

Analyse de grandes données tensorielles couplées

Jérémy E. COHEN, Rodrigo CABRAL FARIAS, Pierre COMON *

GIPSA-Lab, 11 rue des mathématiques, 38000 Grenoble, France

prenom.nom@gipsa-lab.fr

Résumé – Un des sujets phare du traitement du signal moderne est l’analyse de données tensorielles. Les systèmes de mesure actuels permettent en effet l’acquisition de multiples variables d’une même observation, et la multilinéarité intrinsèque de l’algèbre tensorielle en fait un outil de choix pour traiter ce type de données multivariées. Conjointement, la quantité de données à traiter s’est accrue de façon exponentielle. Nous proposons de traiter plusieurs grands jeux de données dont certains observables sont identiques ou proches. Une méthode de compression simultanée est proposée, ainsi qu’un traitement joint des données compressées. Les algorithmes sont présentés sur des données synthétiques.

Abstract – Tensor data analysis is presently one of the hot topics in signal processing. New measurement tools allow huge amount of multivariate complementary data to be collected on the same physical phenomenon, a property very well exploited by tensor algebra. In this paper, we treat the big data problem when tensorial data are coupled along one factor. A joint compression scheme and new coupled decomposition algorithms are introduced and demonstrated on synthetic data.

1 Introduction

Malgré leur intérêt, les outils tensoriels sont encore relativement peu utilisés en traitement du signal. Pourtant, même si de nombreux aspects sont encore mal compris, leur emploi ne requiert que peu de connaissances exotiques, la plupart des outils utilisés provenant de l’algèbre linéaire.

Le premier intérêt des tenseurs réside dans le fait que les phénomènes étudiés sont interprétables comme fonctions de plusieurs variables (au moins trois) de natures différentes. On peut par exemple citer les domaines d’application suivants : la chimiométrie, les réseaux d’antennes, la sociologie, l’imagerie médicale, les statistiques d’ordre supérieur, mais également d’autres domaines comme l’imagerie hyperspectrale ou la mécanique quantique [1, 2, 3, 4, 5]. Dans tous ces domaines, les données sont souvent riches de multiples diversités. Par exemple, dans le cas de données hyperspectrales, on dispose de pixels, à diverses longueurs d’ondes, et potentiellement dans le temps ; l’intensité est donc fonction de trois variables : l’espace, le temps, et la fréquence [5]. Il est bien sûr possible de traiter ces données avec des méthodes usuelles : Analyse en Composantes Principales, régression linéaire, factorisation matricielle ou autres (par exemple, en traitant image par image une vidéo hyperspectrale). Mais alors on perd l’information propre à l’intrication de chacune des variables par rapport aux autres.

Outre leur pertinence et ambivalence, la seconde raison qui fait des tenseurs un outil très puissant est l’existence de résultats théoriques physiquement porteurs de sens, provenant de l’unicité de leur décomposition. Par exemple, si l’on voit un tenseur comme une hypermatrice ayant plus de deux indices, alors on peut être tenté de généraliser la décomposition en valeurs singulières (SVD) bien connue pour les matrices. Une généralisation possible et très usitée est la Décomposition Ca-

nonique Polyadique (CPD) imaginée par Hitchcock en 1927. Contrairement à l’ACP, aucune contrainte (arbitraire) d’orthogonalité n’est nécessaire pour assurer génériquement son unicité [6]. La CPD est décrite dans la section 2.

Parmi les enjeux principaux de l’analyse de données tensorielles, deux thématiques sont récurrentes : grandes dimensions et couplage. En effet, avec l’augmentation de la mémoire des ordinateurs et de la précision des capteurs, la taille des données à traiter a explosé dans la dernière décennie : on est donc confronté au problème des grandes dimensions (Big Data) et de la complexité numérique afférente. D’un autre côté, il est fréquent de mesurer les mêmes phénomènes sous différentes modalités, et d’obtenir des jeux de données complémentaires desquels on cherche à extraire une information commune, qui peut s’interpréter comme le couplage des modèles décrivant les données. Un exemple standard est celui de l’EEG et de la MEG, qui tentent toutes les deux de remonter aux signaux émis par différentes régions du cerveau [7].

