A random matrix perspective on random tensors

Henrique Goulart

Toulouse INP / IRIT

henrique.goulart@irit.fr

Workshop on Tensor Theory and Methods – November 2022

Central object

Model: Rank-1 symmetric spiked tensor (Montanari & Richard, 2014)

$$\underbrace{\mathcal{Y} = \lambda\, x^{\otimes d}}_{\text{signal}} + \overbrace{\frac{1}{\sqrt{N}}\,\mathcal{W}}^{\text{noise}}$$

- SNR $\lambda > 0$
- $x \in \mathbb{S}^{N-1}$
- $\mathcal{W} \in \mathcal{S}^d(N)$:
  $W_{i_1 \ldots i_d} = W_{\pi(i_1 \ldots i_d)}, \ \forall \pi \in \mathfrak{S}_d$

Problem of interest: estimation of x from $\mathcal{Y}$ in the large-N limit ($N \to \infty$).

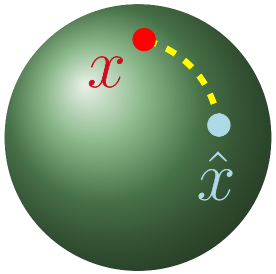
Closely related to rank-1 approximation or approximate rank-1 CPD.

Applications: latent variable analysis, computer vision

Asymptotic performance limits?

Given any estimator $\hat{x} : \mathcal{S}^d(N) \to \mathbb{S}^{N-1}$, a natural performance measure is the alignment (or overlap):

$$\alpha_{d,N}(\lambda) := \langle x, \hat{x}(\boldsymbol{\mathcal{Y}}) \rangle \quad \in [-1, 1]$$

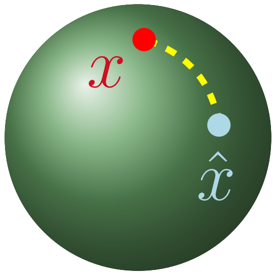


Fact: If $\hat{x} \sim \mathcal{U}(\mathbb{S}^{N-1})$, then asymptotically $x \perp \hat{x}$ a.s.

Asymptotic performance limits?

Given any estimator $\hat{x} : \mathcal{S}^d(N) \to \mathbb{S}^{N-1}$, a natural performance measure is the alignment (or overlap):

$$\alpha_{d,N}(\lambda) := \langle x, \hat{x}(\boldsymbol{\mathcal{Y}}) \rangle \quad \in [-1, 1]$$



Fact: If $\hat{x} \sim \mathcal{U}(\mathbb{S}^{N-1})$, then asymptotically $x \perp \hat{x}$ a.s.

Central questions:

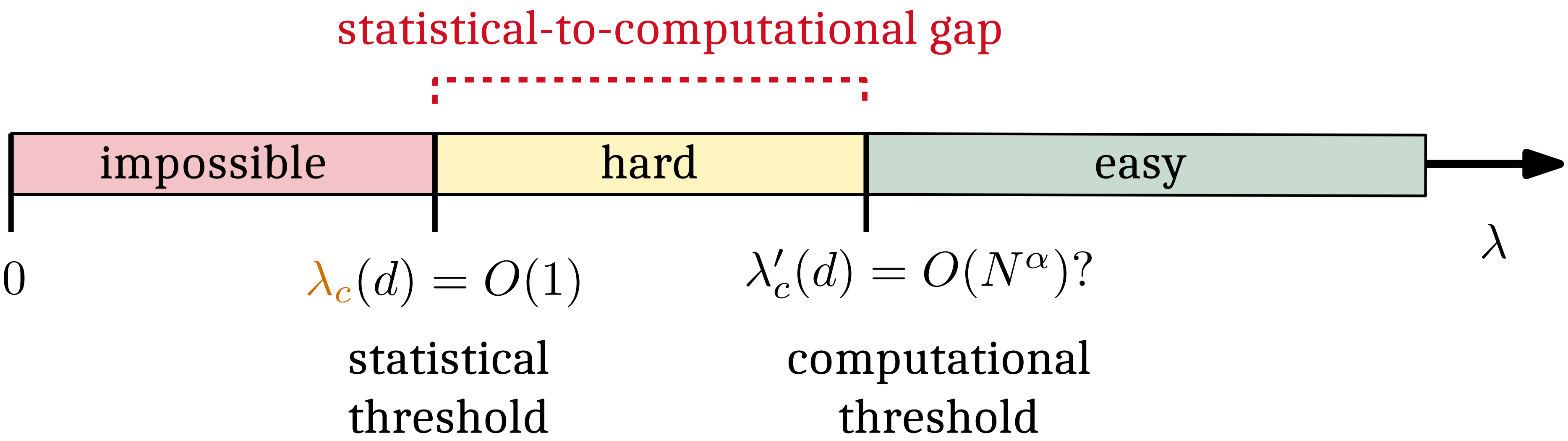
1. Weak recovery: for which λ is there a $\hat{x}$ such that

$$\limsup_{N\to\infty} \mathbb{E}\left\{|\alpha_{d,N}(\lambda)|\right\} > 0 \, ?$$

2. Best asymptotic alignment: what is the largest attainable value of $\limsup_{N\to\infty} \mathbb{E}\left\{\alpha_{d,N}(\lambda)\right\}$ for each λ?

