

Approche algébrique pour la décomposition tensorielle ParaTuck-2

Konstantin USEVICH

Centre de Recherche en Automatique de Nancy, CNRS, Université de Lorraine, 54000 Nancy, France

Résumé – Dans cet article, nous proposons un premier algorithme explicite (algébrique) pour la décomposition ParaTuck-2 des tenseurs d’ordre 3. Notre approche repose sur un plongement des tranches du tenseur dans un espace de dimension élevée, ce qui permet aussi d’améliorer les conditions d’unicité existantes. Notre méthode réduit la décomposition ParaTuck-2 au calcul des noyaux et des décompositions en valeurs propres des matrices. Nos algorithmes sont également applicables dans le cadre de l’approximation, comme le montrent les expériences numériques.

Abstract – In this paper, we propose a first explicit (algebraic) algorithm for the ParaTuck-2 tensor decomposition. We do so by lifting the slices of the tensor to higher-dimensional space, which also allows for refining the existing uniqueness conditions. Our method reduces ParaTuck-2 to computing nullspaces and eigenvalue decompositions of certain matrices. Our algorithms are also applicable in the approximation scenario, as shown by the numerical experiments.

1 Introduction

La décomposition tensorielle ParaTuck-2 (PT2D) est une version « multicouche » de la décomposition canonique polyadique (CPD) des tenseurs d’ordre 3. La PT2D [5] a été proposée par Harshman et Lundy en 1996; elle doit son nom au mélange des propriétés de la décompositions PARAFAC (l’ancien nom de la CPD) et Tucker. La PT2D a démontré son utilité dans plusieurs domaines tels que psychométrie et chimométrie [5, 1], communications sans fil MIMO [4, 3] et, plus récemment, compression des réseaux de neurones [2, 10]. La version symétrique de la PT2D (appelée également DE-DICOM [6]) est étroitement liée à PARAFAC-2 [6, 8], une autre généralisation de la CPD très courante dans l’analyse de données multivariées.

Comme la CPD, la PT2D possède une propriété très pertinente pour des applications : l’unicité. Il a été démontré par Harshman et Lundy [5] que, sous certaines conditions, les matrices facteurs de la PT2D sont uniques à l’échelle et permutations près. Dans le cas symétrique, les conditions de [5] fournissent les meilleures conditions d’unicité connues à ce jour pour PARAFAC-2 [8]. Pourtant il y n’a presque pas de nouveaux résultats théoriques pour la PT2D depuis [5].

Cependant, plusieurs faiblesses se révèlent dans la mise en œuvre de la PT2D. Des algorithmes classiques, tels que ALS (alternating least squares) [1, pp. 68-71], se heurtent à des problèmes de convergence et des minima locaux. Quelques améliorations ont été proposées [7],[11], mais elles manquent encore de preuves de convergence. Une hypothèse simplificatrice est souvent introduite dans des applications, selon laquelle au moins une des matrices facteurs est connue ([3], [1, p. 213]), [10]. Enfin, les résultats d’unicité de [5] ne fournissent pas d’un algorithme explicite pour PT2D, contrairement à la CPD (où il existe des algorithmes [6] qui permettent de calculer la CPD sous les conditions d’unicité de Kruskal).

Cet article vise à combler cette lacune : nous proposons des algorithmes algébriques (basés sur des opérations d’algèbre linéaire), qui permettent de retrouver les facteurs uniques de la PT2D sous les conditions de [5] légèrement améliorées. Les preuves et les détails sur l’implantation se trouvent dans [9].

Notations Les vecteurs, matrices, et tenseurs sont représentés en gras minuscule (\mathbf{u}), gras majuscule (\mathbf{M}) et lettres scriptes (\mathcal{T}). On utilise \boxtimes et \odot pour les produits de Kronecker et de Khatri-Rao. Le symbole \otimes est utilisé pour le produit tensoriel. L’opérateur $\text{vec}\{\cdot\}$ est utilisé pour la vectorisation d’une matrice ou d’un tenseur par colonnes. L’opération \bullet_p désigne la contraction sur mode p , par exemple, $(\mathcal{T} \bullet_p \mathbf{M})_{ijk} = \sum_\ell \mathcal{T}_{\ell jk} M_{i\ell}$. Le troisième dépliement de $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ est noté $\mathbf{T}^{(3)} \in \mathbb{R}^{K \times IJ}$. La notation $\mathcal{T} = \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket$ est utilisée pour la CPD ($\mathcal{T}_{ijk} = \sum_r A_{ir} B_{jr} C_{kr}$). On note par $\text{rank}\{\cdot\}$ et $\text{krank}\{\cdot\}$ le rang et le rang Kruskal d’une matrice.