Pour traiter le problème des grandes dimensions, la compression de données tensorielle est une méthode standard [2]. Cependant, des travaux récents étendent cette compression basée sur de simples changements de base de représentation du tenseur pour des tenseurs contraints par la non négativité [8]. Mais aucun travail n’a encore été publié sur la compression de données tensorielles couplées ; le traitement de telles données en étant lui même à ses balbutiements [9]. Les cas de couplages traités sont simples, mais le sujet s’avère suffisamment complexe pour que l’on s’y attarde. Les couplages plus compliqués font l’objet de travaux en cours.

Après avoir rappelé quelques rudiments d’algèbre tensorielle, nous proposons dans cet article un cadre théorique pour la fouille et la fusion de grandes données multivariées, et des algorithmes pour résoudre les problèmes d’optimisation sous-jacents. Des résultats de simulations viennent confirmer la pertinence de l’approche proposée.

*. Les auteurs sont financés par le projet ERC “DECODA” no. 320594, dans le cadre du programme européen FP7/2007-2013.

2 Décompositions tensorielles

2.1 Définitions

Un tenseur \mathcal{T} est une application multilinéaire d'un produit cartésien de n_1 espaces vectoriels vers un autre produit de n_2 espaces vectoriels. Si l'on fixe des bases de représentation pour ces espaces de définition, \mathcal{T} peut être assimilé à un tableau de coordonnées à $n = n_1 + n_2$ entrées ; on dit alors que \mathcal{T} est d'ordre n . Une matrice est donc un tenseur d'ordre 2 [1].

De la même façon que pour les matrices, un des enjeux principaux du traitement de données tensorielles est d'identifier les bases de représentation du tenseur adaptées à l'information que l'on cherche en extraire. Pour ce faire, il existe deux généralisations aux tenseurs du concept de Décomposition en Valeurs Singulières (SVD), suivant que l'on veuille imposer au tenseur d'être diagonal dans la nouvelle base, ou aux matrices de changement de base d'être orthogonales. Rappelons que pour les matrices, ces deux opérations sont conjointement possibles ; mais cela s'avère impossible de façon générique pour les tenseurs d'ordre supérieur à 2.

Pour changer la base de représentation d'un tenseur, il suffit de changer les bases de chacun des espaces vectoriels qui le définissent. Ainsi pour un tenseur d'ordre 3, 3 matrices définissent ce changement et on utilisera la notation suivante :

$$\mathcal{T} = (\mathbf{U}, \mathbf{V}, \mathbf{W}) \mathcal{T}' \quad (1)$$

Dans le cas d'un tenseur \mathbf{M} d'ordre 2 avec des matrices \mathbf{U} et \mathbf{V} orthonormales bien choisies, on retrouve la SVD : $\mathbf{M} = (\mathbf{U}, \mathbf{V}) \boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$.

2.2 CPD

Si l'on souhaite que le tenseur s'écrive comme la transformation multilinéaire (par des matrices de passage) d'un tenseur diagonal, on aboutit au modèle CPD suivant :

$$\mathcal{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \boldsymbol{\Lambda} = \sum_{r=1}^R \lambda_r \mathbf{A}_{:,r} \otimes \mathbf{B}_{:,r} \otimes \mathbf{C}_{:,r} \quad (2)$$

où $\boldsymbol{\Lambda}$ est un tenseur diagonal. On note R le nombre de valeurs non nulles dans la diagonale de $\boldsymbol{\Lambda}$. R est appelé le rang du tenseur par analogie avec le rang d'une matrice. Les matrices \mathbf{A} , \mathbf{B} et \mathbf{C} sont parfois appelées facteurs de \mathcal{T} , et ne sont généralement pas orthogonales.

