Asymptotic Analysis of Asymmetric Spiked Tensor Models with Random Matrix Theory

Workshop on Tensor Theory and Methods

Mohamed El Amine Seddik
melaseddik.github.io

Joint work with
Maxime Guillaud, Romain Couillet and Alexis Decurninge

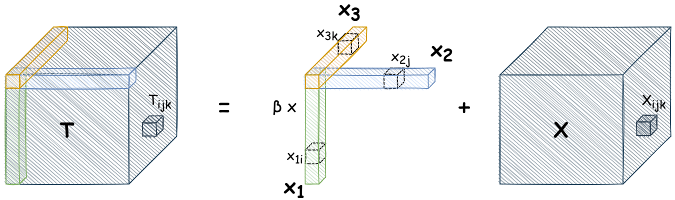
Mathematical and Algorithmic Sciences Laboratory, Huawei Technologies France

Paris, 24th November 2022

HUAWEI

Outline

Introduction: Asymmetric Spiked Tensor Model



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

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

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

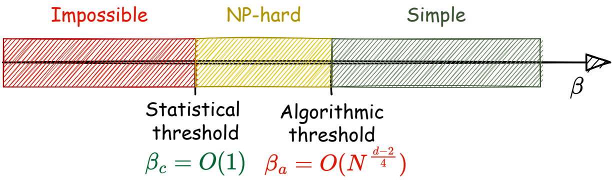
- Is it possible to recover the signal in theory? for which **critical** value of β?
- What **alignment** $\langle \boldsymbol{x}_i, \boldsymbol{u}_i \rangle$ between the signal and an estimator $\boldsymbol{u}_i(\mathbf{T})$?
- Is there an algorithm that can recover the signal in **polynomial time**?

Related Works: Symmetric Case

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

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

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



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 $d = 3$:

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

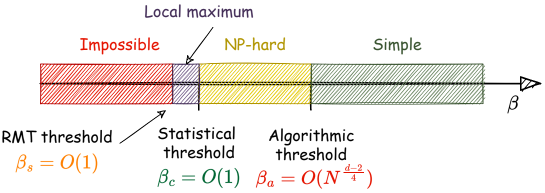
The critical points satisfy **(Lim, 2005)**:

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

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

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

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

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

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 $\boldsymbol{R}(z)$,

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

Associated Random Matrix to T

For an order-d tensor the associated random matrix is $\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ where

$$\Phi_d : (\mathbf{X}, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_d) \longmapsto \begin{bmatrix} \mathbf{0}_{n_1\times n_1} & \mathbf{X}^{12} & \mathbf{X}^{13} & \cdots & \mathbf{X}^{1d} \\ (\mathbf{X}^{12})^\top & \mathbf{0}_{n_2\times n_2} & \mathbf{X}^{23} & \cdots & \mathbf{X}^{2d} \\ (\mathbf{X}^{13})^\top & (\mathbf{X}^{23})^\top & \mathbf{0}_{n_3\times n_3} & \cdots & \mathbf{X}^{3d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (\mathbf{X}^{1d})^\top & (\mathbf{X}^{2d})^\top & (\mathbf{X}^{3d})^\top & \cdots & \mathbf{0}_{n_d\times n_d} \end{bmatrix}$$

with $\mathbf{X}^{ij} \equiv \mathbf{X}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{i-1}, :, \boldsymbol{a}_{i+1}, \ldots, \boldsymbol{a}_{j-1}, :, \boldsymbol{a}_{j+1}, \ldots, \boldsymbol{a}_d) \in \mathbb{R}^{n_i \times n_j}$.

Remark. $(d-1)\lambda$ is an eigenvalue of $\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d)$ with

$$\Phi_d(\mathbf{T}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_d) \begin{bmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_d \end{bmatrix} = (d-1)\lambda \begin{bmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_d \end{bmatrix}$$

since

$$\mathbf{T}\left(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{j-1}, \boldsymbol{I}_{n_j}, \boldsymbol{u}_{j+1}, \ldots, \boldsymbol{u}_d\right) = \lambda \boldsymbol{u}_j$$

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 and denote its resolvent $\boldsymbol{R_S}(z) = (\boldsymbol{S} - z\boldsymbol{I}_n)^{-1}$, the ESM of $\boldsymbol{S}$ and its associated Stieltjes transform are:

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

Definition 1. Let ν by the probability measure with Stieltjes transform $g(z) = \sum_{i=1}^d g_i(z)$ verifying $\Im[g(z)] > 0$ for $\Im[z] > 0$, where $g_i(z)$ satisfies $g_i^2(z) - (g(z) + z)g_i(z) - c_i = 0$, for $z \notin \mathcal{S}(\nu)$.