Paratuck : la définition Soit $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ un tenseur¹ d’ordre 3, avec les tranches frontales qui s’écrivent comme :

$$\mathcal{T}_{::,k} = \mathbf{A} \mathbf{D}_G^{(k)} \mathbf{F} \mathbf{D}_H^{(k)} \mathbf{B}^T, \quad (1)$$

tels que les trois matrices facteurs $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times S}$, $\mathbf{F} \in \mathbb{R}^{R \times S}$ sont les mêmes pour toutes les tranches, où $\mathbf{D}_G^{(k)} \in \mathbb{R}^{R \times R}$ ainsi que $\mathbf{D}_H^{(k)} \in \mathbb{R}^{S \times S}$, $k = 1, \dots, K$ sont des matrices diagonales; on dit qu’un tel tenseur admet une décomposition ParaTuck-2 [5, 6]. Nous proposons une notation plus compacte que celle utilisée d’habitude [6]. Considérons les deux matrices $\mathbf{G} \in \mathbb{R}^{R \times K}$ et $\mathbf{H} \in \mathbb{R}^{S \times K}$ suivantes :

$$\mathbf{G} = [\mathbf{g}_1 \ \cdots \ \mathbf{g}_K], \quad \mathbf{H} = [\mathbf{h}_1 \ \cdots \ \mathbf{h}_K] \quad (2)$$

qui paramétrisent les matrices diagonales $\mathbf{D}_G^{(k)}$, $\mathbf{D}_H^{(k)}$ comme :

$$\mathbf{D}_G^{(k)} = \text{Diag}\{\mathbf{g}_k\}, \quad \mathbf{D}_H^{(k)} = \text{Diag}\{\mathbf{h}_k\}. \quad (3)$$

La notation $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H})$ signifie que \mathcal{T} admet une PT2D (1) (avec \mathbf{G}, \mathbf{H} comme dans (2)).

Quelles propriétés d’unicité pour PT2D? Comme pour la CPD, on peut permuter les colonnes de \mathbf{A} conjointement avec les lignes de \mathbf{F} et les éléments des $\mathbf{D}_G^{(k)}$, ce qui donne une PT2D différente du tenseur $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H})$:

$$\mathcal{T} = \text{PT2D}(\mathbf{A} \mathbf{\Pi}_A, \mathbf{B}, \mathbf{\Pi}_A^T \mathbf{F}, \mathbf{\Pi}_A^T \mathbf{G}, \mathbf{H}).$$

¹Tous les résultats s’étendent facilement au cas complexe.

pour chaque matrice de permutation Π_A . De même, on peut appliquer un changement d'échelle pour les colonnes de A et lignes de F , pour obtenir

$$\mathcal{T} = \text{PT2D}(A\Lambda_A, B, (\Lambda_A\Lambda_G)^{-1}F, \Lambda_G G, H),$$

pour toutes matrices Λ_A, Λ_G diagonales inversibles. On peut aussi permuter les colonnes de B et F conjointement avec les éléments des $D_H^{(k)}$ avec une permutation Π_B (indépendante de Π_A); on peut aussi changer leur échelle de la même façon. Enfin, on peut remplacer G et H par $G\Lambda_{GH}$ et $H\Lambda_{GH}^{-1}$ sans permuter les colonnes. (La description complète des ambiguïtés triviales se trouve dans [9] et aussi dans [5].)

Définition 1.1 La (R, S) -PT2D (1) est dite unique si toute autre (R, S) -PT2D provient des ambiguïtés triviales.

Harshman & Lundy [5] ont montré que la PT2D est unique sous des conditions suivantes.