Le point important de la CPD est qu'elle tente d'expliquer les données contenues dans \mathcal{T} par un modèle trilinéaire ou chacune des diversités, ou modes, interagit avec les autres de façon multiplicative. Ces interactions sont décomposées en R termes qui caractérisent la complexité des interactions entre les modes. Les lois de la physique étant souvent linéaires en première approximation, ce modèle est très souvent utilisé directement pour retrouver les phénomènes ayant généré les données.

Sous des conditions suffisantes souvent vérifiées en pratique, cette décomposition est unique [6, 1]. Les données étant le plus souvent corrompues par du bruit, et le bruit ayant un rang beaucoup plus élevé que les données pures, on obtient un problème d'approximation d'un tenseur par une CPD de faible rang, qui est un problème mal posé, mais extricable en prenant les précautions nécessaires [10].

2.3 High Order Singular Value Decomposition

Si par contre on impose l'orthogonalité des matrices de passage en utilisant le minimum de vecteurs pour décrire les sous-espaces qui définissent \mathcal{T} , on obtient la HOSVD [11]. La HOSVD peut être obtenue avec une très bonne précision par l'algorithme ci-dessous (3) ; en appelant $\mathbf{T}_{(i)}$ la matrice obtenue en dépliant \mathcal{T} selon le $i^{\text{ème}}$ mode,

$$\begin{cases} \mathbf{U}\mathbf{N}_1 = \text{SVD}(\mathbf{T}_{(1)}) \\ \mathbf{V}\mathbf{N}_2 = \text{SVD}(\mathbf{T}_{(2)}) \\ \mathbf{W}\mathbf{N}_3 = \text{SVD}(\mathbf{T}_{(3)}) \end{cases} \quad \mathcal{T} = (\mathbf{U}, \mathbf{V}, \mathbf{W}) \mathcal{G} \quad (3)$$

Le tenseur obtenu \mathcal{G} est appelé le coeur, et les dimensions R_1 , R_2 et R_3 des colonnes de \mathbf{U} , \mathbf{V} et \mathbf{W} sont les rangs multilinéaires de \mathcal{T} . On vérifie aisément que les rangs multilinéaires sont plus petits que le rang du tenseur. Il s'avère que le coeur \mathcal{G} , en l'absence de bruit, contient des zéros pour les indices du tableau plus grand que R , ce qui justifie le recours à la SVD des trois versions dépliées du tenseur. Pour cette raison, il suffit de tronquer le nombre de colonnes des matrices de passage pour obtenir un coeur de dimensions plus petites que \mathcal{T} mais sans perte d'information.

Ainsi la HOSVD n'est souvent pas porteuse d'une interprétation physique comme peut l'être la CPD, mais elle est un outil très efficace de compression de données ou de débruitage. Sa fonction de compression va nous permettre de définir un cadre pour traiter de grandes données couplées.

3 Compression de tenseurs couplés

3.1 Compression pour la CPD

Soit \mathcal{T} un tenseur d'ordre 3 dans $\mathbb{R}^{K \times L \times M}$, considéré ici comme un tableau de dimensions $K \times L \times M$. Supposons qu'il soit de rang R faible devant ses dimensions, donc que les phénomènes sous-jacents aux données sont en réalité peu nombreux.

Un moyen simple de compresser \mathcal{T} est d'en effectuer la HOSVD, puis de tronquer les colonnes des matrices de passage dont la variance expliquée du coeur est négligeable. Si R est connu, on peut prendre ces trois nouvelles dimensions $R_1 \leq K$, $R_2 \leq L$ et $R_3 \leq M$ de l'ordre de R . Alors (3) devient une relation d'approximation, et \mathcal{G} vit dans $\mathbb{R}^{R_1 \times R_2 \times R_3}$.

