A random matrix perspective on random tensors

Henrique Goulart

Toulouse INP / IRIT

henrique.goulart@irit.fr

Random Tensors at CIRM – March 2022

Main ingredients

Optimization problem		Probabilistic model
$\max_{\lVert u \rVert = 1} \sum_{ijk} Y_{ijk}\, u_i u_j u_k$	$\longleftrightarrow$	$Y_{ijk} = \lambda\, x_i x_j x_k + \frac{1}{\sqrt{N}}\, W_{ijk}$
"Tensor PCA"		Rank-1 spiked model

Ingredient # 1: the optimization problem

$$\underbrace{\max_{\|u\|=1} \sum_{ijk} \overbrace{Y_{ijk}}^{\text{components of } \mathcal{Y}} u_i u_j u_k}_{\mathcal{Y}(u,u,u)}, \qquad \text{with } \begin{cases} 1 \leq i, j, k \leq \overbrace{N}^{\text{large}} \\ Y_{ijk} = Y_{p(ijk)}, \ \forall p \in \mathfrak{S}_3 \end{cases}$$

Ingredient # 1: the optimization problem

components of $\mathfrak{Y}$

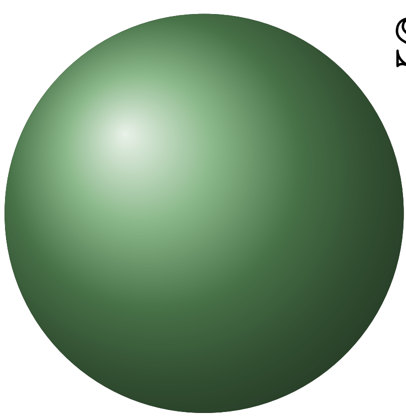
large

$$\max_{\|u\|=1} \sum_{ijk} Y_{ijk}\, u_i u_j u_k, \qquad \text{with} \begin{cases} 1 \leq i, j, k \leq N \\ Y_{ijk} = Y_{p(ijk)}, \ \forall\, p \in \mathfrak{S}_3 \end{cases}$$

$\mathfrak{Y}(u, u, u)$

Homogeneous poly. on

$\mathbb{S}^{N-1}$



- Defines the spectral norm $\|\mathfrak{Y}\|$
- Non-convex
- NP-hard
- Equivalent to:

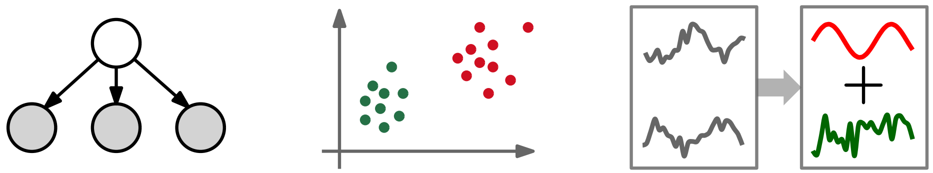
$$\min_{\mu,\, \|u\|=1} \sum_{ijk} (Y_{ijk} - \mu\, u_i u_j u_k)^2$$

Many applications (possibly in higher order)

- Latent variable model learning by decomposition of high-order statistics

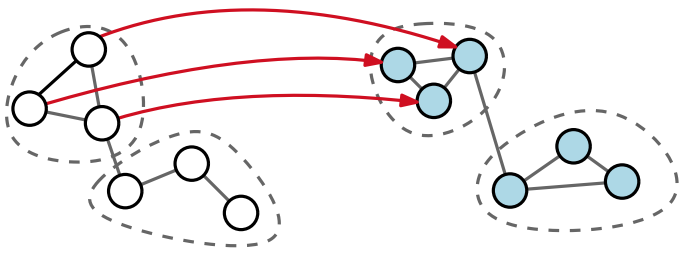
  Naive Bayes, GMM, ICA ...

  (Anandkumar et al., 2014)



- Hypergraph matching

  (Duchenne et al., 2011)



- Statistichal mechanics: spherical p-spin model

  (Crisanti & Sommers, 1992)

$$H(u) = \sum_{ijk} Y_{ijk}\, u_i u_j u_k$$

- . . .

Ingredient # 2: the probabilistic model

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

SNR

normalized, symmetric noise

$$\mathcal{Y} = \lambda \, x \otimes x \otimes x + \frac{1}{\sqrt{N}} \mathcal{W}$$

"signal" ($\|x\| = 1$)

Natural, direct extension of spiked matrix model: $Y = \lambda \, x \, x^\top + \frac{1}{\sqrt{N}} W.$

Ingredient # 2: the probabilistic model

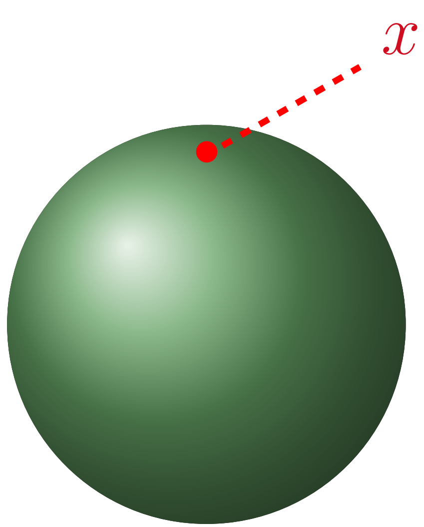
Rank-1 symmetric spiked tensor model (Montanari & Richard, 2014)

SNR — normalized, symmetric noise

$$\mathcal{Y} = \lambda\, x \otimes x \otimes x + \frac{1}{\sqrt{N}} \mathcal{W}$$

"signal" ($\|x\| = 1$)

Natural, direct extension of spiked matrix model: $Y = \lambda\, x\, x^\top + \frac{1}{\sqrt{N}} W.$



$\mathcal{W}$: Gaussian, orthogonally invariant $\Rightarrow$ x on north pole w.l.o.g.

