

Quand les tenseurs aléatoires rencontrent les matrices aléatoires

Mohamed El Amine SEDDIK¹, Maxime GUILLAUD¹, Romain COUILLET²

¹Mathematical and Algorithmic Sciences Laboratory, Huawei Technologies France
18 Quai du Point du Jour, 92100 Boulogne-Billancourt

²Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG
700 avenue Centrale, 38401 Saint Martin d'Hères

mohamed.seddik@huawei.com, maxime.guillaud@huawei.com,
romain.couillet@univ-grenoble-alpes.fr

Résumé – En s'appuyant sur la théorie des matrices aléatoires, cet article étudie les perturbations de rang-1 des tenseurs aléatoires gaussiens *non-symétriques*, également appelés *modèles spike tensoriels*. Ces modèles généralisent aux tenseurs aléatoires la notion de perturbations de rang faible des matrices aléatoires, qui a été largement exploitée dans divers domaines du traitement du signal et de l'apprentissage automatique. Techniquement, notre analyse consiste à associer le tenseur aléatoire étudié à une matrice aléatoire construite à partir de *contractions* de celui-ci avec ses vecteurs singuliers, définis comme étant les composantes de la meilleure approximation rang-1 du tenseur étudié. Cette association apparaît naturellement comme une conséquence de l'utilisation du lemme de Stein, un outil central de notre approche. Enfin, nous caractérisons les limites asymptotiques exactes, lorsque les dimensions du tenseur tendent vers l'infini, de la valeur singulière extrême ainsi que les alignements entre les vecteurs singuliers associés, avec les composantes de la perturbation rang-1 sous-jacente. Nos résultats permettent donc de décrire, pour la première fois, le comportement des modèles spike tensoriels *asymétriques*, ouvrant la voie à l'analyse et à la compréhension du comportement de modèles tensoriels aléatoires plus complexes.

Abstract – Based on random matrix theory, this paper studies rank-1 perturbations of *non-symmetric* Gaussian random tensors, also called *spiked tensor models*. These models generalize the notion of low rank perturbations of random matrices, which have been widely exploited in various fields of signal processing and machine learning, to random tensors. Technically, our analysis consists in associating the studied random tensor to a random matrix constructed from *contractions* of it with its singular vectors, defined as being the components of the best rank-1 approximation of the studied tensor. This association appears naturally as a consequence of the use of Stein's lemma, a central tool of our approach. Finally, we characterize the exact asymptotic limits, when the dimensions of the tensor tend to infinity, of the extreme singular value as well as the alignments between the associated singular vectors, with the components of the underlying rank-1 perturbation. Therefore, our results allow to describe, for the first time, the behavior of *asymmetric* spiked tensor models, opening the way to the analysis and understanding of the behavior of more complex random tensor models.

1 Introduction

L'extraction de structures latentes et de faible dimension à partir de données brutes est une étape clé dans diverses applications d'apprentissage automatique et de traitement du signal. Nous nous intéressons dans cet article aux techniques modernes qui reposent sur l'extraction de telles structures à partir de modèles tensoriels aléatoires de rang faible [AGH⁺14], qui étendent les idées des données de type matriciel aux données structurées en tenseurs. Nous renvoyons le lecteur à [SHL14, RSG17] et aux références qui s'y trouvent, qui présentent un vaste ensemble d'applications des méthodes de décomposition tensorielle à l'apprentissage statistique, notamment la réduction de la dimensionnalité, l'apprentissage supervisé et non supervisé, l'apprentissage de sous-espaces pour l'extraction de caractéristiques, la récupération de tenseurs de rang faible, etc.

Bien que les modèles de matrices aléatoires aient été largement étudiés et bien compris dans la littérature, la compréhension des modèles de tenseurs aléatoires en est encore à ses pre-

miers balbutiements et les idées issues de l'analyse des matrices aléatoires ne s'étendent pas facilement aux tenseurs d'ordre supérieur. En effet, la notion de *résolvante* qui est au cœur des techniques d'analyse des matrices aléatoires ne se généralise pas aux tenseurs ou du moins est difficilement généralisable.

