When Random Tensors meet Random Matrices

Mohamed El Amine Seddik [1]   Maxime Guillaud [1]   Romain Couillet [2]

[1]Huawei Technologies France   [2]Grenoble Alpes University

Abstract

- Study of **asymmetric** spiked tensors with Gaussian noise.
- The tensor model is associated to a random matrix constructed from **contractions** of the tensor with its singular vectors.
- Exact characterization of the almost sure limits of the tensor singular value and alignments are derived in the **high-dimensional regime**.

Asymmetric rank-1 spiked random tensors

Consider the **non-symmetric** spiked tensor model

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

where $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) \in \mathbb{S}^{n_1-1} \times \cdots \times \mathbb{S}^{n_d-1}$, $X_{i_1 \ldots i_d} \sim \mathcal{N}(0,1)$ i.i.d. and $n = \sum_{i=1}^d n_i$. The parameter $\beta > 0$ controls the signal-to-noise ratio (SNR).

(A1) **Growth rate assumptions:** As $n_i \to \infty$, $\frac{n_i}{\sum_{j=1}^d n_j} \to c_i \in (0,1)$ and $d = \mathcal{O}(1)$.

Tensor singular values and vectors

Maximum likelihood estimator (MLE) of best rank-1 approximation of T

$$\underset{\lambda>0,\,(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_d)\in\mathbb{S}^{n_1-1}\times\cdots\times\mathbb{S}^{n_d-1}}{\arg\min} \|\mathsf{T} - \lambda \boldsymbol{u}_1 \otimes \cdots \otimes \boldsymbol{u}_d\|_F^2$$

Equivalent to (Lim, 2005)

$$\max_{\prod_{i=1}^d \|\boldsymbol{u}_i\|=1} |\mathsf{T}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)|$$

The KKT conditions verified by the stationary points $(\lambda, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ are, $\forall i \in [d]$

$$\begin{cases} \mathsf{T}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, :, \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_d) = \lambda \boldsymbol{u}_i \quad \text{and} \quad \|\boldsymbol{u}_i\| = 1 \\ \lambda = \mathsf{T}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d) \end{cases}$$

Goal: Under (A1), how to express the asymptotics of λ and $\langle \boldsymbol{x}_i, \boldsymbol{u}_i \rangle$ of the MLE?

Associated random matrix model ($d = 3$)

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

Under (A1), λ *concentrates* around its expectation and with Stein's Lemma

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

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

Resolvent matrix: $\boldsymbol{R}(z) = (\boldsymbol{\Phi}_3(\mathsf{T}, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3) - z \boldsymbol{I}_n)^{-1}$.

As $n \to \infty$, the non-vanishing terms involve **traces** of $\boldsymbol{R}(z)$,

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

Stieltjes transform

Definition. Given a probability measure ν, the Stieltjes transform of ν is defined as

$$g_\nu(z) = \int \frac{d\nu(\lambda)}{\lambda - z}, \quad z \in \mathbb{C} \setminus \mathcal{S}(\nu)$$

Given a $n \times n$ symmetric matrix $\boldsymbol{S}$ and denote λ_i its eigenvalues, the *empirical spectral measure* (ESM) of $\boldsymbol{S}$ and the corresponding Stieltjes transform express as

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

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

Limiting spectral measure of $\boldsymbol{\Phi}_d(\mathsf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$