Ingredient # 2: the probabilistic model

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

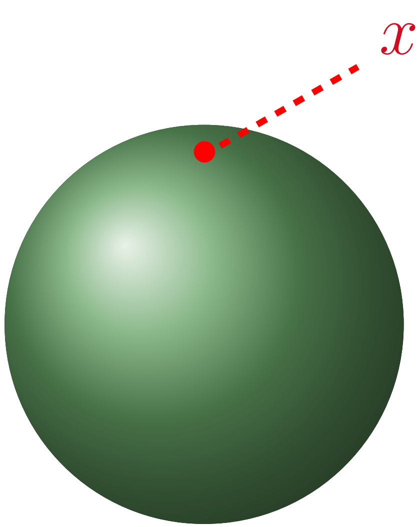
SNR

normalized, symmetric noise

$$\mathcal{Y} = \lambda \, x \otimes x \otimes x + \frac{1}{\sqrt{N}} \mathcal{W}$$

"signal" ($\|x\| = 1$)

Natural, direct extension of spiked matrix model: $Y = \lambda \, x \, x^{\mathsf{T}} + \frac{1}{\sqrt{N}} W.$



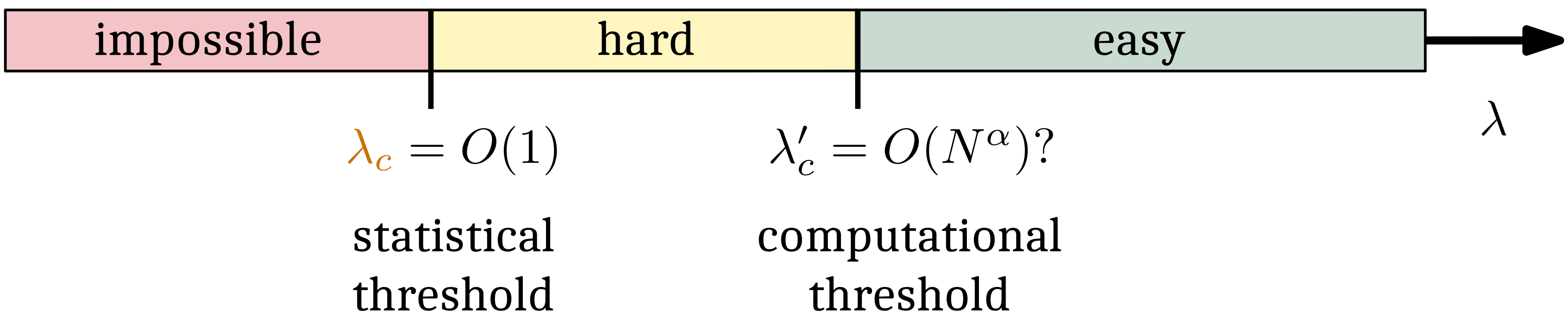
$\mathcal{W}$: Gaussian, orthogonally invariant $\Rightarrow$ x on north pole w.l.o.g.

$$\mathcal{Y}(u,u,u) = \sum_{ijk} \left(\lambda \, x_i x_j x_k + \frac{1}{\sqrt{N}} W_{ijk} \right) u_i u_j u_k$$

$$= \lambda \, \langle u, x \rangle^3 + \frac{1}{\sqrt{N}} \mathcal{W}(u,u,u)$$

Many related results in recent years

In particular, on the thresholds for estimation and detection :



(Richard & Montanari, 2014), (Montanari et al., 2015), (Hopkins et al., 2015), (Kim et al., 2017), (Ben Arous et al., 2019), (Jagannath et al., 2020), (Perry et al., 2020), (Ros et al., 2020)

This talk

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

Noise model: tensor GOE

Tensor Gaussian orthogonal ensemble

$$p(\mathcal{W}) = \frac{1}{Z_3(N)} \exp\left(-\frac{1}{2}\|\mathcal{W}\|_{\mathsf{F}}^2\right)$$

orthogonal

$$\mathcal{W} \overset{\text{dist}}{=} (Q, Q, Q) \cdot \mathcal{W}$$

Noise model: tensor GOE

Tensor Gaussian orthogonal ensemble

$$p(\mathcal{W}) = \frac{1}{Z_3(N)} \exp\left(-\frac{1}{2}\|\mathcal{W}\|_{\mathsf{F}}^2\right)$$

$$\mathcal{W} \overset{\text{dist}}{=} (Q, Q, Q) \cdot \mathcal{W}$$

orthogonal

Consequences:

1. $\mathrm{Var}(W_{ijk})$ depends on the pattern of repetitions in (i, j, k), since:

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

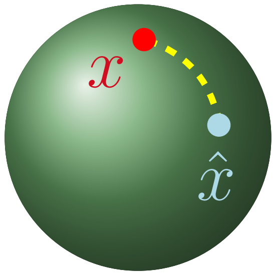
2. Law of $\mathcal{Y}$: $\quad p(\mathcal{Y} \,|\, x) \sim \exp\left(-\frac{N}{2}\|\mathcal{Y} - \lambda\, x \otimes x \otimes x\|_{\mathsf{F}}^2\right)$

Thus: $\hat{x} := \underset{\|u\|=1}{\arg\max} \sum_{ijk} Y_{ijk}\, u_i u_j u_k$ is the MLE of x

MLE performance

As $x, \hat{x} \in \mathbb{S}^{N-1}$, a natural performance measure is the alignment (or overlap) :

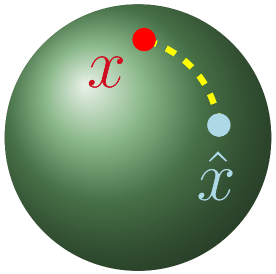
$$\alpha_N(\lambda) := |\langle x, \hat{x} \rangle| \quad \in [0, 1]$$



MLE performance

As $x, \hat{x} \in \mathbb{S}^{N-1}$, a natural performance measure is the alignment (or overlap) :

$$\alpha_N(\lambda) := |\langle x, \hat{x} \rangle| \quad \in [0, 1]$$

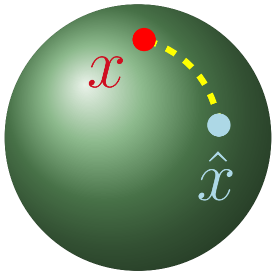


Does $\mathbb{E}\{\alpha_N(\lambda)\} \underset{N\to\infty}{\longrightarrow} \alpha_\infty(\lambda)$? When is $\alpha_\infty(\lambda) > 0$ (weak recovery) ?

MLE performance

As $x, \hat{x} \in \mathbb{S}^{N-1}$, a natural performance measure is the alignment (or overlap) :

$$\alpha_N(\lambda) := |\langle x, \hat{x} \rangle| \quad \in [0, 1]$$



Does $\mathbb{E}\{\alpha_N(\lambda)\} \underset{N\to\infty}{\longrightarrow} \alpha_\infty(\lambda)$? When is $\alpha_\infty(\lambda) > 0$ (weak recovery) ?

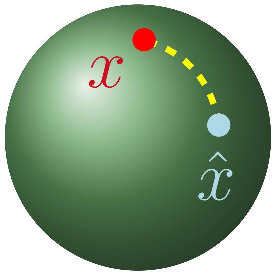
Expected: $\lim\limits_{N\to\infty} \mathbb{E}\{\alpha_N(\lambda)\} \approx \begin{cases} 1 & \text{for "large" } \lambda \\ 0 & \text{for "small" } \lambda \end{cases}$

But how exactly does this quantity behave ?