Dans la présente étude, nous considérons *le modèle tensoriel spike asymétrique*, qui consiste en un tenseur $\mathcal{T} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ d'ordre d et de la forme

$$\mathcal{T} = \beta \mathbf{x}_1 \otimes \dots \otimes \mathbf{x}_d + \frac{1}{\sqrt{n}} \mathcal{X}, \quad (1)$$

où $(\mathbf{x}_1, \dots, \mathbf{x}_d) \in \mathbb{S}^{n_1-1} \times \dots \times \mathbb{S}^{n_d-1}$ avec \mathbb{S}^{p-1} la sphère unité de dimension p , \mathcal{X} est un tenseur aléatoire à entrées gaussiennes indépendantes et identiquement distribuées $X_{i_1 \dots i_d} \sim \mathcal{N}(0, 1)$ et $n = \sum_{i=1}^d n_i$. Typiquement, étant donné \mathcal{T} , on cherche à estimer le tenseur de rang-1 (ou *spike*) $\beta \mathbf{x}_1 \otimes \dots \otimes \mathbf{x}_d$, où β est un paramètre qui contrôle le rapport signal-à-bruit (SNR) et $\mathbf{x}_1 \otimes \dots \otimes \mathbf{x}_d$ est défini comme le produit externe des vecteurs $\mathbf{x}_1, \dots, \mathbf{x}_d$, c'est-à-dire, $(\mathbf{x}_1 \otimes \dots \otimes \mathbf{x}_d)_{i_1 \dots i_d} =$

$\prod_{j=1}^d x_{ji}$ où x_{ji} est l'entrée i_j du vecteur \mathbf{x}_j .

Dans cet article, nous nous intéressons à l'étude de l'estimateur du maximum de vraisemblance de la meilleure approximation de rang 1 du tenseur \mathcal{T} , qui est formellement donné par

$$\arg \min_{\lambda \in \mathbb{R}^+, \mathbf{u}_i \in \mathbb{S}^{n_i-1}} \|\mathcal{T} - \lambda \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d\|_{\text{F}}^2. \quad (2)$$

Étant donné le $(d+1)$ -uplet $(\lambda^*, \mathbf{u}_1^*, \dots, \mathbf{u}_d^*)$ découlant de l'équation (2) à travers les identités dans (5) et vérifiant des hypothèses supplémentaires que nous détaillerons par la suite (Hypothèse 2), nous caractérisons les limites asymptotiques de la valeur singulière λ^* et des alignements $\langle \mathbf{x}_i, \mathbf{u}_i^* \rangle$ lorsque les dimensions du tenseur $n_i \rightarrow \infty$ avec $\frac{n_i}{\sum_{j=1}^d n_j} \rightarrow c_i \in (0, 1)$.

Techniquement, nous montrons tout d'abord que le tenseur aléatoire considéré \mathcal{T} peut être mis en correspondance avec une matrice aléatoire *symétrique* $\mathbf{T}_d^* \in \mathbb{R}^{n \times n}$, construite à partir des contractions de \mathcal{T} avec $d-2$ directions parmi ses vecteurs singuliers $\mathbf{u}_1^*, \dots, \mathbf{u}_d^*$. Ensuite, en nous appuyant sur des outils de la théorie des matrices aléatoires, nous caractérisons d'abord la distribution limite des valeurs propres de \mathbf{T}_d^* (appelée mesure spectrale), puis nous fournissons des expressions exactes des limites asymptotiques de la valeur singulière λ^* et des alignements $\langle \mathbf{x}_i, \mathbf{u}_i^* \rangle$. Nous montrons précisément que :

Théorème 1 (Voir [SGC21]) *Pour tout $d \geq 3$, il existe $\beta_s > 0$ tel que pour $\beta > \beta_s$*

$$\lambda^* \xrightarrow{p.s.} \lambda^\infty, \quad |\langle \mathbf{x}_i, \mathbf{u}_i^* \rangle| \xrightarrow{p.s.} q_i(\lambda^\infty)$$

où λ^∞ satisfait l'équation $f(\lambda^\infty, \beta) = 0$ avec

$$- f(z, \beta) = z + g(z) - \beta \prod_{i=1}^d q_i(z, \beta)$$

$$- q_i(z, \beta) = \left(\frac{\alpha_i(z, \beta)^{d-3}}{\prod_{j \neq i} \alpha_j(z, \beta)} \right)^{\frac{1}{2d-4}}$$

$$- \alpha_i(z, \beta) = \frac{\beta}{z + g(z) - g_i(z)}$$

$$- g_i(z) = \frac{g(z) + z}{2} - \frac{\sqrt{4c_i + (g(z) + z)^2}}{2}$$

$$- g(z) = \sum_{i=1}^d g_i(z)$$

tandis que pour $\beta \in [0, \beta_s]$, λ^* est bornée par une constante avec grande probabilité et $|\langle \mathbf{x}_i, \mathbf{u}_i^* \rangle| \xrightarrow{p.s.} 0$.

