Statistical limits of multi-spiked random tensor models

Yang Qi

joint with Alexis Decurninge (Huawei Paris Research Center)

LoRAINNe'24: workshop on LOw-Rank Approximations and their Interactions with Neural NEtworks

November 26, 2024

Multi-spiked model

Given observed tensor data $\mathcal{T}$, assume

$$\mathcal{T}_{i_1 \dots i_d} = \sum_{j=1}^{r} \beta_j u_{i_1}^{(j)} \cdots u_{i_d}^{(j)} + \frac{1}{\sqrt{N}} \mathcal{X}_{i_1 \dots i_d}$$

$\mathcal{X}$: Gaussian noise

Multi-spiked model

Given observed tensor data $\mathcal{T}$, assume

$$\mathcal{T}_{i_1 \ldots i_d} = \sum_{j=1}^{r} \beta_j u_{i_1}^{(j)} \cdots u_{i_d}^{(j)} + \frac{1}{\sqrt{N}} \mathcal{X}_{i_1 \ldots i_d}$$

$\mathcal{X}$: Gaussian noise

In tensor form:

$$\mathcal{T} = \sum_{j=1}^{r} \beta_j \boldsymbol{u}_j^{\otimes d} + \frac{1}{\sqrt{N}} \mathcal{X}$$

$\boldsymbol{v}_j$: unit vectors in $\mathbb{R}^N$ (N large)
β_j: SNR

Applications

- Latent variable model learning (Anandkumar et al., . . .)
- Video processing
- Collaborative filtering in presence of temporal/context information
- Community detection (Anandkumar et al.)
- Hypergraph matching (Duchenne et al.)
- Statistical mechanics (Crisanti & Sommers)
- Identifying structural properties and information density in neural networks (Martin & Mahoney, Martin et al.)
- Locating feature learning (Thamm et al., Levi & Oz, Staats et al.)
- Low-rank transformer features (Yu & Wu)
- Locating information in LLM (Staats et al.)
- . . .

Matrix case

Principal component analysis (Johnstone & Lu):

$$\mathcal{M} = \sum_{j=1}^{r} \beta_j \boldsymbol{u}_j^{\otimes 2} + \frac{1}{\sqrt{N}} \mathcal{W}$$

Matrix case

Principal component analysis (Johnstone & Lu):

$$\mathcal{M} = \sum_{j=1}^{r} \beta_j \boldsymbol{u}_j^{\otimes 2} + \frac{1}{\sqrt{N}} \mathcal{W}$$

In particular, the rank-one case

$$\mathcal{M} = \beta \boldsymbol{u}^{\otimes 2} + \frac{1}{\sqrt{N}} \mathcal{W}$$

a.k.a. *rank-one deformation of random matrix*

Semicircle law

For symmetric $\frac{1}{\sqrt{N}}\mathcal{W}$, where $\mathcal{W}_{ij} \sim \mathcal{N}(0,1)$ and $\mathcal{W}_{ii} \sim \mathcal{N}(0,2)$, the empirical spectral measure

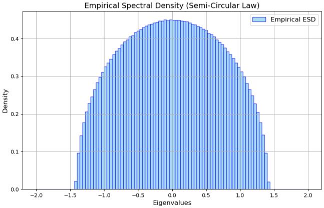
$$\frac{\#\{\text{eigenvalues in } [a,b]\}}{N} \xrightarrow{N} \int_a^b d\mu$$

Semicircle law

For symmetric $\frac{1}{\sqrt{N}}\mathcal{W}$, where $\mathcal{W}_{ij} \sim \mathcal{N}(0,1)$ and $\mathcal{W}_{ii} \sim \mathcal{N}(0,2)$, the empirical spectral measure

$$\frac{\#\{\text{eigenvalues in } [a,b]\}}{N} \xrightarrow{N} \int_a^b d\mu$$

Semi-circular law:



$$d\mu(x) = \frac{1}{2\pi}\sqrt{4 - x^2}dx$$

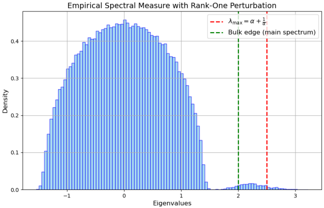
Phase transition phenomenon

For

$$\mathcal{M} = \beta \boldsymbol{u}^{\otimes 2} + \frac{1}{\sqrt{N}} \mathcal{W},$$

the largest eigenvalue satisfies

$$\lim_{N \to \infty} \lambda_1(\mathcal{M}) = \begin{cases} 2 & \text{if } \beta \leqslant 1 \\ \beta + \dfrac{1}{\beta} & \text{if } \beta > 1 \end{cases}$$



Single spiked tensor model

Tensor PCA (Montanari & Richard):

$$\mathcal{T} = \beta \boldsymbol{u}^{\otimes d} + \frac{1}{\sqrt{N}} \mathcal{X},$$

where

$$\mathcal{X}_{i_1 \dots i_d} = \frac{1}{d!} \sum_{\pi \in \mathfrak{S}_d} \mathcal{W}_{i_{\pi(1)} \dots i_{\pi(d)}},$$

$\mathcal{W}$: Gaussian noise with i.i.d. entries $\mathcal{W}_{i_1 \dots i_d} \sim \mathcal{N}(0, 1)$

$\mathfrak{S}_d$: symmetric group on the set $[d]$

Single spiked tensor model

Tensor PCA (Montanari & Richard):

$$\mathcal{T} = \beta \boldsymbol{u}^{\otimes d} + \frac{1}{\sqrt{N}} \mathcal{X},$$

where

$$\mathcal{X}_{i_1 \dots i_d} = \frac{1}{d!} \sum_{\pi \in \mathfrak{S}_d} \mathcal{W}_{i_{\pi(1)} \dots i_{\pi(d)}},$$

- $\mathcal{W}$: Gaussian noise with i.i.d. entries $\mathcal{W}_{i_1 \dots i_d} \sim \mathcal{N}(0,1)$
- $\mathfrak{S}_d$: symmetric group on the set $[d]$
- $\mathbb{E}(\mathcal{X}_{i_1 \dots i_d}) = 0$
- variance $\sigma^2_{i_1 \dots i_d}(\mathcal{X})$:

$$\{i_1, \dots, i_d\} = \{\ell_1, \dots, \ell_1, \dots, \ell_k, \dots, \ell_k\},$$

where $\ell_1, \dots, \ell_k$ are distanct, and $\#\ell_i = m_i$

$$\sigma^2_{i_1 \dots i_d} = \frac{1}{\binom{d}{m_1, \dots, m_k}}$$

Information-theoretic threshold

Given two random tensors $\mathcal{T}_1$ and $\mathcal{T}_2$, recall *total variation distance*:

$$d_{\mathrm{TV}}(\mathcal{T}_1, \mathcal{T}_2) = \sup_A |\mathbb{P}(\mathcal{T}_1 \in A) - \mathbb{P}(\mathcal{T}_2 \in A)|$$

For two sequences $\mathcal{T}_N$ and $\frac{1}{\sqrt{N}}\mathcal{X}_N$,

distinguishable if

$$\lim_{N\to\infty} d_{\mathrm{TV}}(\mathcal{T}_N, \frac{1}{\sqrt{N}}\mathcal{X}_N) = 1$$

indistinguishable if

$$\lim_{N\to\infty} d_{\mathrm{TV}}(\mathcal{T}_N, \frac{1}{\sqrt{N}}\mathcal{X}_N) = 0$$

Statistical threshold

Pick a statistical procedure to estimate $\boldsymbol{u}$

Maximum-likelihood estimator:

$$\boldsymbol{v}^* \in \arg\sup_{\|\boldsymbol{v}\|=1} \langle \mathcal{T}, \boldsymbol{v}^{\otimes d} \rangle$$