MLE performance

As $x, \hat{x} \in \mathbb{S}^{N-1}$, a natural performance measure is the alignment (or overlap) :

$$\alpha_N(\lambda) := |\langle x, \hat{x} \rangle| \quad \in [0, 1]$$



Does $\mathbb{E}\{\alpha_N(\lambda)\} \underset{N\to\infty}{\longrightarrow} \alpha_\infty(\lambda)$? When is $\alpha_\infty(\lambda) > 0$ (weak recovery) ?

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

But how exactly does this quantity behave ?

(Related question : does $\mathbb{E}\{\mathcal{Y}(\hat{x}, \hat{x}, \hat{x})\} = \mathbb{E}\{\|\mathcal{Y}\|\}$ approach a limit ?

Expected: $\lim_{N\to\infty} \mathbb{E}\{\|\mathcal{Y}\|\} \approx \lambda$ for "large" λ (since $\mathcal{Y}(\hat{x}, \hat{x}, \hat{x}) \approx \lambda \langle x, \hat{x} \rangle^3$))

An abrupt phase transition

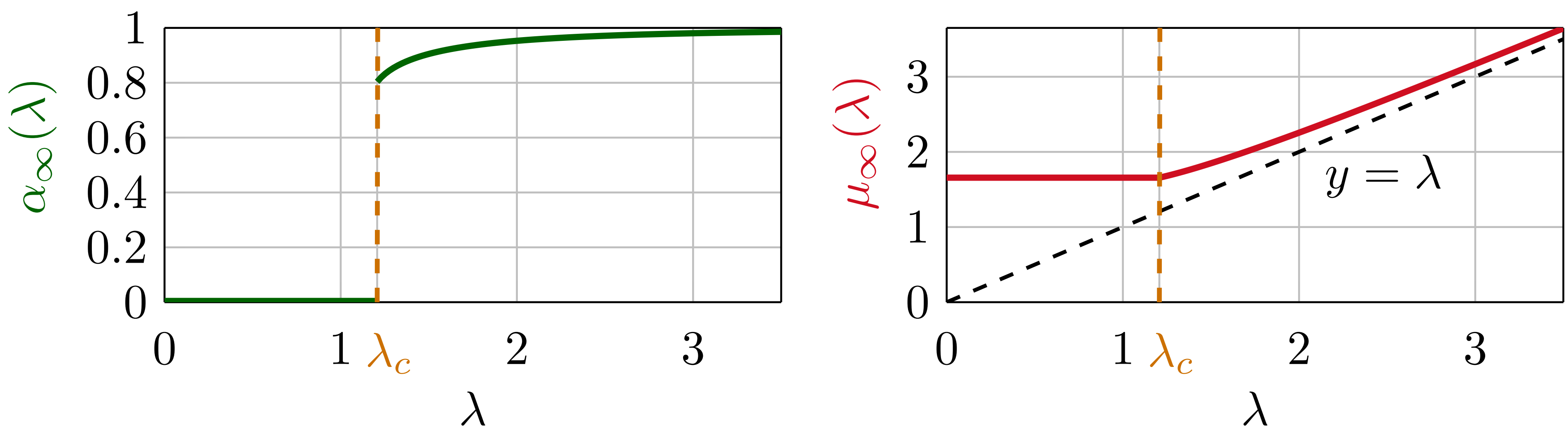
Precise answer by Jagannath–Lopatto–Miolane (2020) based on stat. phys. :

There exists an $O(1)$ threshold λ_c (≈ 1.207) such that

$$\alpha_N(\lambda) \xrightarrow[N\to\infty]{\text{a.s.}} \alpha_\infty(\lambda) = \begin{cases} \sqrt{\frac{1}{2} + \sqrt{\frac{3\lambda^2-4}{12\lambda^2}}}, & \lambda > \lambda_c \\ 0, & \lambda < \lambda_c \end{cases}$$

$$\|\mathcal{Y}\| \xrightarrow[N\to\infty]{\text{a.s.}} \mu_\infty(\lambda) = \begin{cases} \frac{3\lambda^2+\lambda\sqrt{9\lambda^2-12}+4}{\sqrt{18\lambda^2+6\lambda\sqrt{9\lambda^2-12}}}, & \lambda > \lambda_c \\ \mu_0 := 1.657\ldots, & \lambda \leq \lambda_c \end{cases}$$

Moreover, no other estimator can attain a higher alignment.



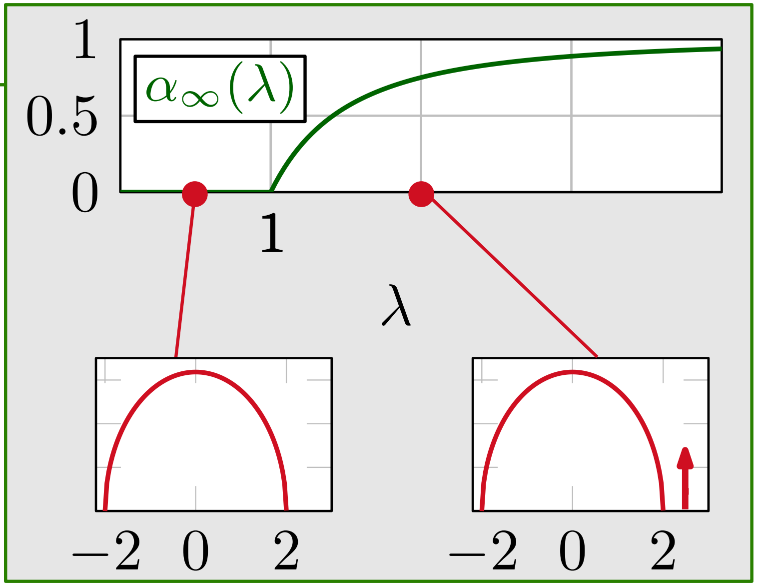
Random optimization landscape

Behavior reminiscent of "BBP phase transition" known for spiked matrix model

$$Y = \lambda \, x \, x^{\mathsf{T}} + \frac{1}{\sqrt{N}} W$$

(Benaych-Georges & Nadakuditi, 2011)

But why the discontinuity ?