2 Travaux liés

Modèles spike tensoriels : Des efforts considérables ont été réalisés pour étudier les performances des méthodes d'approximation de rang faible de tenseurs aléatoires dans le régime des grandes dimensions – lorsque les dimensions du tenseur $n_i \rightarrow \infty$, en considérant toutefois des modèles tensoriels *symétriques* où $n_1 = \dots = n_d = n/d$ (d divise n sans perte de généralité), $\mathbf{x}_1 = \dots = \mathbf{x}_d$, et en supposant que le bruit est symétrique [MR14, PWB20, LML⁺17, Han19, SMBC⁺19, JLM20].

En particulier, pour $d=2$ le modèle spike tensoriel ci-dessus devient un modèle spike matriciel. Il est bien connu que dans le régime de grande dimensions, il existe une valeur critique β_c du SNR en dessous de laquelle il est théoriquement impossible de détecter ou récupérer le signal, alors qu'au-dessus de β_c , il

est possible de détecter le signal et d'estimer ses composantes en un temps polynomial en utilisant la décomposition en valeurs singulières (SVD). Ce phénomène est parfois connu sous le nom de transition de phase BBP (Baik, Ben Arous et Pécché) [BAP05, BGN11, CDMF09, Péc06].

Dans le cas du modèle spike tensoriel (symétrique) pour $d \geq 3$, il existe également une valeur critique¹ $\beta_c(d)$ en dessous de laquelle il est théoriquement impossible de détecter/récupérer le signal, tandis qu'au-dessus de $\beta_c(d)$, la récupération de celui-ci est possible avec l'estimateur du maximum de vraisemblance (MV) en théorie. Le calcul du MV dans le cas d'une matrice correspond au calcul de ses plus grands vecteurs singuliers, qui a une complexité polynomiale, alors que pour $d \geq 3$, le calcul du MV est généralement NP-complexe [MR14, BCRT20]. Ainsi, une transition de phase plus pratique pour les tenseurs consiste à caractériser la valeur critique algorithmique $\beta_a(d, n)$ (qui peut dépendre de n) au-dessus de laquelle la récupération du signal est possible en temps polynomial.

Montanari et Richard [MR14] ont tout d'abord introduit le modèle spike tensoriel *symétrique* (de la forme $\mathcal{Y} = \mu \mathbf{x}^{\otimes d} + \mathcal{W} \in \otimes^d \mathbb{R}^{n/d}$ avec \mathcal{W} aléatoire et symétrique) et ont également considéré les aspects algorithmiques associés. En particulier, ils ont utilisé des heuristiques pour mettre en évidence que la récupération du signal \mathbf{x} est possible, avec l'AMP (Approximate Message Passing) ou la méthode d'itération de puissance, en temps polynomial² à condition que $\mu \geq \mathcal{O}(n^{\frac{d-1}{2}})$. Cette transition de phase a ensuite été prouvée rigoureusement pour l'AMP dans [LML⁺17, JLM20] et plus récemment pour l'algorithme d'itération de puissance dans [HHYC20].

[MR14] ont ensuite introduit une méthode de décomposition tensorielle basée sur le dépliage tensoriel, qui consiste à déplier \mathcal{Y} en une matrice $(n/d) \times (n/d)^{d-1}$, $\text{Mat}(\mathcal{Y}) = \mu \mathbf{x} \mathbf{y}^\top + \text{Mat}(\mathcal{W})$, à laquelle on applique ensuite une SVD classique. Ils ont prédit que cette méthode récupère le signal si $\mu \geq \mathcal{O}(n^{\frac{d-2}{4}})$. Dans un article très récent de Ben Arous et al. [BAHH21], une étude des matrices aléatoires spike rectangulaires *longues*³ a été proposée. Ils ont notamment prouvé l'existence d'un SNR critique pour lequel la valeur singulière extrême et les vecteurs singuliers présentent une transition de phase de type BBP. Ils ont appliqué leur résultat pour le modèle spike tensoriel *asymétrique* de rang-1 dans l'équation (1) en utilisant la méthode de dépliage tensoriel, et ont trouvé le seuil exact obtenu par [MR14], c'est-à-dire, $\beta \geq \mathcal{O}(n^{\frac{d-2}{4}})$ pour que le dépliage tensoriel réussisse à récupérer le signal.

Dans cet article, nous généralisons ces idées au modèle dans l'équation (1) à travers une approche utilisant des outils de la théorie des matrices aléatoires.

