

## **Contrat doctoral au laboratoire LIS (Toulon)**

Décompositions tensorielles couplées pour le démélange de signaux multidimensionnels. Application en spectroscopie de fluorescence.

**Mots clés :** *Démélange de signaux, Décompositions tensorielles, Algèbre multilinéaire, Optimisation numérique, Modèles physiques, Chimiométrie, Spectroscopie de fluorescence*

### **Compétences attendues :**

Le candidat devra avoir de solides connaissances et une appétence pour une majorité des domaines suivants: Traitement du signal, méthodes numériques, programmation (Matlab de préférence), méthodes d'optimisation, séparation de sources, algèbre linéaire, optique, spectroscopie, instrumentation.

### **Personne(s) à contacter (avant le 19 juin):**

Xavier Luciani ([luciani@univ-tln.fr](mailto:luciani@univ-tln.fr)), Roland Redon ([roland.redon@mio.osupytheas.fr](mailto:roland.redon@mio.osupytheas.fr))

## **PHD offer at the LIS laboratory, University of Toulon, France**

Coupled tensor decompositions for the unmixing of multi-dimentional signals. Application in fluorescence spectroscopy

**Key words:** *Signal demixing, Tensor decompositions, Multilinear algebra, Numerical optimisation, Physical models, Chemometrics, Fluorine spectroscopy*

### **Expected Skills :**

The candidate should have sound knowledge of and an interest in most of the following areas: Signal processing, numerical methods, programming (preferably Matlab), optimisation methods, source separation, linear algebra,

Additional knowledge in optics, physics and spectroscopy will be appreciated.

### **For job application and details please contact (before 19th june2024)**

Xavier Luciani ([luciani@univ-tln.fr](mailto:luciani@univ-tln.fr)) or Roland Redon ([roland.redon@mio.osupytheas.fr](mailto:roland.redon@mio.osupytheas.fr))

(Please find the English description of the topic below the French)

## Description du sujet de recherche

Le démélange de signaux est un domaine important du traitement du signal et répond à un problème que l'on rencontre dans de nombreuses applications de la physique et des sciences de l'ingénieur, pour lequel les signaux observés sont vus comme des mélanges de signaux sources d'intérêt que l'on cherche à estimer. Dans le cas des signaux discrets qui nous intéressent ici, les décompositions tensorielles fournissent un cadre théorique naturel pour le démélange de signaux multidimensionnels (c'est-à-dire variant en fonction d'au moins trois variables, on parle également de signaux multivariés) ainsi qu'un ensemble d'outils algorithmiques [1]. Ce type d'approche a notamment été appliqué avec succès en spectroscopie, pour le démélange de signaux de fluorescence [2]. L'équipe SIIM du LIS, en collaboration avec l'équipe CEM du laboratoire MIO a une expertise reconnue dans le domaine depuis plusieurs années [3,4].

Nous appelons ici décomposition tensorielle toute représentation exacte ou approchée d'un tableau de nombres à plus de deux dimensions à l'aide d'un ensemble fini de vecteurs appelés facteurs de la décomposition. Si le modèle de décomposition est adapté (problème direct), alors le calcul effectif des facteurs permet de résoudre le problème inverse d'estimation de signaux sources à partir de leur mélange. Dans le cas multidimensionnel le plus simple, les signaux observés sont tridimensionnels et issus d'un même dispositif de mesure. Ils représentent l'évolution d'un même phénomène  $F(x_1, x_2, x_3)$  en fonction de 3 variables discrètes ( $x_1, x_2, x_3$ ) et sont modélisés comme des mélanges trilinéaires de signaux sources monodimensionnels. La décomposition tensorielle correspondante est la Décomposition Canonique Polyadique (DCP, aussi appelée PARAFAC) d'ordre 3. Elle admet une solution unique qui permet de donner une interprétation physique aux différents facteurs et de nombreux algorithmes permettent de la calculer [5,6]. Ce cas de figure est désormais bien maîtrisé, il est illustré par l'exemple de la figure 1 dans le cas de signaux de fluorescence trivariés (longueur d'onde d'excitation, longueur d'onde d'émission et concentration.)

