# Landscape complexity of the finite-rank spiked tensor model

Journées MAS 2022

VANESSA PICCOLO

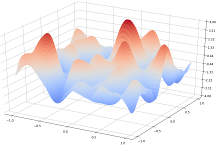
August 29, 2022

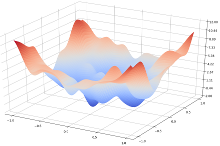
ENS de Lyon

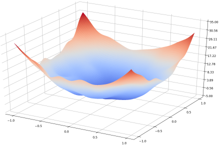
# Overview

# What is the talk about?

## Landscape complexity:

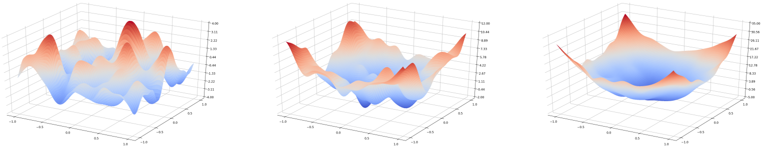






# What is the talk about?

## Landscape complexity:



∗ Goal: For random functions $f_N : \mathbf{R}^N \to \mathbf{R}$, compute the annealed complexity:

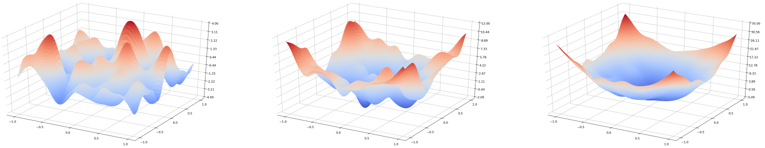
$$\Sigma = \lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}\,\mathrm{Crt}(f_N) = \lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}\left[\#\{\boldsymbol{x} : \nabla f_N(\boldsymbol{x}) = 0\}\right],$$

i.e.,

$$\mathbf{E}\,\mathrm{Crt}(f_N) \approx e^{N\Sigma} \quad \text{for large } N.$$

## What is the talk about?

Landscape complexity:



- Goal: For random functions $f_N : \mathbf{R}^N \to \mathbf{R}$, compute the annealed complexity:

$$\Sigma = \lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}\,\mathrm{Crt}(f_N) = \lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}\left[\#\{\boldsymbol{x} : \nabla f_N(\boldsymbol{x}) = 0\}\right],$$

i.e.,

$$\mathbf{E}\,\mathrm{Crt}(f_N) \approx e^{N\Sigma} \quad \text{for large } N.$$

- We consider a model of the type

$$f_N(x) = \mathrm{signal}_N(x) + \mathrm{noise}_N(x),$$

and we identify a phase transition.

# What is this talk about?

Spiked tensor model:

**Definition**

The spiked tensor model is defined as

$$\boldsymbol{Y} = \lambda \boldsymbol{u}^{\otimes p} + \frac{1}{\sqrt{N}} \boldsymbol{J},$$

where

- $\boldsymbol{Y} \in (\mathbb{R}^N)^{\otimes p}$ is the $p$-th order tensor observation
- $\boldsymbol{J} \in (\mathbb{R}^N)^{\otimes p}$ is a $p$-th order noise tensor whose entries are $\boldsymbol{J}_{i_1 i_2 \ldots i_p} \overset{iid}{\sim} \mathcal{N}(0, 1)$
- $\lambda > 0$ is the signal-to-noise ratio
- $\boldsymbol{u} \in \mathbb{S}^{N-1}$ is an unknown signal vector to be recovered

Introduced by Richard, Montanari 2014.

# What is this talk about?

Recall the rank-1 spiked tensor model: $\boldsymbol{Y} = \lambda \boldsymbol{u}^{\otimes p} + \frac{1}{\sqrt{N}} \boldsymbol{J}$.