Théorème 1.2 Soit $\mathcal{T} = \text{PT2D}(A, B, F, G, H)$ avec des facteurs qui satisfont les hypothèses suivantes :

- A1. $A \in \mathbb{R}^{I \times R}$, $B \in \mathbb{R}^{J \times S}$ sont de rang colonne plein.
- A2. F ne contient pas des zéros ($F_{ij} \neq 0, \forall i, j$);
- A3 : le rang de F est maximal ($= \min(R, S)$);
- A4 : les facteurs G, H ne contiennent pas des zéros;
- A5 : $\text{rank}\{G \odot G \odot H \odot H\} = \frac{R(R+1)}{2} \frac{S(S+1)}{2}$.

Alors la (R, S) -PT2D est unique.

Nous montrerons plus tard que les conditions A3–A4 peuvent être remplacées par une condition plus faible sur le rang de Kruskal de F . La condition la plus contraignante est A5, mais nous n'allons pas la relâcher dans cet article. Mentionnons juste que le rang de $\text{rank}\{G \odot G \odot H \odot H\}$ ne peut pas dépasser $\frac{R(R+1)}{2} \frac{S(S+1)}{2}$, et la condition A5 correspond à une notion de généricité des colonnes de G et H .

Dans le cas symétrique ($I = J, R = S, \mathcal{T} = \text{PT2D}(A, A, F, G, G), F$ symétrique), un résultat similaire a été démontré dans [5] sous A1–A4 et avec A5 remplacée par

$$\bullet \text{ A6 : } \text{rank}\{G \odot G \odot G \odot G\} = \frac{R(R+1)(R+2)(R+3)}{24};$$

cette nouvelle hypothèse est introduite parce que A5 n'est jamais satisfaite si $G = H$.

Tenseur cœur et produit triple Nous finissons cette section par une remarque sur la structure de tenseur cœur ParaTuck-2. La propriété suivante a été démontrée dans [4].

Lemme 1.3 Un tenseur admet une PT2D $\mathcal{T} = \text{PT2D}(A, B, F, G, H)$ si et seulement si

$$\mathcal{T} = \mathbf{C} \bullet_1 A \bullet_2 B \quad \text{avec } \mathbf{C} = \text{TP}(F, G, H);$$

ou la notation $\text{TP}(F, G, H) \in \mathbb{R}^{R \times S \times K}$ désigne le tenseur \mathbf{C} obtenu par la multiplication des éléments de trois matrices $F \in \mathbb{R}^{R \times S}$, $G \in \mathbb{R}^{R \times K}$, $H \in \mathbb{R}^{S \times K}$:

$$C_{ijk} = F_{ij}G_{ik}H_{jk} \quad \begin{array}{c} \text{C} \\ \text{[cube]} \end{array} = \begin{array}{c} \text{G} \\ \text{[cube]} \end{array} \begin{array}{c} \text{H} \\ \text{[cube]} \end{array} \begin{array}{c} \text{F} \\ \text{[cube]} \end{array}$$

Autrement dit, la PT2D est une décomposition Tucker-2 avec le tenseur cœur de structure très particulière. Une autre remarque est que les tranches de \mathbf{C} sont des produits Hadamard de F avec plusieurs matrices de rang 1 :

$$\mathbf{C}_{:, :, k} = (F \otimes \mathbf{g}_k \mathbf{h}_k^T).$$

2 Approche proposée

Par la suite, nous allons considérer uniquement le cas $I = R, J = S$. En effet, le cas $I > R, J > S$ peut être réduit au cas $I = R, J = S$ par une approche standard de compression Tucker avec rangs (R, S) (en utilisant lemme 1.3).

L'idée principale de la méthode proposée est très simple. Le tenseur cœur est un objet très structuré, et en plus est décomposé d'une façon unique.

Lemme 2.1 ([9, §4.4]) Soit F, G, H des matrices qui satisfont A2 et A5. Alors la décomposition du tenseur cœur \mathbf{C} est unique à facteurs d'échelle près et les facteurs peuvent être retrouvés par la résolution d'une suite des systèmes linéaires.

Lemme 2.1 permet de décomposer le problème en deux étapes : (a) trouver les matrices de « rotation » \tilde{A} et \tilde{B} telles que

$$\tilde{\mathbf{C}} = \mathcal{T} \bullet_1 \tilde{A} \bullet_2 \tilde{B}; \quad (4)$$

ait une structure de tenseur cœur, c'est à dire $\tilde{\mathbf{C}} = \text{TP}(\dots)$; (b) retrouver les autres facteurs (F, G, H) à partir du tenseur cœur $\tilde{\mathbf{C}}$ (grâce au lemme 2.1).