Assumption 1. As $n_i \to \infty$ with $\frac{n_i}{\sum_j n_j} \to c_i \in (0, 1)$, there exists a sequence of critical points $(\hat\lambda, \hat{\boldsymbol{u}}_1, \ldots, \hat{\boldsymbol{u}}_d)$ s.t. $\hat\lambda$ and $|\langle \boldsymbol{x}_i, \hat{\boldsymbol{u}}_i\rangle|$ converge to some limits $\lambda \notin \mathcal{S}(\nu)$ and $\rho_i > 0$ respectively.

Theorem 1. Under Assumption 1, the ESM of $\Phi_d(\mathbf{T}, \hat{\boldsymbol{u}}_1, \ldots, \hat{\boldsymbol{u}}_d)$ converges weakly to ν defined in Definition 1 (i.e. $\frac{1}{n}\mathrm{tr}\, \boldsymbol{R}(z) \xrightarrow{\text{a.s.}} g(z)$).

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 $\Phi_d(\mathbf{T}, \hat{\boldsymbol{u}}_1, \ldots, \hat{\boldsymbol{u}}_d)$ converges to a **semi-circle law** ν of support $\left[-2\sqrt{\frac{d-1}{d}}, 2\sqrt{\frac{d-1}{d}}\right]$, where

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

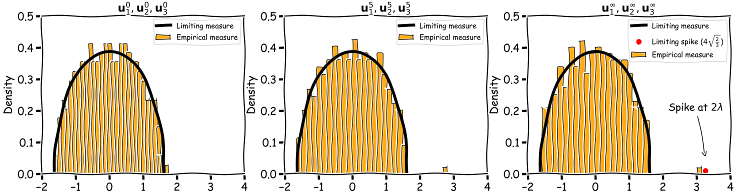


Figure: Spectrum of $\Phi_3(\mathbf{T}, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ at iterations $0, 5, \infty$ of the tensor power iteration algorithm applied on $\mathbf{T}$. $n_1 = n_2 = n_3 = 100$ and $\beta = 0$.

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

Asymptotic Spectral Norm and Alignments

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

Assumption 1. As $n_i \to \infty$ with $\frac{n_i}{\sum_j n_j} \to c_i \in (0,1)$, there exists a sequence of critical points $(\hat{\lambda}, \hat{\boldsymbol{u}}_1, \dots, \hat{\boldsymbol{u}}_d)$ s.t. $\hat{\lambda}$ and $|\langle \boldsymbol{x}_i, \hat{\boldsymbol{u}}_i\rangle|$ converge to some limits $\lambda \notin \mathcal{S}(\nu)$ and $\rho_i > 0$ respectively.

Theorem 2. For all $d \geq 3$, under Assumption 1, there exists $\beta_s > 0$ such that for all $\beta > \beta_s$