L'idée est d'appliquer ensuite la CPD au coeur compressé \mathcal{G} pour obtenir des facteurs \mathbf{A}_c , \mathbf{B}_c and \mathbf{C}_c dans un espace compressé, avant de revenir aux vrais facteurs décompressés par une multiplication matricielle relativement peu coûteuse devant la résolution de la CPD pour le tenseur non compressé \mathcal{T} :

$$\mathcal{T} \approx (\mathbf{U}, \mathbf{V}, \mathbf{W}) \mathcal{G} = (\mathbf{U}\mathbf{A}_c, \mathbf{V}\mathbf{B}_c, \mathbf{W}\mathbf{C}_c) \boldsymbol{\Lambda} \quad (4)$$

La relation entre les facteurs compressés et non compressés est une relation classique de compression où les matrices de la HOSVD jouent le rôle de dictionnaires :

$$\mathbf{A}_c = \mathbf{U}^T \mathbf{A}, \quad \mathbf{B}_c = \mathbf{V}^T \mathbf{B}, \quad \mathbf{C}_c = \mathbf{W}^T \mathbf{C} \quad (5)$$

3.2 Couplage de deux tenseurs

Si l'on dispose maintenant de deux tenseurs (le cas de trois tenseurs ou plus sera traité dans un article long), et que ces tenseurs manifestent en partie la même réalité physique, alors un mode de couplage naturel est de considérer qu'ils ont un facteur similaire dans leur décomposition canonique. Dans ce papier, nous explorerons le cas simple où ces facteurs sont égaux, éventuellement similaires avec des perturbations gaussiennes indépendantes et identiquement distribuées (i.i.d.). Si l'on considère que \mathcal{T} et \mathcal{T}' sont nos deux jeux de données tensorielles mesurés avec du bruit additif gaussien centré, nous avons le modèle suivant :

$$\begin{cases} \mathcal{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \Lambda + \mathcal{E} \\ \mathcal{T}' = (\mathbf{A}', \mathbf{B}', \mathbf{C}') \Lambda + \mathcal{E}' \\ \mathbf{C} = \mathbf{C}' + \Gamma, \Gamma_{ij} \sim \mathcal{N}(0, \sigma_c^2) \end{cases} \quad (6)$$

où \mathcal{E} et \mathcal{E}' sont les bruits de mesure indépendants avec variances σ_n et σ'_n . L'intensité du couplage est fonction de σ_c .

Si l'on applique la HOSVD à chacun des deux tenseurs séparément, le couplage devient donc :

$$\mathbf{W}\mathbf{C}_c = \mathbf{W}'\mathbf{C}'_c + \Gamma \quad (7)$$

Le problème d'optimisation serait alors de résoudre deux petites CPD sous la contrainte que le lien probabiliste (7) soit respecté ; chose tout à fait réalisable, mais dans le contexte présent, il serait dommage de compresser les deux tenseurs de façon indépendante alors que l'on connaît l'existence d'un couplage entre les facteurs. Le problème principal étant que l'estimation des matrices de passage dans la HOSVD peut être périlleuse si les tenseurs présentent un niveau de bruit important, la relation (7) se trouvant alors faussée.

3.3 Compression couplée

Pour réaliser l'estimation jointe des matrices de passage selon le mode couplé, il convient de remarquer en premier lieu que dans la HOSVD, on n'utilise que les vecteurs singuliers à gauche du tenseur déplié sur le mode couplé. Ainsi, n'importe quelle base de l'espace engendré par ces vecteurs singuliers à gauche convient pour effectuer la compression. Autrement dit, si on utilise $\mathbf{W}\mathbf{X}$ au lieu de \mathbf{W} , où \mathbf{X} est une matrice orthonormale, alors le résultat de la compression par HOSVD sera équivalent dans les deux cas.