Answers and conjectured gap

1. Regimes of weak recovery:



2. Maximum likelihood estimation attains the information-theoretic bound on the alignment for all λ.

(Richard & Montanari, 2014), (Montanari et al., 2015), (Ben Arous et al., 2019), (Jagannath et al., 2020), (Perry et al., 2020), (Ros et al., 2020)

This talk

1. Performance of maximum likelihood estimation

2. Tensor eigenpairs and the contraction ensemble (of random matrices)

3. Leveraging random matrix theory tools

4. Summary, extensions and open questions

Noise model: tensor GOE

$\mathcal{W}$ $\in$ tensor Gaussian orthogonal ensemble

$Q \in \mathsf{O}(N)$

$$p(\mathcal{W}) = \frac{1}{Z_d(N)} \exp\left(-\frac{1}{2}\|\mathcal{W}\|_{\mathsf{F}}^2\right) = p((Q, \ldots, Q) \cdot \mathcal{W})$$

Noise model: tensor GOE

$\mathcal{W}$ $\in$ tensor Gaussian orthogonal ensemble

$Q \in \mathrm{O}(N)$

$$p(\mathcal{W}) = \frac{1}{Z_d(N)} \exp\left(-\frac{1}{2}\|\mathcal{W}\|_{\mathsf{F}}^2\right) = p((Q, \ldots, Q) \cdot \mathcal{W})$$

Consequences:

1. $\mathrm{Var}(W_{i_1 \ldots i_d})$ depends on the pattern of repetitions in $(i_1, \ldots, i_d)$, e.g.:

$$\text{for } d = 3, \quad \|\mathcal{W}\|_{\mathsf{F}}^2 = \sum_i W_{iii}^2 + 3 \sum_{i<j} (W_{iij}^2 + W_{ijj}^2) + 6 \sum_{i<j<k} W_{ijk}^2$$

2. Law of $\mathcal{Y}$: $\quad p(\mathcal{Y} \,|\, x) \sim \exp\left(-\frac{N}{2} \left\|\mathcal{Y} - \lambda\, x^{\otimes d}\right\|_{\mathsf{F}}^2\right)$

Maximum likelihood estimator (MLE):

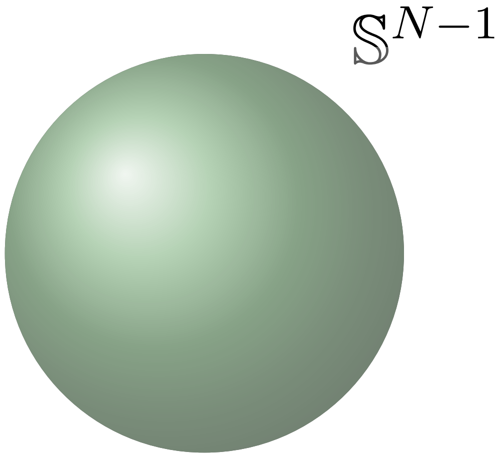
$$\hat{x} := \arg\min_{\|u\|=1} \left\|\mathcal{Y} - \lambda\, u^{\otimes d}\right\|_{\mathsf{F}}^2 = \arg\max_{\|u\|=1} \sum_{i_1 \ldots i_d} Y_{i_1 \ldots i_d} \prod_q u_{i_q}$$

MLE and tensor PCA

$$\max_{\|u\|=1} \sum_{i_1 \ldots i_d} Y_{i_1 \ldots i_d} \prod_q u_{i_q} = \max_{\|u\|=1} \boldsymbol{\mathcal{Y}} \cdot u^d, \qquad \text{with } \boldsymbol{\mathcal{Y}} \in \mathcal{S}^d(N)$$

MLE and tensor PCA

$$\max_{\|u\|=1} \sum_{i_1 \dots i_d} Y_{i_1 \dots i_d} \prod_q u_{i_q} = \max_{\|u\|=1} \boldsymbol{\mathcal{Y}} \cdot u^d, \qquad \text{with } \boldsymbol{\mathcal{Y}} \in \mathcal{S}^d(N)$$



Maximization of homogeneous poly on $\mathbb{S}^{N-1}$

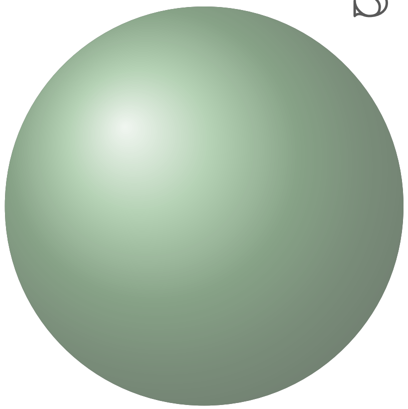
- closely related to the spectral norm $\|\boldsymbol{\mathcal{Y}}\|$
- non-convex
- worst-case NP-hard

MLE and tensor PCA

$$\max_{\|u\|=1} \sum_{i_1 \dots i_d} Y_{i_1 \dots i_d} \prod_q u_{i_q} = \max_{\|u\|=1} \mathcal{Y} \cdot u^d, \qquad \text{with } \mathcal{Y} \in \mathcal{S}^d(N)$$

Maximization of homogeneous poly on $\mathbb{S}^{N-1}$

- closely related to the spectral norm $\|\mathcal{Y}\|$
- non-convex
- worst-case NP-hard



Expected: $\lim_{N\to\infty} \mathbb{E}\{\alpha_{d,N}(\lambda)\} \approx \begin{cases} 1 & \text{for "large" } \lambda \\ 0 & \text{for "small" } \lambda \end{cases}$

But how exactly does $\alpha_{d,N}(\lambda)$ behave ?

(Related Q : does $\mathbb{E}\left\{\mathcal{Y} \cdot \hat{x}_{\mathrm{ML}}^d\right\}$ approach a limit ?

Expected: $\lim_{N\to\infty} \mathbb{E}\left\{\mathcal{Y} \cdot \hat{x}_{\mathrm{ML}}^d\right\} \approx \lambda$ for "large" λ)

