cohen, R. Cabral-Farias, P. Comon, Fast decomposition of large nonnegative tensors, *IEEE Sig. Proc. Letters* 22 (2015) 862–866. Published online in Nov. 2014.
- [27] R. G. Zepp, W. M. Sheldon, M. A. Moran, Dissolved organic fluorophores in southeastern us coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation-emission matrices, *Marine chemistry* 89 (2003) 15–36.
- [28] C. A. Andersson, R. Bro, The N-way toolbox for Matlab, *Chemo. Intel. Lab. Syst.* 52 (2000) 1–4.
- [29] R. Bro, N-way toolbox, 2000. [www.models.life.ku.dk/nwaytoolbox](http://www.models.life.ku.dk/nwaytoolbox).
- [30] X. Luciani, L. Albera, Canonical polyadic decomposition based on joint eigenvalue decomposition, *Chemometrics and Intelligent Laboratory Systems* 132 (2014) 152–167.