De plus, si \mathbf{C} et \mathbf{C}' sont égaux, le sous espace des vecteurs singuliers à gauche des deux tenseurs dépliés selon le 3ème mode sont identiques, donc on peut chercher une matrice de passage commune aux deux tenseurs, qui sera simplement une rotation des bases données par \mathbf{W} et \mathbf{W}' lorsque les décompositions en valeurs singulières sont réalisées indépendamment. Concrètement, on peut chercher une matrice \mathbf{W}'' orthonormale telle que :

$$\left[\frac{\mathcal{T}_{(3)}}{\sigma_n}, \frac{\mathcal{T}'_{(3)}}{\sigma'_n} \right] = \mathbf{W}'' \Sigma'' \mathbf{Q}'' \quad (8)$$

soit la SVD des deux tenseurs dépliés et juxtaposés avec des poids.

Puisque la matrice de passage tronquée est ici identique dans le mode couplé pour les deux tenseurs, le lien de couplage dans

le domaine compressé se simplifie, et le problème d'optimisation dans le domaine compressé devient :

$$\begin{aligned} \text{minimiser} \quad & \Upsilon = (1/\sigma_n^2) \|\mathcal{G} - (\mathbf{A}_c, \mathbf{B}_c, \mathbf{C}_c) \Lambda\|_F^2 + \\ & + (1/\sigma'_n{}^2) \|\mathcal{G}' - (\mathbf{A}'_c, \mathbf{B}'_c, \mathbf{C}'_c) \Lambda'\|_F^2, \\ \text{par rapp. à} \quad & \mathbf{A}_c, \mathbf{B}_c, \mathbf{C}_c, \Lambda, \mathbf{A}'_c, \mathbf{B}'_c, \mathbf{C}'_c, \Lambda' \\ \text{s. contrainte} \quad & \mathbf{C}_c \approx \mathbf{C}'_c \\ \text{où} \quad & \mathcal{T} = (\mathbf{U}, \mathbf{V}, \mathbf{W}'') \mathcal{G} \text{ et } \mathcal{T}' = (\mathbf{U}', \mathbf{V}', \mathbf{W}'') \mathcal{G}' \end{aligned} \quad (9)$$

4 Algorithme de décomposition

4.1 ALS projeté

L'algorithme le plus couramment utilisé pour calculer la CPD est basé sur une résolution successive de problèmes aux moindres carrés qui apparaissent si l'on fixe tous les facteurs de la décomposition sauf un. On itère ensuite en fixant le facteur qui variait et en en faisant varier un autre, jusqu'à convergence, d'où le nom ALS (alternating least squares) de cet algorithme. La convergence de l'ALS peut être démontrée, comme présenté dans [12], soit en garantissant que pour toute trajectoire de l'algorithme, les facteurs sont donnés en résolvant des moindres carrés bien posés, soit par l'ajout d'un terme de régularisation.

Pour tenir compte de la contrainte de couplage, au moins trois solutions viennent assez naturellement à l'esprit. La première serait de remplacer \mathbf{C}' par \mathbf{C} dans (9) comme dans la vaste majorité des méthodes de décomposition couplées [9]. Mais on ne tient alors pas compte de l'existence d'une erreur éventuelle entre les deux facteurs.

Une autre solution est de faire intervenir la relation probabiliste de couplage comme une pénalisation dans Υ . On obtient alors une formation compatible avec le maximum *a posteriori*. Par souci de concision, la formulation d'un algorithme dans ce cas étant fastidieuse et le format plutôt concis, nous ne présenterons pas cette solution très efficace dans le présent article.

Enfin, il est possible de considérer une version projetée de l'ALS. A chaque itération, après résolution par moindres carrés du problème convexe pour un facteur, on projète les facteurs couplés \mathbf{C} et \mathbf{C}' sur le noyau d'une matrice $[\mathbf{W} \ \mathbf{W}']$. Mais puisque nous réalisons une compression jointe, on cherche la projection des facteurs couplés sur le noyau suivant :

$$[\mathbf{W}'', -\mathbf{W}''] \left[\hat{\mathbf{C}}^\top \ \hat{\mathbf{C}}'^\top \right]^\top = 0 \quad (10)$$