MLE performance

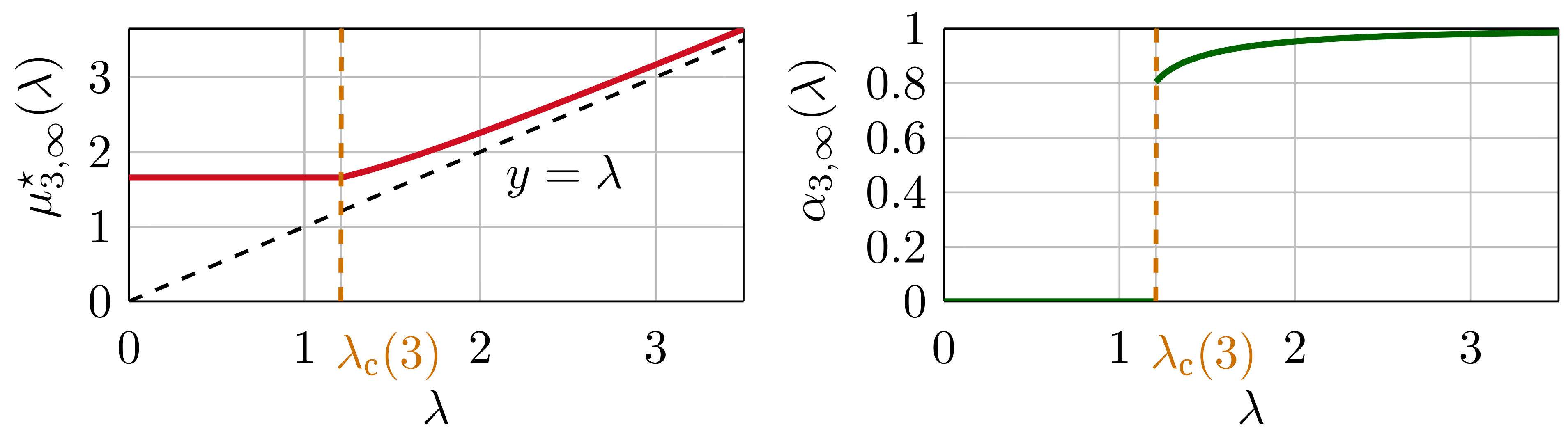
Settled by Jagannath–Lopatto–Miolane (2020), thanks to spin glass theory:

$$\mu^\star_{d,N}(\lambda) = \max_{\|u\|=1} \left\{ \lambda \langle x, u \rangle^d + \frac{1}{\sqrt{N}} \mathcal{W} \cdot u^d \right\} \xrightarrow[N\to\infty]{\text{a.s.}} \mathsf{GS}_d + \int_0^\lambda q_d^*(t)^{d/2}\, dt$$

$$|\alpha_{d,N}(\lambda)| = |\langle x, \hat{x}_{\mathrm{ML}}(\lambda)\rangle| \xrightarrow[N\to\infty]{\text{a.s.}} \sqrt{q_d^*(\lambda)}$$

Explicit expressions exist for $d = 3, 4, 5$.

For all d, these quantites undergo a phase transition at a threshold $\lambda_{\mathrm{c}}(d) = O(1)$.



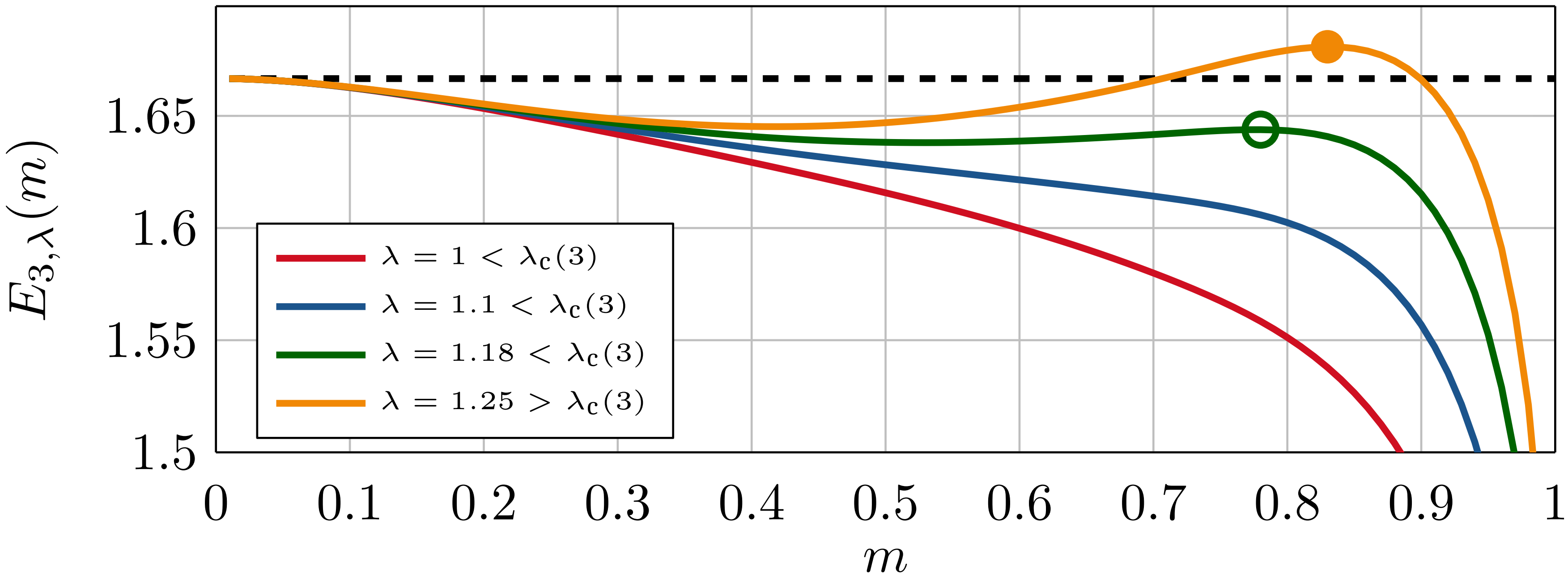
Furthermore, the MLE attains the bound $\limsup \mathbb{E}\left\{\langle x, \hat{x}\rangle^d\right\} \leq q_d^*(\lambda)^{d/2}$

Why the discontinuity?