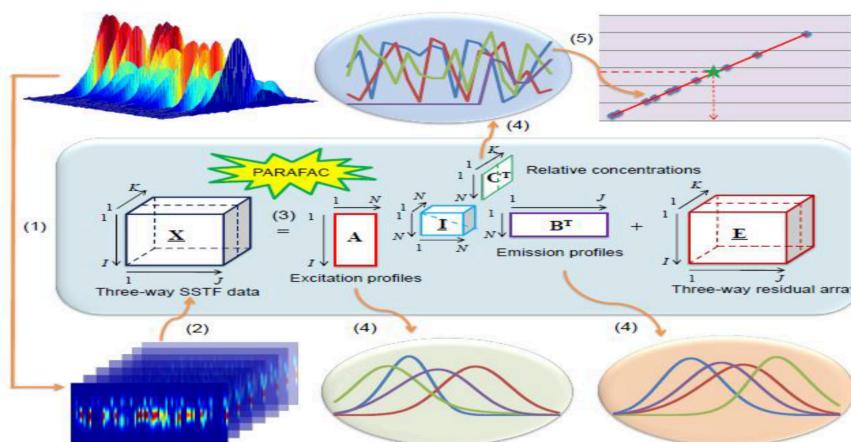


Figure 1: Décomposition CP (PARAFAC) de signaux de fluorescence

Cette thèse sera donc l'occasion 1) d'envisager des scénarios de mélange plus généraux et d'en déduire les modèles de décomposition tensorielle sous-jacents 2)

développer des algorithmes de calcul spécifiques à ces décompositions 3) les appliquer à des données réelles issues de la spectroscopie de fluorescence.

Nous considérerons ainsi l'analyse conjointe d'ensembles de signaux multivariés d'ordres quelconques  $F(x_1, \dots, x_p)$ ,  $F(y_1, \dots, y_q)$ ,  $F(z_1, \dots, z_r)$ ... regroupés dans différents tenseurs et pour lesquels l'une des variétés au moins est commune (justifiant ainsi le traitement conjoint). Le problème sera donc réécrit sous la forme de décompositions tensorielles couplées, permettant de prendre en compte cette information commune. Ce type d'approche s'est développé durant la dernière décennie, en particulier pour la fusion de données. Nous pourrons ainsi nous appuyer sur un état de l'art conséquent dont un exemple est donné dans [7]. Cependant, il nous sera probablement nécessaire d'introduire de nouveaux modèles de décompositions couplées répondant aux contraintes suivantes.

Tout d'abord les valeurs prises par les variables communes pourront être différentes d'un ensemble de signaux à un autre, par exemple du fait de l'utilisation de capteurs ayant des résolutions très différentes. A cette occasion, nous nous poserons également la question de savoir jusqu'à quel point un tenseur peut être sous-échantillonné suivant certaines de ses dimensions sans dégrader l'estimation des facteurs liés aux autres dimensions. Deuxièmement, deux cas de figures seront envisagés lors du calcul de la décomposition : avec ou sans connaissance a priori sur les facteurs. Dans le premier cas nous étudierons la possibilité d'injecter pour certains facteurs des modèles paramétriques pertinents dans un cadre applicatif donné. Par ailleurs, jusqu'ici, nous avons considéré implicitement que les signaux sources présents dans le modèle de mélange sont identiques en forme pour tous les mélanges. Cette hypothèse, très classique, n'est pas toujours réalisée en pratique. Nous serons donc amenés à proposer des modèles de décomposition tenant compte de déformations possibles des signaux sources d'un mélange à l'autre sans que ceux-ci soient identifiés comme résultats de sources différentes lors de la résolution du problème inverse. Ces signaux sources sont supposés unimodaux ou bien bimodaux. Une première transformation des signaux sources consiste à leur appliquer une translation sans modification de forme. Cette approche modélisée par Harshman, dont la résolution numérique (PARAFAC2) a été proposée par Kiers [8] constituera une des méthodes de référence dans ce contexte. La deuxième transformation possible des signaux sources sera la dilatation autour du maximum de leur profil, dilatation qui sera fonction de chaque signal source. Enfin, il est envisagé une troisième transformation qui consisterait à déformer chaque signal source en lui appliquant un profil multiplicatif proche de un, profil qu'il serait donc nécessaire d'optimiser en plus des amplitudes de translation et des facteurs de dilatation liés aux deux transformations précédentes. Enfin, au-delà des modèles pré-existants ou non, nous proposerons des solutions algorithmiques originales pour le calcul effectif des décompositions. Nous privilierons ainsi les approches de résolution semi-algébriques, basées sur la diagonalisation conjointe de matrices [6,9].