Si la PT2D est unique, ces matrices doivent être égales à $\tilde{A} = A^{-1}$ et $\tilde{B} = B^{-1}$. Nous montrerons par la suite que les matrices A, B peuvent être trouvées à l'aide d'une matrice structurée que l'on peut construire à partir du tenseur \mathcal{T} .

Matrice de lifting Pour retrouver les matrices A, B , nous allons plonger le tenseur dans l'espace de plus grand dimension. A cette fin, nous introduisons la matrice dite « lifting » suivante : $\Phi(\mathcal{T}) \in \mathbb{R}^{(RS)^2 \times K}$, $\Phi(\mathcal{T}) =$

$$[\text{vec}\{\mathcal{T}_{:, :, 1}\} \boxtimes \text{vec}\{\mathcal{T}_{:, :, 1}\} \cdots \text{vec}\{\mathcal{T}_{:, :, K}\} \boxtimes \text{vec}\{\mathcal{T}_{:, :, K}\}];$$

cette matrice n'est rien d'autre que le produit Khatri-Rao de troisième dépliement de \mathcal{T} : $\Phi(\mathcal{T}) = (\mathbf{T}^{(3)})^T \odot (\mathbf{T}^{(3)})^T$.

Nous allons utiliser le noyau à gauche de cette matrice, mais juste la partie symétrique (les colonnes de $\Phi(\mathcal{T})$ sont des vectorisations des matrices symétriques). Notons $\sigma_2(\text{lker}\{\Phi(\mathcal{T})\})$ un sous-espace des vectorisations des $RS \times RS$ matrices symétriques qui est orthogonal à l'espace engendré par les colonnes de $\Phi(\mathcal{T})$.

Prenons un exemple concret de $I = J = R = S = 2$ et $K = 10$, pour \mathcal{T} avec les facteurs suivants :

$$F = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 1 & -3 \end{bmatrix},$$

$$G = \begin{bmatrix} -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 2 & 1 & -3 & 2 & -2 & -1 & 0 & 1 \end{bmatrix}$$

$$H = \begin{bmatrix} -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Dans ce dernier cas, $\Phi(\mathcal{T}) \in \mathbb{R}^{16 \times 10}$ et $\text{rank}\{\Phi(\mathcal{T})\} = 9$. Mais l'espace des matrices symétriques 4×4 est de dimension 10. Cela veut dire que l'espace $\sigma_2(\text{lker}\{\Phi(\mathcal{T})\})$ est de dimension $1 = 10 - 9$ est, après vérification, est engendré par un seul vecteur p qui est la vectorisation de la matrice suivante :

$$p = \text{vec}\left\{ \begin{bmatrix} -12 & -3 & 4 & 3 \\ -3 & 6 & -1 & -2 \\ 4 & -1 & 4 & 1 \\ 3 & -2 & 1 & -2 \end{bmatrix} \right\} \quad (5)$$

Cette observation peut être généralisée aux rangs quelconques.

Lemme 2.2 ([9, §4.4]) Soit $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H})$ avec les facteurs qui satisfont (A1 & A2 & A5). Alors

$$\text{rank}\{\Phi(\mathcal{T})\} = \frac{R(R+1)}{2} \frac{S(S+1)}{2}$$

$$\text{et } \text{rank}\{\mathcal{T}^{(1)}\} = R, \text{rank}\{\mathcal{T}^{(2)}\} = S$$

Lemme 2.2 montre que dans le cas des \mathbf{A} et \mathbf{B} carrées, l'espace $\sigma_2(\text{lker}\{\Phi(\mathcal{T})\})$ est engendré par

$$M = \frac{R(R-1)}{2} \frac{S(S-1)}{2} = \frac{RS(RS+1)}{2} - \frac{R(R+1)}{2} \frac{S(S+1)}{2}$$

vecteurs $\mathbf{p}_1, \dots, \mathbf{p}_M$ (donc $M = 1$ pour $(R, S) = (2, 2)$).

