Quand les tenseurs aléatoires rencontrent les matrices aléatoires

Gretsi 2022

MEA. Seddik[1], **M. Guillaud**[1] **& R. Couillet**[2]

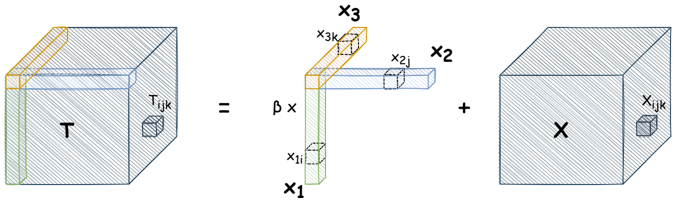
https://melaseddik.github.io/

[1] Mathematical and Algorithmic Sciences Laboratory, Huawei Technologies France
[2] Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG

Nancy, 9 septembre 2022

Sommaire

Introduction : Modèle tensoriel *spike* asymétrique



Nous considérons le modèle suivant : $(\boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \otimes \boldsymbol{x}_3)_{ijk} = x_{1i} x_{2j} x_{3k}$

$$\mathbf{T} = \underbrace{\beta \boldsymbol{x}_1 \otimes \cdots \otimes \boldsymbol{x}_d}_{\text{signal}} + \frac{1}{\sqrt{n}} \underbrace{\mathbf{X}}_{\text{bruit}} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$$

où $\beta \geq 0$, $\|\boldsymbol{x}_i\| = 1$, $X_{i_1 \ldots i_d} \sim \mathcal{N}(0,1)$ i.i.d. et $n = \sum_{i=1}^{d} n_i$.

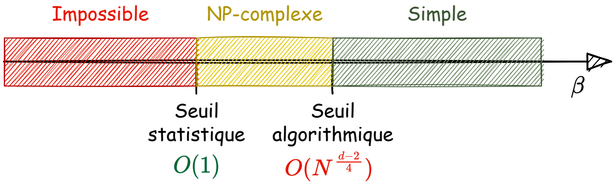
- Peut-on recouvrir le signal en théorie ? pour quelle valeur **critique** de β ?
- Quel est **l'alignement** $\langle \boldsymbol{x}_i, \boldsymbol{u}_i \rangle$ entre le signal et un estimateur $\boldsymbol{u}_i(\mathbf{T})$?
- Existe-t-il un algorithme pour recouvrir le signal en **temps polynomial** ?

Résultats dans la littérature : cas symétrique

Initialement introduit par **(Montanari & Richard, 2014)**

$$\mathbf{Y} = \beta \boldsymbol{x}^{\otimes d} + \frac{1}{\sqrt{N}} \mathbf{W} \in \mathbb{R}^{N \times \cdots \times N}$$

où $\|\boldsymbol{x}\| = 1$ et $\mathbf{W}$ à entrées aléatoires Gaussiennes et **symétrique**.Ce modèle est une extension naturelle du modèle spike matriciel $\boldsymbol{Y} = \beta \boldsymbol{x}\boldsymbol{x}^\top + \frac{1}{\sqrt{N}} \boldsymbol{W}$.



D'autres travaux dans la littérature : **(Montanari et al., 2015)**, **(Hopkins et al., 2020)**, **(Kim et al., 2017)**, **(Ben Arous et al., 2019)**, **(Jagannath et al, 2020)**, **(Perry et al., 2020)**, **(Ros et al., 2020)**, **(Goulart et al., 2021)**.

Dont **Goulart et al. "*A random matrix perspective on random tensors*", 2021.**

Approche matrices aléatoires (Goulart et al., 2021)

Problème d'optimisation du maximum de vraisemblance (pour $d = 3$) :

$$\min_{\lambda>0,\, \|\boldsymbol{u}\|=1} \left\| \mathbf{Y} - \lambda \boldsymbol{u}^{\otimes 3} \right\|_F^2 \quad \Leftrightarrow \quad \max_{\|\boldsymbol{u}\|=1} \langle \mathbf{Y}, \boldsymbol{u} \otimes \boldsymbol{u} \otimes \boldsymbol{u} \rangle$$