**Tensor PCA**

The maximum likelihood estimation requires solving

$$\begin{aligned} \text{maximize} \quad & f_p(\boldsymbol{\sigma}) = \langle \boldsymbol{Y}, \boldsymbol{\sigma}^{\otimes p} \rangle \\ \text{subject to} \quad & \boldsymbol{\sigma} \in \mathbf{S}^{N-1}. \end{aligned} \tag{1}$$

# What is this talk about?

Recall the rank-1 spiked tensor model: $\boldsymbol{Y} = \lambda \boldsymbol{u}^{\otimes p} + \frac{1}{\sqrt{N}} \boldsymbol{J}$.

**Tensor PCA**

The maximum likelihood estimation requires solving

$$\begin{aligned} \text{maximize} \quad & f_p(\boldsymbol{\sigma}) = \langle \boldsymbol{Y}, \boldsymbol{\sigma}^{\otimes p} \rangle \\ \text{subject to} \quad & \boldsymbol{\sigma} \in \mathbf{S}^{N-1}. \end{aligned} \tag{1}$$

The landscape of the optimization problem (1) is the function $f_p : \mathbf{S}^{N-1} \to \mathbf{R}$:

$$f_p(\boldsymbol{\sigma}) = \langle \boldsymbol{Y}, \boldsymbol{\sigma}^{\otimes p} \rangle = \lambda \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle^p + \frac{1}{\sqrt{N}} \sum_{i_1, \ldots, i_p = 1}^{N} \boldsymbol{J}_{i_1 \ldots i_p} \boldsymbol{\sigma}_{i_1} \cdots \boldsymbol{\sigma}_{i_p}.$$

# What is this talk about?

Recall the rank-1 spiked tensor model: $\boldsymbol{Y} = \lambda \boldsymbol{u}^{\otimes p} + \frac{1}{\sqrt{N}} \boldsymbol{J}$.

**Tensor PCA**

The maximum likelihood estimation requires solving

$$\begin{aligned} \text{maximize} \quad & f_p(\boldsymbol{\sigma}) = \langle \boldsymbol{Y}, \boldsymbol{\sigma}^{\otimes p} \rangle \\ \text{subject to} \quad & \boldsymbol{\sigma} \in \mathbf{S}^{N-1}. \end{aligned} \tag{1}$$

The landscape of the optimization problem (1) is the function $f_p : \mathbf{S}^{N-1} \to \mathbf{R}$:

$$f_p(\boldsymbol{\sigma}) = \langle \boldsymbol{Y}, \boldsymbol{\sigma}^{\otimes p} \rangle = \lambda \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle^p + \frac{1}{\sqrt{N}} \sum_{i_1, \ldots, i_p = 1}^{N} \boldsymbol{J}_{i_1 \ldots i_p} \boldsymbol{\sigma}_{i_1} \cdots \boldsymbol{\sigma}_{i_p}.$$

* More general things we can look at: The set of all critical points, the set of all local maxima, etc.
* Questions we can ask: Where are the critical points located? How far are they from $\boldsymbol{u}$? What is the energy value of the critical points?

# Landscape complexity of the rank-1 spiked tensor model

# Landscape complexity

Recall:

$$f_p(\boldsymbol{\sigma}) = \lambda \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle^p + \frac{1}{\sqrt{N}} \sum_{i_1,\ldots,i_p=1}^{N} \boldsymbol{J}_{i_1 \ldots i_p} \boldsymbol{\sigma}_{i_1} \cdots \boldsymbol{\sigma}_{i_p} .$$

# Landscape complexity

Recall:

$$f_p(\boldsymbol{\sigma}) = \lambda \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle^p + \frac{1}{\sqrt{N}} \sum_{i_1,\ldots,i_p=1}^{N} \boldsymbol{J}_{i_1 \ldots i_p} \boldsymbol{\sigma}_{i_1} \cdots \boldsymbol{\sigma}_{i_p}.$$

For $M \subset (-1, 1)$ and $E \subset \mathbf{R}$, define

$$\mathrm{Crt}_{N,*}(M, E) := \sum_{\boldsymbol{\sigma} : \mathrm{grad}\, f_p(\boldsymbol{\sigma}) = 0} \mathbf{1}\{\langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \in M\} \mathbf{1}\{f_p(\boldsymbol{\sigma}) \in E\}.$$