Jagannath et al. (2020) derived a variational expression for the limiting expected max of $\mathcal{Y} \cdot u^d$ at a fixed allignment $\langle x, u \rangle = m$:

$$E_{d,\lambda}(m) = \lim_{N\to\infty} \mathbb{E} \left\{ \max_{\substack{\|u\|=1 \\ \langle u,x\rangle = m}} \lambda \langle u, x \rangle^d + \frac{1}{\sqrt{N}} \mathcal{W} \cdot u^d \right\}$$

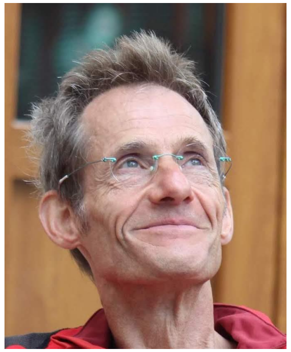
Numerical evaluation for $d = 3$:



This talk

1. Performance of maximum likelihood estimation
2. Tensor eigenpairs and the contraction ensemble (of random matrices)
3. Leveraging random matrix theory tools
4. Summary, extensions and open questions

From here on : joint work with Romain Couillet and Pierre Comon (JMLR '22)



Tensor eigenpairs and MLE

ML problem

$$\max_{\|u\|=1} \mathcal{Y} \cdot u^d$$

Lagrangian

$$L(\mu, u) = \frac{1}{d} \mathcal{Y} \cdot u^d - \frac{\mu}{2} \left(\|u\|^2 - 1\right)$$

Tensor eigenpairs and MLE

ML problem

$$\max_{\|u\|=1} \boldsymbol{\mathcal{Y}} \cdot u^d$$

Lagrangian

$$L(\mu, u) = \frac{1}{d} \boldsymbol{\mathcal{Y}} \cdot u^d - \frac{\mu}{2} \left(\|u\|^2 - 1 \right)$$

Critical points satisfy

$$\frac{\partial}{\partial u} L(\mu, u) = \boldsymbol{\mathcal{Y}} \cdot u^{d-1} - \mu\, u = 0, \qquad \|u\| = 1$$

Tensor eigenpairs and MLE

ML problem

$$\max_{\|u\|=1} \boldsymbol{\mathcal{Y}} \cdot u^d$$

Lagrangian

$$L(\mu, u) = \frac{1}{d} \boldsymbol{\mathcal{Y}} \cdot u^d - \frac{\mu}{2} \left(\|u\|^2 - 1\right)$$

Critical points satisfy

$$\frac{\partial}{\partial u} L(\mu, u) = \boldsymbol{\mathcal{Y}} \cdot u^{d-1} - \mu\, u = 0, \qquad \|u\| = 1$$

Tensor ℓ_2-eigenvalue equations : (Lim, 2005)

$$\boldsymbol{\mathcal{Y}} \cdot u^{d-1} = \mu\, u, \qquad \|u\| = 1$$

In particular :

$$\boldsymbol{\mathcal{Y}} \cdot \hat{x}_{\mathrm{ML}}^{d-1} = \mu_{\max}\, \hat{x}_{\mathrm{ML}}$$

Tensor and matrix eigenpairs

Another characterization of tensor eigenpairs (assuming $\|u\| = 1$):

$$(\mu, u) \text{ eigenpair of } \boldsymbol{\mathcal{Y}} \quad \Leftrightarrow \quad (\mu, u) \text{ eigenpair of } \boldsymbol{\mathcal{Y}} \cdot u^{d-2}$$

Tensor and matrix eigenpairs

Another characterization of tensor eigenpairs (assuming $\|u\| = 1$):

$$(\mu, u) \text{ eigenpair of } \boldsymbol{\mathcal{Y}} \quad \Leftrightarrow \quad (\mu, u) \text{ eigenpair of } \boldsymbol{\mathcal{Y}} \cdot u^{d-2}$$

Proof: $\mu\, u = \boldsymbol{\mathcal{Y}} \cdot u^{d-1} = (\boldsymbol{\mathcal{Y}} \cdot u^{d-2})\, u$

Tensor and matrix eigenpairs

Another characterization of tensor eigenpairs (assuming $\|u\| = 1$):

$$(\mu, u) \text{ eigenpair of } \mathcal{Y} \quad \Leftrightarrow \quad (\mu, u) \text{ eigenpair of } \mathcal{Y} \cdot u^{d-2}$$

Proof: $\mu\, u = \mathcal{Y} \cdot u^{d-1} = (\mathcal{Y} \cdot u^{d-2})\, u$

In particular, if (μ, u) is a local max, then $\operatorname{Sp}(\mathcal{Y} \cdot u^{d-2}) \setminus \{\mu\} \subset \;]-\infty, \frac{\mu}{d-1}]$

Proof: Apply the second-order necessary condition

$$\langle \nabla^2_{uu} L(\mu, u)\, w, w \rangle \leq 0, \qquad \forall\, w \in u^\perp$$

with $\nabla^2_{uu} L(\mu, u) = \frac{\partial}{\partial u}\left[\mathcal{Y} \cdot u^{d-1} - \mu u\right] = (d-1)\, \mathcal{Y} \cdot u^{d-2} - \mu I$ to get

$$\max_{\|w\|=1,\, w \in u^\perp} \langle (\mathcal{Y} \cdot u^{d-2})\, w, w \rangle \leq \frac{\mu}{d-1}$$

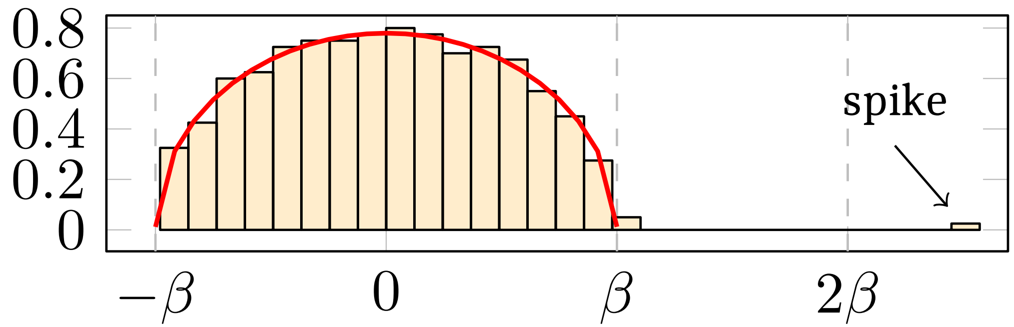
From spiked tensor model to matrix models

Idea : study spiked rank-one matrix models at critical points (μ, u)

$$\mathcal{Y} \cdot u^{d-2} = \lambda \, \langle x, u \rangle^{d-2} \, x \, x^\mathsf{T} + \frac{1}{\sqrt{N}} \mathcal{W} \cdot u^{d-2}$$

