

HOSVD Tronquée : Analyse d'une Approximation Tensorielle Rapide

Hugo LEBEAU^{1,2} Romain COUILLET¹ Florent CHATELAIN²

¹Université Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France

²Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

Résumé – Les outils de la théorie des matrices aléatoires nous permettent d'étudier un modèle spike tensoriel modélisant l'observation bruitée d'un signal de petit rang multilinéaire. Une caractérisation précise du comportement spectral des dépliements du tenseur révèle un phénomène de transition de phase, en-dessous de laquelle l'information ne peut pas être reconstruite. Cette analyse donne les performances exactes de l'approximation par HOSVD tronquée. Il s'agit d'une première étape vers l'étude précise de l'approximation tensorielle de petit rang.

Abstract – Tools from random matrix theory allow us to study a tensor spiked model describing the noisy observation of a low multilinear-rank signal. A precise characterization of the spectral behavior of the tensor unfoldings reveals a phase-transition phenomenon below which no information can be recovered. This analysis gives the exact performance of the approximation with truncated HOSVD. It paves the way towards a precise study of low-rank tensor approximation.

1 Introduction

L'extraction d'information à partir d'une large base de données est une tâche fondamentale en traitement du signal, comme dans le domaine plus récent de l'apprentissage machine. L'explosion de la masse de données disponibles nous invite en outre à en traiter de nouvelles formes, telles que les données multidimensionnelles. L'imagerie cérébrale, par exemple, produit des données $\mathcal{A}_i \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ représentées sous la forme d'un tableau (un tenseur) d'ordre 3, qui devient d'ordre 4 avec un ensemble de n images. De l'imagerie hyperspectrale à la génomique, en passant par la détection de communautés, les tenseurs interviennent naturellement dans de nombreuses applications récentes du traitement du signal [9].

L'information sous-jacente étant bien souvent de petit rang, lorsque les données peuvent se mettre sous forme matricielle, une décomposition en valeurs singulières suffit pour extraire la meilleure approximation de rang faible en temps polynomial. De plus, de nombreux travaux basés sur la théorie des matrices aléatoires indiquent avec quelle précision cette information peut être reconstruite [2]. Le cas tensoriel se révèle en revanche plus épineux : s'il est possible d'étendre la décomposition en valeurs singulières aux tenseurs [3], la simple suppression des composantes les moins "énergétiques" n'aboutit pas, en général, à la meilleure approximation de petit rang¹ — dont le calcul est en réalité NP-difficile [6]. Cependant, cette procédure extrêmement simple² est *quasi-optimale* [5] : étant donnée une meilleure approximation de petit rang \mathcal{A}_* d'un tenseur \mathcal{A} d'ordre $d \geq 3$, l'approximation $\hat{\mathcal{A}}$ de même petit rang obtenue par troncation de la HOSVD³ vérifie $\|\mathcal{A} - \hat{\mathcal{A}}\|_F \leq \sqrt{d} \|\mathcal{A} - \mathcal{A}_*\|_F$. La simplicité de cette procédure ainsi que sa quasi-optimalité, ajouté au volume conséquent des données tensorielles, en font un bon choix d'approximation à moindre coût ainsi qu'une très bonne initialisation pour un algorithme d'estimation de \mathcal{A}_* [4].

L'étude de modèles tensoriels dits *spikes* — c'est-à-dire une

observation $\mathcal{J} = \mathcal{P} + \mathcal{N}$ avec \mathcal{P} une information de petit rang et \mathcal{N} un bruit — s'est largement développée au cours de la dernière décennie. Les outils de la théorie des matrices aléatoires ont notamment permis d'étudier précisément les limites statistique et algorithmique⁴ du modèle dans le cas d'un signal \mathcal{P} de rang 1 [7, 1]. Au-delà, deux notions de *rang* coexistent. Le rang, proprement dit, c'est-à-dire le nombre minimal de tenseurs de rang 1 nécessaires pour représenter \mathcal{P} comme leur somme, amène à considérer la meilleure approximation de rang R , qui est un problème mal posé⁵ dès que $R \geq 2$. Il est néanmoins possible de traiter ce problème au moyen d'une méthode de déflation récemment étudiée sous le prisme des tenseurs aléatoires [8]. Par ailleurs, c'est la notion de *rang multilinéaire* qui découle directement de la HOSVD : il s'agit du d -uplet (r_1, \dots, r_d) où r_ℓ correspond au rang de la *matrice* $\mathbf{P}^{(\ell)}$ obtenue en dépliant \mathcal{P} selon le mode ℓ . Le problème de la meilleure approximation de petit rang multilinéaire est NP-difficile en général. Elle peut être estimée algorithmiquement [4] et des bornes sur l'erreur d'estimation existent [10].