# Landscape complexity

Recall:

$$f_p(\boldsymbol{\sigma}) = \lambda \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle^p + \frac{1}{\sqrt{N}} \sum_{i_1,\ldots,i_p=1}^{N} \boldsymbol{J}_{i_1\ldots i_p} \boldsymbol{\sigma}_{i_1} \cdots \boldsymbol{\sigma}_{i_p}.$$

For $M \subset (-1,1)$ and $E \subset \mathbf{R}$, define

$$\mathrm{Crt}_{N,*}(M,E) := \sum_{\boldsymbol{\sigma}:\mathrm{grad}\, f_p(\boldsymbol{\sigma})=0} \mathbf{1}\{\langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \in M\} \mathbf{1}\{f_p(\boldsymbol{\sigma}) \in E\}.$$

**Theorem (Ben Arous et al. 2019)**

Let $M \subset (-1,1)$ and let $E \subset \mathbf{R}$. Then:

$$\mathbf{E}\, \mathrm{Crt}_{N,*}(M,E) \approx \exp\left\{ N \sup_{(m,x)\in M\times E} S_*(m,x) \right\},$$

where $S_*$ is a nasty but explicit function.

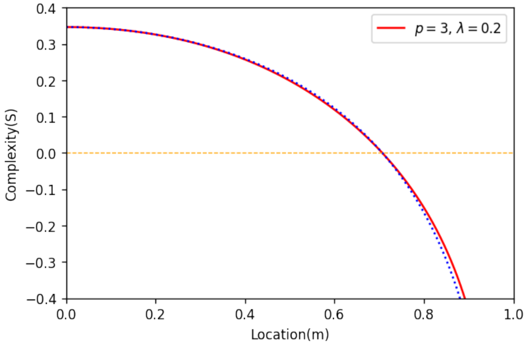
- There is another function $S_0$ satisfying an analogue for local maxima.

# Location of critical points for $r = 1$

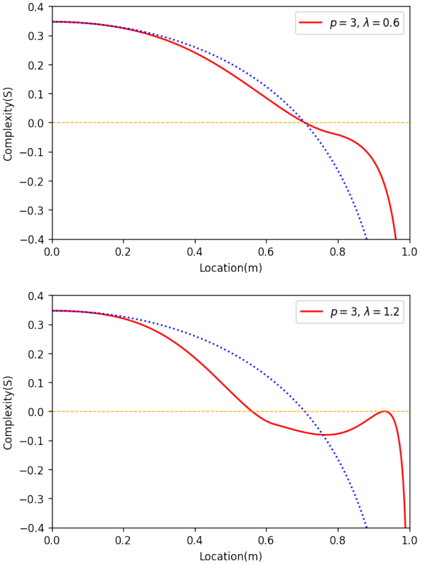
Consider

$$S_*(m) = \max_x S_*(m, x),$$

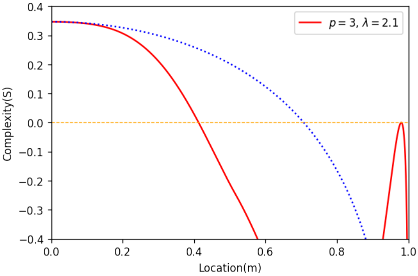
which gives the exponential growth rate of the number of critical points at $\langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle = m \in [0, 1]$.



# Location of critical points for $r = 1$



## Location of critical points for $r = 1$




**Proposition (Ben Arous et al. 2019)**

Let

$$\lambda_c := \sqrt{\frac{1}{2p}\frac{(p-1)^{p-1}}{(p-2)^{p-2}}}.$$

Then:

- If $\lambda < \lambda_c$, then there are no “good” critical points.
- If $\lambda \geq \lambda_c$, then $S_*(m) = 0$ at the point where

$$m^{2p-4}(1-m^2) = \frac{1}{2p\lambda^2}.$$

# Rank-$r$ spiked tensor model

# Rank-$r$ spiked tensor model

**Definition**

The rank-$r$ spiked tensor model is defined as

$$\boldsymbol{Y} = \sum_{i=1}^{r} \lambda_i \boldsymbol{u_i}^{\otimes p} + \frac{1}{\sqrt{N}} \boldsymbol{J},$$

where

- $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ are the signal-to-noise ratios
- $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_r} \in \mathbf{S}^{N-1}$ are $r$ unknown orthogonal signal vectors

We consider the random function $f_p : \mathbf{S}^{N-1} \to \mathbf{R}$:

$$f_p(\boldsymbol{\sigma}) = \langle \boldsymbol{Y}, \boldsymbol{\sigma}^{\otimes p} \rangle = \sum_{i=1}^{r} \lambda_i \langle \boldsymbol{u_i}, \boldsymbol{\sigma} \rangle^p + \frac{1}{\sqrt{N}} \sum_{i_1, \ldots, i_p = 1}^{N} \boldsymbol{J}_{i_1 \ldots i_p} \boldsymbol{\sigma}_{i_1} \cdots \boldsymbol{\sigma}_{i_p}.$$

# Landscape complexity

For any $M_1, \ldots, M_r \subset (-1, 1)$ and $E \subset \mathbf{R}$, define

$$\mathrm{Crt}_{N,*}(M, E) := \sum_{\boldsymbol{\sigma}:\mathrm{grad}\, f_p(\boldsymbol{\sigma})=0} \mathbf{1}\{\langle \boldsymbol{\sigma}, \boldsymbol{u_i} \rangle \in M_i, 1 \le i \le r\} \mathbf{1}\{f_p(\boldsymbol{\sigma}) \in E\}.$$

For any $M_1, \ldots, M_r \subset (-1,1)$ and $E \subset \mathbf{R}$, define

$$\mathrm{Crt}_{N,*}(M,E) := \sum_{\boldsymbol{\sigma}:\mathrm{grad}\, f_p(\boldsymbol{\sigma})=0} \mathbf{1}\{\langle \boldsymbol{\sigma}, \boldsymbol{u_i}\rangle \in M_i, 1 \le i \le r\}\mathbf{1}\{f_p(\boldsymbol{\sigma}) \in E\}.$$

**Theorem (P. 2022)**

Let $M = M_1 \times \cdots \times M_r \subset (-1,1)^r$ and $E \subset \mathbf{R}$. Then:

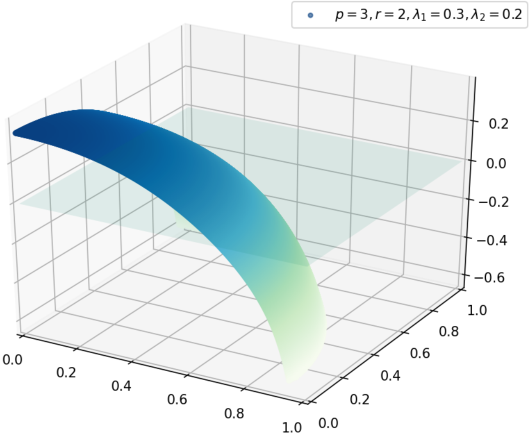
$$\mathbf{E}\,\mathrm{Crt}_{N,*}(M,E) \approx \exp\left\{N \sup_{(m,x)\in M\times E} S_*(m,x)\right\},$$

where $S_*$ is a nasty but explicit function.

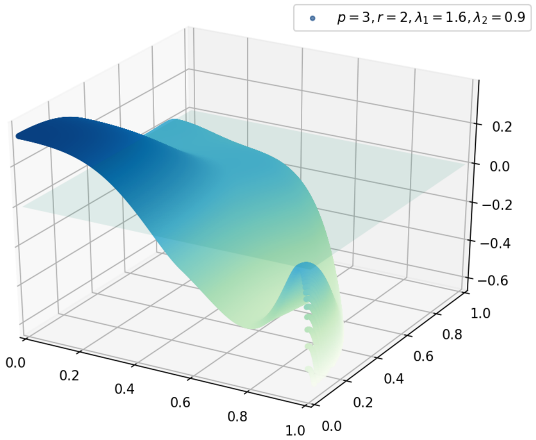
- There is another function $S_0$ satisfying an analogue for local maxima.
- For the special case $r = 1$, the theorem reduces to the result proved in Ben Arous et al. 2019.