L'application considérée est la spectroscopie de fluorescence. Cette technique est employée depuis plusieurs années au laboratoire MIO pour caractériser la matière organique dissoute dans les eaux de mer et de rivière. Celle-ci consiste à mesurer l'intensité de fluorescence d'un échantillon liquide en fonction des longueurs d'onde d'excitation et d'émission. Dans le cas particulier des molécules fluorescentes en milieu liquide, le spectre d'émission de fluorescence est décalé vers des longueurs d'onde plus élevées que les longueurs d'onde du spectre d'excitation et la forme du spectre d'émission ne dépend pas

de la longueur d'onde en excitation. De plus, si l'absorbance de l'échantillon est suffisamment faible, le modèle physique de l'intensité de fluorescence devient trilinéaire : il dépend alors uniquement des spectres d'excitation, des spectres d'émission et des concentrations liées aux molécules fluorescentes (fluorophores) ainsi que d'autres paramètres physiques et instrumentaux. Le laboratoire MIO dispose d'un spectrofluorimètre Fluoromax (Horiba) permettant de mesurer l'intensité de fluorescence en mode stationnaire (source d'excitation continue) ou bien résolue en temps (source d'excitation impulsionnelle).

A l'aide de ce système de mesure, il sera donc possible d'obtenir des intensités de fluorescence  $F$  d'un échantillon  $k$  en fonction de plusieurs variables : les longueurs d'onde d'excitation et d'émission  $\lambda_{ex}$  et  $\lambda_{em}$  et le temps  $t$  de décroissance de la fluorescence. A ces premières variables liées au système de mesure, on pourra aussi ajouter les variables physico-chimiques suivantes : la concentration en métal ajouté à l'échantillon (notée  $C_M$ ) qui provoque une atténuation de l'intensité de fluorescence (appelé quenching), et le pH (potentiel hydrogène) qui peut engendrer un décalage en longueurs d'onde des spectres d'excitation et d'émission. Les différentes combinaisons de ces variables permettront de constituer des tenseurs d'ordre 3, 4 ou 5 ayant des variétés communes par exemple:

Tenseurs d'ordre 3	Tenseurs d'ordre 4	Tenseur d'ordre 5
$F(\lambda_{ex}, \lambda_{em}, k)$ ,	$F(\lambda_{ex}, \lambda_{em}, t, k)$ , $F(\lambda_{ex}, \lambda_{em}, t, C_M)$ ,	$F(\lambda_{ex}, \lambda_{em}, C_M, pH, k)$
$F(\lambda_{ex}, \lambda_{em}, t)$ ,	$F(\lambda_{ex}, \lambda_{em}, C_M, k)$ , $F(\lambda_{ex}, \lambda_{em}, t, pH)$ ,	
$F(\lambda_{ex}, \lambda_{em}, C_M)$	$F(\lambda_{ex}, \lambda_{em}, pH, k)$	

Ces tenseurs nous serviront ainsi à valider les différents modèles et algorithmes de décomposition proposés dans la partie méthodologique de la thèse sur des signaux réels.