Cet article présente une étude du modèle spike $\mathcal{J} = \mathcal{P} + \mathcal{N}$ où \mathcal{P} est une information de petit rang multilinéaire. Nos résultats décrivent le comportement asymptotique du grand tenseur aléatoire \mathcal{J} et sous quel régime l'information peut être reconstruite. Ces conclusions sont en lien direct avec l'approximation par HOSVD tronquée et sont une première étape vers la compréhension des limites de l'approximation tensorielle de petit rang.

Notations Les scalaires, vecteurs, matrices et tenseurs sont respectivement distingués par une écriture minuscule (n), gras minuscule (\mathbf{u}), gras majuscule (\mathbf{A}) et gras script majuscule (\mathcal{A}). En particulier $\mathbf{A}^{(\ell)}$ désigne la *matrice* obtenue par dépliement du tenseur \mathcal{A} selon le mode ℓ (voir Définition 2).

⁴La limite *statistique*, à partir de laquelle il existe une solution au problème du maximum de vraisemblance, diffère de la limite *algorithmique*, à partir de laquelle l'information peut être reconstruite en temps polynomial. Lorsque toutes les dimensions du tenseur sont d'ordre $N \rightarrow +\infty$, cette première est $\mathcal{O}(1)$ alors que la seconde croît avec N en $\mathcal{O}(N^{\frac{d-2}{4}})$.

⁵L'ensemble des tenseurs de rang R n'est pas fermé [9].

¹La notion de *rang multilinéaire* considérée ici sera définie plus bas.

²Elle se ramène à d SVD matricielles.

³Higher-Order Singular Value Decomposition.

2 La HOSVD et sa troncature

Un tenseur \mathcal{A} d'ordre d est représenté par un tableau dans $\mathbb{R}^{n_1 \times \dots \times n_d}$. Afin de ne pas alourdir l'exposé et simplifier sa compréhension, nous nous restreindrons au cas des tenseurs d'ordre $d = 3$; les résultats présentés se généralisent naturellement aux ordres supérieurs.

2.1 Définition et propriétés de la HOSVD

Définition 1 (ℓ -fibre). Pour $\ell \in \{1, 2, 3\}$, une ℓ -fibre de \mathcal{A} est le vecteur de \mathbb{R}^{n_ℓ} obtenu en fixant tous les indices de \mathcal{A} sauf selon le mode ℓ . Il s'agit de l'extension du concept de ligne et colonne des matrices.

Définition 2 (Dépliement). Le dépliement du tenseur \mathcal{A} selon le mode 1 est la matrice $\mathbf{A}^{(1)} \in \mathbb{R}^{n_1 \times n_2 n_3}$ dont les colonnes sont les 1-fibres de \mathcal{A} . Les dépliements $\mathbf{A}^{(2)}$ et $\mathbf{A}^{(3)}$ selon les autres modes sont définis de la même manière.

Remarque 1. L'ordre des colonnes de $\mathbf{A}^{(\ell)}$ n'importe pas ici.

Définition 3 (Rang multilinéaire). Le rang multilinéaire de \mathcal{A} est (r_1, r_2, r_3) où r_ℓ est le rang de la matrice $\mathbf{A}^{(\ell)}$.

Dans toute la suite, le rang d'un tenseur désignera toujours son rang multilinéaire.

Remarque 2. On a nécessairement l'inégalité $\max(r_1, r_2, r_3) \leq \min(r_1 r_2, r_2 r_3, r_1 r_3)$.

Définition 4 (HOSVD). La HOSVD (*Higher-Order Singular Value Decomposition*) de \mathcal{A} est la donnée d'un tenseur $\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ et des matrices $\mathbf{U} \in \mathbb{R}^{n_1 \times r_1}$, $\mathbf{V} \in \mathbb{R}^{n_2 \times r_2}$, $\mathbf{W} \in \mathbb{R}^{n_3 \times r_3}$ à colonnes orthonormées vérifiant

- $\mathcal{S}^{(\ell)} \mathcal{S}^{(\ell)\top} = \mathbf{\Lambda}^{(\ell)}$ où $\mathbf{\Lambda}^{(\ell)}$ est une matrice diagonale telle que $\Lambda_{1,1}^{(\ell)} \geq \dots \geq \Lambda_{r_\ell, r_\ell}^{(\ell)} > 0$,⁶

- $\mathcal{A}_{i,j,k} = \sum_{a,b,c=1}^{r_1, r_2, r_3} \mathcal{S}_{a,b,c} \mathbf{U}_{i,a} \mathbf{V}_{j,b} \mathbf{W}_{k,c} \quad \begin{matrix} 1 \leq i \leq n_1 \\ 1 \leq j \leq n_2 \\ 1 \leq k \leq n_3 \end{matrix}$.