# Location of critical points for $r = 2$

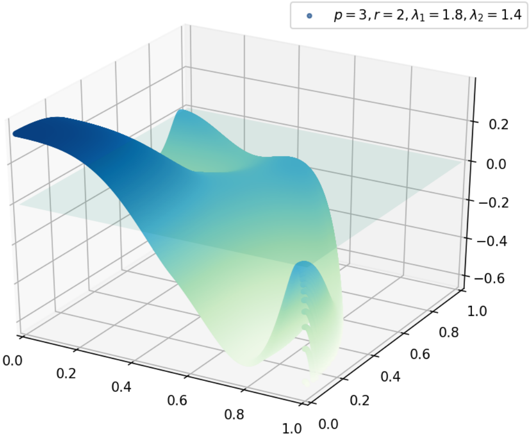
$S_*(m) = \max_x S_*(m, x)$ gives the exponential growth rate of the number of critical points at $\langle \boldsymbol{u_1}, \boldsymbol{\sigma} \rangle = m_1$ and $\langle \boldsymbol{u_2}, \boldsymbol{\sigma} \rangle = m_2$.



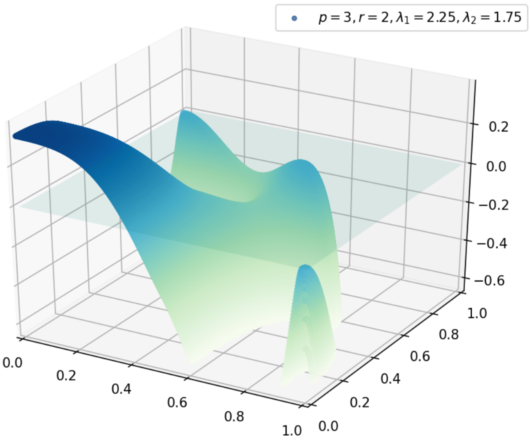
# Location of critical points for $r = 2$



# Location of critical points for $r = 2$



# Location of critical points for $r = 2$



# Location of critical points for finite $r$

**Proposition (P. 2022)**

Let

$$\eta(m) := \sum_{i=1}^{r} \lambda_i^{-\frac{2}{p-2}} \mathbf{1}\{m_i \neq 0\} \quad \text{and} \quad \eta_c := (p-2)\left(\frac{2p}{(p-1)^{p-1}}\right)^{\frac{1}{p-2}}.$$

Then:

* If $\eta(m) > \eta_c$, then the "good" critical points are exponentially rare.
* If $\eta(m) \leq \eta_c$, then $S_*(m) = 0$ at the points where

$$\sum_{i=1}^{r} \lambda_i m_i^p = \frac{1}{\sqrt{2p}} \frac{\sum_{i=1}^{r} m_i^2}{\sqrt{1 - \sum_{i=1}^{r} m_i^2}}.$$

# Proof Sketch

# The Kac-Rice formula

**Kac-Rice formula (Kac '43, Rice '44)**

For Gaussian processes $f_N : \mathbf{R}^N \to \mathbf{R}$,

$$\mathbf{E}\,\mathrm{Crt}(f_N) = \int_{\mathbf{R}^N} \mathbf{E}\left[\left|\det(\nabla^2 f_N(\boldsymbol{x}))\right| \,|\nabla f_N(\boldsymbol{x}) = 0\right] \varphi_{\boldsymbol{x}}(\mathbf{0})\mathrm{d}\boldsymbol{x},$$

where $\varphi_{\boldsymbol{x}}(\mathbf{0})$ is the density of the Gaussian vector $\nabla f_N(\boldsymbol{x})$ at $\mathbf{0}$.