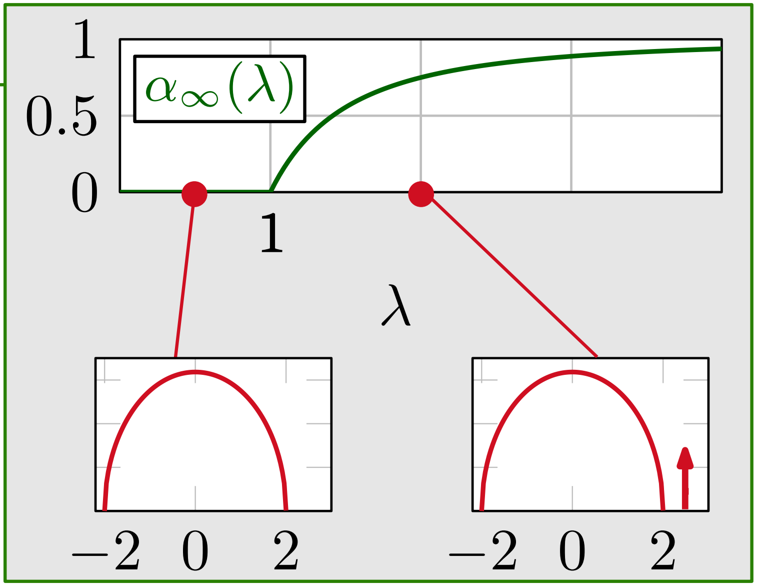
Random optimization landscape

Behavior reminiscent of "BBP phase transition" known for spiked matrix model

$$Y = \lambda\, x\, x^\top + \frac{1}{\sqrt{N}}\, W$$

(Benaych-Georges & Nadakuditi, 2011)

But why the discontinuity ?

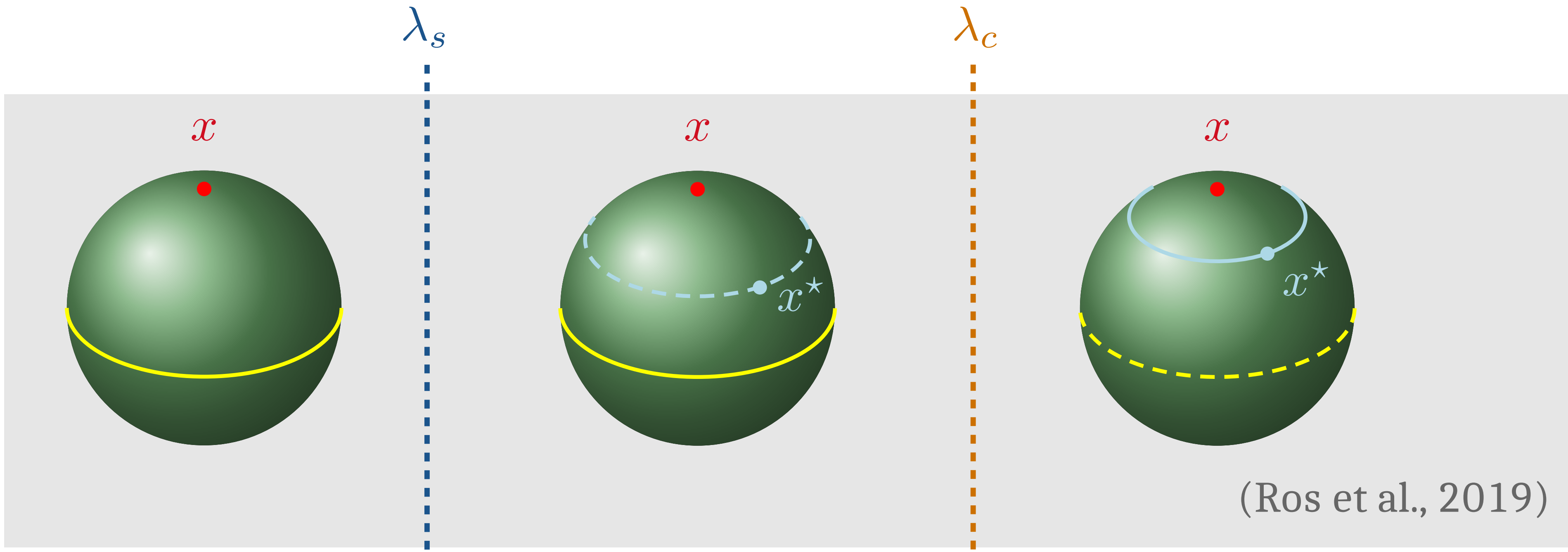


Insight found in study of the (random) ML landscape
(Ros et al., 2019) (Ben Arous et al., 2019)

- Quantification of "landscape complexity" (# of critical pts/local max)
- Connection with (spin) glasses and "rough energy landscapes"
- Configuration encoding signal competes with random ones

$$\mathcal{Y}(u,u,u) = \lambda\,\langle u, x\rangle^3 + \frac{1}{\sqrt{N}}\mathcal{W}(u,u,u)$$

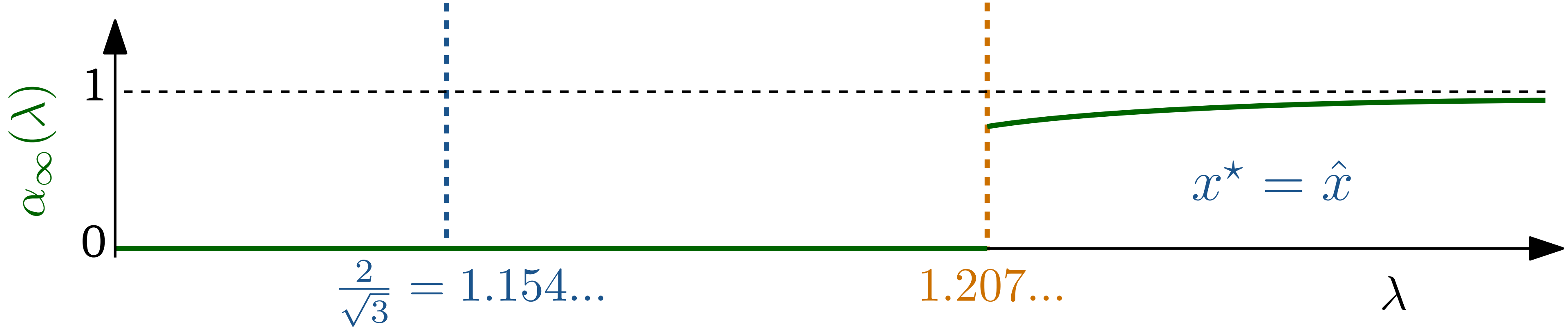
Geometric phase transitions



All max around equator

Global max on equator, an "informative" local max $x^\star$ "escapes" towards north