On note cela $\mathcal{A} = \llbracket \mathcal{S}; \mathbf{U}, \mathbf{V}, \mathbf{W} \rrbracket$.

Bien que le tenseur \mathcal{S} ne soit pas diagonal en général, la HOSVD est l'extension naturelle de la SVD matricielle : $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top = \llbracket \mathbf{\Sigma}; \mathbf{U}, \mathbf{V} \rrbracket$. Elle a de nombreuses propriétés intéressantes [3]. Notamment, les colonnes de \mathbf{U} (resp. \mathbf{V} , \mathbf{W}) sont les r_1 (resp. r_2 , r_3) vecteurs singuliers à gauche de $\mathbf{A}^{(1)}$ (resp. $\mathbf{A}^{(2)}$, $\mathbf{A}^{(3)}$). De plus, en notant $s_1(\mathbf{A}^{(\ell)}) \geq \dots \geq s_{r_\ell}(\mathbf{A}^{(\ell)}) > 0$ les valeurs singulières de $\mathbf{A}^{(\ell)}$, on a $\Lambda_{\kappa, \kappa}^{(\ell)} = s_{\kappa}^2(\mathbf{A}^{(\ell)})$ pour tout $\kappa \in \{1, \dots, r_\ell\}$.

2.2 Quasi-optimalité de la HOSVD tronquée

Le théorème d'Eckart-Young-Mirsky assure que la meilleure approximation de rang R d'une matrice est obtenue en tronquant sa SVD aux R composantes ayant les valeurs singulières les plus grandes. Ce résultat peut-il être étendu aux tenseurs en terme de HOSVD? La réponse est malheureusement négative [9] et le calcul de la meilleure approximation de petit rang multilinéaire d'un tenseur est un problème NP-difficile [6].

⁶Cela revient à dire que les tranches de \mathcal{S} sont orthogonales entre elles et ordonnées par énergie décroissante.

Étant donné un tenseur quelconque $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, il est néanmoins très facile d'en calculer l'approximation $\hat{\mathcal{A}}$ de rang (r'_1, r'_2, r'_3) résultant de la troncature de sa HOSVD (Algorithme 1).

Algorithme 1 : HOSVD tronquée de \mathcal{A}

$\mathbf{U}' \leftarrow r'_1$ premiers vecteurs singuliers à gauche de $\mathbf{A}^{(1)}$
 $\mathbf{V}' \leftarrow r'_2$ premiers vecteurs singuliers à gauche de $\mathbf{A}^{(2)}$
 $\mathbf{W}' \leftarrow r'_3$ premiers vecteurs singuliers à gauche de $\mathbf{A}^{(3)}$
 $\mathcal{S}'_{a,b,c} \leftarrow \sum_{i,j,k=1}^{n_1, n_2, n_3} \mathcal{A}_{i,j,k} \mathbf{U}'_{i,a} \mathbf{V}'_{j,b} \mathbf{W}'_{k,c}$
 $\hat{\mathcal{A}} \leftarrow \llbracket \mathcal{S}'; \mathbf{U}', \mathbf{V}', \mathbf{W}' \rrbracket$

L'approximation $\hat{\mathcal{A}}$ ainsi obtenue est qualifiée de *quasi-optimale* [5] : si \mathcal{A}_* est solution du problème NP-difficile $\min_{\mathcal{B} \text{ de rang } (r'_1, r'_2, r'_3)} \|\mathcal{A} - \mathcal{B}\|_F$ alors

$$\|\mathcal{A} - \mathcal{A}_*\|_F \leq \|\mathcal{A} - \hat{\mathcal{A}}\|_F \leq \sqrt{d} \|\mathcal{A} - \mathcal{A}_*\|_F.$$

3 Analyse du modèle spike tensoriel

On considère un tenseur $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ résultant de l'observation bruitée d'une information \mathcal{P} de rang (r_1, r_2, r_3) :

$$\mathcal{T} = \mathcal{P} + \frac{1}{\sqrt{N}} \mathcal{N}, \quad \mathcal{P} = \llbracket \mathcal{H}; \mathbf{X}, \mathbf{Y}, \mathbf{Z} \rrbracket$$

avec $\mathcal{N}_{i,j,k} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ et $N = n_1 + n_2 + n_3$.

