When Random Tensors meet Random Matrices

Gretsi 2022

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

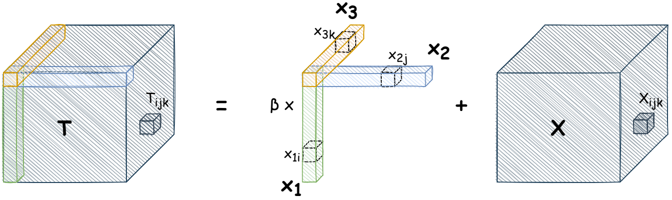
https://melaseddik.github.io/

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

Nancy, 9 September 2022

Outline

Introduction: Asymmetric Spiked Tensor Model



We consider the following model: $(\boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \otimes \boldsymbol{x}_3)_{ijk} = x_{1i} x_{2j} x_{3k}$

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

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

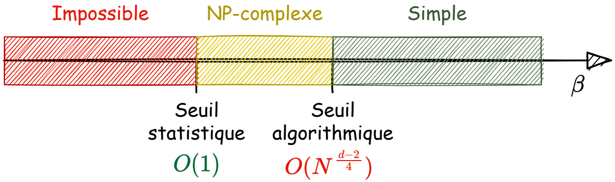
- Is it possible to recover the signal in theory? for which **critical** value of β?
- What **alignment** $\langle \boldsymbol{x}_i, \boldsymbol{u}_i \rangle$ between the signal and an estimator $\boldsymbol{u}_i(\mathbf{T})$?
- Is there an algorithm that can recover the signal in **polynomial time**?

Related Works: Symmetric Case

Introduced initially by **(Montanari & Richard, 2014)**

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

where $\|x\| = 1$ and **W** has random Gaussian entries and is **symmetric**.This is a natural extension of the classical spiked matrix model $Y = \beta x x^\top + \frac{1}{\sqrt{N}} W$.



Other works in the literature include: **(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)**.

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

Random Matrix Approach (Goulart et al., 2021)

The optimization problem of maximum likelihood estimator (MLE) for $d = 3$:

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

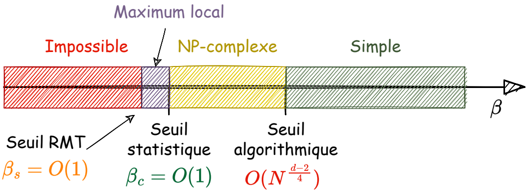
The critical points satisfy **(Lim, 2005)** :

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

where $(\mathbf{Y}(\boldsymbol{u}, \boldsymbol{u}))_i = \sum_{jk} u_j u_k Y_{ijk}$ et $(\mathbf{Y}(\boldsymbol{u}))_{ij} = \sum_k u_k Y_{ijk}$. The MLE $\hat{\boldsymbol{x}}$ corresponds to the dominant eigenvector of $\mathbf{Y}(\hat{\boldsymbol{x}})$: $\mathbf{Y}(\hat{\boldsymbol{x}})\hat{\boldsymbol{x}} = \|\mathbf{Y}\|\hat{\boldsymbol{x}}$.

Hence, the approach from **(Goulart et al., 2021)** consists in studying:

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



Outline

Introduction
- Asymmetric Spiked Tensor Model
- Related Works
- Random Matrix Approach

Analysis of the Asymmetric Spiked Tensor Model
- **Tensors Singular Values and Vectors**
- **Associated Random Matrix**
- **Asymptotic Spectral Norm and Alignments**

Decomposition Algorithms and Complexity

Tensors Singular Values and Vectors

The optimization problem of MLE for $d = 3$:

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

The critical points satisfy **(Lim, 2005)** :

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

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

- In contrast to the symmetric case, the choice of the associated *contraction* matrix is not straightforward. For instance:

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

Objectives:

- Evaluate the asymptotic limits of λ^* and $\langle \boldsymbol{x}_i, \boldsymbol{u}_i^* \rangle$ associated (a priori) to the MLE when $n_i \to \infty$.
- Define a *symmetric* random matrix that is equivalent to $\mathbf{T}$.

Associated Random Matrix to T

Stein's Lemma: Let $X \sim \mathcal{N}(0,1)$, then $\mathbb{E}[X f(X)] = \mathbb{E}[f'(X)]$.

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

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

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

The resolvent matrix: $\boldsymbol{R}(z) = (\boldsymbol{\Phi}_3(\mathbf{T}, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3) - z \boldsymbol{I}_n)^{-1}$.
When $n_i \to \infty$, the non-vanishing terms involve the **trace** of $\boldsymbol{R}(z)$,

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

Spectral Measure of $\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$

Stieltjes Transform: The Stieltjes transform of a probability measure ν is $g_\nu(z) = \int \frac{d\nu(\lambda)}{\lambda - z}$, $z \in \mathbb{C} \setminus \mathcal{S}(\nu)$.