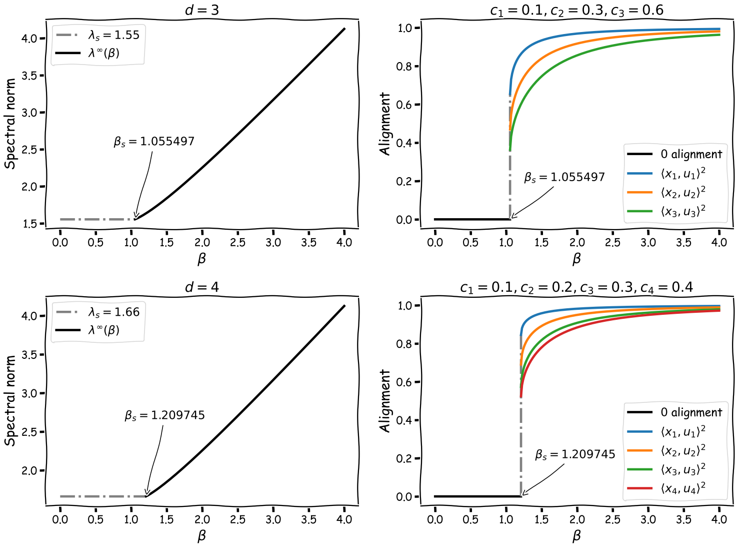
$$\hat{\lambda} \xrightarrow{\text{a.s.}} \lambda, \quad |\langle \boldsymbol{x}_i, \hat{\boldsymbol{u}}_i\rangle| \xrightarrow{\text{a.s.}} q_i(\lambda)$$

where λ satisfies $f(\lambda, \beta) = 0$ with

$$f(z, \beta) = z + g(z) - \beta \prod_{i=1}^{d} q_i(z), \quad q_i(z) = \sqrt{1 - \frac{g_i^2(z)}{c_i}}$$

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

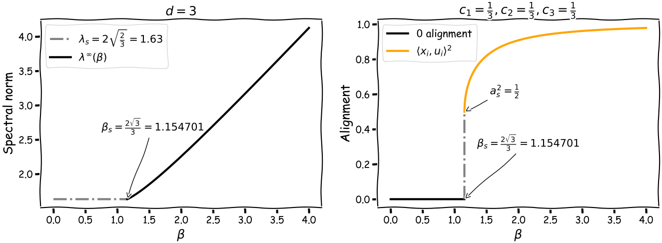
Asymptotic Spectral Norm and Alignments



Cubic Tensors

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

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



For hyper-cubic tensors of order d, we have

$$\beta_s = \sqrt{\frac{d-1}{d}\left(\frac{d-2}{d-1}\right)^{1-\frac{d}{2}}}, \quad \lim_{\beta \to \beta_s} \rho_i(\beta) = \sqrt{\frac{d-2}{d-1}}$$

Spiked Matrix Model

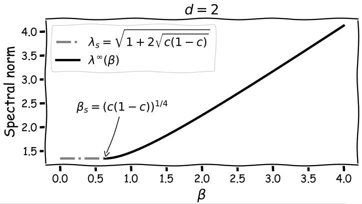
For $d = 3, n_3 = 1 \quad \Rightarrow \quad \boldsymbol{M} = \beta \boldsymbol{x}\boldsymbol{y}^\top + \frac{1}{\sqrt{n_1 + n_2}} \boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$

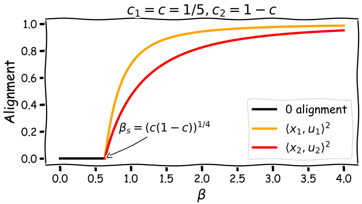
Corollary 3. If $d = 3$ with $c_1 = c$ et $c_2 = 1 - c$ for $c \in [0, 1]$, the spiked tensor model becomes a **spiked matrix model** (i.e. $c_3 = 0$).

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 $\beta > \beta_s = \sqrt[4]{c(1-c)}$

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

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



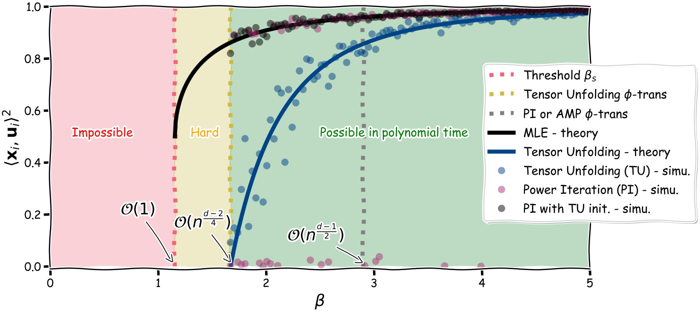


Outline

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)**.