Se pose alors la question suivante : étant donné $\hat{\mathcal{T}} = \llbracket \mathcal{G}; \mathbf{U}, \mathbf{V}, \mathbf{W} \rrbracket$ l'approximation de rang (r_1, r_2, r_3) de \mathcal{T} par troncature de sa HOSVD, avec quelle précision \mathbf{U} , \mathbf{V} et \mathbf{W} reconstruisent-ils \mathbf{X} , \mathbf{Y} et \mathbf{Z} ? Autrement dit, quelle est la perte d'information induite par l'ajout du bruit $\frac{1}{\sqrt{N}} \mathcal{N}$? Les résultats présentés ci-dessous répondent à cette question dans le régime asymptotique $n_1, n_2, n_3 \rightarrow +\infty$ avec $0 < c_\ell = \lim n_\ell / N < +\infty$, dénoté plus simplement " $N \rightarrow +\infty$ ".

On s'intéresse, en somme, aux vecteurs singuliers à gauche de $\mathbf{T}^{(\ell)} = \mathbf{P}^{(\ell)} + \frac{1}{\sqrt{N}} \mathcal{N}^{(\ell)}$, qu'on reconnaît comme un modèle *spike matriciel*. Nous pouvons alors l'étudier grâce aux outils de la théorie des grandes matrices aléatoires [2].

Remarque 3. Il faut cependant faire attention au fait que nous ne sommes pas dans une situation "classique" de grandes matrices aléatoires puisque les deux dimensions de la matrice $\mathbf{T}^{(\ell)}$ ne sont pas du même ordre ($\mathcal{O}(N) \times \mathcal{O}(N^{d-1})$).

Afin de considérer des situations pertinentes en pratique, on va supposer un rapport signal sur bruit tel que le problème ne soit "ni trop dur ni trop simple". Pour cela, on introduit la condition de non-trivialité suivante.

Hypothèse 1 (Condition de non-trivialité).

$$0 < \liminf_{N \rightarrow +\infty} \frac{\|\mathcal{P}\|_F}{N^{\frac{d-2}{4}}} \leq \limsup_{N \rightarrow +\infty} \frac{\|\mathcal{P}\|_F}{N^{\frac{d-2}{4}}} < +\infty.$$

Cette condition fixe l'ordre de grandeur du signal $\|\mathcal{P}\|_F = \mathcal{O}(N^{\frac{d-2}{4}})$, pour que celui-ci soit "d'énergie comparable à celle du bruit"⁷.

⁷En comparaison, $\mathbb{E} \left\| \frac{1}{\sqrt{N}} \mathcal{N} \right\|_F = \sqrt{\frac{n_1 n_2 n_3}{N}} = \mathcal{O}(N^{\frac{d-1}{2}})$ mais, le signal étant de petit rang, son énergie va se "concentrer" tandis que celle du bruit sera plus diffuse. Ainsi, malgré la différence d'ordre entre $\|\mathcal{P}\|_F$ et $\left\| \frac{1}{\sqrt{N}} \mathcal{N} \right\|_F$, c'est dans ce régime que l'on va pouvoir observer une transition entre la présence, ou non, d'un signal discernable du bruit (Théorème 2).

3.1 Distribution spectrale limite

Afin de suivre le comportement des valeurs et vecteurs *sin-guliers* de $\mathbf{T}^{(\ell)}$, nous étudions, de manière équivalente, les valeurs et vecteurs *propres* de $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$. Notre premier résultat concerne la distribution de ses valeurs propres lorsque $N \rightarrow +\infty$. L'information $\mathbf{P}^{(\ell)}$ étant de petit rang $r_\ell = \mathcal{O}(1)$, il est attendu que celle-ci ait un impact négligeable sur la distribution asymptotique des valeurs propres de $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$. Le Théorème 1 caractérise donc l'influence du bruit $\frac{1}{\sqrt{N}}\mathbf{N}^{(\ell)}$ sur celle-ci.