- SNR weighted by alignment $\langle x, u \rangle$
- $\mathcal{W}$ and u are correlated $\Rightarrow$ "spike" at every local max u regardless of λ

$\mathrm{Sp}(\mathcal{Y} \cdot u^{d-2}) \quad (d = 3)$



- Special matrices from contraction ensemble

$$\mathcal{M}_{\mathcal{Y}} := \{\mathcal{Y} \cdot v^{d-2} \ : \ v \in \mathbb{S}^{N-1}\}$$

This talk

1. Performance of maximum likelihood estimation
2. Tensor eigenpairs and the contraction ensemble (of random matrices)
3. Leveraging random matrix theory tools
4. Summary, extensions and open questions

Contractions $\mathcal{Y} \cdot u^{d-2}$ at critical points

Recall the tensor eigenvalue equation:

$$\mu\, u = \mathcal{Y} \cdot u^{d-1} = \lambda \, \langle x, u \rangle^{d-1} \, x + \frac{1}{\sqrt{N}} \mathcal{W} \cdot u^{d-1}$$

Our quantities of interest are obtained by taking scalar products with u and x :

$$\begin{aligned} \mu &= \lambda \, \langle x, u \rangle^{d} + \frac{1}{\sqrt{N}} \mathcal{W} \cdot u^{d}, \\ \langle x, u \rangle &= \frac{\lambda}{\mu} \, \langle x, u \rangle^{d-1} + \frac{1}{\mu \sqrt{N}} \left\langle x, \mathcal{W} \cdot u^{d-1} \right\rangle \qquad (\mu \neq 0) \end{aligned}$$

Contractions $\mathcal{Y} \cdot u^{d-2}$ at critical points

Recall the tensor eigenvalue equation:

$$\mu\, u = \mathcal{Y} \cdot u^{d-1} = \lambda \, \langle x, u \rangle^{d-1} \, x + \frac{1}{\sqrt{N}} \mathcal{W} \cdot u^{d-1}$$

Our quantities of interest are obtained by taking scalar products with u and x :

$$\mu = \lambda \, \langle x, u \rangle^{d} + \frac{1}{\sqrt{N}} \mathcal{W} \cdot u^{d},$$

$$\langle x, u \rangle = \frac{\lambda}{\mu} \, \langle x, u \rangle^{d-1} + \frac{1}{\mu \sqrt{N}} \left\langle x, \mathcal{W} \cdot u^{d-1} \right\rangle \qquad (\mu \neq 0)$$

Q : Assuming $\mu \xrightarrow[N \to \infty]{\text{a.s.}} \mu_{d,\infty}$ and $\langle x, u \rangle \xrightarrow[N \to \infty]{\text{a.s.}} \alpha_{d,\infty}$, which solutions do we get for $(\mu_{d,\infty}, \alpha_{d,\infty})$?

Contractions $\mathcal{Y} \cdot u^{d-2}$ at critical points

Recall the tensor eigenvalue equation:

$$\mu\, u = \mathcal{Y} \cdot u^{d-1} = \lambda\, \langle x, u \rangle^{d-1}\, x + \frac{1}{\sqrt{N}} \mathcal{W} \cdot u^{d-1}$$

Our quantities of interest are obtained by taking scalar products with u and x :

$$\mathbb{E}\{\mu\} = \lambda\, \mathbb{E}\left\{\langle x, u \rangle^{d}\right\} + \frac{1}{\sqrt{N}}\, \mathbb{E}\left\{\mathcal{W} \cdot u^{d}\right\}$$

$$\mathbb{E}\{\langle x, u \rangle\} = \lambda\, \mathbb{E}\left\{\frac{1}{\mu}\, \langle x, u \rangle^{d-1}\right\} + \frac{1}{\sqrt{N}}\, \mathbb{E}\left\{\frac{1}{\mu} \left\langle x, \mathcal{W} \cdot u^{d-1} \right\rangle\right\} \qquad (\mu \neq 0)$$

Q : Assuming $\mu \xrightarrow[N\to\infty]{\text{a.s.}} \mu_{d,\infty}$ and $\langle x, u \rangle \xrightarrow[N\to\infty]{\text{a.s.}} \alpha_{d,\infty}$, which solutions do we get for $(\mu_{d,\infty}, \alpha_{d,\infty})$?

Contractions $\mathcal{Y} \cdot u^{d-2}$ at critical points

Recall the tensor eigenvalue equation:

$$\mu\, u = \mathcal{Y} \cdot u^{d-1} = \lambda \, \langle x, u \rangle^{d-1} \, x + \frac{1}{\sqrt{N}} \mathcal{W} \cdot u^{d-1}$$

Our quantities of interest are obtained by taking scalar products with u and x :

$$\mathbb{E}\{\mu\} = \lambda \, \mathbb{E}\left\{\langle x, u \rangle^{d}\right\} + \frac{1}{\sqrt{N}} \mathbb{E}\left\{\mathcal{W} \cdot u^{d}\right\}$$

$$\mathbb{E}\{\langle x, u \rangle\} = \lambda \, \mathbb{E}\left\{\frac{1}{\mu} \langle x, u \rangle^{d-1}\right\} + \frac{1}{\sqrt{N}} \mathbb{E}\left\{\frac{1}{\mu} \langle x, \mathcal{W} \cdot u^{d-1} \rangle\right\} \qquad (\mu \neq 0)$$

Q : Assuming $\mu \xrightarrow[N\to\infty]{\text{a.s.}} \mu_{d,\infty}$ and $\langle x, u \rangle \xrightarrow[N\to\infty]{\text{a.s.}} \alpha_{d,\infty}$, which solutions do we get for $(\mu_{d,\infty}, \alpha_{d,\infty})$?