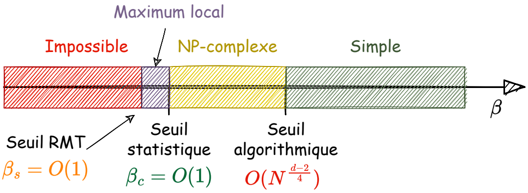
Les points critiques satisfont **(Lim, 2005)** :

$$\mathbf{Y}(\boldsymbol{u}, \boldsymbol{u}) = \lambda \boldsymbol{u} \quad \Leftrightarrow \quad \mathbf{Y}(\boldsymbol{u})\boldsymbol{u} = \lambda \boldsymbol{u}, \quad \|\boldsymbol{u}\| = 1$$

où $(\mathbf{Y}(\boldsymbol{u}, \boldsymbol{u}))_i = \sum_{jk} u_j u_k Y_{ijk}$ et $(\mathbf{Y}(\boldsymbol{u}))_{ij} = \sum_k u_k Y_{ijk}$. L'estimateur du MV $\hat{\boldsymbol{x}}$ est le vecteur propre dominant de $\mathbf{Y}(\hat{\boldsymbol{x}})$: $\mathbf{Y}(\hat{\boldsymbol{x}})\hat{\boldsymbol{x}} = \|\mathbf{Y}\|\hat{\boldsymbol{x}}$.

Ainsi, l'approche de **(Goulart et al., 2021)** est d'étudier la matrice suivante :

$$\mathbf{Y}(\boldsymbol{u}) = \beta \langle \boldsymbol{x}, \boldsymbol{u} \rangle \boldsymbol{x}\boldsymbol{x}^\top + \frac{1}{\sqrt{N}} \mathbf{W}(\boldsymbol{u}) \in \mathbb{R}^{N \times N}$$



Sommaire

Valeurs singulières et vecteurs singuliers de tenseurs

Problème d'optimisation du maximum de vraisemblance (pour $d = 3$) :

$$\min_{\lambda>0,\, \|\boldsymbol{u}_i\|=1} \|\mathbf{T} - \lambda \boldsymbol{u}_1 \otimes \boldsymbol{u}_2 \otimes \boldsymbol{u}_3\|_F^2 \quad \Leftrightarrow \quad \max_{\prod_{i=1}^3 \|\boldsymbol{u}_i\|=1} \langle \mathbf{T}, \boldsymbol{u}_1 \otimes \boldsymbol{u}_2 \otimes \boldsymbol{u}_3 \rangle$$

Les points critiques satisfont **(Lim, 2005)** :

$$\mathbf{T}(\boldsymbol{I}_{n_1}, \boldsymbol{u}_2, \boldsymbol{u}_3) = \lambda \boldsymbol{u}_1,\ \mathbf{T}(\boldsymbol{u}_1, \boldsymbol{I}_{n_2}, \boldsymbol{u}_3) = \lambda \boldsymbol{u}_2,\ \mathbf{T}(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{I}_{n_3}) = \lambda \boldsymbol{u}_3$$

avec $\|\boldsymbol{u}_i\| = 1$ pour tout $i \in [3]$ et $(\mathbf{T}(\boldsymbol{I}_{n_1}, \boldsymbol{u}_2, \boldsymbol{u}_3))_i = \sum_{jk} u_{2j} u_{3k} T_{ijk}$.

- Contrairement au cas symétrique, le choix de la matrice de *contraction* à étudier n'est pas immédiat. Par exemple :

$$\mathbf{T}(\boldsymbol{u}_3) \equiv \mathbf{T}(\boldsymbol{I}_{n_1}, \boldsymbol{I}_{n_2}, \boldsymbol{u}_3) = \beta \langle \boldsymbol{x}_3, \boldsymbol{u}_3 \rangle \boldsymbol{x}_1 \boldsymbol{x}_2^\top + \frac{1}{\sqrt{n}} \mathbf{X}(\boldsymbol{I}_{n_1}, \boldsymbol{I}_{n_2}, \boldsymbol{u}_3) \in \mathbb{R}^{n_1 \times n_2}$$

Objectifs :

- Évaluer les limites asymptotiques de λ^* et $\langle \boldsymbol{x}_i, \boldsymbol{u}_i^* \rangle$ associés (à priori) à l'estimateur MV quand les $n_i \to \infty$.
- Définir une matrice aléatoire **symétrique** équivalente à $\mathbf{T}$ à étudier.

Matrice aléatoire associée à T

Lemme de Stein : Soit $X \sim \mathcal{N}(0,1)$, alors $\mathbb{E}[X f(X)] = \mathbb{E}[f'(X)]$.