Theorem (Jagannath et al., Chen)

$$\beta_{\text{IT}} = \beta_{\text{stat,MLE}} = \sup \left\{ \beta \geqslant 0 \mid \sup_{t \in [0,1]} \left[\beta t^d + \log(1-t) + t \right] \leqslant 0 \right\}$$

Algorithmic threshold

Pick a polynomial-time algorithm to optimize

$$\beta_{\text{algo}} = \begin{cases} N^{\frac{d-2}{2}} & \text{(gradient descent, SGD)} \\ N^{\frac{d-2}{4}} & \text{(sum of squares, tensor unfolding)} \end{cases}$$

big gap

$$\frac{\beta_{\text{algo}}}{\beta_{\text{stat}}} \sim N^{\frac{d-2}{4}}$$

Multi-spiked model

$$\mathcal{T} = \sum_{j=1}^{r} \beta_j \boldsymbol{u}_j^{\otimes d} + \frac{1}{\sqrt{N}} \mathcal{X}$$

Very little is known.

- Information-theoretic threshold:
  - assuming $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ are sampled independently (Lesieur et al., Chen et al.)
- Algorithmic threshold:
  - power iteration assuming $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ orthogoanl (Huang et al.)
  - online SGD, gradient flow assuming $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ orthogoanl (Ben Arous et al.)

Local methods for MLE

Maximum-likelihood estimation:

$$\inf_{\substack{\gamma_1,\ldots,\gamma_r\in\mathbb{R}\\ \boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\in\mathbb{R}^N}} \quad \|\mathcal{T} - \sum_{j=1}^{r} \gamma_j \boldsymbol{v}_j^{\otimes d}\|^2$$

$$\text{subject to} \quad \|\boldsymbol{v}_1\| = \cdots = \|\boldsymbol{v}_r\| = 1$$

Existence and uniqueness of approximation:

appearance of noise $\mathcal{X}$

N sufficiently large

Local methods:

gradient descent (steepest, conjugate, ...)

stochastic gradient descent

Newton (quasi-Newton)

...

Rank-one case

Local methods find critical points instead of global optima. So study the phase transition phenomenon for critical points of MLE.

Tensor eigenvalue equation:

$$\langle \mathcal{T}, \boldsymbol{v}^{\otimes(d-1)} \rangle = \gamma \boldsymbol{v}, \quad \langle \boldsymbol{v}, \boldsymbol{v} \rangle = 1$$

Theorem (Goulart et al., Seddik et al.)

Assume $\gamma \to \gamma_*$, when $\gamma_* > \sqrt{\frac{d-1}{d}}$, detection of critical points is possible.

Critical points

Maximum-likelihood estimation:

$$\begin{aligned} \inf_{\substack{\gamma_1,\ldots,\gamma_r\in\mathbb{R} \\ \boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\in\mathbb{R}^N}} \quad & \|\mathcal{T} - \sum_{j=1}^{r} \gamma_j \boldsymbol{v}_j^{\otimes d}\|^2 \\ \text{subject to} \quad & \|\boldsymbol{v}_1\| = \cdots = \|\boldsymbol{v}_r\| = 1 \end{aligned}$$

Critical points

Maximum-likelihood estimation:

$$\inf_{\substack{\gamma_1,\ldots,\gamma_r\in\mathbb{R}\\ \boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\in\mathbb{R}^N}} \quad \|\mathcal{T}-\sum_{j=1}^{r}\gamma_j\boldsymbol{v}_j^{\otimes d}\|^2$$

$$\text{subject to} \quad \|\boldsymbol{v}_1\|=\cdots=\|\boldsymbol{v}_r\|=1$$

KKT conditions:

$$\begin{cases}\langle\mathcal{T}-\sum_{j=1}^{r}\gamma_j\boldsymbol{v}_j^{\otimes d},\boldsymbol{v}_i^{\otimes(d-1)}\rangle=0\\ \langle\boldsymbol{v}_i,\boldsymbol{v}_i\rangle=1\end{cases}$$