### Références :

- [1] P. Comon, "Tensors: a brief introduction", IEEE Sig. Proc. Magazine, 31(3):44–53, May 2014. special issue on BSS.
- [2] R. Bro, PARAFAC. Tutorial and applications, Chemometrics and Intelligent Laboratory Systems, Volume 38, Issue 2, 1997, Pages 149-171,
- [3] X. Luciani, R. Redon, S. Mounier et A. Bois, "A simple correction method of inner filter effects affecting FEEM and its application to the PARAFAC decomposition" Chemometrics and Intelligent Laboratory Systems 96.2 (2009)
- [4] I. W. Sanou, R. Redon, X. Luciani, S. Mounier, "Online Nonnegative and Sparse Canonical Polyadic Decomposition of Fluorescence Tensors", Chemometrics and Intelligent Laboratory Systems 225 (2022)
- [5] P. Comon, X. Luciani et A. L. F. De Almeida. "Tensor Decompositions, Alternating, Least Squares and other Tales", Journal of Chemometrics 23, 2009.
- [6] X. Luciani et L. Albera. "Canonical polyadic decomposition based on joint eigenvalue decomposition", Chemometrics and Intelligent Laboratory Systems 132, 2014,
- [7] M. Sørensen, I. Domanov, L. De Lathauwer, Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- (lr,n, lr,n, 1) terms – part II: Algorithms, SIAM J. Matrix Anal. Appl., vol. 36, no. 3, 2015
- [8] H. Kiers, J. ten Berge, R. Bro, PARAFAC2 Part I. A direct fitting algorithm for the PARAFAC2 model, Special Issue:Proceedings of the International Chemometrics Research Meeting, 1998
- [9] R. André, X. Luciani et E. Moreau. "Joint EigenValue Decomposition Algorithms

## Subject description in English

Signal unmixing is an important area of signal processing and addresses a problem encountered in many applications of physics and engineering sciences, in which the observed signals are seen as mixtures of source signals of interest that we are trying to estimate. In the case of discrete signals, tensor decompositions provide a natural theoretical framework for the unmixing of multi-dimensional signals (i.e. those varying as a function of at least three variables, also known as multivariate signals) as well as a set of algorithmic tools [1]. This type of approach has been successfully applied to fluorescence spectroscopy, for unmixing fluorescence signals [2]. The LIS lab, in collaboration with the MIO laboratory, has had a recognised expertise in this field for several years [3,4].

Here we refer to tensor decomposition as any exact or approximate representation of a multi-dimensional array of numbers using a finite set of vectors. These vectors are called factors of the decomposition. If the decomposition model is correct (direct problem), then the actual calculation of the factors enables us to solve the inverse problem of estimating source signals from their mixture. In the simplest multi-dimensional case, the signals observed are three-dimensional and come from the same measurement device. They represent the evolution of the same phenomenon with respect to three discrete variables ( $x_1, x_2, x_3$ ) and are represented as a trilinear mixture of the decomposition factors. The corresponding tensor decomposition is the Canonical Polyadic Decomposition (CPD, also known as PARAFAC) of order 3. It admits a unique solution which makes it possible to give a physical interpretation to the various factors. In addition, many algorithms were proposed for calculating it [5,6]. This case is now well understood, and is illustrated by the example in figure 1 for trivariate fluorescence signals (excitation wavelength, emission wavelength and concentration).