For $\boldsymbol{S} \in \mathrm{Sym}_n$ with λ_i its eigenvalues, *the empirical spectral measure* (ESM) of $\boldsymbol{S}$ and its associated Stieltjes transform are:

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

where $\boldsymbol{R}_{\boldsymbol{S}}(z) = (\boldsymbol{S} - z\boldsymbol{I}_n)^{-1}$ is the resolvent of $\boldsymbol{S}$.

Theorem 1. When $n_i \to \infty$ with $\frac{n_i}{\sum_j n_j} \to c_i \in [0, 1]$, the ESM of $\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ converges to a **deterministic measure** ν having as Stieltjes transform $\frac{1}{n}\operatorname{tr} \boldsymbol{R}(z) \xrightarrow{\text{a.s.}} g(z) = \sum_{i=1}^d g_i(z)$ verifying $\Im[g(z)] > 0$ for $\Im[z] > 0$, where

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

Remark: $(\lambda, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ must satisfy $\lambda \notin \mathcal{S}(\nu)$ and $|\langle \boldsymbol{x}_i, \boldsymbol{u}_i \rangle| > 0$.

Spectral Measure of $\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$

Corollary 1. When $c_i = \frac{1}{d}$ for all $i \in [d]$, the ESM of $\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ converges to a **semi-circle law** ν of compact support $\left[-2\sqrt{\frac{d-1}{d}}, 2\sqrt{\frac{d-1}{d}}\right]$, where

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

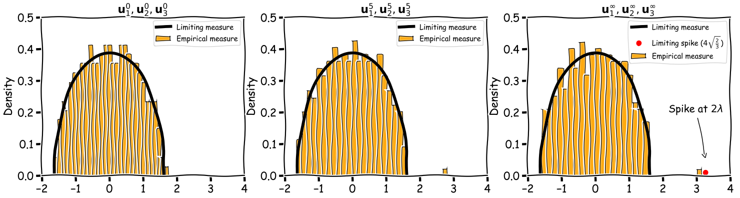


Figure: Spectrum of $\Phi_3(\mathbf{T}, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ at iterations $0, 5, \infty$ of the tensor power iteration algorithm applied on $\mathbf{T}$. $n_1 = n_2 = n_3 = 100$ and $\beta = 0$.

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

Asymptotic Spectral Norm and Alignments

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

Theorem 2. For all $d \geq 3$, when $n_i \to \infty$ with $\frac{n_i}{\sum_j n_j} \to c_i \in (0, 1)$, there exists $\beta_s > 0$ such that for all $\beta > \beta_s$

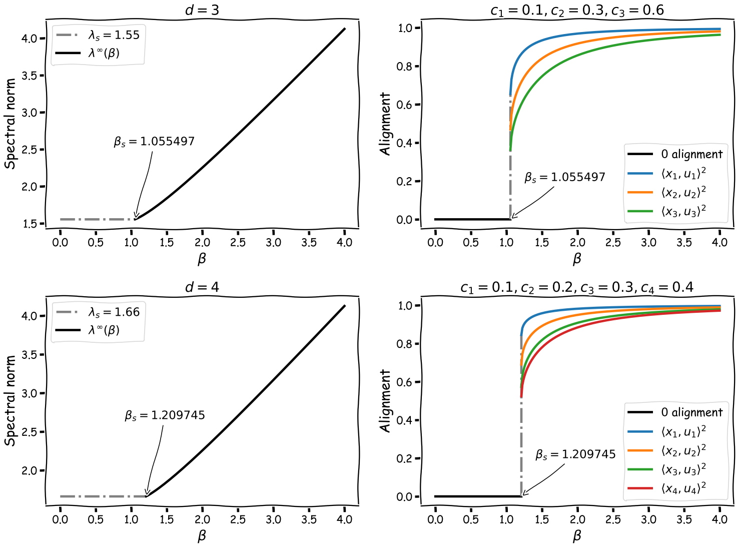
$$\lambda^* \xrightarrow{\text{a.s.}} \lambda^\infty, \quad |\langle \boldsymbol{x}_i, \boldsymbol{u}_i^* \rangle| \xrightarrow{\text{a.s.}} q_i(\lambda^\infty, \beta) = \sqrt{1 - \frac{g_i^2(\lambda^\infty)}{c_i}}$$

where λ^∞ satisfies $f(\lambda^\infty, \beta) = 0$ with $f(z, \beta) = z + g(z) - \beta \prod_{i=1}^d q_i(z, \beta)$, and

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

for $\beta \in [0, \beta_s]$, λ^∞ is bounded and $|\langle \boldsymbol{x}_i, \boldsymbol{u}_i^* \rangle| \xrightarrow{\text{a.s.}} 0$.