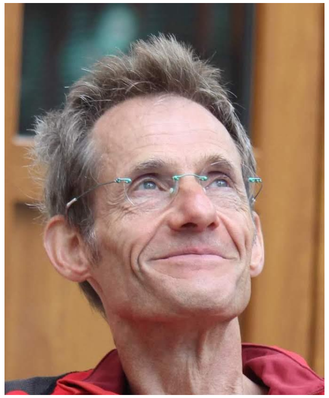
Local max on equator, $x^\star$ becomes global and approaches north



This talk

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

From here on : joint work with Romain Couillet and Pierre Comon



Tensor eigenpairs and MLE

ML problem

$$\max_{\|u\|=1} \sum_{ijk} Y_{ijk}\, u_i u_j u_k$$

Lagrangian

$$L(\mu, u) = \frac{1}{3}\,\boldsymbol{\mathcal{Y}}(u,u,u) - \frac{\mu}{2}\left(\|u\|^2 - 1\right)$$

Tensor eigenpairs and MLE

ML problem

$$\max_{\|u\|=1} \sum_{ijk} Y_{ijk}\, u_i u_j u_k$$

Lagrangian

$$L(\mu, u) = \frac{1}{3}\,\mathcal{Y}(u,u,u) - \frac{\mu}{2}\left(\|u\|^2 - 1\right)$$

Critical points satisfy

$$\frac{\partial}{\partial u} L(\mu, u) = \mathcal{Y}(u,u) - \mu\, u = 0, \qquad \text{with} \quad \Big(\mathcal{Y}(u,u)\Big)_i = \sum_{jk} Y_{ijk}\, u_j\, u_k$$

Tensor eigenpairs and MLE

ML problem

$$\max_{\|u\|=1} \sum_{ijk} Y_{ijk}\, u_i u_j u_k$$

Lagrangian

$$L(\mu, u) = \frac{1}{3}\boldsymbol{\mathcal{Y}}(u,u,u) - \frac{\mu}{2}\left(\|u\|^2 - 1\right)$$

Critical points satisfy

$$\frac{\partial}{\partial u} L(\mu, u) = \boldsymbol{\mathcal{Y}}(u,u) - \mu\, u = 0, \qquad \text{with} \qquad \Big(\boldsymbol{\mathcal{Y}}(u,u)\Big)_i = \sum_{jk} Y_{ijk}\, u_j\, u_k$$

Tensor ℓ_2-eigenvalue equations : (Lim, 2005)

$$\boldsymbol{\mathcal{Y}}(u,u) = \mu\, u, \qquad \|u\| = 1$$

In particular, MLE sol'n $\hat{x}$ = dominant eigenvec. : $\boldsymbol{\mathcal{Y}}(\hat{x},\hat{x}) = \|\boldsymbol{\mathcal{Y}}\|\, \hat{x}$

Tensor and matrix eigenpairs

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

$$(\mu, u) \text{ eigenpair of } \mathcal{Y} \quad \Leftrightarrow \quad (\mu, u) \text{ eigenpair of } \mathcal{Y}(u)$$

where $\big(\mathcal{Y}(u)\big)_{i\,j} = \sum_k Y_{ijk}\, u_k$

Tensor and matrix eigenpairs

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

$$(\mu, u) \text{ eigenpair of } \boldsymbol{\mathcal{Y}} \quad \Leftrightarrow \quad (\mu, u) \text{ eigenpair of } \boldsymbol{\mathcal{Y}}(u)$$

where $\left(\boldsymbol{\mathcal{Y}}(u)\right)_{i\,j} = \sum_k Y_{ijk}\, u_k$

Proof: $\mu\, u = \boldsymbol{\mathcal{Y}}(u, u) = \boldsymbol{\mathcal{Y}}(u)u$

Tensor and matrix eigenpairs

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

$$(\mu, u) \text{ eigenpair of } \mathcal{Y} \quad \Leftrightarrow \quad (\mu, u) \text{ eigenpair of } \mathcal{Y}(u)$$

where $\Big(\mathcal{Y}(u)\Big)_{i\,j} = \sum_k Y_{ijk}\, u_k$

Proof: $\mu\, u = \mathcal{Y}(u, u) = \mathcal{Y}(u)u$

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

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

$$\max_{\|w\|=1,\, \langle w, u\rangle = 0} \langle \mathcal{Y}(u)\, w, w \rangle \leq \frac{\mu}{2}$$

From spiked tensor model to matrix models

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

$$\mathcal{Y}(u) = \lambda \, \langle x, u \rangle \, x \, x^\mathsf{T} + \frac{1}{\sqrt{N}} \mathcal{W}(u)$$

- SNR weighted by alignment $\langle x, u \rangle$
- $\mathcal{W}$ and u are correlated $\Rightarrow$ "spike" at every local max u regardless of λ
- Special matrices from contraction ensemble $\mathcal{M}_{\mathcal{Y}} := \{\mathcal{Y}(v) \; : \; v \in \mathbb{S}^{N-1}\}$

From spiked tensor model to matrix models

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

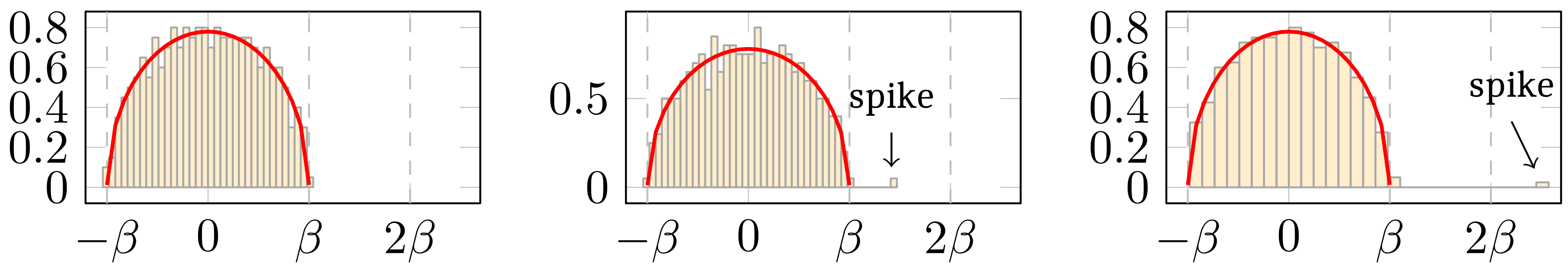
$$\boldsymbol{\mathcal{Y}}(u) = \lambda \, \langle x, u \rangle \; x \, x^\mathsf{T} + \frac{1}{\sqrt{N}} \boldsymbol{\mathcal{W}}(u)$$

- SNR weighted by alignment $\langle x, u \rangle$
- $\boldsymbol{\mathcal{W}}$ and u are correlated ⇒ "spike" at every local max u regardless of λ
- Special matrices from contraction ensemble $\mathcal{M}_{\boldsymbol{\mathcal{Y}}} := \{\boldsymbol{\mathcal{Y}}(v) \; : \; v \in \mathbb{S}^{N-1}\}$