En fait, de par la différence d'ordre entre les dimensions de $\mathbf{T}^{(\ell)}$, les valeurs propres de $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ vont diverger autour d'une valeur moyenne $\mu_N^{(\ell)} = \mathcal{O}(N^{d-2})$ et s'étaler sur un segment de largeur $4\sigma_N = \mathcal{O}(N^{\frac{d-2}{2}})$. Afin d'étudier proprement sa distribution spectrale limite, on s'intéresse donc plutôt à la matrice "centrée réduite" $\frac{1}{\sigma_N}\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top} - \frac{\mu_N^{(\ell)}}{\sigma_N}\mathbf{I}_{n_\ell}$ dont le spectre reste $\mathcal{O}(1)$.

Notre objet d'étude étant maintenant convenablement défini, les outils de matrices aléatoires nous permettent d'identifier sa distribution spectrale limite $\tilde{\nu}$ via sa transformée de Stieltjes $\tilde{m} : \tilde{z} \in \mathbb{C} \setminus \text{supp } \tilde{\nu} \mapsto \int_{\mathbb{R}} \frac{\tilde{\nu}(dt)}{t - \tilde{z}}$.

Théorème 1 (Distribution spectrale limite). *Pour $\ell \in \{1, 2, 3\}$, soient les quantités suivantes*

$$\begin{aligned} \mu_N^{(\ell)} &= \frac{n_\ell^2 + n_1 n_2 n_3}{n_\ell N} \underset{N \rightarrow +\infty}{\sim} \frac{c_1 c_2 c_3}{c_\ell} N^{d-2}, \\ \sigma_N &= \frac{\sqrt{n_1 n_2 n_3}}{N} \underset{N \rightarrow +\infty}{\sim} \sqrt{c_1 c_2 c_3} N^{d-2}. \end{aligned}$$

La matrice $\frac{1}{\sigma_N}\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top} - \frac{\mu_N^{(\ell)}}{\sigma_N}\mathbf{I}_{n_\ell}$ admet une distribution spectrale limite $\tilde{\nu}$ lorsque $N \rightarrow +\infty$ dont la transformée de Stieltjes \tilde{m} vérifie

$$\tilde{m}^2(\tilde{z}) + \tilde{z}\tilde{m}(\tilde{z}) + 1 = 0 \quad \tilde{z} \in \mathbb{C} \setminus \text{supp } \tilde{\nu}.$$

L'équation vérifiée par \tilde{m} est en réalité bien connue et nous indique que $\tilde{\nu}$ n'est autre que la loi du demi-cercle de support $[-2, +2]$. Ainsi, la distribution des valeurs propres de $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ se rapproche d'une loi du demi-cercle sur $[\mu_N^{(\ell)} \pm 2\sigma_N]$.

La première ligne de la Figure 1 montre la distribution spectrale de $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ pour $\ell \in \{1, 2, 3\}$ ainsi que la loi du demi-cercle associée. On observe effectivement que les valeurs propres sont distribuées selon la loi de Wigner hormis quelques-unes qui s'isolent à droite du *bulk*. Ces valeurs propres isolées, appelées *spikes*, ne viennent pas modifier la distribution spectrale limite et sont l'expression de l'information de petit rang $\mathbf{P}^{(\ell)}$. Il convient maintenant de les étudier plus précisément ainsi que leurs vecteurs propres associés.

Remarque 4 (Entre Marchenko-Pastur et Wigner). À horizon *fini*, on constate que la forme du *bulk* des valeurs propres de $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ est entre celle de la distribution de Marchenko-Pastur et du demi-cercle. En effet, si $n_2 n_3$ n'est pas "suffisamment grand" devant n_1 alors les deux dimensions peuvent être considérées du même ordre et on retrouve un modèle de covariance empirique standard.

3.2 Comportement des spikes

Notre second résultat caractérise donc le comportement des spikes, qui déterminent comment l'information $\mathbf{P}^{(\ell)}$ pourra être reconstruite à partir de $\mathbf{T}^{(\ell)}$.

Théorème 2 (Comportement des spikes). *Pour $\ell \in \{1, 2, 3\}$ et $\kappa \in \{1, \dots, r_\ell\}$, soient les quantités suivantes⁸*

$$\begin{aligned} \rho_{\kappa, N}^{(\ell)} &= \frac{s_\kappa^2(\mathbf{P}^{(\ell)})}{\sigma_N}, & \tilde{\zeta}_{\kappa, N}^{(\ell)} &= \rho_{\kappa, N}^{(\ell)} + \frac{1}{\rho_{\kappa, N}^{(\ell)}}, \\ \zeta_{\kappa, N}^{(\ell)} &= 1 - \frac{1}{\rho_{\kappa, N}^{(\ell)}}. \end{aligned}$$