Rappelons $\lambda = \mathbf{T}(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3) = \frac{1}{\sqrt{n}} \sum_{ijk} u_{1i} u_{2j} u_{3k} X_{ijk} + \beta \prod_{i=1}^{3} \langle \boldsymbol{x}_i, \boldsymbol{u}_i \rangle$.

$$\mathbb{E}[\lambda] = \frac{1}{\sqrt{n}} \sum_{ijk} \mathbb{E}\left[u_{2j} u_{3k} \frac{\partial u_{1i}}{\partial X_{ijk}}\right] + \mathbb{E}\left[u_{1i} u_{3k} \frac{\partial u_{2j}}{\partial X_{ijk}}\right] + \mathbb{E}\left[u_{1i} u_{2j} \frac{\partial u_{3k}}{\partial X_{ijk}}\right] + \cdots$$

$$\begin{bmatrix} \frac{\partial \boldsymbol{u}_1}{\partial X_{ijk}} \\ \frac{\partial \boldsymbol{u}_2}{\partial X_{ijk}} \\ \frac{\partial \boldsymbol{u}_3}{\partial X_{ijk}} \end{bmatrix} \simeq -\frac{1}{\sqrt{n}} \left(\underbrace{\begin{bmatrix} \mathbf{0}_{n_1 \times n_1} & \mathbf{T}(\boldsymbol{u}_3) & \mathbf{T}(\boldsymbol{u}_2) \\ \mathbf{T}(\boldsymbol{u}_3)^\intercal & \mathbf{0}_{n_2 \times n_2} & \mathbf{T}(\boldsymbol{u}_1) \\ \mathbf{T}(\boldsymbol{u}_2)^\intercal & \mathbf{T}(\boldsymbol{u}_1)^\intercal & \mathbf{0}_{n_3 \times n_3} \end{bmatrix}}_{\boldsymbol{\Phi}_3(\mathbf{T}, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)} - \lambda \boldsymbol{I}_n \right)^{-1} \begin{bmatrix} u_{2j} u_{3k} \boldsymbol{e}_i \\ u_{1i} u_{3k} \boldsymbol{e}_j \\ u_{1i} u_{2j} \boldsymbol{e}_k \end{bmatrix}$$

La matrice résolvante: $\boldsymbol{R}(z) = (\boldsymbol{\Phi}_3(\mathbf{T}, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3) - z \boldsymbol{I}_n)^{-1}$.
Quand $n_i \to \infty$, les termes prépondérants dépendent de la **trace** de $\boldsymbol{R}(z)$,

$$\boxed{\lambda + \frac{1}{n} \operatorname{tr} \boldsymbol{R}(\lambda) = \beta \prod_{i=1}^{3} \langle \boldsymbol{x}_i, \boldsymbol{u}_i \rangle}$$

Mesure spectrale de $\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$

Transformée de Stieltjes : La transformée de Stieltjes d'une mesure de probabilité ν est $g_\nu(z) = \int \frac{d\nu(\lambda)}{\lambda - z}$, $z \in \mathbb{C} \setminus \mathcal{S}(\nu)$.

Pour $\boldsymbol{S} \in \mathrm{Sym}_n$ avec λ_i ses valeurs propres, *la mesure spectrale empirique* (MeSE) de $\boldsymbol{S}$ et la transformée de Stieltjes associée sont :

$$\nu_{\boldsymbol{S}} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i}, \; g_{\nu_{\boldsymbol{S}}}(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i - z} = \frac{1}{n}\mathrm{tr}\, \boldsymbol{R}_{\boldsymbol{S}}(z), \; z \in \mathbb{C} \setminus \mathcal{S}(\nu_{\boldsymbol{S}})$$

où $\boldsymbol{R}_{\boldsymbol{S}}(z) = (\boldsymbol{S} - z\boldsymbol{I}_n)^{-1}$ est la résolvante de $\boldsymbol{S}$.