Here, we consider the random variable:

$$\mathrm{Crt}_{N,*}(M,E) := \sum_{\boldsymbol{\sigma}:\mathrm{grad}\, f_p(\boldsymbol{\sigma})=0} \mathbf{1}\{\langle \boldsymbol{\sigma}, \boldsymbol{u_i}\rangle \in M_i, 1 \le i \le r\}\mathbf{1}\{f_p(\boldsymbol{\sigma}) \in E\}.$$

* The Kac-Rice formula gives

$$\begin{aligned}
&\mathbf{E}\,\mathrm{Crt}_{N,*}(M,E) \\
&= \int_{\{\boldsymbol{\sigma}:\ \langle \boldsymbol{\sigma}, \boldsymbol{u_i}\rangle \in M_i\ \forall i \in [r]\}} \mathbf{E}\left[\left|\det(\nabla^2 f_p(\boldsymbol{\sigma}))\right| \cdot \mathbf{1}\{f_p(\boldsymbol{\sigma}) \in E\} \middle| \nabla f_p(\boldsymbol{\sigma}) = \mathbf{0}\right] \varphi_{\boldsymbol{\sigma}}(\mathbf{0})\mathrm{d}\boldsymbol{\sigma}
\end{aligned}$$

# The Kac-Rice formula

Here, we consider the random variable:

$$\mathrm{Crt}_{N,*}(M,E) := \sum_{\boldsymbol{\sigma}:\mathrm{grad}\, f_p(\boldsymbol{\sigma})=0} \mathbf{1}\{\langle \boldsymbol{\sigma}, \boldsymbol{u_i}\rangle \in M_i, 1 \le i \le r\}\mathbf{1}\{f_p(\boldsymbol{\sigma}) \in E\}.$$

* The Kac-Rice formula gives

$$\begin{aligned} &\mathbf{E}\,\mathrm{Crt}_{N,*}(M,E) \\ &= \int_{\{\boldsymbol{\sigma}:\ \langle\boldsymbol{\sigma},\boldsymbol{u_i}\rangle\in M_i\ \forall i\in[r]\}} \mathbf{E}\left[\left|\det(\nabla^2 f_p(\boldsymbol{\sigma}))\right| \cdot \mathbf{1}\{f_p(\boldsymbol{\sigma}) \in E\} \middle| \nabla f_p(\boldsymbol{\sigma}) = \mathbf{0}\right] \varphi_{\boldsymbol{\sigma}}(\mathbf{0})\mathrm{d}\boldsymbol{\sigma} \end{aligned}$$

* The conditioned Hessian is distributed as a spiked random matrix:

$$\nabla^2 f_p(\boldsymbol{\sigma}) \sim H_N = \sum_{i=1}^r \gamma_i(m)\boldsymbol{e_i}\boldsymbol{e_i}^T + W_N,$$

where $W_N \sim \mathrm{GOE}(N)$.

# Random determinants

Let $H_N = \sum_{i=1}^{r} \gamma_i(m) \boldsymbol{e_i} \boldsymbol{e_i}^T + W_N$ and let $\hat{\mu}_{H_N} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(H_N)}$ be the empirical spectral measure of $H_N$.

* Then:

$$\mathbf{E}\left[|\det H_N|\right] = \mathbf{E}\left[\prod_{i=1}^{N} |\lambda_i|\right] = \mathbf{E}\left[e^{\sum_{i=1}^{N} \log|\lambda_i|}\right] = \mathbf{E}\left[e^{N \int_{\mathbf{R}} \log|\lambda| \hat{\mu}_{H_N}(\mathrm{d}\lambda)}\right]$$

Let $H_N = \sum_{i=1}^{r} \gamma_i(m)\boldsymbol{e_i e_i^T} + W_N$ and let $\hat{\mu}_{H_N} = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(H_N)}$ be the empirical spectral measure of $H_N$.