Valeurs et vecteurs singuliers de tenseurs : Dans l'équation (2), λ et les \mathbf{u}_i 's peuvent être interprétés comme la généralisation au cas tensoriel des concepts de valeur singulière dominante et

1. Nous omettons parfois la dépendance en d s'il n'y a pas d'ambiguïté.

2. Pour l'itération de puissance ou l'AMP avec initialisation aléatoire.

3. On permet au nombre de lignes m de croître de façon polynomiale en fonction du nombre de colonnes n , c'est-à-dire, $\frac{m}{n} = n^\alpha$ avec $\alpha > 0$.

$$\begin{bmatrix} \frac{\partial \mathbf{u}_1}{\partial X_{ijk}} \\ \frac{\partial \mathbf{u}_2}{\partial X_{ijk}} \\ \frac{\partial \mathbf{u}_3}{\partial X_{ijk}} \end{bmatrix} = -\frac{1}{\sqrt{n}} \left(\underbrace{\begin{bmatrix} \mathbf{0}_{n_1 \times n_1} & \mathcal{T} \times_3 \mathbf{u}_3 & \mathcal{T} \times_2 \mathbf{u}_2 \\ (\mathcal{T} \times_3 \mathbf{u}_3)^\top & \mathbf{0}_{n_2 \times n_2} & \mathcal{T} \times_1 \mathbf{u}_1 \\ (\mathcal{T} \times_2 \mathbf{u}_2)^\top & (\mathcal{T} \times_1 \mathbf{u}_1)^\top & \mathbf{0}_{n_3 \times n_3} \end{bmatrix}}_{\Phi_3(\mathcal{T}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)} - \lambda \mathbf{I}_n \right)^{-1} \begin{bmatrix} u_{2j} u_{3k} (e_i^{n_1} - u_{1i} \mathbf{u}_1) \\ u_{1i} u_{3k} (e_j^{n_2} - u_{2j} \mathbf{u}_2) \\ u_{1i} u_{2j} (e_k^{n_3} - u_{3k} \mathbf{u}_3) \end{bmatrix} \in \mathbb{R}^n \quad (3)$$

de vecteurs singuliers associés [Lim05]. En suivant les arguments variationnels de [Lim05], l'équation (2) peut être reformulée à l'aide de contractions de \mathcal{T} comme suit

$$\max_{\prod_{i=1}^d \|\mathbf{u}_i\|=1} |\mathcal{T} \times_1 \mathbf{u}_1 \times_2 \mathbf{u}_2 \cdots \times_d \mathbf{u}_d|, \quad (4)$$

avec $[\mathcal{T} \times_k \mathbf{u}_k]_{i_1 \cdots i_{k-1} i_{k+1} \cdots i_d} \equiv \sum_{i_k=1}^{n_k} u_{ki_k} T_{i_1 \cdots i_{k-1} i_k i_{k+1} \cdots i_d}$. Le Lagrangien de (4) est $\mathcal{T} \times_1 \mathbf{u}_1 \cdots \times_d \mathbf{u}_d - \lambda \left(\prod_{i=1}^d \|\mathbf{u}_i\| - 1 \right)$ avec $\lambda > 0$. Ainsi, les points stationnaires $(\lambda, \mathbf{u}_1, \dots, \mathbf{u}_d)$, avec les \mathbf{u}_i 's étant des vecteurs unitaires, doivent satisfaire les conditions de Karush-Kuhn-Tucker, pour $i \in [d]$

$$\begin{cases} \mathcal{T} \times_1 \mathbf{u}_1 \cdots \times_{i-1} \mathbf{u}_{i-1} \times_{i+1} \mathbf{u}_{i+1} \cdots \times_d \mathbf{u}_d = \lambda \mathbf{u}_i, \\ \lambda = \mathcal{T} \times_1 \mathbf{u}_1 \times_2 \mathbf{u}_2 \cdots \times_d \mathbf{u}_d \text{ et } \|\mathbf{u}_i\| = 1. \end{cases} \quad (5)$$

Une question intéressante concerne le calcul du nombre de points stationnaires (optima locaux ou points selle) satisfaisant les identités dans l'équation (5). [BAMMN19] a notamment étudié le *paysage* d'un modèle spike tensoriel *symétrique*, quand ses dimensions tendent vers l'infini, et a constaté qu'il existe $\beta_c > 0$ tel que pour $\beta < \beta_c$ les valeurs de la fonction objective de tous les maxima locaux (y compris le global) ont tendance à se concentrer sur un petit intervalle, alors que pour $\beta > \beta_c$ la valeur atteinte par le maximum global sort de cet intervalle et augmente avec β . Plus récemment, Goulart et al. [GCC21] ont étudié un tenseur aléatoire spike *symétrique* \mathcal{Y} d'ordre 3 par matrices aléatoires, où il a été montré qu'il existe un seuil $0 < \beta_s < \beta_c$ tel que pour $\beta \in [\beta_s, \beta_c]$ il existe un optimum local du problème du maximum de vraisemblance qui est corrélé avec le signal et un tel optimum local coïncide avec l'optimum global pour $\beta > \beta_c$.