\mathbf{W}'' étant orthogonale à gauche (on a supprimé uniquement des colonnes dans la compression), la solution de la projection orthogonale est donnée par

$$\hat{\mathbf{C}} = \hat{\mathbf{C}}' := (\mathbf{C} + \mathbf{C}')/2 \quad (11)$$

Cependant, toute autre projection sur le sous espace défini par (10) satisfait la contrainte de couplage. D'un autre côté, il faut choisir la projection qui si possible minimise Υ . La relation de couplage étant par ailleurs faussée par l'erreur finale dans l'ALS et le bruit de couplage, il est préférable, si l'on dispose de cette information, d'introduire dans (11) les niveaux de bruit σ_n et σ'_n des deux tenseurs. On obtient alors

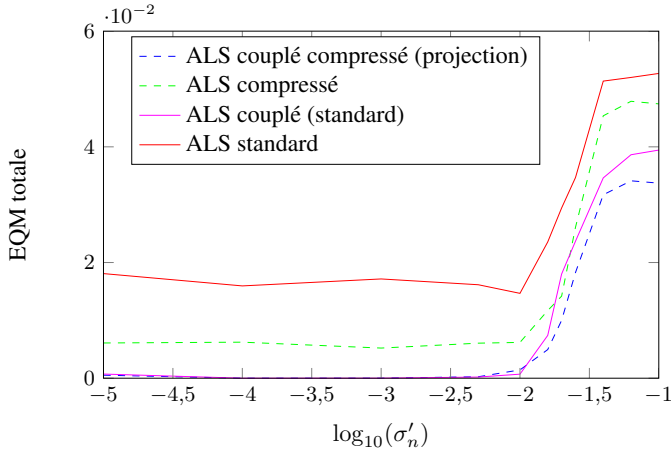


FIGURE 1 – Quatre algorithmes de décomposition appliqués sur 50 réalisations à deux tenseurs avec un couplage sans bruit sur le 3^{ème} mode. En abscisses, l'écart type du bruit sur le second tenseur en échelle logarithmique. En ordonnées, la somme des erreurs quadratiques moyennes (EQM) sur \mathbf{C} et \mathbf{C}' .

$$\hat{\mathbf{C}} = \hat{\mathbf{C}}' := \left(\sigma_n^2 \mathbf{C} + \sigma_n'^2 \mathbf{C}' \right) / \left(\sigma_n^2 + \sigma_n'^2 \right) \quad (12)$$

L'ALS projeté est certes sous-optimal, et moins pertinent en terme d'erreur sur les facteurs couplés qu'un ALS pénalisé qui inclut la connaissance de l'intensité du couplage. Mais il possède deux qualités qui en font un bon candidat pour la CPD couplée. Il est tout d'abord rapide car il travaille exclusivement dans le domaine compressé. De plus, il est possible de paralléliser les étapes de moindres carrés sur chacun des tenseurs, puis de voir la projection comme un consensus. Enfin au vu des nombreuses sources d'erreur sur les facteurs couplés en plus de l'erreur présente dans le modèle, il est naïf de croire que l'on puisse imposer la contrainte de couplage de façon optimale ; une projection ad-hoc est somme tout très satisfaisante.

4.2 Simulations sur données synthétiques

On tire les trois matrices facteurs des deux tenseurs chacune selon une loi normale i.i.d. centrée réduite, un des facteurs étant strictement commun aux deux tenseurs. Sur les deux tenseurs normalisés de tailles $50 \times 50 \times 50$, on va appliquer quatre algorithmes de décomposition CPD. Deux algorithmes sont dans le domaine compressé de taille $10 \times 10 \times 10$, la compression ayant été réalisé de façon conjointe, mais un seul des deux (le projeté) tient compte du couplage dans la CPD, son détail est donné dans le paragraphe précédent. Deux autres algorithmes ne compressent pas les données, et sont donc plus lents mais théoriquement plus précis. Un seul de ces deux algorithmes tient compte du couplage, en remplaçant simplement \mathbf{C}' par \mathbf{C} dans (6) ; c'est l'état de l'art en factorisation couplée [9]. L'erreur de reconstruction sur les deux facteurs couplés est donnée par la Figure 1, le bruit d'un des deux tenseurs étant fixé à $\sigma_n = 10^{-1.6}$.