Extraire les matrices \mathbf{A} et \mathbf{B} à partir du noyau Pour cela on aura besoin d'une operation supplémentaire $\pi : \mathbb{R}^{RSRS} \rightarrow \mathbb{R}^{R^2 \times S^2}$ de permutation-symétrisation-matricisation :

$$\pi : \text{vec}\{\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{y} \otimes \mathbf{z}\} \mapsto \frac{1}{2}(\mathbf{a} \boxtimes \mathbf{y} + \mathbf{y} \boxtimes \mathbf{a}) \cdot \frac{1}{2}(\mathbf{b} \boxtimes \mathbf{z} + \mathbf{z} \boxtimes \mathbf{b})^\top, \quad (6)$$

pour tous $\mathbf{y}, \mathbf{a} \in \mathbb{R}^R, \mathbf{z}, \mathbf{b} \in \mathbb{R}^S$. L'operation (6) fait la permutation des dimensions du tenseur sous-jacent ($R \times S \times R \times S \rightarrow R \times R \times S \times S$) et symétrise ensuite les modes de la même dimension.

Reprenons l'exemple du vecteur \mathbf{p} dans (5). La permutation des dimensions du tenseur $2 \times 2 \times 2 \times 2$ sous-jacent donne :

$$\left[\begin{array}{cc|cc} -12 & 4 & 4 & 4 \\ -3 & 3 & -1 & 1 \\ \hline -3 & -1 & 3 & 1 \\ 6 & -2 & -2 & -2 \end{array} \right],$$

qui, après la symétrisation et matricisation, devient

$$\pi(\mathbf{p}) = \left[\begin{array}{cc|cc} -12 & 4 & 4 & 4 \\ -3 & 1 & 1 & 1 \\ \hline -3 & 1 & 1 & 1 \\ 6 & -2 & -2 & -2 \end{array} \right].$$

On peut remarquer que $\mathbf{P} = \pi(\mathbf{p})$ et une matrice de rang 1 est se factorise, par exemple, comme

$$\mathbf{P} = \begin{bmatrix} -4 \\ -1 \\ -1 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & -1 & -1 & -1 \end{bmatrix}.$$

De plus, on peut vérifier que $\mathbf{P}^\top(\mathbf{A} \odot \mathbf{A}) = 0$ ainsi que $\mathbf{P}(\mathbf{B} \odot \mathbf{B}) = 0$. Cela implique ensuite que si on connaît \mathbf{P} , alors on peut retrouver \mathbf{A} et \mathbf{B} à partir du noyau à gauche symétrique de \mathbf{P} et \mathbf{P}^\top .

En effet, soient $\{\mathbf{q}_{A,1}, \mathbf{q}_{A,2}\} \subset \mathbb{R}^4$ et $\{\mathbf{q}_{B,1}, \mathbf{q}_{B,2}\} \subset \mathbb{R}^4$ des bases de $\sigma_2(\text{lker}\{\mathbf{P}\})$ et $\sigma_2(\text{lker}\{\mathbf{P}^\top\})$, respectivement². Il suffit alors de retrouver des bases structurées $\{\mathbf{a}_1 \boxtimes \mathbf{a}_1, \mathbf{a}_2 \boxtimes \mathbf{a}_2\}$ et $\{\mathbf{b}_1 \boxtimes \mathbf{b}_1, \mathbf{b}_2 \boxtimes \mathbf{b}_2\}$ de $\sigma_2(\text{lker}\{\mathbf{P}\})$ et $\sigma_2(\text{lker}\{\mathbf{P}^\top\})$, ce qui est équivalent à la CPD de deux tenseurs $\hat{\mathbf{Q}}_A, \hat{\mathbf{Q}}_B \in \mathbb{R}^{2 \times 2 \times 2}$, construites à partir des vecteurs $\{\mathbf{q}_{A,1}, \mathbf{q}_{A,2}\} \subset \mathbb{R}^4$ et $\{\mathbf{q}_{B,1}, \mathbf{q}_{B,2}\} \subset \mathbb{R}^4$, comme ceci

$$\text{vec}\{(\hat{\mathbf{Q}}_A)_{:, :, k}\} = \mathbf{q}_{A,k}, \text{vec}\{(\hat{\mathbf{Q}}_B)_{:, :, k}\} = \mathbf{q}_{B,k}. \quad (7)$$