Example : v^k produced by k iterations of power method, random init, $N = 500$

$$\tilde{v}^k = \boldsymbol{\mathcal{Y}}(v^k, v^k) + \gamma \, v^k, \qquad v^k = \tilde{v}^k / \left\| \tilde{v}^k \right\|$$

Spectrum of $\boldsymbol{\mathcal{Y}}(v^0)$ — Spectrum of $\boldsymbol{\mathcal{Y}}(v^5)$ — Spectrum of $\boldsymbol{\mathcal{Y}}(v^{20})$



This talk

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

RMT to the rescue

Problem: Given a local max (μ, u) and

spectral dec.

$$\boldsymbol{\mathcal{Y}}(u) = \lambda \langle x, u \rangle \, x \, x^\top + \frac{1}{\sqrt{N}} \boldsymbol{\mathcal{W}}(u) \quad = \sum_i \nu_i \, v_i v_i^\top$$

compute the limiting values of $\langle x, u \rangle$ and μ (if any).

RMT to the rescue

Problem: Given a local max (μ, u) and

spectral dec.

$$\mathcal{Y}(u) = \lambda \langle x, u\rangle \, x\, x^\top + \frac{1}{\sqrt{N}} \mathcal{W}(u) \qquad = \sum_i \nu_i \, v_i v_i^\top$$

compute the limiting values of $\langle x, u\rangle$ and μ (if any).

Key tool: Resolvent of $\mathcal{Y}(u)$

$$R(z) := (\mathcal{Y}(u) - zI)^{-1} \qquad = \sum_i \frac{1}{\nu_i - z} v_i v_i^\top$$

- Analytic on $\mathbb{C} \setminus \mathrm{Sp}(\mathcal{Y}(u))$
- For ν_i of multiplicity one, $\langle v_i, x\rangle^2 = -\frac{1}{2\pi i} \oint_{C_{\nu_i}} x^\top R(z)\, x \, dz$
- Encodes (random) spectral measure of $\mathcal{Y}(u)$

$$\frac{1}{N} \operatorname{tr} R(z) = \int \frac{1}{\nu - z} \rho_{\mathcal{Y}(u)}(d\nu), \qquad \rho_{\mathcal{Y}(u)} = \frac{1}{N} \sum_i \delta_{\nu_i}$$

Overall strategy and main technical tools

Target quantities

$x^\top R(z)\, x$ ⟶ $\bar{\alpha}_\infty(\lambda)$

$\langle x, u \rangle$

$\underbrace{u^\top R(z)\, u}_{= \frac{1}{\mu - z}}$ ⟶ $\bar{\mu}_\infty(\lambda)$

Overall strategy and main technical tools

$x^\top R(z)\, x \longrightarrow \lim_{N\to\infty} \mathbb{E}\left\{x^\top R(z)\, x\right\}$

$\underbrace{u^\top R(z)\, u}_{= \frac{1}{\mu - z}} \longrightarrow \lim_{N\to\infty} \mathbb{E}\left\{u^\top R(z)\, u\right\}$

Target quantities

$\bar{\alpha}_\infty(\lambda)$

$\langle x, u\rangle$

$\bar{\mu}_\infty(\lambda)$

Gaussian integration by parts

$$\mathbb{E}\left\{W_{im\ell}\, f(W_{im\ell})\right\} = \mathbb{E}\left\{f'(W_{im\ell})\right\} \sigma^2_{W_{im\ell}}$$

Overall strategy and main technical tools

Nash-Poincaré inequality
+ Markov inequality
+ Borel-Cantelli lemma

$x^\mathsf{T} R(z)\, x$ → almost surely, $N \to \infty$ → $\lim_{N\to\infty} \mathbb{E}\left\{x^\mathsf{T} R(z)\, x\right\}$

$u^\mathsf{T} R(z)\, u = \frac{1}{\mu - z}$ → almost surely, $N \to \infty$ → $\lim_{N\to\infty} \mathbb{E}\left\{u^\mathsf{T} R(z)\, u\right\}$

Gaussian integration by parts

$$\mathbb{E}\left\{W_{im\ell}\, f(W_{im\ell})\right\} = \mathbb{E}\left\{f'(W_{im\ell})\right\} \sigma^2_{W_{im\ell}}$$