* Then:

$$\mathbf{E}\left[|\det H_N|\right] = \mathbf{E}\left[\prod_{i=1}^{N} |\lambda_i|\right] = \mathbf{E}\left[e^{\sum_{i=1}^{N} \log|\lambda_i|}\right] = \mathbf{E}\left[e^{N\int_{\mathbf{R}} \log|\lambda|\hat{\mu}_{H_N}(\mathrm{d}\lambda)}\right]$$

* LDP result: the spectrum of the deformed GOE concentrates around the semi-circle law:

$$\mathbf{P}_N^{\gamma}(\hat{\mu}_{H_N} \notin \mathrm{B}(\sigma_{sc}, \delta)) \approx e^{-N^2 C(\delta)},$$

where $\sigma_{sc}(\mathrm{d}x) = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbf{1}_{x\in[-2,2]}\mathrm{d}x$.

Let $H_N = \sum_{i=1}^r \gamma_i(m)\boldsymbol{e_i}\boldsymbol{e_i}^T + W_N$ and let $\hat{\mu}_{H_N} = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i(H_N)}$ be the empirical spectral measure of $H_N$.

* Then:

$$\mathbf{E}\left[|\det H_N|\right] = \mathbf{E}\left[\prod_{i=1}^N |\lambda_i|\right] = \mathbf{E}\left[e^{\sum_{i=1}^N \log|\lambda_i|}\right] = \mathbf{E}\left[e^{N\int_{\mathbf{R}} \log|\lambda|\hat{\mu}_{H_N}(\mathrm{d}\lambda)}\right]$$

* LDP result: the spectrum of the deformed GOE concentrates around the semi-circle law:

$$\mathbf{P}_N^\gamma(\hat{\mu}_{H_N} \notin \mathrm{B}(\sigma_{sc}, \delta)) \approx e^{-N^2 C(\delta)},$$

where $\sigma_{sc}(\mathrm{d}x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{x\in[-2,2]}\mathrm{d}x$.

Therefore,

$$\mathbf{E}[|\det H_N|] \approx e^{N\int_{\mathbf{R}} |\lambda|\sigma_{sc}(\mathrm{d}\lambda)}.$$

# Summary

- We studied the loss function $f_p : \mathbf{S}^{N-1} \to \mathbf{R}$:

$$f_p(\boldsymbol{\sigma}) = \sum_{i=1}^{r} \lambda_i \langle \boldsymbol{u_i}, \boldsymbol{\sigma} \rangle + \frac{1}{\sqrt{N}} \sum_{i_1,\dots,i_p=1}^{N} \boldsymbol{J}_{i_1\dots i_p} \boldsymbol{\sigma}_{i_1} \cdots \boldsymbol{\sigma}_{i_p}$$

- We computed the annealed complexity

$$\Sigma = \lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}[\mathrm{Crt}(f_p)]$$

  and we found that there is a threshold that separates $\Sigma > 0$ from $\Sigma = 0$.
- The proof relied on the Kac-Rice formula and on LDPs.

# Extra slide

The function $S_* : (-1,1)^r \times \mathbf{R} \to \overline{\mathbf{R}}$ is defined by

$$S_*(m,x) := S(m,x) + \Phi_*\left(\sqrt{\frac{2p}{p-1}}x\right),$$

where

$$\begin{aligned} S(m,x) = \frac{1}{2}(\log(p-1)+1) + \frac{1}{2}\log\left(1-\sum_{i=1}^{r} m_i^2\right) - p\sum_{i=1}^{r}\lambda_i^2 m_i^{2p-2}(1-m_i^2) \\ + 2p\sum_{i<j}\lambda_i\lambda_j m_i^p m_j^p - \left(x - \sum_{i=1}^{r}\lambda_i m_i^p\right)^2 \end{aligned} \tag{2}$$

and

$$\Phi_*(x) = \begin{cases} \frac{x^2}{4} - \frac{1}{2} & \text{if } |x| \le 2, \\ \frac{x^2}{4} - \frac{1}{2} - \frac{|x|}{4}\sqrt{x^2-4} + \log\left(\sqrt{\frac{x^2}{4}-1} + \frac{|x|}{2}\right) & \text{if } |x| > 2. \end{cases} \tag{3}$$