Nous conjecturons que ces observations s'étendent aux tenseurs spike *asymétriques*, à savoir qu'il existe une valeur critique $\beta_c > 0$ au-dessus de laquelle le problème du maximum de vraisemblance dans l'équation (2) admet un maximum global. Comme pour [GCC21], nos résultats actuels ne permettent pas d'exprimer un tel β_c et sa caractérisation exacte est laissée à une étude future. Cependant, pour les tenseurs aléatoires spike *asymétriques*, nous trouvons également un seuil β_s tel que pour $\beta > \beta_s$ il existe un optimum local de l'objectif du maximum de vraisemblance qui est aligné avec le signal sous-jacent mais β_s ne coïncide a priori pas avec β_c .

3 Approche et résultats principaux

3.1 Matrice aléatoire associée ($d = 3$)

Notre approche consiste en l'application du *lemme de Stein* qui est largement utilisé pour l'étude des matrices aléatoires

gaussiennes [MP67, BAP05, BGN11]. Pour rappel, le lemme de Stein permet de lier l'espérance du produit d'une variable aléatoire gaussienne avec une fonction différentiable f de celle-ci au produit de sa variance avec l'espérance de f' .

Lemme 1 (Lemme de Stein [Ste81]) *Soit $X \sim \mathcal{N}(0, \sigma^2)$ et $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continuellement différentiable ayant une croissance au plus polynomiale, $\mathbb{E}[X f(X)] = \sigma^2 \mathbb{E}[f'(X)]$ lorsque les espérances existent.*

En effet, lorsque les dimensions du tenseur tendent vers l'infini, les fonctionnelles scalaires dépendant de celui-ci ont tendance à se concentrer autour de leurs espérances. Ce phénomène est connu sous le nom de *la concentration de la mesure* [LC18, Ver20]. En particulier, à titre d'illustration, une valeur singulière λ de \mathcal{T} est une fonctionnelle scalaire de \mathcal{T} se concentrant autour de son espérance $\mathbb{E}[\lambda]$, sous l'hypothèse suivante de grandes dimensions.

Hypothèse 1 *Pour tout $i \in [d]$, lorsque $n_i \rightarrow \infty$, le ratio des dimensions $\frac{n_i}{\sum_j n_j} \rightarrow c_i \in (0, 1)$ et $d = \mathcal{O}(1)$.*

Il reste donc à évaluer l'espérance $\mathbb{E}[\lambda]$ qui se développe grâce au lemme de Stein 1 comme suit (pour $d = 3$).

$$\begin{aligned} \mathbb{E}[\lambda] &= \beta \mathbb{E} \left[\prod_{\ell=1}^3 \langle \mathbf{x}_\ell, \mathbf{u}_\ell \rangle \right] + \frac{1}{\sqrt{n}} \sum_{ijk} \mathbb{E} \left[u_{2j} u_{3k} \frac{\partial u_{1i}}{\partial X_{ijk}} \right] \\ &+ \mathbb{E} \left[u_{1i} u_{3k} \frac{\partial u_{2j}}{\partial X_{ijk}} \right] + \mathbb{E} \left[u_{1i} u_{2j} \frac{\partial u_{3k}}{\partial X_{ijk}} \right] \end{aligned}$$

où les dérivées $\frac{\partial u_{\ell i}}{\partial X_{ijk}}$ sont données par l'équation (3). Ainsi, la matrice aléatoire associée au tenseur \mathcal{T} est précisément la matrice $\mathbf{T}_3 = \Phi_3(\mathcal{T}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$. En particulier, notons que les dérivées $\frac{\partial u_{\ell i}}{\partial X_{ijk}}$ s'expriment en fonction de la *résolvante* $\mathbf{R}(z) = (\mathbf{T}_3 - z \mathbf{I}_n)^{-1}$ évaluée en λ . Plus précisément, sous l'hypothèse 1, les termes prépondérants dans l'expression de $\mathbb{E}[\lambda]$ s'expriment à l'aide de *traces* de la résolvante $\mathbf{R}(z)$ car