Théorème 1. Quand $n_i \to \infty$ avec $\frac{n_i}{\sum_j n_j} \to c_i \in [0, 1]$, la MeSE de $\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ converge vers une **mesure déterministe** ν dont la transformée de Stieltjes est $\frac{1}{n}\mathrm{tr}\, \boldsymbol{R}(z) \xrightarrow{\text{p.s.}} g(z) = \sum_{i=1}^{d} g_i(z)$ vérifiant $\Im[g(z)] > 0$ pour $\Im[z] > 0$, avec

$$\frac{1}{n}\mathrm{tr}\, \boldsymbol{R}^{ii}(z) \xrightarrow{\text{p.s.}} g_i(z) = \frac{g(z) + z}{2} - \frac{\sqrt{4c_i + (g(z) + z)^2}}{2}, \quad z \in \mathbb{C} \setminus \mathcal{S}(\nu)$$

Remarque : $(\lambda, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ doit satisfaire $\lambda \notin \mathcal{S}(\nu)$ et $|\langle \boldsymbol{x}_i, \boldsymbol{u}_i \rangle| > 0$.

Mesure spectrale de $\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$

Corollaire 1. Pour $c_i = \frac{1}{d}$ pour tout $i \in [d]$, la MeSE de $\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ converge vers une loi en **demi-cercle** ν de support $\left[-2\sqrt{\frac{d-1}{d}}, 2\sqrt{\frac{d-1}{d}}\right]$, avec

$$\nu(dx) = \frac{d}{2(d-1)\pi}\sqrt{\left(\frac{4(d-1)}{d} - x^2\right)^+}, \quad g(z) = \frac{-zd + d\sqrt{z^2 - \frac{4(d-1)}{d}}}{2(d-1)}$$

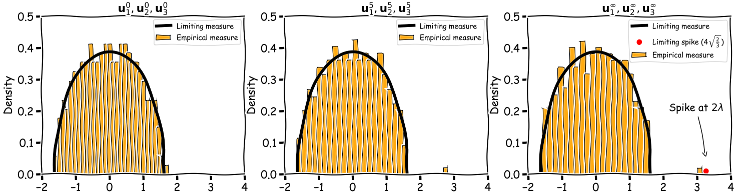


Figure: Spectre de $\Phi_3(\mathbf{T}, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ aux itérations $0, 5, \infty$ de l'algorithme d'itération de puissance appliqué sur $\mathbf{T}$. $n_1 = n_2 = n_3 = 100$ et $\beta = 0$.

$$\boldsymbol{u}_1 \leftarrow \frac{\mathbf{T}(I_{n_1}, \boldsymbol{u}_2, \boldsymbol{u}_3)}{\|\mathbf{T}(I_{n_1}, \boldsymbol{u}_2, \boldsymbol{u}_3)\|}, \quad \boldsymbol{u}_2 \leftarrow \frac{\mathbf{T}(\boldsymbol{u}_1, I_{n_2}, \boldsymbol{u}_3)}{\|\mathbf{T}(\boldsymbol{u}_1, I_{n_2}, \boldsymbol{u}_3)\|}, \quad \boldsymbol{u}_3 \leftarrow \frac{\mathbf{T}(\boldsymbol{u}_1, \boldsymbol{u}_2, I_{n_3})}{\|\mathbf{T}(\boldsymbol{u}_1, \boldsymbol{u}_2, I_{n_3})\|}$$

Norme spectrale et alignements asymptotiques

$$\mathbf{T}(\boldsymbol{x}_1, \boldsymbol{u}_2, \boldsymbol{u}_3) = \lambda\langle \boldsymbol{x}_1, \boldsymbol{u}_1\rangle \underbrace{\Rightarrow}_{\text{Stein}} \boxed{\left[\lambda + g_2(\lambda) + g_3(\lambda)\right] \langle \boldsymbol{x}_1, \boldsymbol{u}_1\rangle = \beta \prod_{i=2}^{3} \langle \boldsymbol{x}_i, \boldsymbol{u}_i\rangle}$$

Théorème 2. Pour tout $d \geq 3$, lorsque $n_i \to \infty$ avec $\frac{n_i}{\sum_j n_j} \to c_i \in (0,1]$, il existe $\beta_s > 0$ tel que pour tout $\beta > \beta_s$

