Topological complexity of polynomials with Gaussian coefficients and its implications for Tensor PCA

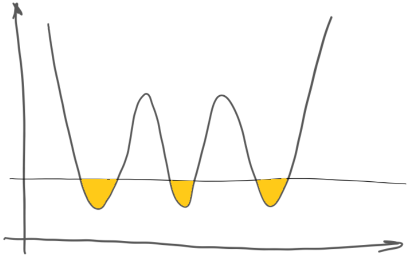
Igor Molybog

Topological complexity

- $f : \mathcal{X} \to \mathbb{R}$
- subgraph $(\alpha) : \{(x, y) | f(x) \leq y \leq \alpha\}$
- Complexity is low if the number of connected components in subgraph (α) monotonically decreasing on $[\inf_{x \in \mathcal{X}} f(x), \infty]$

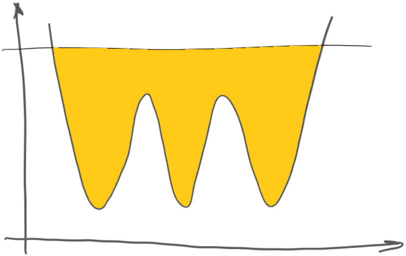
Topological complexity

- $f : \mathcal{X} \to \mathbb{R}$
- subgraph $(\alpha) : \{(x, y) | f(x) \leq y \leq \alpha\}$
- Complexity is low if the number of connected components in subgraph (α) monotonically decreasing on $[\inf_{x \in \mathcal{X}} f(x), \infty]$



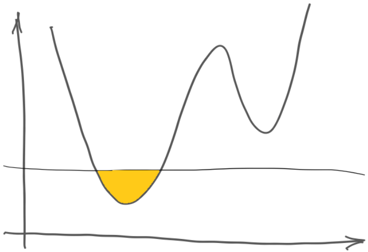
Topological complexity

- $f : \mathcal{X} \to \mathbb{R}$
- subgraph (α) : $\{(x, y) | f(x) \leq y \leq \alpha\}$
- Complexity is low if the number of connected components in subgraph (α) monotonically decreasing on $[\inf_{x \in \mathcal{X}} f(x), \infty]$



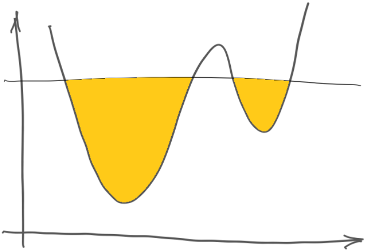
Topological complexity

- $f : \mathcal{X} \to \mathbb{R}$
- subgraph $(\alpha) : \{(x, y)|f(x) \leq y \leq \alpha\}$
- Complexity is low if the number of connected components in subgraph (α) monotonically decreasing on $[\inf_{x\in\mathcal{X}} f(x), \infty]$



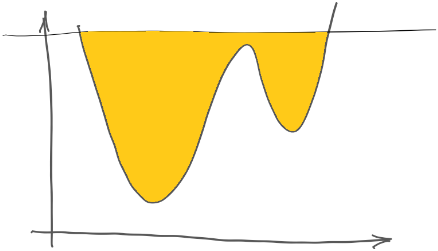
Topological complexity

- $f : \mathcal{X} \to \mathbb{R}$
- subgraph (α) : $\{(x, y) | f(x) \leq y \leq \alpha\}$
- Complexity is low if the number of connected components in subgraph (α) monotonically decreasing on $[\inf_{x \in \mathcal{X}} f(x), \infty]$



Topological complexity

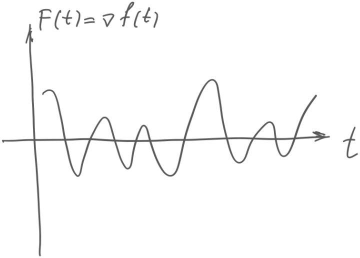
- $f : \mathcal{X} \to \mathbb{R}$
- subgraph (α) : $\{(x, y) | f(x) \leq y \leq \alpha\}$
- Complexity is low if the number of connected components in subgraph (α) monotonically decreasing on $[\inf_{x \in \mathcal{X}} f(x), \infty]$