Si $\rho_{\kappa, N}^{(\ell)} > 1$ alors $\tilde{\zeta}_{\kappa, N}^{(\ell)}$ est presque sûrement la position asymptotique d'une valeur propre isolée de $\frac{1}{\sigma_N}\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top} - \frac{\mu_N^{(\ell)}}{\sigma_N}\mathbf{I}_{n_\ell}$. De plus, étant donné $\mathbf{u}_{\kappa, N}^{(\ell)}$, son vecteur propre associé, et $\mathbf{x}_\kappa^{(\ell)}$, le vecteur propre de $\mathbf{P}^{(\ell)}\mathbf{P}^{(\ell)\top}$ associé à la valeur propre $s_\kappa^2(\mathbf{P}^{(\ell)})$, on a

$$\left\langle \mathbf{u}_{\kappa, N}^{(\ell)}, \mathbf{x}_\kappa^{(\ell)} \right\rangle^2 \xrightarrow[N \rightarrow +\infty]{p.s.} \zeta_{\kappa, N}^{(\ell)}.$$

Remarque 5. De par la condition de non trivialité (Hypothèse 1), $\sum_{\kappa=1}^{r_\ell} \frac{s_\kappa^2(\mathbf{P}^{(\ell)})}{\sigma_N} = \mathcal{O}(1)$ donc les quantités $\rho_{\kappa, N}^{(\ell)}$, $\tilde{\zeta}_{\kappa, N}^{(\ell)}$ et $\zeta_{\kappa, N}^{(\ell)}$ n'explorent ni ne disparaissent lorsque $N \rightarrow +\infty$.

La quantité $\rho_{\kappa, N}^{(\ell)}$ s'interprète comme un rapport signal sur bruit associé à la valeur $s_\kappa^2(\mathbf{P}^{(\ell)})$. Si ce rapport est suffisamment grand ($\rho_{\kappa, N}^{(\ell)} > 1$), une valeur propre de $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ s'extirpe du *bulk* et s'isole du reste du spectre à la position $\mu_N^{(\ell)} + \sigma_N \tilde{\zeta}_{\kappa, N}^{(\ell)}$. Le vecteur propre $\mathbf{u}_{\kappa, N}^{(\ell)}$ associé à ce spike se retrouve alors porteur de l'information $\mathbf{x}_\kappa^{(\ell)}$ — qui est la κ -ème colonne de \mathbf{X} (resp. \mathbf{Y} , \mathbf{Z}) si $\ell = 1$ (resp. 2, 3) — avec un alignement quantifié par $\zeta_{\kappa, N}^{(\ell)} \in [0, 1]$.

Les positions asymptotiques des spikes $\mu_N^{(\ell)} + \sigma_N \tilde{\zeta}_{\kappa, N}^{(\ell)}$ sont représentées par des tirets verts dans les distributions spectrales de la Figure 1. La seconde ligne de la figure compare les alignements $\langle \mathbf{u}_{\kappa, N}^{(\ell)}, \mathbf{x}_\kappa^{(\ell)} \rangle^2$ observés à leur valeur asymptotique $\zeta_{\kappa, N}^{(\ell)}$. Alors qu'on aurait pu avoir jusqu'à $r_3 = 5$ valeurs propres isolées dans le spectre de $\mathbf{T}^{(3)}\mathbf{T}^{(3)\top}$, on en observe uniquement quatre, le signal n'étant pas assez puissant pour en arracher une cinquième du *bulk*. De plus, on peut constater que l'alignement est d'autant plus élevé que le spike est éloigné du *bulk*. Cela se devine aisément des équations de $\tilde{\zeta}_{\kappa, N}^{(\ell)}$ et $\zeta_{\kappa, N}^{(\ell)}$ mais on l'interprète également en terme d'énergie $s_\kappa^2(\mathbf{P}^{(\ell)})$: plus celle-ci est grande, plus l'information portée par le spike se distingue du bruit.

Remarque 6 (Similitude avec le modèle spike de Wigner). Les expressions de la position des spikes $\tilde{\zeta}_{\kappa, N}^{(\ell)}$ ainsi que des alignements $\zeta_{\kappa, N}^{(\ell)}$ sont les mêmes que pour celles d'un modèle spike de Wigner avec un rapport signal sur bruit $\rho_{\kappa, N}^{(\ell)}$.

⁸On rappelle que la notation $s_\kappa(\mathbf{A})$ désigne la κ -ème valeur singulière de la matrice \mathbf{A} dans l'ordre décroissant.