Higher rank case

Assume $\beta_1 \geqslant |\beta_2| > 0$. Let

$$\boldsymbol{A} = \langle \mathcal{T}, \boldsymbol{v}_1^{\otimes(d-2)} \rangle, \quad \boldsymbol{B} = \langle \mathcal{T}, \boldsymbol{v}_2^{\otimes(d-2)} \rangle, \quad \lambda = \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle^{d-1}, \quad \nu = \frac{\gamma_2}{\gamma_1}.$$

Equations of critical points:

$$\flat(\mathcal{T}) := \frac{1}{1-\lambda^2} \begin{bmatrix} \nu \boldsymbol{A} & -\lambda\nu \boldsymbol{B} \\ -\lambda \boldsymbol{A} & \boldsymbol{B} \end{bmatrix} = \gamma_2 \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{bmatrix}$$

Tools from random matrix theory

Resolvent of $\flat(\mathcal{T})$:

$$\boldsymbol{Q}(z) = (\flat(\mathcal{T}) - z\mathbf{Id})^{-1} = \begin{bmatrix} \boldsymbol{Q}^{11}(z) & \boldsymbol{Q}^{12}(z) \\ \boldsymbol{Q}^{21}(z) & \boldsymbol{Q}^{22}(z) \end{bmatrix}.$$

The eigenvalues of $\flat(\mathcal{T})$ are real, say $\lambda_1 \geqslant \cdots \geqslant \lambda_{2N}$. Define the *empirical spectral measure* of $\flat(\mathcal{T})$ by

$$\mu = \frac{1}{2N} \sum_i \delta_{\lambda_i},$$

and *Cauchy-Stieltjes transform* S_μ by

$$S_\mu(z) = \int_{\mathbb{R}} \frac{1}{t-z}\, d\mu(t) \quad \text{for } z \in \mathbb{C} \setminus \mathbb{R}.$$

Then the Stieltjes transform of μ satisfies

$$S_\mu(z) = \frac{1}{2N} \operatorname{Tr}(\boldsymbol{Q}(z)) \ .$$

Technical assumptions

Given a sequence of tensors $(\mathcal{T})_*$ such that each member $\mathcal{T}$ satisfies

$$\mathcal{T} = \sum_{j=1}^{2} \beta_j \boldsymbol{u}_j^{\otimes d} + \frac{1}{\sqrt{N}} \mathcal{X},$$

assume that there is a sequence $\left(\gamma_1, \gamma_2, \boldsymbol{v}_1, \boldsymbol{v}_2\right)_*$ satisfying the following equations

$$\begin{cases} \langle \mathcal{T} - \sum_{j=1}^{2} \gamma_j \boldsymbol{v}_j^{\otimes d}, \boldsymbol{v}_1^{\otimes(d-1)} \rangle = \langle \mathcal{T} - \sum_{j=1}^{2} \gamma_j \boldsymbol{v}_j^{\otimes d}, \boldsymbol{v}_2^{\otimes(d-1)} \rangle = 0 \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle = \langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle = 1 \end{cases}, \qquad (1)$$

such that $(\gamma_i)_* \xrightarrow{\text{a.s.}} \gamma_i^\infty, (\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle)_* \xrightarrow{\text{a.s.}} \alpha_{ij}, (\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle)_* \xrightarrow{\text{a.s.}} \tau$.

Rank-two case

Theorem (Q.-Decurninge)

Under Assumption, the limit

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}[\operatorname{Tr} \boldsymbol{Q}^{ij}(z)] \to g_{ij}(z) \tag{2}$$

The empirical spectral measure of $\flat(\mathcal{T})$ *converges weakly almost surely to a deterministic measure* μ *whose Stieltjes transform is*

$$g(z) = g_{11}(z) + g_{22}(z),$$

which is a complex analytic function on $\mathbb{C} \setminus \operatorname{Supp}(\mu)$.