Kac-Rice formula

- $F : U \to \mathbb{R}^d$ — a.s. C^1, centered Gaussian field (non-degenerate)
- $u \in \mathbb{R}^d$, Borel $B \subset$ open $U \subset \mathbb{R}^d$
- $N_u(B) = |\{t \in B, F(t) = u\}|$
- $\mathbb{P}[\exists t \in U \;\; F(t) = u, det\nabla F(t) = 0] = 0$

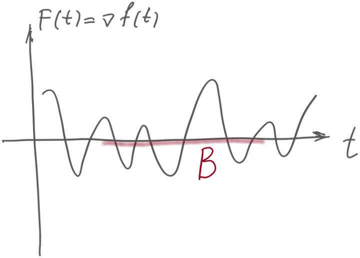
$$\mathbb{E}[N_u(B)] = \int_B \mathbb{E}\left[|det\nabla F(t)|\middle|F(t) = u\right] \mathbb{P}_{F(t)}(u)\mathbf{d}t$$



Kac-Rice formula

- $F : U \to \mathbb{R}^d$ — a.s. C^1, centered Gaussian field (non-degenerate)
- $u \in \mathbb{R}^d$, Borel $B \subset$ open $U \subset \mathbb{R}^d$
- $N_u(B) = |\{t \in B, F(t) = u\}|$
- $\mathbb{P}[\exists t \in U \;\; F(t) = u, det\nabla F(t) = 0] = 0$

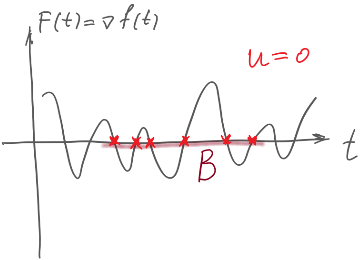
$$\mathbb{E}[N_u(B)] = \int_B \mathbb{E}\left[|det\nabla F(t)|\middle| F(t) = u\right] \mathbb{P}_{F(t)}(u)\mathbf{d}t$$



Kac-Rice formula

- $F : U \to \mathbb{R}^d$ — a.s. C^1, centered Gaussian field (non-degenerate)
- $u \in \mathbb{R}^d$, Borel $B \subset$ open $U \subset \mathbb{R}^d$
- $N_u(B) = |\{t \in B, F(t) = u\}|$
- $\mathbb{P}[\exists t \in U \ \ F(t) = u, det\nabla F(t) = 0] = 0$

$$\mathbb{E}[N_u(B)] = \int_B \mathbb{E}\left[|det\nabla F(t)| \middle| F(t) = u\right] \mathbb{P}_{F(t)}(u)\mathbf{d}t$$



Kac-Rice formula

- $F : U \to \mathbb{R}^d$ — a.s. C^1, centered Gaussian field (non-degenerate)
- $u \in \mathbb{R}^d$, Borel $B \subset$ open $U \subset \mathbb{R}^d$
- $N_u(B) = |\{t \in B, F(t) = u\}|$
- $\mathbb{P}[\exists t \in U \;\; F(t) = u, det\nabla F(t) = 0] = 0$

$$\mathbb{E}[N_u(B)] = \int_B \mathbb{E}\left[|det\nabla F(t)|\middle|F(t) = u\right] \mathbb{P}_{F(t)}(u)\mathbf{d}t$$

- Can consider $B = f^{-1}(V)$, $B = \{t|\nabla^2 f(t) \succeq 0)\}$, etc.
- Generalizations to non-Gaussian fields are available
- Generalizations to higher moments are available

Spherical p-spin glass Hamiltonian

$J_{i_1 \dots i_p}$ are i.i.d $\mathcal{N}(0,1)$; $x \in \mathbb{S}^{N-1}(\sqrt{N})$

$$H_{N,p}(x) = N^{\frac{1-p}{2}} \sum_{i_1,\dots,i_p=1}^{N} J_{i_1 \dots i_p} x_{i_1} \dots x_{i_p}$$

- any isotropic Gaussian field on a sphere is a conic combination of p-spin glasses for different p
- worst case num of crit points of $H_{N,p}$ is $2((p-1)^{N-1} + (p-1)^{N-2} + \ldots + 1)$
- average case is half of that

The average number of critical points

$Cr_{N,k}(B)$ is the number of critical points of $H_{N,p}$ on $\mathbb{S}^{N-1}(\sqrt{N})$ with index k (num. of descent directions) and values in B.