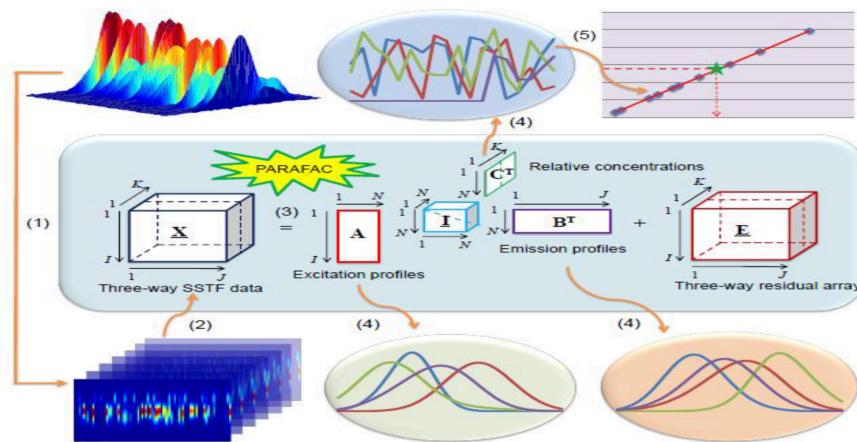


Figure 1: CPD (PARAFAC) of fluorescence signals

This thesis will therefore provide an opportunity to i.) consider more general mixing scenarios

and deduce the underlying tensor decomposition models ii.) develop calculation algorithms specific to these decompositions iii.) apply them to real data from fluorescence spectroscopy. We will thus consider the joint analysis of sets of multivariate signals of any order  $F(x_1, \dots, x_P)$   $F(y_1, \dots, y_Q)$   $F(z_1, \dots, z_R)$  grouped in several tensors for which at least one of the varieties is common (thus justifying the joint treatment). The problem will therefore be rewritten in the form of coupled tensor decompositions, allowing this common information to be taken into account. This type of approach has been developed over the last decade, particularly for data fusion. We can thus draw on a substantial state of the art, an example of which is given in [7]. However, it will probably be necessary to introduce new coupled decomposition models that meet the following constraints.

First of all, the values taken by the common variables may be different from one set of signals to another, for example due to the use of sensors with very different resolutions. On this occasion, we will also ask ourselves the question how far a tensor can be downsampled in some dimensions without affecting the factor estimation in the other dimensions. Secondly, two scenarios will be considered when calculating the decomposition: with or without a priori knowledge of the factors. In the first case, we will study the possibility of injecting parametric models for certain factors that are relevant to a given application. Furthermore, until now, we have implicitly considered that the source signals present in the mixture model are identical in form for all mixtures. This is a very conventional assumption, but it is not always true in practice. We will therefore have to propose decomposition models that take account of possible distortions in the source signals from one mixture to another without these being identified as resulting from different sources when the inverse problem is solved. These source signals are assumed to be unimodal or bimodal. A first transformation of the source signals consists of applying a translation to them without modifying their shape. This approach, introduced by Harshman, whose numerical resolution (PARAFAC2) was proposed by Kiers [8], will be one of the reference methods in this context. The second possible transformation of the source signals will be the dilation around the maximum of their profile, a dilation that will be a function of each source signal. Finally, a third transformation is envisaged, which would consist of deforming each source signal by applying a multiplicative profile close to one, a profile that would therefore need to be optimized in addition to the translation amplitudes and dilation factors associated with the two previous transformations. Finally, in addition to pre-existing or non-existing models, we will propose original algorithmic solutions for the actual calculation of decompositions. We will thus favor semi-algebraic resolution approaches, based on the joint diagonalization of matrices [6,9].

The application under consideration is fluorescence spectroscopy. This technique has been used for several years in the MIO laboratory to characterize dissolved organic matter in sea and river water. It involves measuring the fluorescence intensity of a liquid sample as a function of excitation and emission wavelengths. In the particular case of fluorescent molecules in a liquid medium, the fluorescence emission spectrum is shifted towards wavelengths higher than the wavelengths of the excitation spectrum and the shape of the emission spectrum does not depend on the excitation wavelength. Furthermore, if the absorbance of the sample is sufficiently low, the physical model of fluorescence intensity becomes trilinear: it then depends solely on the excitation spectra, the emission spectra and the concentrations associated with the fluorescent molecules (fluorophores), as well as other physical and instrumental parameters. The MIO laboratory has a Fluoromax spectrofluorometer (Horiba) for measuring fluorescence intensity in stationary mode (continuous excitation source) or time-resolved mode (pulsed excitation source).