Rank-two case continued

Theorem (continued)

Let

$$\begin{cases} Z_1 = \nu_\infty g_{11}(z) - \lambda_\infty g_{12}(z), & Z_2 = -\lambda_\infty \nu_\infty g_{11}(z) + g_{12}(z), \\ Y_1 = \nu_\infty g_{21}(z) - \lambda_\infty g_{22}(z), & Y_2 = -\lambda_\infty \nu_\infty g_{21}(z) + g_{22}(z), \end{cases}$$

where $\lambda_\infty = \tau^{d-1}$ *and* $\nu_\infty = \frac{\gamma_2^\infty}{\gamma_1^\infty}$. *Let* $\boldsymbol{H} = \begin{bmatrix} Z_1 & Y_1 \\ Z_2 & Y_2 \end{bmatrix}$.

Then there is a unique solution to the following equation

$$\boldsymbol{H}^2 + d(d-1)(1-\lambda_\infty^2) z \begin{bmatrix} \frac{1}{\nu_\infty} & \frac{\lambda_\infty}{\nu_\infty} \\ \lambda_\infty & 1 \end{bmatrix} \boldsymbol{H} + d(d-1)(1-\lambda_\infty^2)^2 \mathbf{Id} = 0,$$

such that $\Im[g(z)] > 0$ *for all* z *satisfying* $\Im[z] > 0$.

Rank-two case continued

Theorem (continued)

When $\lambda_\infty \neq 0$, μ has the form

$$\mu(dx) = \frac{d(d-1)}{4\nu_\infty(\kappa_1-\kappa_2)\pi}\left[\kappa_1\xi_1\sqrt{\left(\frac{4}{d(d-1)\kappa_1^2}-x^2\right)_+} + \kappa_2\xi_2\sqrt{\left(\frac{4}{d(d-1)\kappa_2^2}-x^2\right)_+}\right] dx,$$

which is supported on $[-\beta_d^0, \beta_d^0]$, where

$$\begin{cases} \kappa_1 = & \frac{1}{2}+\frac{1}{2\nu_\infty}+\frac{1}{2}\sqrt{\left[1-\frac{1-2\lambda_\infty^2}{\nu_\infty}\right]^2+\frac{1-(1-2\lambda_\infty^2)^2}{\nu_\infty^2}} \\ \kappa_2 = & \frac{1}{2}+\frac{1}{2\nu_\infty}-\frac{1}{2}\sqrt{\left[1-\frac{1-2\lambda_\infty^2}{\nu_\infty}\right]^2+\frac{1-(1-2\lambda_\infty^2)^2}{\nu_\infty^2}} \\ \xi_1 = & (\kappa_1-1)+\lambda_\infty^2-(\kappa_1-1)(\kappa_2-1)\nu_\infty-(\kappa_2-1)\nu_\infty \\ \xi_2 = & -(\kappa_2-1)-\lambda_\infty^2+(\kappa_1-1)(\kappa_2-1)\nu_\infty+(\kappa_1-1)\nu_\infty \end{cases}, \quad \beta_d^0 = \begin{cases} \frac{2}{\sqrt{d(d-1)}\kappa_1} & \text{if } 0<\nu_\infty\leqslant 1, \\ -\frac{2}{\sqrt{d(d-1)}\kappa_2} & \text{if } -1\leqslant\nu_\infty<0. \end{cases}$$

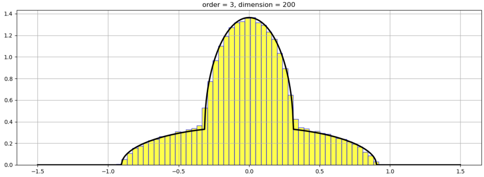
Rank-two case continued

Theorem (continued)

When $\lambda_\infty = 0$, μ has the form

$$\mu(dx) = \frac{d(d-1)}{4\pi}\left[\frac{1}{\nu_\infty}\sqrt{\left(\frac{4\nu_\infty^2}{d(d-1)} - x^2\right)_+} + \sqrt{\left(\frac{4}{d(d-1)} - x^2\right)_+}\right] dx.$$

Figure: order $= 3$, dimension $= 200$



Limiting alignments

Would like to find $\alpha_{ij} := \langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle$

Let

$$z = \frac{\gamma_2^\infty}{d-1}, \quad \boldsymbol{N} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}, \quad \boldsymbol{L} = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{K} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Denote by $\boldsymbol{N}_d = \boldsymbol{N} \odot \cdots \odot \boldsymbol{N}$ the dth iterated Hadamard product of $\boldsymbol{N}$, i.e., the (i,j) entry of $\boldsymbol{N}_d$ is α_{ij}^d.