$$\begin{aligned} \frac{\partial u_{1i}}{\partial X_{ijk}} &\simeq -\frac{1}{\sqrt{n}} u_{2j} u_{3k} R_{ii}^{11}(\lambda), & \frac{\partial u_{2j}}{\partial X_{ijk}} &\simeq -\frac{1}{\sqrt{n}} u_{1i} u_{3k} R_{jj}^{22}(\lambda), \\ \frac{\partial u_{3k}}{\partial X_{ijk}} &\simeq -\frac{1}{\sqrt{n}} u_{1i} u_{2j} R_{kk}^{33}(\lambda) \end{aligned}$$

où $\mathbf{R}^{\ell\ell}(z) \in \mathbb{R}^{n_\ell \times n_\ell}$ sont les matrices bloc-diagonales de $\mathbf{R}(z)$. Pour la suite, nous avons besoin d'une hypothèse additionnelle sur $(\lambda, \mathbf{u}_1, \dots, \mathbf{u}_d)$ pour que $\mathbf{R}(\lambda)$ existe.

Hypothèse 2 *Il existe $(\lambda^*, \mathbf{u}_1^*, \dots, \mathbf{u}_d^*)$ tel que $\lambda^* n$ n'est pas valeur propre de $\Phi_d(\mathcal{T}, \mathbf{u}_1^*, \dots, \mathbf{u}_d^*)$ et $\forall i \in [d], |\langle \mathbf{x}_i, \mathbf{u}_i^* \rangle| > 0$.*

L'hypothèse 2 n'est pas très restrictive dans le sens où elle est vérifiée par tous les maxima locaux de la fonction objective dans (4). Ceci découle des conditions nécessaires vérifiées par ces maxima comme détaillé dans [SGC21, Remarque 1].

La résolvante est un objet fondamental en théorie des matrices aléatoires car celle-ci est intrinsèquement liée à la notion de la transformée de Stieltjes, qui est au coeur de l'étude des mesures spectrales de grandes matrices aléatoires.

Définition 1 *Étant donnée une mesure de probabilité ν , la transformée de Stieltjes de ν est définie par $g_\nu(z) = \int \frac{d\nu(\lambda)}{\lambda - z}$, $z \in \mathbb{C} \setminus \mathcal{S}(\nu)$ où $\mathcal{S}(\nu)$ est le support de ν .*

En effet, étant donné une matrice *symétrique* \mathbf{S} de dimension $n \times n$ avec λ_i ses valeurs propres, la transformée de Stieltjes de la mesure spectrale empirique (MeSE) de \mathbf{S} , définie par $\nu_{\mathbf{S}} = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}$, est $g_{\nu_{\mathbf{S}}}(z) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i - z} = \frac{1}{n} \text{tr} \mathbf{R}_{\mathbf{S}}(z)$ où $\mathbf{R}_{\mathbf{S}}(z) = (\mathbf{S} - z\mathbf{I}_n)^{-1}$ est la résolvante de \mathbf{S} . Naturellement, la notion de la transformée de Stieltjes, s'exprimant en fonction de la *trace* de $\mathbf{R}_{\mathbf{S}}(z)$, apparaît comme un concept adapté pour évaluer les fonctionnelles d'intérêt ci-dessus.

3.2 Mesure spectrale limite

En exploitant des identités algébriques supplémentaires de la résolvante, nous parvenons au résultat suivant qui décrit la mesure spectrale limite de la matrice $\mathbf{T}_d^* = \Phi_d(\mathcal{T}, \mathbf{u}_1^*, \dots, \mathbf{u}_d^*)$.

Théorème 2 *Sous les hypothèses 1 et 2, la MeSE de \mathbf{T}_d^* converge vers une mesure déterministe ν dont la transformée de Stieltjes est donnée par $g(z) = \sum_{i=1}^d g_i(z)$ telle que $\Im[g(z)] > 0$ pour $\Im[z] > 0$, où $\frac{1}{n} \text{tr} \mathbf{R}^{ii}(z) \xrightarrow{p.s.} g_i(z) = \frac{g(z)+z}{2} - \frac{\sqrt{4c_i+(g(z)+z)^2}}{2}$.*

Dans le cas particulier où toutes les dimensions de \mathcal{T} sont égales, i.e. $c_i = \frac{1}{d}$, $g(z)$ décrit une loi en demi-cercle de support $\mathcal{S}(\nu) = [-2\sqrt{\frac{d-1}{d}}, 2\sqrt{\frac{d-1}{d}}]$ et dont la fonction de densité est donnée par $\nu(dx) = \frac{d}{2(d-1)\pi} \sqrt{\left(\frac{4(d-1)}{d} - x^2\right)^+}$.