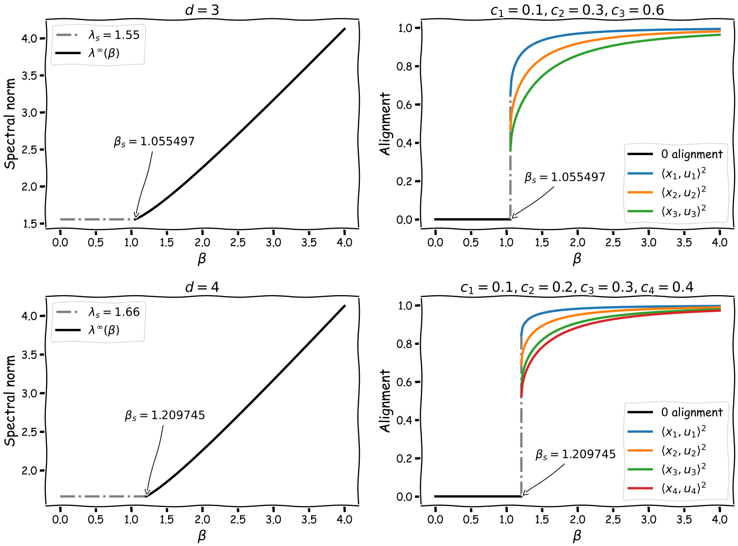
$$\lambda^* \xrightarrow{\text{p.s.}} \lambda^\infty, \quad |\langle \boldsymbol{x}_i, \boldsymbol{u}_i^* \rangle| \xrightarrow{\text{p.s.}} q_i(\lambda^\infty, \beta) = \sqrt{1 - \frac{g_i^2(\lambda^\infty)}{c_i}}$$

où λ^∞ satisfait $f(\lambda^\infty, \beta) = 0$ avec $f(z, \beta) = z + g(z) - \beta \prod_{i=1}^d q_i(z, \beta)$, et

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

pour $\beta \in [0, \beta_s]$, λ^∞ est bornée et $|\langle \boldsymbol{x}_i, \boldsymbol{u}_i^* \rangle| \xrightarrow{\text{p.s.}} 0$.

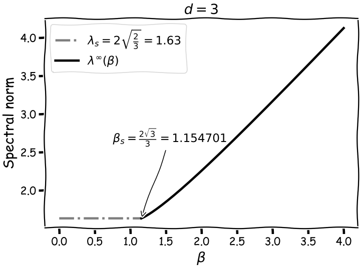
Norme spectrale et alignements asymptotiques

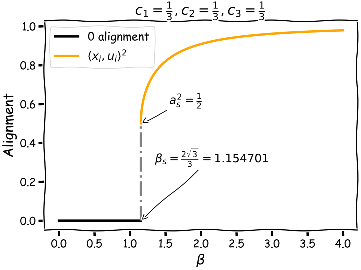


Tenseurs cubiques

Corollaire 2. Si $d = 3$ avec $c_i = \frac{1}{3}$, alors pour tout $\beta > \frac{2\sqrt{3}}{3}$

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





Modèle spike matriciel

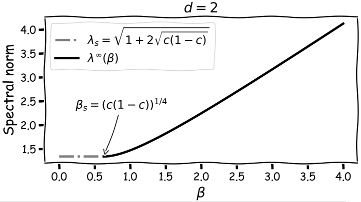
$$\text{Pour } d = 3, n_3 = 1 \quad \Rightarrow \quad \boldsymbol{M} = \beta \boldsymbol{x}\boldsymbol{y}^\top + \frac{1}{\sqrt{n_1 + n_2}} \boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$$

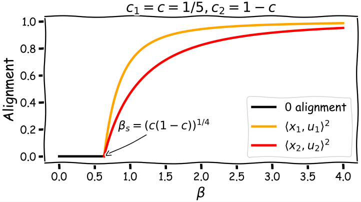
Corollaire 3. Si $d = 3$ avec $c_1 = c$ et $c_2 = 1 - c$ pour $c \in [0, 1]$, le modèle spike tensoriel devient un **modèle spike matriciel** (i.e. $c_3 = 0$).

Soit $\kappa(\beta, c) = \beta \sqrt{\frac{\beta^2(\beta^2+1)-c(c-1)}{(\beta^4+c(c-1))(\beta^2+1-c)}}$, pour $\beta > \beta_s = \sqrt[4]{c(1-c)}$

$$\lambda^* \xrightarrow{\text{p.s.}} \sqrt{\beta^2 + 1 + \frac{c(1-c)}{\beta^2}}, \quad |\langle \boldsymbol{x}_i, \boldsymbol{u}_i^* \rangle| \xrightarrow{\text{p.s.}} \frac{1}{\kappa(\beta, c_i)}, \ i \in \{1, 2\}$$

tandis que pour $\beta \in [0, \beta_s]$, $\lambda^* \xrightarrow{\text{p.s.}} \sqrt{1 + 2\sqrt{c(1-c)}}$ et $\left|\langle \boldsymbol{x}_i, \boldsymbol{u}_i^* \rangle\right| \xrightarrow{\text{p.s.}} 0$.