Using this measurement system, it will therefore be possible to obtain fluorescence intensities  $F$  from a sample  $k$  as a function of several variables: the excitation and emission wavelengths  $\lambda_{ex}, \lambda_{em}$  and the time  $t$  for the fluorescence to decay. In addition to these initial variables linked to the measurement system, the following physico-chemical variables can also be added: the concentration of metal added to the sample (noted  $C_M$ ), which causes an attenuation of the fluorescence intensity (known as quenching), and the pH (hydrogen potential), which can cause a shift in the wavelengths of the excitation and emission spectra. The different combinations of these variables will make it possible to form tensors of order 3, 4 or 5 with common varieties, for example:

Tensors of order 3	Tensors of order 4	Tensors of order 5
$F(\lambda_{ex}, \lambda_{em}, k)$ ,	$F(\lambda_{ex}, \lambda_{em}, t, k)$ , $F(\lambda_{ex}, \lambda_{em}, t, C_M)$ ,	$F(\lambda_{ex}, \lambda_{em}, C_M, pH, k)$
$F(\lambda_{ex}, \lambda_{em}, t)$ ,	$F(\lambda_{ex}, \lambda_{em}, C_M, k)$ , $F(\lambda_{ex}, \lambda_{em}, t, pH)$ ,	
$F(\lambda_{ex}, \lambda_{em}, C_M)$	$F(\lambda_{ex}, \lambda_{em}, pH, k)$	

These tensors will thus be used to validate on real signals the various models and decomposition algorithms proposed in the methodological part of the thesis .

### Références :

- [1] P. Comon, "Tensors: a brief introduction", IEEE Sig. Proc. Magazine, 31(3):44–53, May 2014. special issue on BSS.
- [2] R. Bro, PARAFAC. Tutorial and applications, Chemometrics and Intelligent Laboratory Systems, Volume 38, Issue 2, 1997, Pages 149-171,
- [3] X. Luciani, R. Redon, S. Mounier et A. Bois, "A simple correction method of inner filter effects affecting FEEM and its application to the PARAFAC decomposition" Chemometrics and Intelligent Laboratory Systems 96.2 (2009)
- [4] I. W. Sanou, R. Redon, X. Luciani, S. Mounier, "Online Nonnegative and Sparse Canonical Polyadic Decomposition of Fluorescence Tensors", Chemometrics and Intelligent Laboratory Systems 225 (2022)
- [5] P. Comon, X. Luciani, A. L. F. De Almeida. "Tensor Decompositions, Alternating, Least Squares and other Tales", Journal of Chemometrics 23, 2009.
- [6] X. Luciani et L. Albera. "Canonical polyadic decomposition based on joint eigenvalue decomposition", Chemometrics and Intelligent Laboratory Systems 132, 2014,
- [7] M. Sørensen, I. Domanov, L. De Lathauwer, Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- (lr,n, lr,n, 1) terms – part II: Algorithms, SIAM J. Matrix Anal. Appl., vol. 36, no. 3, 2015
- [8] H. Kiers, J. ten Berge, R. Bro, PARAFAC2 Part I. A direct fitting algorithm for the PARAFAC2 model, Special Issue:Proceedings of the International Chemometrics Research Meeting, 1998
- [9] R. André, X. Luciani, E. Moreau. "Joint EigenValue Decomposition Algorithms Based on First-Order Taylor Expansion". In : IEEE Transactions on Signal Processing, 68 (2020)