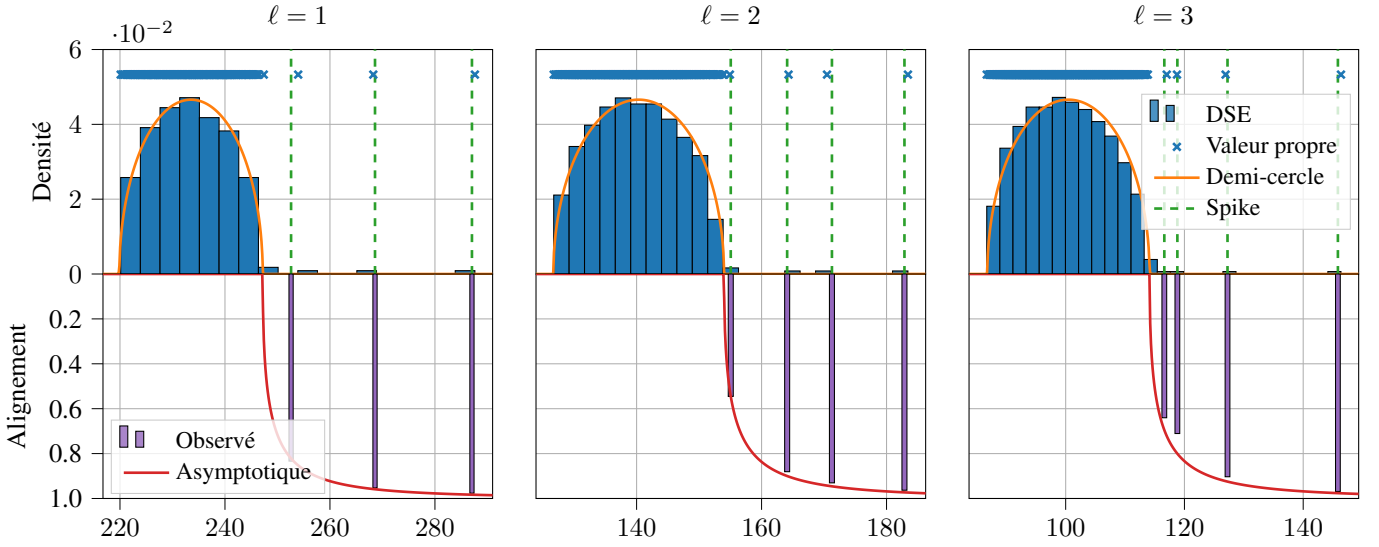


FIGURE 1 : **Haut** : Distribution spectrale empirique (DSE) de $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$. La densité du demi-cercle $t \mapsto \frac{1}{\sigma_N} \tilde{\nu}\left(\frac{t - \mu_N^{(\ell)}}{\sigma_N}\right)$ est représentée par la courbe orange. Les positions asymptotiques des spikes $\mu_N^{(\ell)} + \sigma_N \tilde{\zeta}_{\kappa, N}^{(\ell)}$ sont représentées en tirets verts. **Bas** : Alignements observés et asymptotiques ($\zeta_{\kappa, N}^{(\ell)}$) entre les premiers vecteurs propres de $\mathbf{T}^{(\ell)}\mathbf{T}^{(\ell)\top}$ et $\mathbf{P}^{(\ell)}\mathbf{P}^{(\ell)\top}$. **Paramètres** : $(n_1, n_2, n_3) = (300, 500, 700)$, $(r_1, r_2, r_3) = (3, 4, 5)$ et $\|\mathcal{P}\|_{\text{F}}^2 = 15\sigma_N$.

3.3 Conséquences pour la HOSVD tronquée

Le Théorème 2 ci-dessus révèle la présence d’une *transition de phase* : l’information associée à la κ -ème valeur singulière de $\mathbf{P}^{(\ell)}$ est discernable du bruit uniquement si $s_{\kappa}(\mathbf{P}^{(\ell)}) > \sqrt{\sigma_N} = \mathcal{O}(N^{\frac{d-2}{4}})$. Soit alors $\hat{\mathcal{J}} = \llbracket \mathcal{G}; \mathbf{U}, \mathbf{V}, \mathbf{W} \rrbracket$ l’approximation résultant de la troncation de la HOSVD de \mathcal{J} au rang (r_1, r_2, r_3) . La qualité de la reconstruction de \mathbf{X} , \mathbf{Y} et \mathbf{Z} par \mathbf{U} , \mathbf{V} et \mathbf{W} dépend de la position de chaque $s_{\kappa}(\mathbf{P}^{(\ell)})$ par rapport au seuil $\sqrt{\sigma_N}$. On peut alors quantifier les “alignements” entre les images de \mathbf{U} et \mathbf{X} , \mathbf{V} et \mathbf{Y} , \mathbf{W} et \mathbf{Z} (ce sont donc des alignements entre sous-espaces) par⁹

$$\frac{1}{r_{\ell}} \sum_{\kappa=1}^{r_{\ell}} \zeta_{\kappa, N}^{(\ell)} \mathbf{1}_{\rho_{\kappa, N}^{(\ell)} > 1} \quad \text{pour } \ell \in \{1, 2, 3\}.$$