$$\mathbb{E}[Cr_{N,k}(B)] = 2\sqrt{\frac{2}{p}}(p-1)^{\frac{N}{2}}\mathbb{E}_{GOE(N)}\left[e^{-N\frac{p-2}{2p}\lambda_k^2}\mathbb{I}\{\lambda_k \in \sqrt{\frac{p}{2(p-1)}}B\}\right]$$

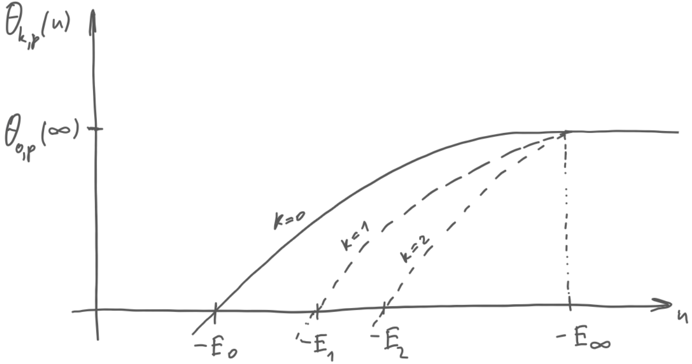
If we take $B = (-\infty, Nu]$ then

$$\lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}[Cr_{N,k}(B)] = \Theta_{k,p}(u)$$

The average number of critical points

$Cr_{N,k}(B)$ is the number of critical points of $H_{N,p}$ on $\mathbb{S}^{N-1}(\sqrt{N})$ with index k (num. of descent directions) and values in $B = (-\infty, u]$.

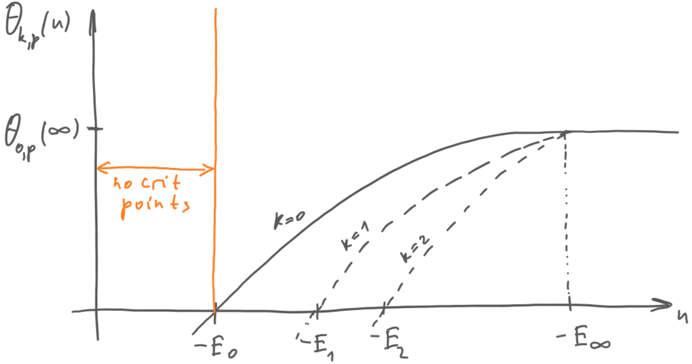
$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[Cr_{N,k}(B)] = \Theta_{k,p}(u)$$



The average number of critical points

$Cr_{N,k}(B)$ is the number of critical points of $H_{N,p}$ on $\mathbb{S}^{N-1}(\sqrt{N})$ with index k (num. of descent directions) and values in $B = (-\infty, u]$.

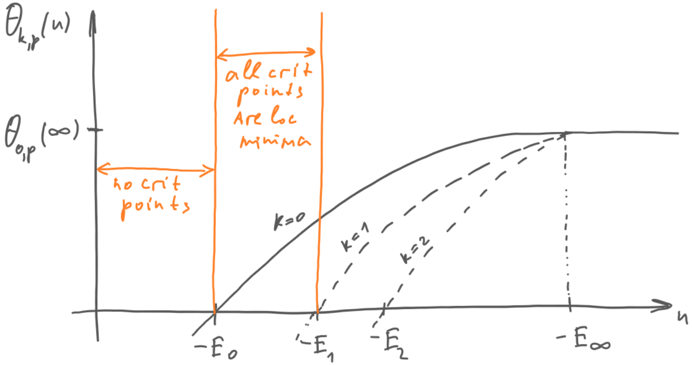
$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[Cr_{N,k}(B)] = \Theta_{k,p}(u)$$



The average number of critical points

$Cr_{N,k}(B)$ is the number of critical points of $H_{N,p}$ on $\mathbb{S}^{N-1}(\sqrt{N})$ with index k (num. of descent directions) and values in $B = (-\infty, u]$.

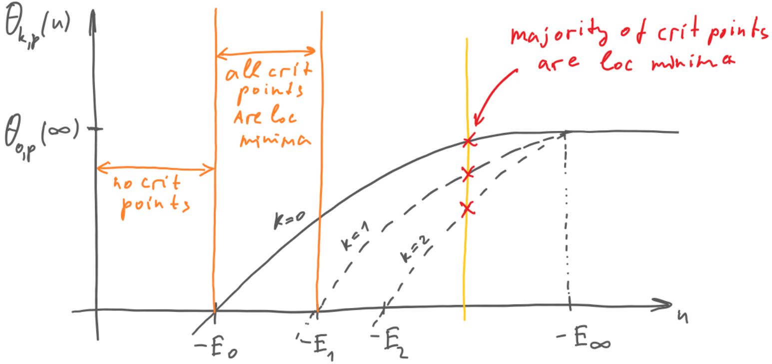
$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[Cr_{N,k}(B)] = \Theta_{k,p}(u)$$



The average number of critical points

$Cr_{N,k}(B)$ is the number of critical points of $H_{N,p}$ on $\mathbb{S}^{N-1}(\sqrt{N})$ with index k (num. of descent directions) and values in $B = (-\infty, u]$.