3.3 Valeur singulière et alignements

De manière similaire les alignements $\langle \mathbf{x}_i, \mathbf{u}_i^* \rangle$ se concentrent autour de leurs espérances et à l'aide du lemme de Stein 1 nous obtenons le résultat du Théorème 1 qui exploite également le Théorème 2. En particulier, pour $d = 3$ avec $c_i = \frac{1}{3}$, on retrouve le corollaire suivant exprimant *explicitement* les limites asymptotiques en fonction du SNR β .

Corollaire 1 *Si $d = 3$ et $c_i = \frac{1}{3}$, alors pour $\beta > \beta_s = \frac{2\sqrt{3}}{3}$*

$$\left\{ \begin{array}{l} \lambda^* \xrightarrow{p.s.} \sqrt{\frac{\beta^2}{2} + 2} + \frac{\sqrt{3}\sqrt{(3\beta^2-4)^3}}{18\beta} \\ |\langle \mathbf{x}_i, \mathbf{u}_i^* \rangle| \xrightarrow{p.s.} \frac{\sqrt{9\beta^2-12} + \frac{\sqrt{3}\sqrt{(3\beta^2-4)^3}}{\beta}}{6\sqrt{2}\beta} + \frac{\sqrt{9\beta^2+36} + \frac{\sqrt{3}\sqrt{(3\beta^2-4)^3}}{\beta}}{6\sqrt{2}\beta} \end{array} \right.$$

Notre résultat permet également de décrire le comportement des modèles spike matriciels. En effet, même si nos formules

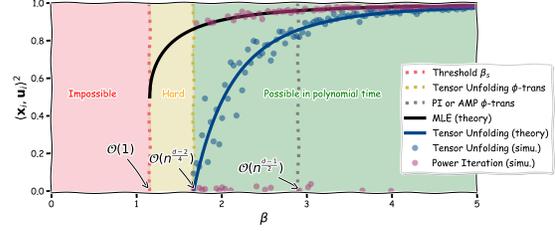


FIGURE 1 – Les différents seuils du SNR représentés en différentes régions. Les alignements asymptotiques pour un tenseur cubique en noir et sa version dépliée en bleu. Les simulations sont obtenues avec l'algorithme d'itération de puissance et le dépliage tensoriel respectivement, appliqués sur un tenseur cubique de dimensions $n_i = 70$.

ne sont pas définies pour $d = 2$ (voir Théorème 1), le cas matriciel peut être obtenu en prenant $d = 3$ avec $c_3 = 0$, auquel cas le modèle spike tensoriel devient un modèle spike matriciel. Nous retrouvons donc le corollaire suivant qui décrit le comportement des modèles spike matriciels.

Corollaire 2 *Soit $c \in (0, 1)$ et $\kappa(\beta, c) = \beta \sqrt{\frac{\beta^2(\beta^2+1)-c(c-1)}{(\beta^4+c(c-1))(\beta^2+1-c)}}$, si $d = 3$ avec $c_1 = c$ et $c_2 = 1 - c$, alors pour $\beta > \beta_s = \sqrt{4c(1-c)}$ on a $\lambda^* \xrightarrow{p.s.} \sqrt{\beta^2 + 1 + \frac{c(1-c)}{\beta^2}}$ et pour tout $i \in \{1, 2\}$, $|\langle \mathbf{x}_i, \mathbf{u}_i^* \rangle| \xrightarrow{p.s.} \frac{1}{\kappa(\beta, c_i)}$. Tandis que pour $\beta \in [0, \beta_s]$, $\lambda^* \xrightarrow{p.s.} \sqrt{1 + 2\sqrt{c(1-c)}}$.*

4 Discussion

Nos résultats permettent de décrire *pour la première fois* le comportement asymptotique de modèles spike tensoriels *non-symétriques*. Nous illustrons ces résultats dans la Figure 1 qui représente les différents seuils du SNR ainsi que les alignements correspondants. En effet, notre théorie exhibe un seuil β_s d'ordre 1 en dessous duquel il est impossible de détecter et de retrouver le signal. Nous précisons toutefois que β_s représente la valeur à partir de laquelle les équations sont algébriquement définies et donc ne représente pas forcément la transition de phase du maximum global de vraisemblance comme discuté dans [GCC21, SGC21]. En particulier, le Corollaire 2 permet de décrire le comportement de la méthode du dépliage tensoriel et nous retrouvons le même seuil que Ben Arous et al. [BAHH21]. Nous invitons le lecteur à découvrir [SGC21] la version longue du présent article pour plus de détails et discussions sur nos résultats. Notons toutefois que nos résultats s'applique également à un modèle de rang faible en imposant des conditions d'orthogonalité sur les composantes du signal. Ceci étant décrit plus en détail dans [SGC21, Section 6].