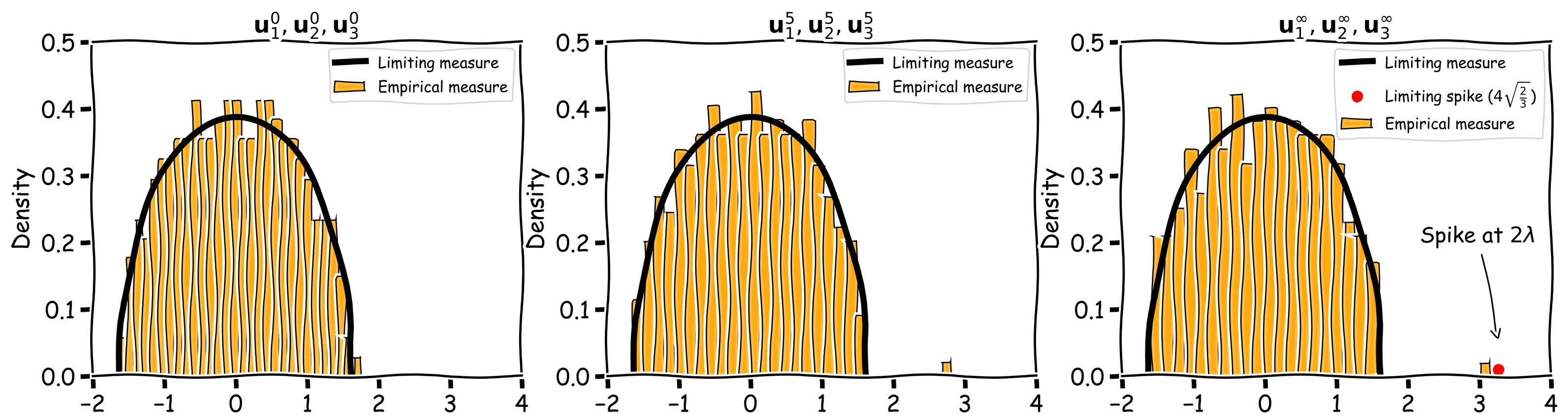

Figure 1. Spectrum of $\boldsymbol{\Phi}_3(\mathsf{T}, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ at iterations $0, 5, \infty$ of power iteration applied on T. $n_1 = n_2 = n_3 = 100$ and $\beta = 0$.

(A2) $\exists (\lambda^*, \boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_d^*)$ such that λ^* is not an eigenvalue of $\boldsymbol{\Phi}_d(\mathsf{T}, \boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_d^*)$.

Theorem 1. Under (A1) and (A2), the ESM of $\boldsymbol{\Phi}_d(\mathsf{T}, \boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_d^*)$ converges to a *deterministic measure* ν whose Stieltjes transform is given as $g(z) = \sum_{i=1}^d g_i(z)$ such that $\Im[g(z)] > 0$ for $\Im[z] > 0$, with

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

Corollary 1. If $c_i = \frac{1}{d}$ for all $i \in [d]$, then the ESM of $\boldsymbol{\Phi}_d(\mathsf{T}, \boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_d^*)$ converges to a **semi-circle** distribution supported on $\mathcal{S}(\nu) = \left[-2\sqrt{\frac{d-1}{d}}, 2\sqrt{\frac{d-1}{d}}\right]$, with

$$\nu(dx) = \frac{d}{2(d-1)\pi} \sqrt{\left(\frac{4(d-1)}{d} - x^2\right)^+}, \quad g(z) = \frac{-zd + d\sqrt{z^2 - \frac{4(d-1)}{d}}}{2(d-1)}, \; z \in \mathbb{C} \setminus \mathcal{S}(\nu)$$

Asymptotic singular value and alignments

Theorem 2. Under (A1) and (A2), for $d \geq 3$, there exists $\beta_s > 0$ such that for $\beta > \beta_s$

$$\lambda^* \xrightarrow{\text{a.s.}} \lambda^\infty, \quad |\langle \boldsymbol{x}_i, \boldsymbol{u}_i^* \rangle| \xrightarrow{\text{a.s.}} \left(\frac{\alpha_i(\lambda^\infty)^{d-3}}{\prod_{j \neq i} \alpha_j(\lambda^\infty)} \right)^{\frac{1}{2d-4}}$$

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

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

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

Corollary 2 (Cubic tensors). If $d = 3$ with $c_i = \frac{1}{3}$, for $\beta > \frac{2\sqrt{3}}{3}$

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

Corollary 3 (Spiked random matrices). If $d = 3$ with $c_1 = c$ and $c_2 = 1 - c$ for some $c \in (0,1)$, 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 $\beta > \beta_s = \sqrt[4]{c(1-c)}$