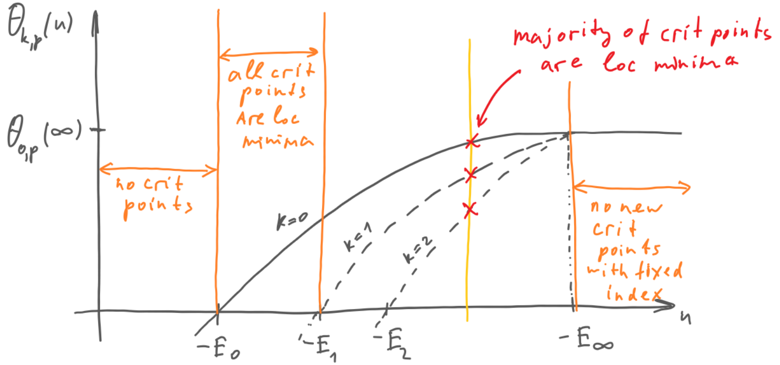
$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[Cr_{N,k}(B)] = \Theta_{k,p}(u)$$



The average number of critical points

$Cr_{N,k}(B)$ is the number of critical points of $H_{N,p}$ on $\mathbb{S}^{N-1}(\sqrt{N})$ with index k (num. of descent directions) and values in $B = (-\infty, u]$.

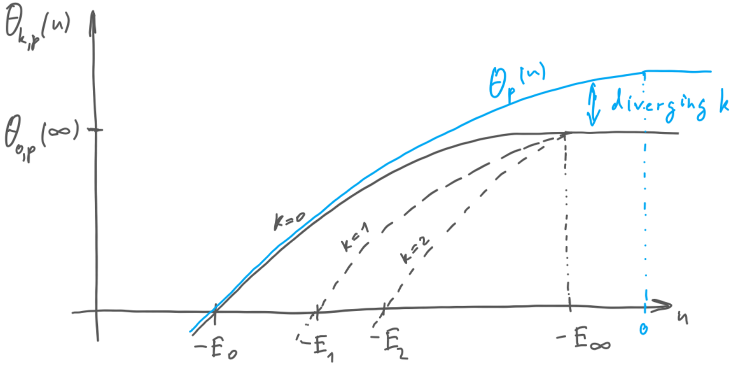
$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[Cr_{N,k}(B)] = \Theta_{k,p}(u)$$



The average number of critical points

$Cr_{N,k}(B)$ is the number of critical points of $H_{N,p}$ on $\mathbb{S}^{N-1}(\sqrt{N})$ with index k (num. of descent directions) and values in $B = (-\infty, u]$.

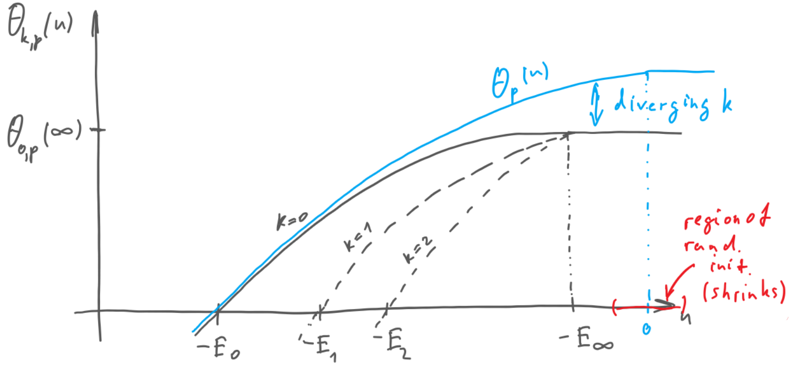
$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[Cr_{N,k}(B)] = \Theta_{k,p}(u)$$



The average number of critical points

$Cr_{N,k}(B)$ is the number of critical points of $H_{N,p}$ on $\mathbb{S}^{N-1}(\sqrt{N})$ with index k (num. of descent directions) and values in $B = (-\infty, u]$.