Outline

Hotteling-type Tensor Deflation

We consider the following rank-r order-d spiked tensor model

$$\mathbf{T}_1 = \sum_{i=1}^{r} \beta_i \boldsymbol{x}_{i,1} \otimes \cdots \otimes \boldsymbol{x}_{i,d} + \frac{1}{\sqrt{n}} \mathbf{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$$

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

Tensor Deflation. Compute $\mathbf{T}_2, \mathbf{T}_3, \dots$ as

$$\mathbf{T}_{i+1} = \mathbf{T}_i - \hat{\lambda}_i \hat{\boldsymbol{u}}_{i,1} \otimes \cdots \otimes \hat{\boldsymbol{u}}_{i,d} \quad \text{for} \quad i \in [r]$$

where $\hat{\lambda}_i \hat{\boldsymbol{u}}_{i,1} \otimes \cdots \otimes \hat{\boldsymbol{u}}_{i,d}$ is a critical point of

$$\operatorname*{arg\,min}_{\lambda_i > 0, \|\boldsymbol{u}_{i,j}\| = 1} \left\| \mathbf{T}_i - \lambda_i \boldsymbol{u}_{i,1} \otimes \cdots \otimes \boldsymbol{u}_{i,d} \right\|_F^2$$

Such a critical point satisfy

$$\mathbf{T}_i \left(\hat{\boldsymbol{u}}_{i,1}, \dots, \hat{\boldsymbol{u}}_{i,j-1}, \boldsymbol{I}_{n_j}, \hat{\boldsymbol{u}}_{i,j+1}, \dots, \hat{\boldsymbol{u}}_{i,d} \right) = \hat{\lambda}_i \hat{\boldsymbol{u}}_{i,j}$$

for $(i,j) \in [r] \times [d]$ with $\|\hat{\boldsymbol{u}}_{i,j}\| = 1$.

Associated Random Matrices

$$\mathbf{T}_i \quad \to \quad \mathbf{\Phi}_d(\mathbf{T}_i, \hat{\boldsymbol{u}}_{i,1}, \ldots, \hat{\boldsymbol{u}}_{i,d}) \quad \to \quad \text{Stieltjes transform } g(z)$$

Assumption 2. Assume that as $n_i \to \infty$

- r, d are fixed and $\frac{n_i}{\sum_{j=1}^d n_j} \to c_i \in (0,1)$.
- There exists a sequence of critical points $(\hat{\lambda}_i, \hat{\boldsymbol{u}}_{i,1}, \ldots, \hat{\boldsymbol{u}}_{i,d})$ s.t. $\hat{\lambda}_i \xrightarrow{\text{a.s.}} \lambda_i$, $|\langle \boldsymbol{x}_{i,k}, \hat{\boldsymbol{u}}_{j,k}\rangle| \xrightarrow{\text{a.s.}} \rho_{ijk}$ and $|\langle \hat{\boldsymbol{u}}_{i,k}, \hat{\boldsymbol{u}}_{j,k}\rangle| \xrightarrow{\text{a.s.}} \eta_{ijk}$ with $\lambda_i \notin \mathcal{S}(\nu)$ and $\rho_{ijk}, \eta_{ijk} > 0$.

Theorem 3. Under Assumption 2, the ESM of $\mathbf{\Phi}_d(\mathbf{T}_i, \hat{\boldsymbol{u}}_{i,1}, \ldots, \hat{\boldsymbol{u}}_{i,d})$ converges to the deterministic measure ν defined in Definition 1.

When $c_i = \frac{1}{d}$ for all i, ν describes a semi-circle law of compact support $\mathcal{S}(\nu) = \left[-2\sqrt{\frac{d-1}{d}}, 2\sqrt{\frac{d-1}{d}}\right]$, with Stieltjes transform

$$g(z) = \frac{-zd + d\sqrt{z^2 - \frac{4(d-1)}{d}}}{2(d-1)}, \quad z \notin \mathcal{S}(\nu)$$

Asymptotic Spectral Norms and Alignments

Let $\alpha_{ijk} = \lim |\langle \boldsymbol{x}_{i,k}, \boldsymbol{x}_{j,k} \rangle|$, $f(z) = z + g(z)$ and $h_i(z) = -\frac{c_i}{g_i(z)}$.