Des résultats des simulations, il apparaît que le couplage est très important au moment de la compression ; on notera ainsi que l'erreur est plus faible sur le cas compressé sans couplage au moment de la CPD que sur le cas non compressé. Puisque la compression elle-même utilise une SVD jointe, le jeu de don-

nées le moins bruité permet en fait d'effectuer un débruitage sur le jeu le plus bruité, ce qui explique cette amélioration surprenante des performances, la compression étant sensé introduire une erreur supplémentaire. De plus, l'ALS projeté est plus efficace que l'ALS qui égalise les deux facteurs couplés dans l'espace décompressé. En plus du débruitage, deux causes peuvent être à l'origine de ce comportement : un nombre réduit de minima locaux de la fonction de coût dans l'espace compressé et une convergence finale plus rapide et moins sensible aux erreurs d'arrondi aussi pour l'algorithme compressé. Ces deux causes semblent être liées à la réduction de dimension du problème. D'autres simulations ont été effectuées en dimension plus grande (de 100 à 100000) et ont conduit à des constats de même nature.

5 Conclusion

Nous proposons de traiter des grandes données tensorielles où un des facteurs de la décomposition canonique est commun à deux jeux de données. Pour ce faire, nous compressons de façon jointe les deux tenseurs, puis nous effectuons une CPD en tenant compte de la proximité des facteurs compressés. Il est possible d'utiliser tous les résultats présentés ici dans les cas d'une relation de couplage bruitée et du couplage de plus de deux tenseurs. De plus, Nous étudions actuellement des interactions plus complexes entre les facteurs, d'autres algorithmes de décomposition, ainsi qu'une application en fouille de données.

Références

- [1] P. Comon. *Tensors : a brief survey*. IEEE Sig. Proc. Mag., 2014.
- [2] R. Bro. *PARAFAC. Tutorial and applications*. Chemometr. Intell. Lab., 1997.
- [3] N. D. Sidiropoulos, R. Bro, G. B. Giannakis (2000). *Parallel factor analysis in sensor array processing*. IEEE Trans. Sig. Proc., 2000.
- [4] L.Gauvin, A. Panisson, C. Cattuto. *Detecting the community structure and activity patterns of temporal networks : a non-negative tensor factorization approach*. PloS one, 2014.
- [5] M. A. Veganzones, J.E. Cohen, R. Cabral Farias, R. Marrero, J. Chanussot, P. Comon. *Multilinear spectral unmixing of hyperspectral multiangle images*. Soumis à European Signal Processing Conference (EUSIPCO 2015), 2015.
- [6] J. B. Kruskal. *Three-way arrays : rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*. Linear Algebra Appl., 1977.
- [7] H. Becker, L. Albera, P. Comon, M. Haardt, G. Birot, F. Wendling, I. Merlet. *EEG extended source localization : tensor-based vs. conventional methods*. NeuroImage, 2014.
- [8] J. E. Cohen, R. Cabral Farias, P. Comon. *Fast Decomposition of Large Nonnegative Tensors*. IEEE Sig. Proc. letters, 2015.
- [9] E. Acar, T. G. Kolda, D. M. Dunlavy, *All-at-once optimization for coupled matrix and tensor factorizations*. arXiv :1105.3422, 2011
- [10] V. De Silva, L.H.Lim *Tensor rank and the ill-posedness of the best low-rank approximation problem*. SIAM J. Matrix Anal. Appl., 2008
- [11] L. De Lathauwer, B. De Moor, J. Vandewalle. *A multilinear Singular Value Decomposition*. SIAM J. Matrix Anal. Appl., 2000.
- [12] A. Uschmajew *A new convergence proof for the high-order power method and generalizations* arXiv :1407.4586 2014