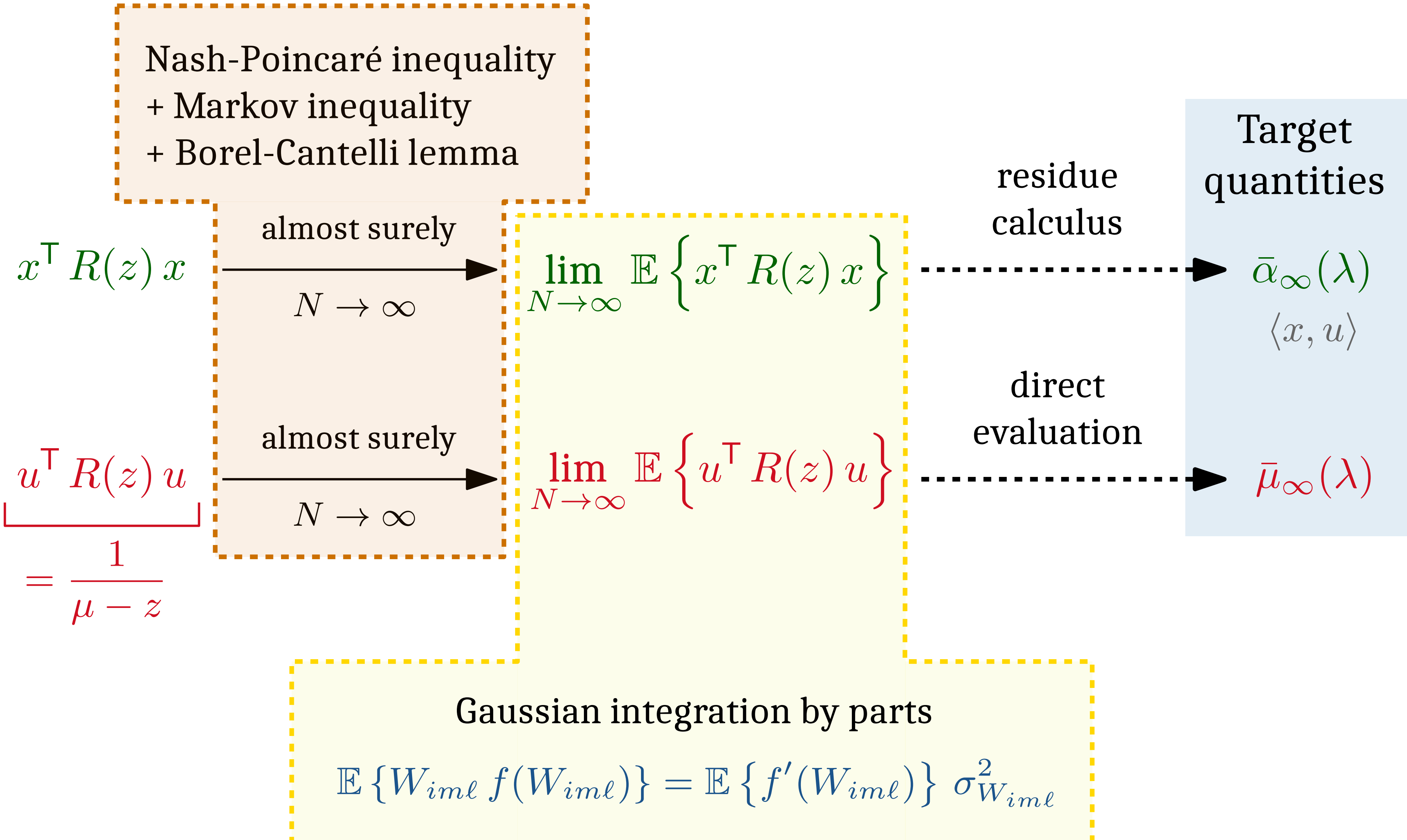
Target quantities

$\bar{\alpha}_\infty(\lambda)$

$\langle x, u\rangle$

$\bar{\mu}_\infty(\lambda)$

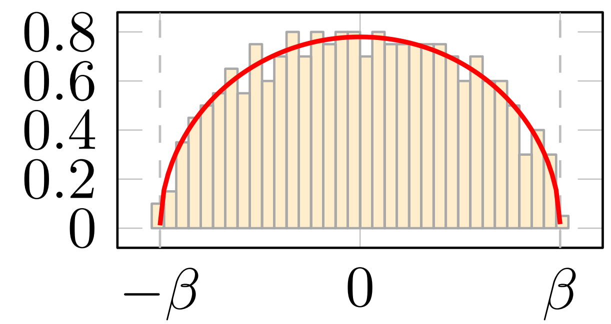
Overall strategy and main technical tools



Spectral measure of contraction ensemble $\{\mathcal{Y}(v)\}$

"Byproduct": limiting spectrum of $\mathcal{Y}(v)$, $v \in \mathbb{S}^{N-1}$

$$\rho(dx) = \frac{3}{\pi}\sqrt{\left[\frac{2}{3} - x^2\right]_+}\,dx$$



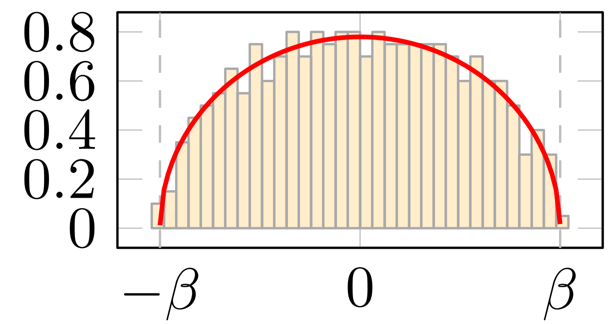
$\beta = \sqrt{2/3} = 0.8164\ldots$

Seems trivial (Gaussian model), but symmetry induces dependencies

Spectral measure of contraction ensemble $\{\mathcal{Y}(v)\}$

"Byproduct": limiting spectrum of $\mathcal{Y}(v)$, $v \in \mathbb{S}^{N-1}$

$$\rho(dx) = \frac{3}{\pi}\sqrt{\left[\frac{2}{3} - x^2\right]_+}\,dx$$

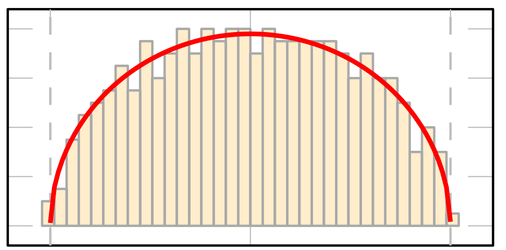


$\beta = \sqrt{2/3} = 0.8164\ldots$

Seems trivial (Gaussian model), but symmetry induces dependencies

Consequences :

- At critical points, Hessian $2\,\mathcal{Y}(u) - \mu I$ behaves as a shifted GOE (Ros et al., 2019)
- At local maxima : $\mu \geq 2\beta$ (and $\mu \leq 1.657\ldots$ for $\lambda < \lambda_c$)



Limiting fixed-point equation

Bottom line: Solution characterized by

$$\bar{\mu}_{\infty}(\lambda) = \phi(\bar{\mu}_{\infty}(\lambda), \lambda), \qquad \bar{\alpha}_{\infty}(\lambda) = \alpha(\bar{\mu}_{\infty}(\lambda), \lambda)$$

with

$$\phi(z, \lambda) = \lambda \left(\alpha(z, \lambda)\right)^3 + \frac{3}{4} z - \frac{3}{2} h(z/2),$$

$$\alpha(z, \lambda) = \frac{1}{\lambda} \frac{(h(z) + z)(h(z/2) + z/2) - 2/3}{z + h(z) - z/2 + h(z/2)}, \qquad h(z) = \sqrt{z^2 - 2/3}$$

Limiting fixed-point equation

Bottom line: Solution characterized by

$$\bar{\mu}_\infty(\lambda) = \phi(\bar{\mu}_\infty(\lambda), \lambda), \qquad \bar{\alpha}_\infty(\lambda) = \alpha(\bar{\mu}_\infty(\lambda), \lambda)$$

with

$$\phi(z, \lambda) = \lambda\,(\alpha(z, \lambda))^3 + \frac{3}{4} z - \frac{3}{2} h(z/2),$$

$$\alpha(z, \lambda) = \frac{1}{\lambda} \frac{(h(z) + z)(h(z/2) + z/2) - 2/3}{z + h(z) - z/2 + h(z/2)}, \qquad h(z) = \sqrt{z^2 - 2/3}$$

Solution: For $\lambda \geq \lambda_s = 2/\sqrt{3}$, the only positive solution for $\bar{\mu}_\infty(\lambda)$ is

$$\bar{\mu}_\infty(\lambda) = \frac{3\lambda^2 + \lambda\sqrt{9\lambda^2 - 12} + 4}{\sqrt{18\lambda^2 + 6\lambda\sqrt{9\lambda^2 - 12}}}, \qquad \bar{\alpha}_\infty(\lambda) = \sqrt{\frac{1}{2} + \sqrt{\frac{3\lambda^2 - 4}{12\lambda^2}}}$$

which precisely matches that of Jagannath et al. (2020), and thus seems to describe the "informative" local max $x^\star$ (=MLE for $\lambda > \lambda_c$)

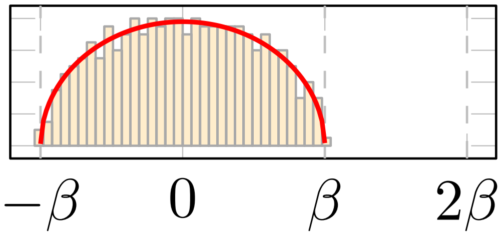
Open question : but why ?