Enfin, nous précisons à travers cet article la capacité des outils de la théorie des matrices aléatoires à décrire de manière simple et surprenante le comportement de modèles tensoriels asymétriques inconnus jusqu'alors, ouvrant ainsi la brèche à l'analyse et la compréhension des comportements de modèles tensoriels aléatoires plus complexes.

Références

- [AGH⁺14] Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of machine learning research*, 15 :2773–2832, 2014.
- [BAHH21] Gérard Ben Arous, Daniel Zhengyu Huang, and Jiaoyang Huang. Long random matrices and tensor unfolding. *arXiv preprint arXiv :2110.10210*, 2021.
- [BAMMN19] Gerard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11) :2282–2330, 2019.
- [BAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5) :1643–1697, 2005.
- [BCRT20] Giulio Biroli, Chiara Cammarota, and Federico Ricci-Tersenghi. How to iron out rough landscapes and get optimal performances : averaged gradient descent and its application to tensor pca. *Journal of Physics A : Mathematical and Theoretical*, 53(17) :174003, 2020.
- [BGN11] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1) :494–521, 2011.
- [CDMF09] Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. The largest eigenvalues of finite rank deformation of large wigner matrices : convergence and nonuniversality of the fluctuations. *The Annals of Probability*, 37(1) :1–47, 2009.
- [GCC21] José Henrique Goulart, Romain Couillet, and Pierre Comon. A random matrix perspective on random tensors. *stat*, 1050 :2, 2021.
- [Han19] Madeline Curtis Handschy. *Phase Transition in Random Tensors with Multiple Spikes*. PhD thesis, University of Minnesota, 2019.
- [HHYC20] Jiaoyang Huang, Daniel Z Huang, Qing Yang, and Guang Cheng. Power iteration for tensor pca. *arXiv preprint arXiv :2012.13669*, 2020.
- [JLM20] Aukosh Jagannath, Patrick Lopatto, and Leo Miolane. Statistical thresholds for tensor PCA. *The Annals of Applied Probability*, 30(4) :1910–1933, 2020.
- [LC18] Cosme Louart and Romain Couillet. Concentration of measure and large random matrices with an application to sample covariance matrices. *arXiv preprint arXiv :1805.08295*, 2018.
- [Lim05] Lek-Heng Lim. Singular values and eigenvalues of tensors : a variational approach. In *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, pages 129–132, 2005.
- [LML⁺17] Thibault Lesieur, Léo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, pages 511–515, 2017.
- [MP67] Vladimir A Marčenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4) :457, 1967.
- [MR14] Andrea Montanari and Emile Richard. A statistical model for tensor PCA. *arXiv preprint arXiv :1411.1076*, 2014.
- [Péc06] Sandrine Péché. The largest eigenvalue of small rank perturbations of hermitian random matrices. *Probability Theory and Related Fields*, 134(1) :127–173, 2006.
- [PWB20] Amelia Perry, Alexander S Wein, and Afonso S Bandeira. Statistical limits of spiked tensor models. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 230–264. Institut Henri Poincaré, 2020.
- [RSG17] Stephan Rabanser, Oleksandr Shchur, and Stephan Günnemann. Introduction to tensor decompositions and their applications in machine learning. *arXiv preprint arXiv :1711.10781*, 2017.
- [SGC21] Mohamed El Amine Seddik, Maxime Guillaud, and Romain Couillet. When random tensors meet random matrices. *arXiv preprint arXiv :2112.12348*, 2021.
- [SHL14] Will Wei Sun, Botao Hao, and Lexin Li. Tensors in modern statistical learning. *Wiley StatsRef : Statistics Reference Online*, pages 1–25, 2014.
- [SMBC⁺19] Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, and Lenka Zdeborová. Who is afraid of big bad minima ? analysis of gradient-flow in spiked matrix-tensor models. *Advances in Neural Information Processing Systems*, 32, 2019.
- [Ste81] Charles M Stein. Estimation of the mean of a multivariate normal distribution. *The annals of Statistics*, pages 1135–1151, 1981.
- [Ver20] Roman Vershynin. Concentration inequalities for random tensors. *Bernoulli*, 26(4) :3139–3162, 2020.