Limiting alignments continued

Theorem (Q.-Decurninge)

Assumption $(\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle)_* \xrightarrow{a.s.} \rho$, *when*

$$\gamma_1^\infty \geqslant |\gamma_2^\infty| > (d-1)\beta_d^0,$$

the limiting alignments α_{ij} *and the estimators* $\gamma_1^\infty, \gamma_2^\infty, \tau$ *satisfy the following equations.*

$$\boldsymbol{H}^2 + d(d-1)(1-\lambda_\infty^2)z \begin{bmatrix} \frac{1}{\nu_\infty} & \frac{\lambda_\infty}{\nu_\infty} \\ \lambda_\infty & 1 \end{bmatrix} \boldsymbol{H} + d(d-1)(1-\lambda_\infty^2)^2 \mathbf{Id} = 0$$

$$\boldsymbol{N}_{d-1}^\top \boldsymbol{L}\boldsymbol{K} = \begin{bmatrix} \gamma_1^\infty & \gamma_2^\infty \lambda_\infty \\ \gamma_1^\infty \lambda_\infty & \gamma_2^\infty \end{bmatrix} \boldsymbol{N}^\top + \frac{1}{d(1-\lambda_\infty^2)} \begin{bmatrix} Z_1 & \tau^{d-2} Z_2 \\ \tau^{d-2} Y_2 & Y_2 \end{bmatrix} \boldsymbol{N}^\top$$

$$\boldsymbol{N}_{d-1}^\top \boldsymbol{L}\boldsymbol{N} = \begin{bmatrix} \gamma_1^\infty & \gamma_2^\infty \lambda_\infty \\ \gamma_1^\infty \lambda_\infty & \gamma_2^\infty \end{bmatrix} \begin{bmatrix} 1 & \tau \\ \tau & 1 \end{bmatrix} + \frac{1}{d(1-\lambda_\infty^2)} \begin{bmatrix} Z_1 & \tau^{d-2} Z_2 \\ \tau^{d-2} Y_1 & Y_2 \end{bmatrix} \begin{bmatrix} 1 & \tau \\ \tau & 1 \end{bmatrix} + \frac{1}{d(d-1)(1-\lambda_\infty^2)} \begin{bmatrix} 1 & \lambda_\infty \\ \lambda_\infty & 1 \end{bmatrix} \boldsymbol{H}$$

Phase transition phenomenon

Intuitively, we would like to distinguish

$$\frac{1}{\sqrt{N}}\flat(\mathcal{X}) \quad \text{and} \quad \flat(\mathcal{T}) = \sum_{i=1}^{2} \boldsymbol{U}_i \boldsymbol{A}_i \boldsymbol{U}_i^\top + \frac{1}{\sqrt{N}}\flat(\mathcal{X}) \ ,$$

where

$$\boldsymbol{U}_i = \begin{bmatrix} \boldsymbol{u}_i & 0 \\ 0 & \boldsymbol{u}_i \end{bmatrix} \text{ and } \boldsymbol{A}_i = \frac{\beta_i}{1-\lambda^2} \begin{bmatrix} \nu\langle \boldsymbol{u}_i, \boldsymbol{v}_1\rangle^{d-2} & -\lambda\nu\langle \boldsymbol{u}_i, \boldsymbol{v}_2\rangle^{d-2} \\ -\lambda\langle \boldsymbol{u}_i, \boldsymbol{v}_1\rangle^{d-2} & \langle \boldsymbol{u}_i, \boldsymbol{v}_2\rangle^{d-2} \end{bmatrix} \ ,$$

by comparing the largest eigenvalues.