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[Cr_{N,k}(B)] = \Theta_{k,p}(u)$$

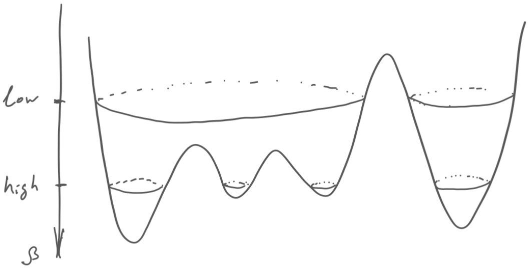


Implications on Langevin dynamics

w — Brownian motion on the sphere

$$\begin{cases} \mathbf{d}x_t = \mathbf{d}w_t - \beta\nabla H_N(x_t)\mathbf{d}t \\ x\big|_{t=0} = x_0 \end{cases}$$

- Convergence is fast if β is small and exponentially slow if β is low.
- Multiple interesting time scales in high β regime, extensively studied by physicists

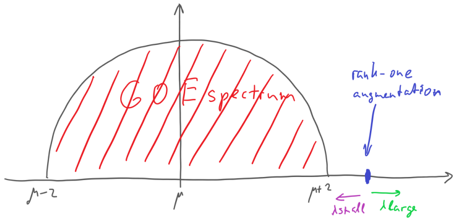


Tensor PCA problem formulation

Equivalent reformulations :

- $\max_{x \in \mathbb{S}^{N-1}} \sum_{i=1}^{M} \langle Y_i, \sigma^{\otimes p} \rangle$;
  $Y_i = \mu u^{\otimes p} + \frac{1}{\sqrt{2N}} W_i$; $W_i = \sum_{\pi \in \Sigma_p} \frac{G^{\pi}}{p!}$; G — i.i.d. Gaussian
- $\max_{x \in \mathbb{S}^{N-1}} \langle Y, \sigma^{\otimes p} \rangle$; $Y = \lambda u^{\otimes p} + \frac{1}{\sqrt{2N}} W$
- $\max_{x \in \mathbb{S}^{N-1}} f(x)$; $f(x) = \lambda \langle u, x \rangle^p - H_{N,p}(x)$

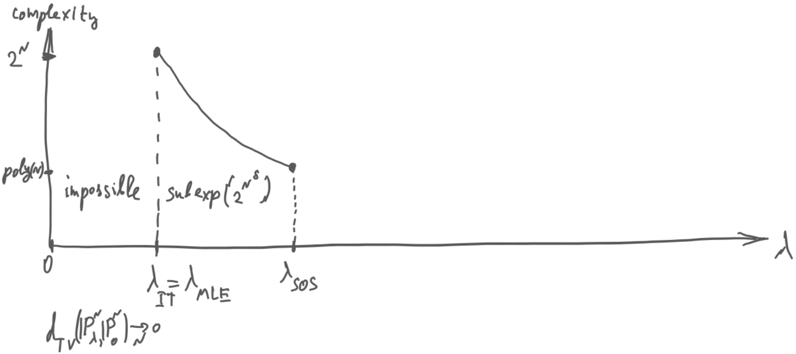
Hessian looks like $\underbrace{GOE_N + \lambda I_N}_{\text{like p-spin glass}} + c_N \lambda A_N$