Theorem 4. Under Assumption 2, λ_i, ρ_{ijk} and η_{ijk} satisfy the following system of equations

$$\begin{cases} \bullet \quad f(\lambda_j) + \sum_{i=1}^{j-1} \lambda_i \prod_{k=1}^{d} \eta_{ijk} - \sum_{i=1}^{r} \beta_i \prod_{k=1}^{d} \rho_{ijk} = 0 \ , \ 1 \leq j \leq r \\ \bullet \quad h_\ell(\lambda_j) \rho_{kj\ell} + \sum_{i=1}^{j-1} \lambda_i \rho_{ki\ell} \prod_{m \neq \ell}^{d} \eta_{ijm} - \sum_{i=1}^{r} \beta_i \alpha_{ik\ell} \prod_{m \neq \ell}^{d} \rho_{ijm} = 0 \\ 1 \leq \ell \leq d, 1 \leq j, k \leq r \\ \bullet \quad h_\ell(\lambda_j) \eta_{kj\ell} + g_\ell(\lambda_k) \prod_{m \neq \ell}^{d} \eta_{kjm} + \sum_{i=1}^{j-1} \lambda_i \eta_{ik\ell} \prod_{m \neq \ell}^{d} \eta_{ijm} + \dots \\ - \sum_{i=1}^{r} \beta_i \rho_{ik\ell} \prod_{m \neq \ell}^{d} \rho_{ijm} = 0, \ 1 \leq \ell \leq d, 1 \leq j < k \leq r \end{cases}$$

Particular case of rank-2 order-3 tensors

For simplicity, let $r = 2$, $d = 3$ and $n_1 = n_2 = n_3 = N$, i.e.

$$\mathbf{T}_1 = \sum_{i=1}^{2} \beta_i \boldsymbol{x}_{i,1} \otimes \boldsymbol{x}_{i,2} \otimes \boldsymbol{x}_{i,3} + \frac{1}{\sqrt{n}} \mathbf{X}$$

Denote $\alpha = \lim |\langle \boldsymbol{x}_{1,k}, \boldsymbol{x}_{2,k} \rangle|$, $\eta = \lim |\langle \hat{\boldsymbol{u}}_{1,k}, \hat{\boldsymbol{u}}_{2,k} \rangle|$, $\rho_{ij} = \lim |\langle \boldsymbol{x}_{i,k}, \hat{\boldsymbol{u}}_{j,k} \rangle|$

$$\underbrace{\boldsymbol{\beta} = (\beta_1, \beta_2, \alpha)}_{\text{Parameters}}, \quad \underbrace{\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \eta)}_{\text{Measurements}}, \quad \underbrace{\boldsymbol{\rho} = (\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22})}_{\text{Alignments}}$$

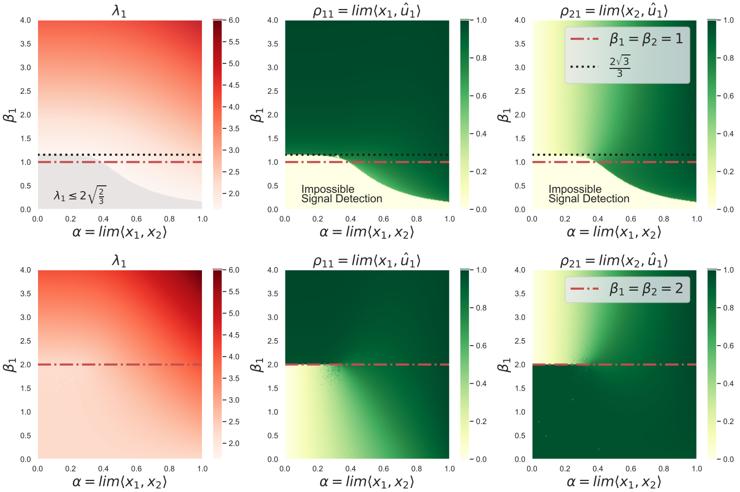
Corollary 4. Fixing β, under Assumption 2, $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ satisfy $\psi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\rho}) = 0$