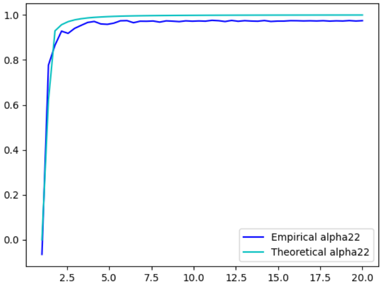
Theorem (Q.-Decurninge)

There exists some $\beta_{\text{cri}}(\rho) > 0$ depending on ρ, when

$$\beta_1 \geqslant |\beta_2| > \beta_{\text{cri}}(\rho),$$

the limiting alignments α_{ij} and the estimators $\gamma_1^\infty, \gamma_2^\infty, \tau$ satisfy the above equations, i.e., the detection of two critical points is possible.

Numerical examples



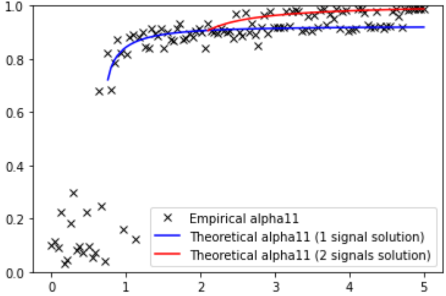


Figure: $\langle \boldsymbol{u}_2, \boldsymbol{v}_2 \rangle$ on the left when $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are not correlated, $\langle \boldsymbol{u}_1, \boldsymbol{v}_1 \rangle$ on the right when $\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle \approx 1$.

Exact threshold

(a) By elimination theory, γ_2^∞ is a solution of some

$$\mathcal{G}(X, \rho, \beta_1, \beta_2) = 0,$$

where X is the variable, and ρ, β_1, β_2 are parameters.

(b) Let $\beta_1 = \beta_2$ if $\beta_2 > 0$ or let $\beta_1 = -\beta_2$ if $\beta_2 < 0$.

(c) Express β_2 by a function of X where ρ is a parameter, say

$$\beta_2 = \widetilde{\mathcal{G}}_\rho(X).$$

(d) Take $\beta_{\mathrm{cri}}(\rho)$ as

$$\beta_{\mathrm{cri}}(\rho) = \lim_{X \to \beta_d^0} \widetilde{\mathcal{G}}_\rho(X)$$

Better estimator

From α_{11}, α_{12}, $\boldsymbol{v}_1$, and $\boldsymbol{v}_2$, we obtain $\boldsymbol{u}_1^*$.

$$\mathcal{H}(\gamma_1, \gamma_2, \boldsymbol{v}_1, \boldsymbol{v}_2) := \|\mathcal{T} - \gamma_1 \boldsymbol{v}_1^{\otimes d} - \gamma_2 \boldsymbol{v}_2^{\otimes d}\| - \frac{1}{N}\|\mathcal{X}\|^2$$

Note that

$$\lim_{N\to\infty} \mathcal{H}(\gamma_1, \gamma_2, \boldsymbol{v}_1, \boldsymbol{v}_2) = \beta_1^2 + \beta_2^2 + 2\beta_1\beta_2\rho^d + (\gamma_1^\infty)^2 + (\gamma_2^\infty)^2 + 2\gamma_1^\infty\gamma_2^\infty\tau^d$$

Pick the algebraic solution such that

$$(\gamma_1^\infty, \gamma_2^\infty, \tau) \in \underset{\gamma_1^\infty, \gamma_2^\infty, \tau}{\arg\inf} \left[\lim_{N\to\infty} \mathcal{H}(\gamma_1^\infty, \gamma_2^\infty, \boldsymbol{v}_1, \boldsymbol{v}_2)\right]$$

Plug in β_1^* and β_2^* get $\beta_1^*\boldsymbol{u}_1^*$ and $\beta_2^*\boldsymbol{u}_2^*$

Conclusion

- Phase transition phenomenon of detecting critical points
- Limiting spectral measure
- Limiting alignments
- Use these alignments to correct MLE

Thank you for your attention!