Technical tools

Computation of $\mathbb{E}\left\{\boldsymbol{\mathcal{W}} \cdot u^d\right\}$ and $\mathbb{E}\left\{\frac{1}{\mu}\left\langle x, \boldsymbol{\mathcal{W}} \cdot u^{d-1}\right\rangle\right\}$:

1. Gaussian integration-by-parts

$$z \sim \mathcal{N}(0, \sigma^2) \quad \Rightarrow \quad \mathbb{E}\left\{z\, f(z)\right\} = \sigma^2\, \mathbb{E}\left\{f'(z)\right\}$$

Technical tools

Computation of $\mathbb{E}\left\{\mathcal{W}\cdot u^{d}\right\}$ and $\mathbb{E}\left\{\frac{1}{\mu}\left\langle x, \mathcal{W}\cdot u^{d-1}\right\rangle\right\}$:

1. Gaussian integration-by-parts

$$z \sim \mathcal{N}(0,\sigma^2) \quad \Rightarrow \quad \mathbb{E}\left\{z\, f(z)\right\} = \sigma^2\, \mathbb{E}\left\{f'(z)\right\}$$

$$\mathbb{E}\left\{\mathcal{W}\cdot u^{d}\right\} = \sum_{\boldsymbol{i}} \mathbb{E}\left\{W_{\boldsymbol{i}}\, u_{i_1}\dots u_{i_d}\right\} = \sum_{\boldsymbol{i}} \sigma_{\boldsymbol{i}}^2 \sum_{j=1}^{d} \mathbb{E}\left\{\frac{\partial u_{i_j}}{\partial W_{\boldsymbol{i}}} \prod_{k\neq j}^{d} u_{i_k}\right\}$$

$$\mathbb{E}\left\{\frac{1}{\mu}\left\langle x, \mathcal{W}\cdot u^{d-1}\right\rangle\right\} = \sum_{\boldsymbol{i}} \sigma_{\boldsymbol{i}}^2\, \mathbb{E}\left\{-\frac{1}{\mu^2}\frac{\partial \mu}{\partial W_{\boldsymbol{i}}}\left(\prod_{k=1}^{d-1} u_{i_k}\right) x_{i_d} + \frac{1}{\mu} x_{i_d} \sum_{j=1}^{d-1} \frac{\partial u_{i_j}}{\partial W_{\boldsymbol{i}}} \prod_{k\neq j}^{d-1} u_{i_k}\right\}.$$

Technical tools

Computation of $\mathbb{E}\left\{\mathcal{W}\cdot u^d\right\}$ and $\mathbb{E}\left\{\frac{1}{\mu}\left\langle x, \mathcal{W}\cdot u^{d-1}\right\rangle\right\}$:

1. Gaussian integration-by-parts

$$z \sim \mathcal{N}(0,\sigma^2) \quad \Rightarrow \quad \mathbb{E}\left\{z\, f(z)\right\} = \sigma^2\, \mathbb{E}\left\{f'(z)\right\}$$

$$\mathbb{E}\left\{\mathcal{W}\cdot u^d\right\} = \sum_{\boldsymbol{i}} \mathbb{E}\left\{W_{\boldsymbol{i}}\, u_{i_1}\dots u_{i_d}\right\} = \sum_{\boldsymbol{i}} \sigma_{\boldsymbol{i}}^2 \sum_{j=1}^{d} \mathbb{E}\left\{\frac{\partial u_{i_j}}{\partial W_{\boldsymbol{i}}} \prod_{k\neq j}^{d} u_{i_k}\right\}$$

$$\mathbb{E}\left\{\frac{1}{\mu}\left\langle x, \mathcal{W}\cdot u^{d-1}\right\rangle\right\} = \sum_{\boldsymbol{i}} \sigma_{\boldsymbol{i}}^2\, \mathbb{E}\left\{-\frac{1}{\mu^2}\frac{\partial \mu}{\partial W_{\boldsymbol{i}}}\left(\prod_{k=1}^{d-1} u_{i_k}\right) x_{i_d} + \frac{1}{\mu} x_{i_d} \sum_{j=1}^{d-1}\frac{\partial u_{i_j}}{\partial W_{\boldsymbol{i}}}\prod_{k\neq j}^{d-1} u_{i_k}\right\}.$$

2. Implicit function theorem to compute the required derivatives :

$$\begin{pmatrix} \dfrac{\partial u}{\partial W_{\boldsymbol{i}}} \\[2ex] \dfrac{\partial \mu}{\partial W_{\boldsymbol{i}}} \end{pmatrix} = \begin{pmatrix} -\dfrac{1}{(d-1)\sqrt{N}} R\left(\dfrac{\mu}{d-1}\right)\phi + \dfrac{1}{(d-2)\,\mu}\dfrac{\partial \mu}{\partial W_{\boldsymbol{i}}}\, u \\[2ex] \dfrac{1}{\sigma_{\boldsymbol{i}}^2\sqrt{N}} \displaystyle\prod_{j=1}^{d} u_{i_j} \end{pmatrix}$$

Byproduct: limiting spectrum of contractions

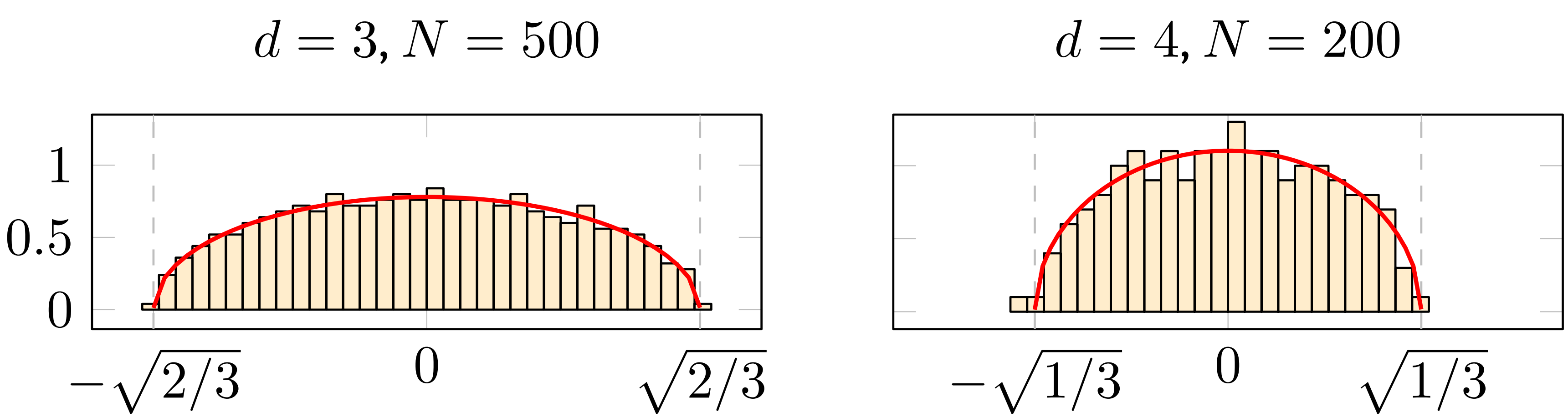
Theorem. Take a sequence of vectors $v \in \mathbb{S}^{N-1}$ and of Gaussian tensors $\mathcal{W}$.