Solution obtained under the technical conditions :

1. $\bar{\alpha}_\infty(\lambda) > 0$: otherwise no positive solution $\bar{\mu}_\infty(\lambda)$ can possibly exist

2. $\bar{\mu}_\infty(\lambda) > 2\beta$: Gaussian integration by parts requires $\frac{\partial u}{\partial W_{ijk}}$, derived from $\mathbf{\mathcal{Y}}(u,u) = \mu\, u$ and $\|u\|^2 = 1$ (by the implicit function thm) :

$$\frac{\partial u}{\partial W_{ijk}} = -\frac{1}{2\sqrt{N}}\, \underbrace{R\left(\frac{\mu}{2}\right)}_{\frac{\mu}{2} \notin \mathrm{Sp}(\mathbf{\mathcal{Y}}(u))}\, \phi + \frac{1}{\mu}\, \frac{\partial \mu}{\partial W_{im\ell}}\, u$$

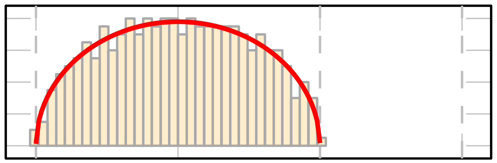


Open question : but why ?

Solution obtained under the technical conditions :

1. $\bar{\alpha}_\infty(\lambda) > 0$: otherwise no positive solution $\bar{\mu}_\infty(\lambda)$ can possibly exist
2. $\bar{\mu}_\infty(\lambda) > 2\beta$: Gaussian integration by parts requires $\frac{\partial u}{\partial W_{ijk}}$, derived from $\mathbf{Y}(u,u) = \mu\, u$ and $\|u\|^2 = 1$ (by the implicit function thm) :

$$\frac{\partial u}{\partial W_{ijk}} = -\frac{1}{2\sqrt{N}} \underbrace{R\left(\frac{\mu}{2}\right)}_{\frac{\mu}{2} \notin \mathrm{Sp}(\mathbf{Y}(u))} \phi + \frac{1}{\mu} \frac{\partial \mu}{\partial W_{im\ell}} u$$



... which do not rule out all other local maxima (Ben Arous et al., 2019)

Possible explanation : $x^\star$ the only "polarized" max, all others being purely due to fluctuations $\Rightarrow$ only $\langle x, x^\star \rangle$ converges

This talk

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

Summary

Rank-one symmetric tensor model : simple but quite rich

$$Y_{ijk} = \lambda\, x_i x_j x_k + \frac{1}{\sqrt{N}}\, W_{ijk}$$

Statistical thresholds, MLE landscape and performance now well understood, largely thanks to statistical physics tools.

Standard RMT tools can be leveraged by studying contractions and

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

Possible extensions

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

$$\mathcal{Y} = \lambda\, x \otimes y \otimes z + \frac{1}{\sqrt{N_1 + N_2 + N_3}}\, \mathcal{W}$$

with $W_{ijk} \sim \mathcal{N}(0,1)$ via study of

$$\begin{pmatrix} 0 & \mathcal{Y}(\cdot,\cdot,w) & \mathcal{Y}(\cdot,v,\cdot) \\ \mathcal{Y}(\cdot,\cdot,w)^{\mathsf{T}} & 0 & \mathcal{Y}(u,\cdot,\cdot) \\ \mathcal{Y}(\cdot,v,\cdot)^{\mathsf{T}} & \mathcal{Y}(u,\cdot,\cdot)^{\mathsf{T}} & 0 \end{pmatrix}$$

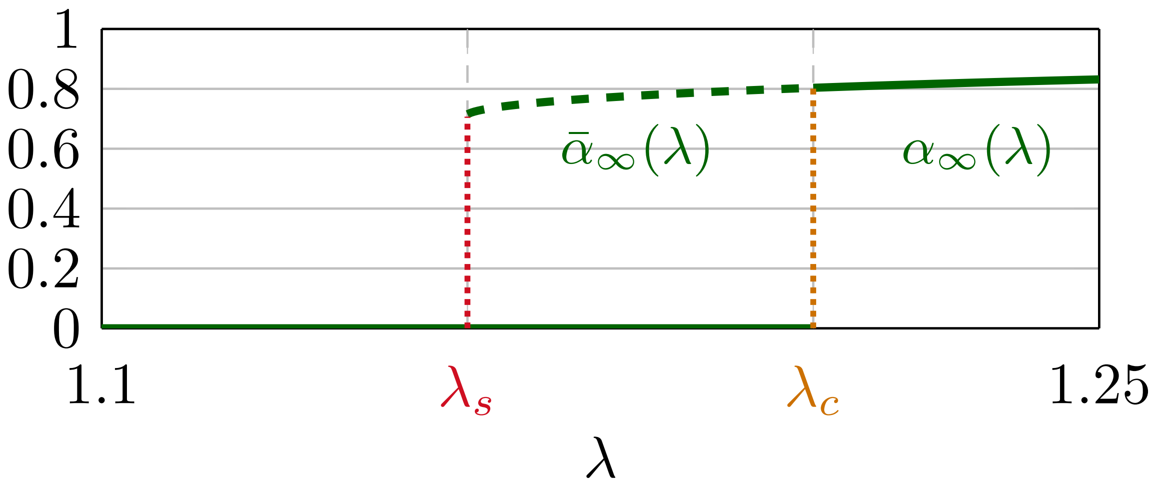
at a singular triplet (u, v, w) (critical point of ML problem)

- Higher orders d : work in progress
- Orthogonal rank-R model : boils down to R “local” rank-one models
- (Non-orthogonal) rank-R case is hard

Open questions

- Why does the obtained fixed-point equation describe only $x^\star$?

- Can we "see" the phase transition (critical value λ_c) with an RMT approach ?



For more info : arXiv:2108.00774

henrique.goulart @ irit.fr