Algorithme 1 : Algorithme algébrique pour la décomposition ParaTuck-2 non symétrique

- 1 **Input** : $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$, (R, S) — rangs cible
 - 2 Calculer la décomposition Tucker-2 :
 $\mathcal{T} = \mathcal{T}_c \bullet_1 \mathbf{U} \bullet_2 \mathbf{V}$ with $\mathbf{U} \in \mathbb{R}^{I \times R}$, $\mathbf{V} \in \mathbb{R}^{J \times S}$
 - 3 Trouver $M = \binom{R}{2} \binom{S}{2}$ vecteurs $\mathbf{p}_1, \dots, \mathbf{p}_M$ d'une base du noyau symétrique à gauche $\sigma_2(\text{lker}\{\Phi(\mathcal{T}_c)\})$
 - 4 Trouver les $\mathbf{P}_k = \pi(\mathbf{p}_k) \in \mathbb{R}^{R^2 \times S^2}$ et les ranger dans deux matrices $\mathbf{P}_A = [\mathbf{P}_1 \ \dots \ \mathbf{P}_M] \in \mathbb{R}^{R^2 \times S^2 K}$ et $\mathbf{P}_B = [\mathbf{P}_1^\top \ \dots \ \mathbf{P}_M^\top] \in \mathbb{R}^{S^2 \times R^2 K}$.
 - 5 Trouver des bases $\{\mathbf{q}_{A,1}, \dots, \mathbf{q}_{A,R}\}$, $\{\mathbf{q}_{B,1}, \dots, \mathbf{q}_{B,R}\}$ de $\sigma_2(\text{lker}\{\mathbf{P}_A\})$, $\sigma_2(\text{lker}\{\mathbf{P}_B\})$.
 - 6 Ranger ces vecteurs dans tenseurs $\hat{\mathbf{Q}}_A \in \mathbb{R}^{R \times R \times R}$ et $\hat{\mathbf{Q}}_B \in \mathbb{R}^{S \times S \times S}$ suivant (7) et calculer leur CPD :
 $\hat{\mathbf{Q}}_A = [\hat{\mathbf{A}}_c, \hat{\mathbf{A}}_c, \mathbf{W}_A]$ et $\hat{\mathbf{Q}}_B = [\hat{\mathbf{B}}_c, \hat{\mathbf{B}}_c, \mathbf{W}_B]$.
 - 7 Récupérer les facteurs $\hat{\mathbf{A}} = \mathbf{U} \hat{\mathbf{A}}_c, \hat{\mathbf{B}} = \mathbf{V} \hat{\mathbf{B}}_c$.
 - 8 Retrouver le tenseur cœur $\hat{\mathbf{C}} = \mathcal{T}_c \bullet_1 \hat{\mathbf{A}}_c^\dagger \bullet_2 \hat{\mathbf{B}}_c^\dagger$.
 - 9 Retrouver les facteurs $\hat{\mathbf{F}}, \hat{\mathbf{G}}$ et $\hat{\mathbf{H}}$ à partir de $\hat{\mathbf{C}}$ à l'aide du lemme 2.1
 - 10 **Output** : Matrices $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{F}}, \hat{\mathbf{G}}, \hat{\mathbf{H}}$, telles que
 $\mathcal{T} = \text{PT2D}(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{F}}, \hat{\mathbf{G}}, \hat{\mathbf{H}})$
-

L'algorithme pour le cas non symétrique L'algorithme complète est présenté dans Algorithme 1. Nous donnons quelques remarques importantes :

- Algorithme 1 donne le résultat correct sous A1-A2, A5 et une condition supplémentaire suivante :

$$A7 : \text{krank}\{\mathbf{F}^\top\} \geq 2 \text{ et } \text{krank}\{\mathbf{F}\} \geq 2;$$

A7 garantit que $\sigma_2(\text{lker}\{\mathbf{P}_A\})$ et $\sigma_2(\text{lker}\{\mathbf{P}_B\})$ ont des dimensions R et S , respectivement.