Define the empirical spectral measure $\rho_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\nu_i}$, where

$$\{\nu_i\}_{i=1}^{N} = \mathrm{Sp}\left(\frac{1}{\sqrt{N}}\mathcal{W} \cdot v^{d-2}\right)$$

Then, ρ_N converges a.s. weakly to a semicircle law ρ with density

$$\rho(dx) = \frac{2}{\pi\,\beta_d^2}\sqrt{(\beta_d^2 - x^2)^+}\,dx, \qquad \beta_d := \frac{2}{\sqrt{d(d-1)}}.$$



Main result

Theorem. Let (μ, u) be a sequence of critical points of the MLE problem s.t.

$$\langle x, u \rangle \xrightarrow[N\to\infty]{\text{a.s.}} \alpha_{d,\infty}(\lambda) > 0,$$

$$\mu \xrightarrow[N\to\infty]{\text{a.s.}} \mu_{d,\infty}(\lambda) > (d-1)\beta_d.$$

Then, $\mu_{d,\infty}(\lambda)$ satisfies the fixed-point equation

$$\mu_{d,\infty}(\lambda) = \phi_d(\mu_{d,\infty}(\lambda), \lambda),$$

where

$$\phi_d(z,\lambda) := \lambda\, \omega_d^d(z,\lambda) - \frac{1}{d-1}\, m_d\left(\frac{z}{d-1}\right),$$

$$\omega_d(z,\lambda) := \left[\frac{1}{\lambda}\left(z + \frac{1}{d}\, m_d\left(\frac{z}{d-1}\right)\right)\right]^{\frac{1}{d-2}},$$

$$m_d(z) = \frac{2}{\beta_d^2}\left(-z + z\sqrt{1 - \frac{\beta_d^2}{z^2}}\right) \qquad \text{(Stieltjes transform of } \rho\text{)}$$

Furthermore, $\alpha_{d,\infty}(\lambda) = \omega_d(\mu_{d,\infty}(\lambda), \lambda)$.

Relation to known results

For $d = 3$, the only positive solution to the fixed-point equation is:

$$\mu_{3,\infty}(\lambda) = \frac{3\lambda^2 + \lambda\sqrt{9\lambda^2 - 12} + 4}{\sqrt{18\lambda^2 + 6\lambda\sqrt{9\lambda^2 - 12}}}, \qquad \lambda \geq 2/\sqrt{3},$$

which gives

$$\alpha_{3,\infty}(\lambda) = \sqrt{\frac{1}{2} + \sqrt{\frac{3\lambda^2 - 4}{12\lambda^2}}}.$$

These equations precisely describe the local max of $E_{d,\lambda}(m)$ that becomes global (MLE) for $\lambda > \lambda_{\mathrm{c}}(d)$.

Relation to known results

For $d = 3$, the only positive solution to the fixed-point equation is:

$$\mu_{3,\infty}(\lambda) = \frac{3\lambda^2 + \lambda\sqrt{9\lambda^2 - 12} + 4}{\sqrt{18\lambda^2 + 6\lambda\sqrt{9\lambda^2 - 12}}}, \qquad \lambda \geq 2/\sqrt{3},$$

which gives

$$\alpha_{3,\infty}(\lambda) = \sqrt{\frac{1}{2} + \sqrt{\frac{3\lambda^2 - 4}{12\lambda^2}}}.$$

These equations precisely describe the local max of $E_{d,\lambda}(m)$ that becomes global (MLE) for $\lambda > \lambda_c(d)$.

By numerical verification: also true beyond $\lambda_c(d)$ for $d = 4, 5$.

We conjecture that this holds for all $d \geq 3$.

But why ?

Ingredients: tensor eigenvalue equation satsfied by all critical points and a.s. convergence

$$\langle x, u \rangle \xrightarrow[N\to\infty]{\text{a.s.}} \alpha_{d,\infty}(\lambda) > 0$$
$$\mu \xrightarrow[N\to\infty]{\text{a.s.}} \mu_{d,\infty}(\lambda) > (d-1)\beta_d$$

Output: description of global maximizer(s) for $\lambda > \lambda_{\mathrm{c}}(d)$.

But why ?

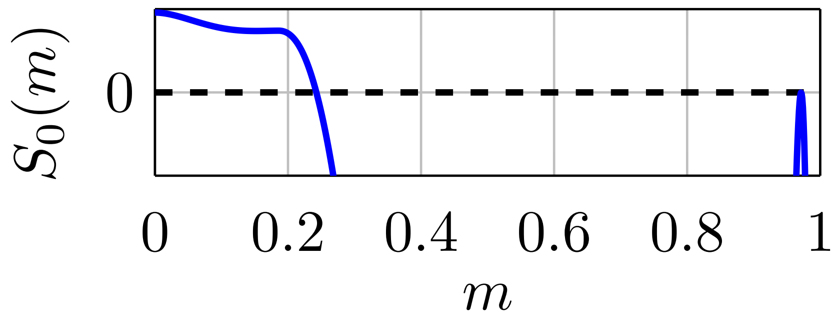
Ingredients: tensor eigenvalue equation satsfied by all critical points and a.s. convergence

$$\langle x, u \rangle \xrightarrow[N\to\infty]{\text{a.s.}} \alpha_{d,\infty}(\lambda) > 0$$
$$\mu \xrightarrow[N\to\infty]{\text{a.s.}} \mu_{d,\infty}(\lambda) > (d-1)\beta_d$$

Output: description of global maximizer(s) for $\lambda > \lambda_c(d)$.