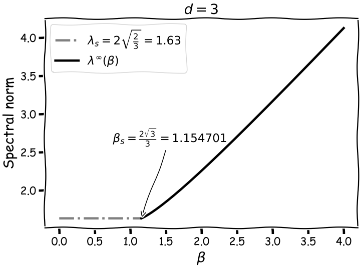
Asymptotic Spectral Norm and Alignments

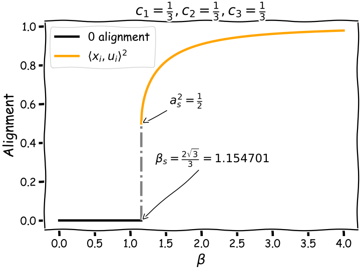


Cubic Tensors

Corollary 2. If $d = 3$ with $c_i = \frac{1}{3}$, then for all $\beta > \frac{2\sqrt{3}}{3}$

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





Spiked Matrix Model

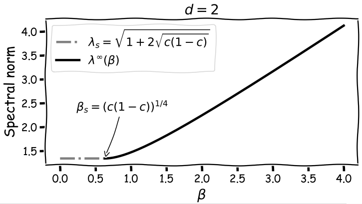
$$\text{For } d=3, n_3=1 \quad \Rightarrow \quad \boldsymbol{M} = \beta \boldsymbol{x}\boldsymbol{y}^\top + \frac{1}{\sqrt{n_1+n_2}}\boldsymbol{X} \in \mathbb{R}^{n_1\times n_2}$$

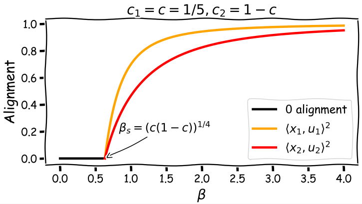
Corollary 3. If $d=3$ with $c_1 = c$ et $c_2 = 1-c$ for $c \in [0,1]$, the spiked tensor model becomes a **spiked matrix model** (i.e. $c_3 = 0$).

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

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

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



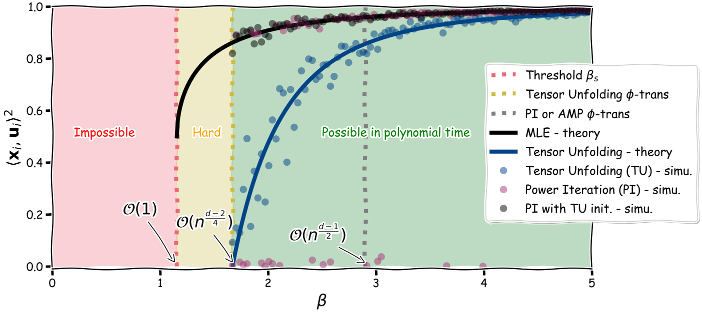


Outline

Decomposition Algorithms and Complexity

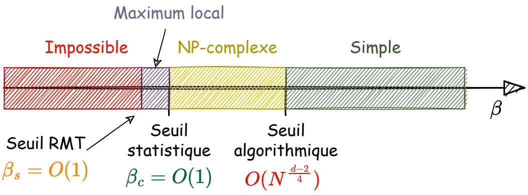
$$\min_{\lambda>0,\, \|\boldsymbol{u}_i\|=1} \|\mathbf{T} - \lambda \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_d\|_F^2 \Rightarrow$$ NP-hard **(Hillar et al., 2013)**

- Tensor unfolding: $\mathcal{M}_i(\mathbf{T}) = \beta \boldsymbol{x}_i \boldsymbol{y}_i^\top + \frac{1}{\sqrt{n}} \mathcal{M}_i(\mathbf{X}) \in \mathbb{R}^{n_i \times \prod_{j \neq i} n_j}$.
- Using Corollary 3, we find $\beta_a = \left(\prod_i n_i\right)^{1/4} / \sqrt{\sum_i n_i}$.
- Coincides with $O\left(N^{\frac{d-2}{4}}\right)$ of **(Ben Arous et al, 2021)** for $n_i = N$.
- Same threshold for tensor power iteration initialized with tensor unfolding **(Auddy et al., 2021)**.



Take Away Messages

- The RMT approach allows the study of asymmetric spiked tensor models.
- The obtained results characterize the performance of the MLE for β large enough (i.e., $\beta \geq \beta_c$).



Open questions:

- Still unclear how to characterize the **phase transition** of the MLE with the RMT approach.
- Is it possible to find a **polynomial time algorithm** that is consistent below the computational threshold β_a?
- **Universality** and generalization to **higher-ranks**.

Thank you for your attention!
https://arxiv.org/abs/2112.12348

References

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)