$$\psi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\rho}) \equiv \begin{pmatrix} \boxed{\begin{array}{c} f(\lambda_1) - \beta_1 {\rho_{11}}^3 - \beta_2 {\rho_{21}}^3 \\ h(\lambda_1)\rho_{11} - \beta_1 {\rho_{11}}^2 - \beta_2 \alpha {\rho_{21}}^2 \\ h(\lambda_1)\rho_{21} - \beta_1 \alpha {\rho_{11}}^2 - \beta_2 {\rho_{21}}^2 \end{array}} \\ f(\lambda_2) + \lambda_1 \eta^3 - \beta_1 {\rho_{12}}^3 - \beta_2 {\rho_{22}}^3 \\ h(\lambda_2)\rho_{12} + \lambda_1 \rho_{11} \eta^2 - \beta_1 {\rho_{12}}^2 - \beta_2 \alpha {\rho_{22}}^2 \\ h(\lambda_2)\rho_{22} + \lambda_1 \rho_{21} \eta^2 - \beta_1 \alpha {\rho_{12}}^2 - \beta_2 {\rho_{22}}^2 \\ h(\lambda_2)\eta + q(\lambda_1)\eta^2 - \beta_1 \rho_{11} {\rho_{12}}^2 - \beta_2 \rho_{21} {\rho_{22}}^2 \end{pmatrix}$$

where $h(z) = \frac{-1}{g(z)}$ and $q(z) = z + \frac{g(z)}{3}$.

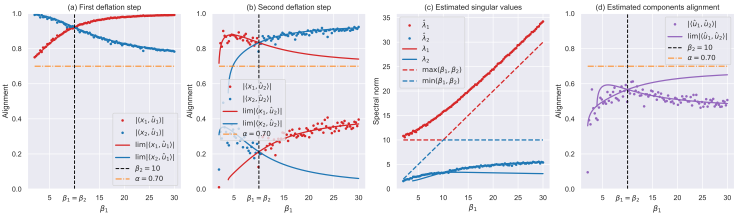
Phase Diagram

- Solving the three first equations in λ_1, ρ_{11} and ρ_{21}.
- Random initialisation, take solutions with $\rho_{ij} \in [0,1]$ and $\lambda_1 > 2\sqrt{\frac{2}{3}}$.

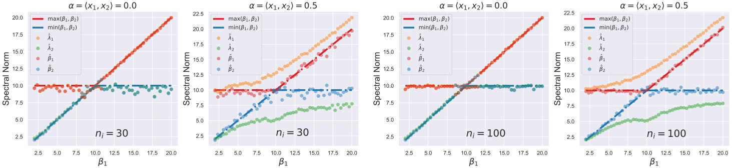


Theory versus Simulation

- Fixing $\boldsymbol{\beta}$ and solving $\psi(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\rho}) = 0$ in $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$.
- Random initialisation, solutions with $\eta, \rho_{ij} \in [0, 1]$ and $\lambda_1, \lambda_2 > 2\sqrt{\frac{2}{3}}$.
- Simulations with deflation using tensor power iteration initialized with tensor-SVD on a tensor of shape $(50, 50, 50)$.

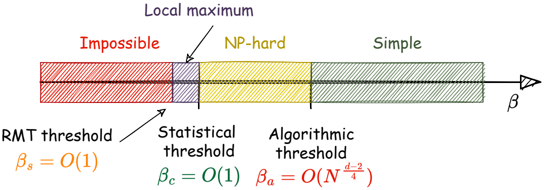


- Compute $\hat{\boldsymbol{\lambda}} = (\hat\lambda_1, \hat\lambda_2, \hat\eta)$ with deflation and solve $\psi(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\rho}}) = 0$ in $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\rho}}$.



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?
- Study the **existence** and **uniqueness** of the solutions of the deflation case.
- **Universality** and generalization to other **higher-rank** decomposition methods.

Thank you for your attention!

melaseddik.github.io

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

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)

Harold Hotelling. "Analysis of a complex of statistical variables into principal components". In: *Journal of Educational Psychology* (1933)

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

Mohamed El Amine Seddik, Maxime Guillaud, and Alexis Decurninge. "On the Accuracy of Hotelling-Type Tensor Deflation: A Random Tensor Analysis". In: *arXiv preprint arXiv:2211.09004* (2022)