The conditions on $\alpha_{d,\infty}$ and $\mu_{d,\infty}$ could be met by some strictly local max beyond $\lambda_c(d)$, as per Ben Arous–Mei–Montanari (2019)

expected # local max with alignment m $\sim \exp(N\, S_0(m))$



This selectivity probably comes from the a.s. convergence assumption.

This talk

1. Performance of maximum likelihood estimation
2. Tensor eigenpairs and the contraction ensemble (of random matrices)
3. Leveraging random matrix theory tools
4. Summary, extensions and open questions

Summary

Rank-one symmetric tensor model: simple but quite rich

$$Y_{i_1 \ldots i_d} = \lambda \, x_{i_1} \ldots x_{i_d} + \frac{1}{\sqrt{N}} W_{i_1 \ldots i_d}$$

Statistical thresholds, MLE landscape and performance now well understood, largely thanks to spin glass theory.

Standard RMT tools can be leveraged by studying contractions and

- bring additional insights
- provide more elementary means of reaching some of those results
- are flexible and accessible for extensions/generalization

Extensions

- Extension to asymmetric models by Seddik–Guillaud–Couillet (2022):

$$\mathcal{Y} = \lambda\, x^{(1)} \otimes \cdots \otimes x^{(d)} + \frac{1}{\sqrt{N_1 + \cdots + N_d}} \mathcal{W}$$

  with $W_{i_1 \ldots i_d} \sim \mathcal{N}(0, 1)$

- Joint spectral law of contracted ensembles for Wigner-type tensors: Au & Garza-Vargas (2021)

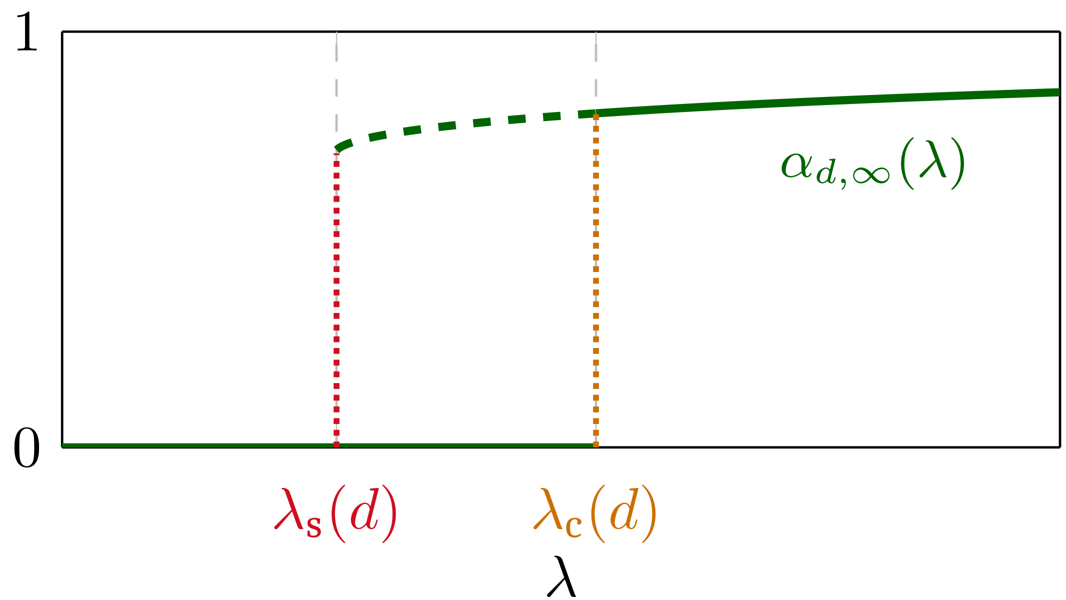
- Higher rank: straightforward for (all-)orthogonal rank-1 terms, but hard otherwise

$$\mathcal{Y} = \sum_{r=1}^{R} \lambda_r\, x_r^{\otimes d} + \frac{1}{\sqrt{N}} \mathcal{W}$$

- Other noise models: possible with interpolation tools of Pastur & Shchrebina (2011)

Open questions

- Proof that limiting equations describe MLE?
- Can we determine the phase transition with this approach?



- Can we extend it to higher-rank tensors?

For more details: JMLR 21-1038

henrique.goulart @ irit.fr

Last but not least, the bibliography (1/2)

- Montanari, A., & Richard, E. A statistical model for tensor PCA. NIPS, 2014.
- Montanari, A., Reichman, D., & Zeitouni, O. On the Limitation of Spectral Methods: From the Gaussian Hidden Clique Problem to Rank One Perturbations of Gaussian Tensors. NIPS, 2015.
- Perry, A., Wein, A. S., & Bandeira, A. S. Statistical limits of spiked tensor models. AIHP-PS, 2020.
- Ben Arous, G., Mei, S., Montanari, A., & Nica, M. The Landscape of the Spiked Tensor Model. CPAM, 2019.
- Ros, V., Ben Arous, G., Biroli, G., & Cammarota, C. Complex Energy Landscapes in Spiked-Tensor and Simple Glassy Models: Ruggedness, Arrangements of Local Minima, and Phase Transitions. Physical Review X, 2019.
- Jagannath, A., Lopatto, P., & Miolane, L. Statistical thresholds for Tensor PCA. AAP, 2020.

Last but not least, the bibliography (2/2)

- Seddik, M. E. A., Guillaud, M., & Couillet, R. When Random Tensors meet Random Matrices. arXiv:2112.12348, 2021.
- Lim, L. H. Singular values and eigenvalues of tensors: A variational approach. CAMSAP, 2005.
- Pastur, L. & Shchrebina, M. Eigenvalue distribution of large random matrices. AMS, 2011.
- Au, B. & Garza-Vargas, J. Spectral asymptotics for contracted tensor ensembles. arXiv:2110.01652, 2021.