- Comme le noyau symétrique est calculée dans étape 5, $\pi(\mathbf{p})$ peut être remplacé par une opération plus simple :

$$\text{reshape}(\text{permute}(\text{reshape}(\mathbf{p}, [R, S, R, S]), [1, 3, 2, 4]), R^2, S^2);$$

- Dans étape 6, on doit calculer la CPD de rang R d'un tenseur $R \times R \times R$. Cela peut être fait, par exemple, avec la décomposition en valeurs propres généralisée.
- Des algorithmes peuvent être utilisées pour l'approximation, où le noyau approché est obtenu par la SVD.

3 Exemples numériques

Pour des exemples dans cette section, nous utilisons la notation $MSE(\mathcal{T}, \hat{\mathcal{T}}) = \|\mathcal{T} - \hat{\mathcal{T}}\|_F^2$ pour l'erreur quadratique moyenne, ainsi que la métrique suivante pour mesurer la distance entre les facteurs :

$$SSS(\mathbf{A}, \hat{\mathbf{A}}) = \sum_{k=1}^R 1 - \left(\frac{\langle \mathbf{a}_k, \hat{\mathbf{a}}_k \rangle}{\|\mathbf{a}_k\| \|\hat{\mathbf{a}}_k\|} \right)^2;$$

dans la formule ci-dessus, les colonnes d'un facteur estimé $\hat{\mathbf{A}}$ sont permutés au préalable avec l'algorithme hongrois. Ainsi cette métrique mesure la somme des sinus carrés des angles entre les vrais facteurs et les facteurs estimés.

²La dimension de l'espace des matrices symétriques 2×2 est 3.

Exemple 2×2 La première illustration concerne le comportement d'un algorithme ALS pour un exemple le plus simple possible de la section 2. Les facteurs dans cette exemple satisfont les conditions d'unicité de [5], mais l'ALS converge dans une petite fraction des initialisations uniquement. En revanche, Algorithme 1 algébrique trouve une solution avec $MSE(\mathcal{T}, \hat{\mathcal{T}}) = 1.7 \cdot 10^{-26}$ et distances entre les facteurs $SSS(\mathbf{A}, \hat{\mathbf{A}}) = 6.6 \cdot 10^{-26}$ et $SSS(\mathbf{B}, \hat{\mathbf{B}}) = 2.2 \cdot 10^{-26}$

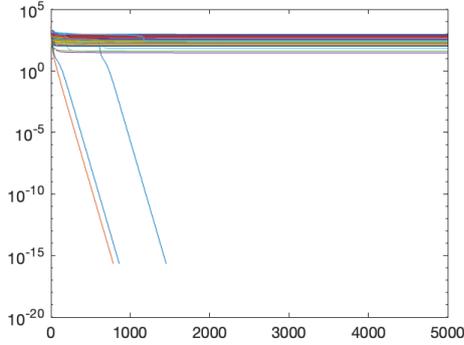


FIGURE 1 : Performances de l'ALS [1] pour $(R, S) = (2, 2)$, convergence pour 100 initialisations.

Exemple 4×4 On prend un exemple généré aléatoirement, avec $I = J = 10$, $R = S = 4$, $K = 100$. Les éléments des matrices \mathbf{A} , \mathbf{B} , \mathbf{H} , \mathbf{G} sont indépendants et suivent la loi normale centrée réduite; les éléments de \mathbf{F} sont tirés de la loi uniforme $[0.5, 1.5]$.

Pour l'illustration, une seule réalisation des facteurs est considérée (des résultats pour plusieurs réalisations sont disponibles dans [9]). On ajoute ainsi du bruit blanc normal avec plusieurs niveaux de RSB, pour 100 réalisations. Les résultats sont démontrés dans Fig. 2 (HOOI signifié les résultats pour une méthode itérative [9] qui utilise la structure fine du noyau de $\Phi(\mathcal{T})$ et utilise Algorithme 1 comme initialisation). On peut voir dans Fig. 2, que l'algorithme donne des bons résultats seulement pour forts RSB, mais il reste encore plusieurs pistes d'amélioration (par exemple, dans l'implantation actuelle une méthode très basique est utilisé pour la CPD de $\hat{\mathcal{Q}}_A$ et $\hat{\mathcal{Q}}_B$).