On caractérise ainsi la qualité de la reconstruction des sous-espaces singuliers du signal \mathcal{P} via une HOSVD tronquée de son observation bruitée \mathcal{J} .

4 Conclusions et perspectives

À l’aide d’outils de la théorie des matrices aléatoires, nous avons décrit comment la HOSVD d’un signal \mathcal{P} de petit rang multilinéaire est perturbée par l’addition d’un bruit $\frac{1}{\sqrt{N}}\mathcal{N}$. L’étude de la distribution des valeurs singulières des dépliements du tenseur \mathcal{J} résultant de l’observation du signal bruité se ramène en fait à celle de la familière loi du demi-cercle. Le comportement des spikes nous révèle alors exactement avec quelle précision l’information \mathcal{P} peut être reconstruite à partir de l’observation \mathcal{J} . En particulier, cette reconstruction n’est possible qu’au-delà d’un seuil $\mathcal{O}(N^{\frac{d-2}{4}})$. On précise ainsi, dans le régime non-trivial, les résultats de Zhang et Xia [10].

Notre étude caractérise pleinement la qualité de l’approximation de \mathcal{J} par sa HOSVD tronquée. Cette méthode est

simple à mettre en œuvre et quasi-optimale. De plus, il s’agit d’une très bonne initialisation pour un algorithme d’estimation de la *meilleure* approximation de petit rang. Ces résultats sont donc une première étape vers l’étude précise de cette dernière.

Références

- [1] Gérard BEN AROUS, Daniel Zhengyu HUANG et Jiaoyang HUANG : Long Random Matrices and Tensor Unfolding. *arXiv :2110.10210 [cs, math, stat]*, octobre 2021. arXiv : 2110.10210.
- [2] Romain COUILLET et Zhenyu LIAO : *Random Matrix Methods for Machine Learning*. Cambridge University Press, Cambridge, 2022.
- [3] Lieven DE LATHAUWER, Bart DE MOOR et Joos VANDEWALLE : A Multilinear Singular Value Decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, janvier 2000. Publisher : Society for Industrial and Applied Mathematics.
- [4] Lieven DE LATHAUWER, Bart DE MOOR et Joos VANDEWALLE : On the Best Rank-1 and Rank-(R1, R2, ..., RN) Approximation of Higher-Order Tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324–1342, janvier 2000. Publisher : Society for Industrial and Applied Mathematics.
- [5] Lars GRASEDYCK, Daniel KRESSNER et Christine TOBLER : A literature survey of low-rank tensor approximation techniques, février 2013. arXiv :1302.7121 [quant-ph].
- [6] Christopher HILLAR et Lek-Heng LIM : Most tensor problems are NP-hard, juin 2013. arXiv :0911.1393 [cs, math].
- [7] Mohamed El Amine SEDDIK, Maxime GUILLAUD et Romain COUILLET : When Random Tensors meet Random Matrices. *arXiv :2112.12348 [math, stat]*, janvier 2022. arXiv : 2112.12348.
- [8] Mohamed El Amine SEDDIK, Maxime GUILLAUD et Alexis DECURRINGE : On the Accuracy of Hotelling-Type Tensor Deflation : A Random Tensor Analysis, novembre 2022. arXiv :2211.09004 [math, stat].
- [9] Nicholas D. SIDIROPOULOS, Lieven DE LATHAUWER, Xiao FU, Kejun HUANG, Evangelos E. PAPALEXAKIS et Christos FALOUTSOS : Tensor Decomposition for Signal Processing and Machine Learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, juillet 2017. Conference Name : IEEE Transactions on Signal Processing.
- [10] Anru ZHANG et Dong XIA : Tensor SVD : Statistical and Computational Limits, janvier 2020. arXiv :1703.02724 [cs, math, stat].

⁹Cette quantité est en fait liée à la notion de distance $\sin \Theta$ [10].