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

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

Theory versus simulations

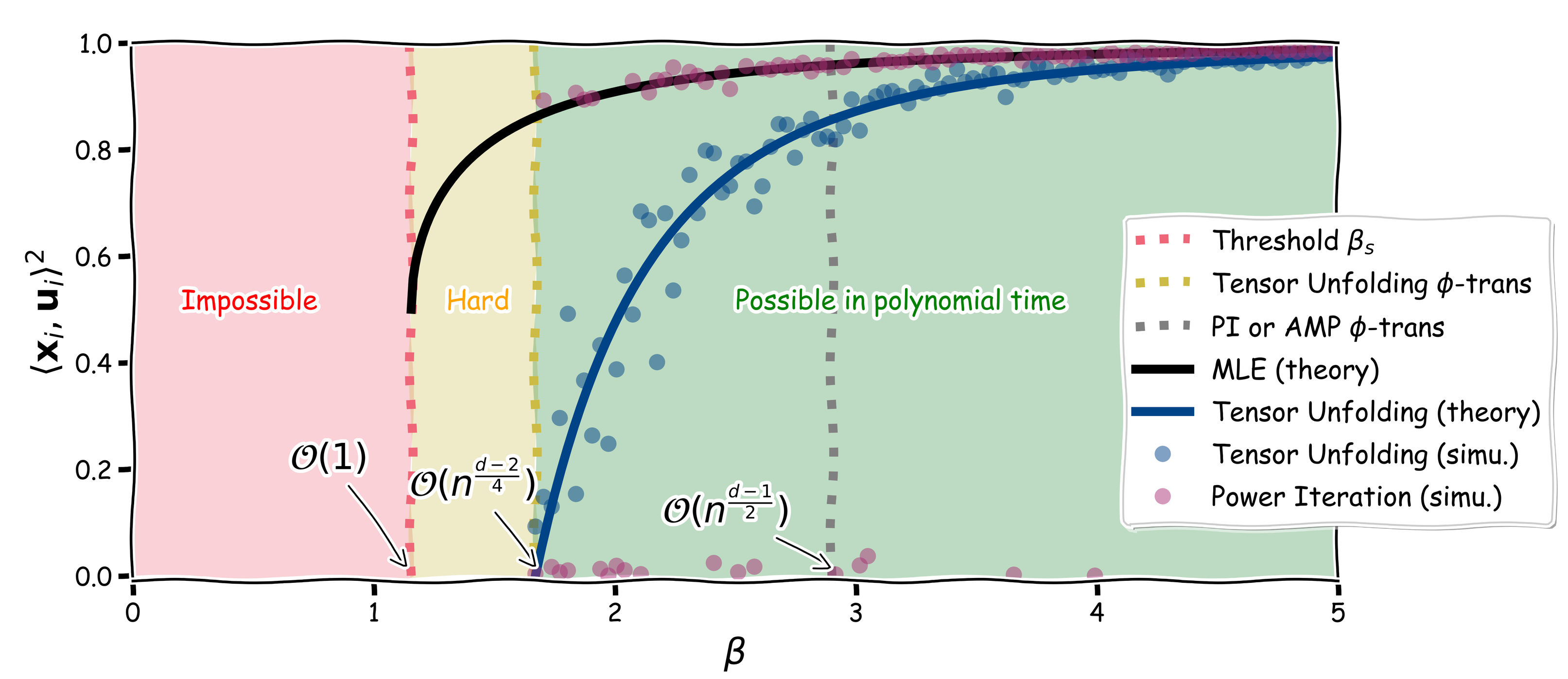

Figure 2. Thresholds and asymptotic alignments for a cubic spiked random tensor: MLE **in black** and tensor unfolding (Ben Arous, 2021) in blue. Simulations with power iteration and tensor unfolding applied on a cubic tensor with $n_i = 70$.

Conclusion

- The derived result **seems** to describe the behavior of the MLE.
- Still unclear how to characterize the **phase transition** for the MLE with our approach, as also mentioned in (Goulart, 2021) for symmetric tensors.
- **Universality** and generalization to **higher-ranks** remain to investigate.

References

Lim, Lek-Heng. "Singular values and eigenvalues of tensors: a variational approach." 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005.. IEEE, 2005.

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

Ben Arous, Gérard, Daniel Zhengyu Huang, and Jiaoyang Huang. "Long Random Matrices and Tensor Unfolding." arXiv e-prints (2021): arXiv-2110.

Workshop on random tensors at CIRM - March 2022
Link to article: `https://arxiv.org/abs/2112.12348`
Get in touch: `melaseddik.github.io`