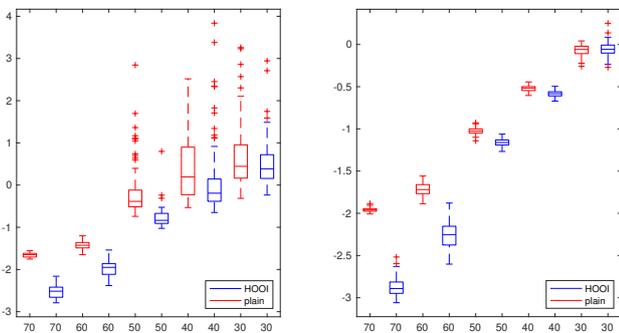


FIGURE 2 : Performance de l'Algorithme 1 (échelle logarithmique). A gauche : erreur relative $\log_{10}(MSE(\mathcal{T}, \hat{\mathcal{T}})/\|\mathcal{T}\|^2)$. A droite : $\log_{10}(SSS(\mathbf{A}, \hat{\mathbf{A}}))$.

Les performances dans le cas bruitée peuvent être aussi expliquées par la taille de matrice $\Phi(\mathcal{T})$ qui explose quand on augmente R et S . Par exemple, dans Fig. 3 on voit bien que les valeurs singulières du bruit sont très proches aux valeurs singulières du signal (les $100 = \frac{R(R+1)}{2} \frac{S(S+1)}{2}$ premières valeurs singulières de $\Phi(\mathcal{T})$).

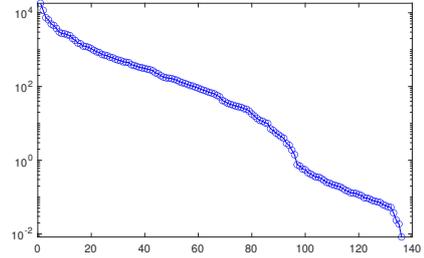


FIGURE 3 : Valeurs singulières de $\Phi(\mathcal{T})$ pour 60dB.

Références

- [1] Rasmus BRO : *Multi-way Analysis in the Food Industry*. Thèse de doctorat, Vrije Universiteit Brussel (VUB), 1998.
- [2] Joppe DE JONGHE, Konstantin USEVICH, Philippe DRESEN et Mariya ISHTEVA : Compressing neural networks with two-layer decoupling. In *2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 226–230, 2023.
- [3] Pedro Marinho R. de OLIVEIRA, C. Alexandre Rolim FERNANDES, Gérard FAVIER et Remy BOYER : PARATUCK Semi-Blind Receivers for Relaying Multi-Hop MIMO Systems. *Digital Signal Processing*, 92:127–138, 2019.
- [4] Gérard FAVIER et André LF de ALMEIDA : Overview of constrained PARAFAC models. *EURASIP Journal on Advances in Signal Processing*, 2014(1):142, 2014.
- [5] Richard A HARSHMAN et Margaret E LUNDY : Uniqueness proof for a family of models sharing features of Tucker's three-mode factor analysis and parafac/candecomp. *Psychometrika*, 61(1):133–154, 1996.
- [6] Tamara G KOLDA et Brett W BADER : Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.
- [7] Kristina NASKOVSKA : *Advanced tensor based signal processing techniques for wireless communication systems and biomedical signal processing*. Thèse de doctorat, Ilmenau, Jan 2020. Dissertation, Technische Universität Ilmenau, 2019.
- [8] Marie ROALD, Carla SCHENKER, Vince CALHOUN, Tülay ADALI, Rasmus BRO, Jeremy COHEN et Evrim ATAMAN : An AO-ADMM approach to constraining PARAFAC2 on all modes. *SIAM Journal on Mathematics of Data Science*, 4(3):1191–1222, 2022.
- [9] Konstantin USEVICH : A lifting approach to ParaTuck-2 tensor decompositions. <https://hal.science/hal-03966869v3>, mars 2025.
- [10] Konstantin USEVICH, Yassine ZNIYED, Mariya ISHTEVA, Philippe DRESEN et André de ALMEIDA : Tensor-based two-layer decoupling of multivariate polynomial maps. In *31st European Signal Processing Conference, EUSIPCO 2023*, Helsinki, Finland, septembre 2023.
- [11] Yassine ZNIYED et André LF de ALMEIDA : A stochastic algorithm for the ParaTuck decomposition. *Digital Signal Processing*, 156:104767, 2025.