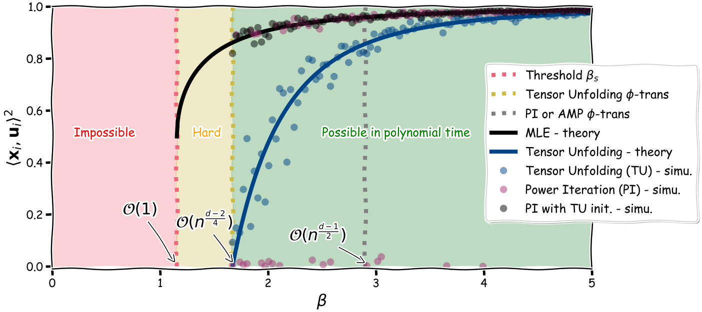


Sommaire

Algorithmes de décomposition et complexité

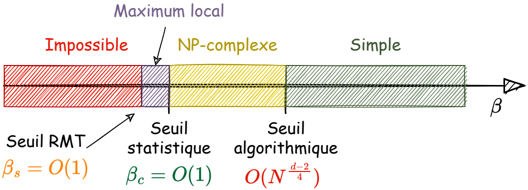
$$\min_{\lambda>0,\,\|\boldsymbol{u}_i\|=1} \|\mathbf{T} - \lambda \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_d\|_F^2 \Rightarrow \text{NP-complexe } \textbf{(Hillar et al., 2013)}$$

- Dépliage tensoriel : $\mathcal{M}_i(\mathbf{T}) = \beta \boldsymbol{x}_i \boldsymbol{y}_i^\top + \frac{1}{\sqrt{n}} \mathcal{M}_i(\mathbf{X}) \in \mathbb{R}^{n_i \times \prod_{j\neq i} n_j}$.
- A l'aide du Corollaire 3, on trouve $\beta_a = \left(\prod_i n_i\right)^{1/4} / \sqrt{\sum_i n_i}$.
- Coïncide avec $O\left(N^{\frac{d-2}{4}}\right)$ de **(Ben Arous et al, 2021)** pour $n_i = N$.
- Même seuil pour l'algorithme d'itération de puissance initialisé avec le dépliage tensoriel **(Auddy et al., 2021)**.



Messages à emporter

- L'approche RMT permet d'étudier le comportement des modèles tensoriels *asymétriques*.
- Les résultats obtenus caractérisent les performances de **l'estimateur de MV** pour β assez grand (i.e., $\beta \geq \beta_c$).



Questions ouvertes :

- Encore peu clair comment caractériser **la transition de phase** de l'estimateur de MV avec l'approche RMT.
- Peut-on trouver un **algorithme de complexité polynomiale** qui est consistant en dessous de β_a?
- Questions **d'universalité** et de généralisation à des **rangs supérieurs** à explorer.

Merci pour votre attention!

https://arxiv.org/abs/2112.12348

Références

Andrea Montanari and Emile Richard. "A statistical model for tensor PCA". In: *arXiv preprint arXiv:1411.1076* (2014)

Aukosh Jagannath, Patrick Lopatto, and Leo Miolane. "Statistical thresholds for tensor PCA". In: *The Annals of Applied Probability* 30.4 (2020), pp. 1910–1933

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*. Vol. 56. 1. Institut Henri Poincaré. 2020, pp. 230–264

José Henrique Goulart, Romain Couillet, and Pierre Comon. "A Random Matrix Perspective on Random Tensors". In: *stat* 1050 (2021), p. 2

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*. 2005, pp. 129–132

Christopher J Hillar and Lek-Heng Lim. "Most tensor problems are NP-hard". In: *Journal of the ACM (JACM)* 60.6 (2013), pp. 1–39

Gérard Ben Arous, Daniel Zhengyu Huang, and Jiaoyang Huang. "Long Random Matrices and Tensor Unfolding". In: *arXiv preprint arXiv:2110.10210* (2021)

Arnab Auddy and Ming Yuan. "On Estimating Rank-One Spiked Tensors in the Presence of Heavy Tailed Errors". In: *arXiv preprint arXiv:2107.09660* (2021)

Mohamed El Amine Seddik, Maxime Guillaud, and Romain Couillet. "When Random Tensors meet Random Matrices". In: *arXiv preprint arXiv:2112.12348* (2021)