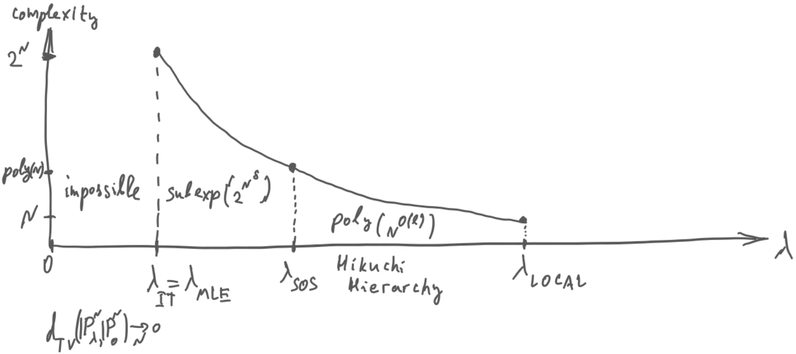
Complexity of Tensor PCA

Complexity of Tensor PCA

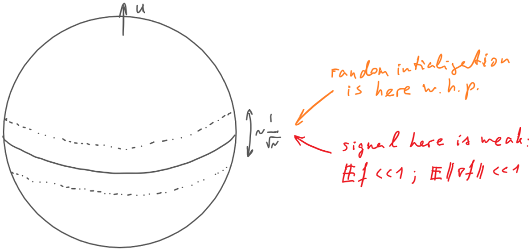
Complexity of Tensor PCA

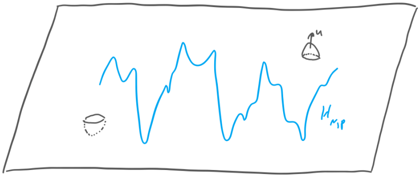


Complexity of Tensor PCA



MLE landscape

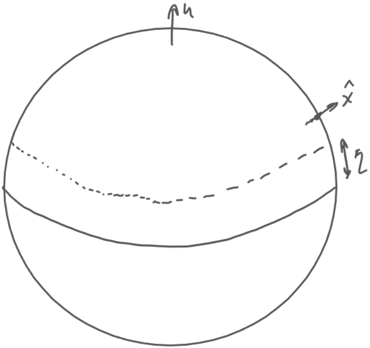




Note : weak recovery

An estimate $\hat{x}_N$ achieves weak recovery if for large N and some fixed $\eta > 0$

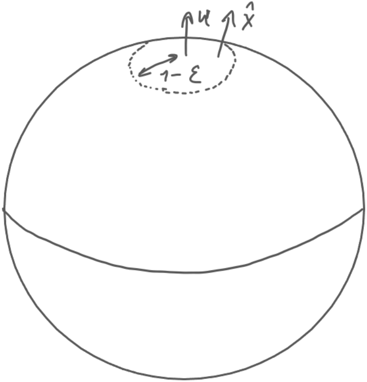
$$\langle \hat{x}_N, u \rangle > \eta$$



Note : strong recovery

An estimate $\hat{x}_N$ achieves strong recovery if there exists $\varepsilon_N \to 0$

$$\langle \hat{x}_N, u \rangle > 1 - \varepsilon$$



Complexity of MLE

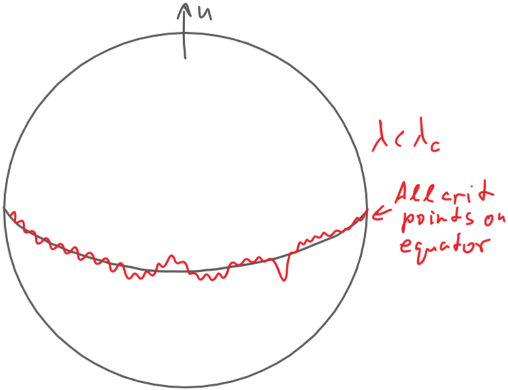
$Cr_f(M, E)$ num of crit points x of f with $f(x) \in E$ and $\langle x, u\rangle \in M$

$$\lim_{N\to\infty} \sup \frac{1}{N} \log \mathbb{E}[Cr_f(M, E)] \leq \sup_{m\in\bar{M}, e\in\bar{E}} S_x(m, e)$$

$$\lim_{N\to\infty} \inf \frac{1}{N} \log \mathbb{E}[Cr_f(M, E)] \geq \inf_{m\in\bar{M}, e\in\bar{E}} S_x(m, e)$$

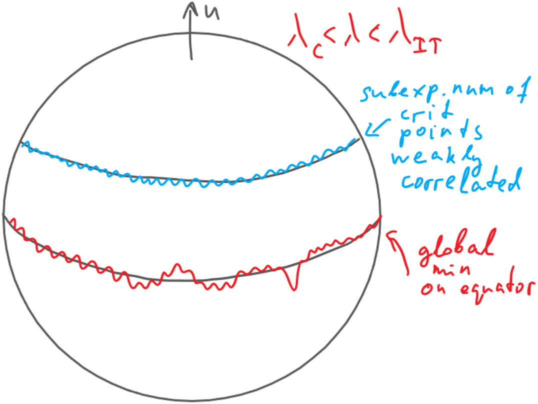
Complexity of MLE

Considering λ around $\lambda_{IT} = \lambda_{MLE}$, introduce $\lambda_c < \lambda_{IT}$



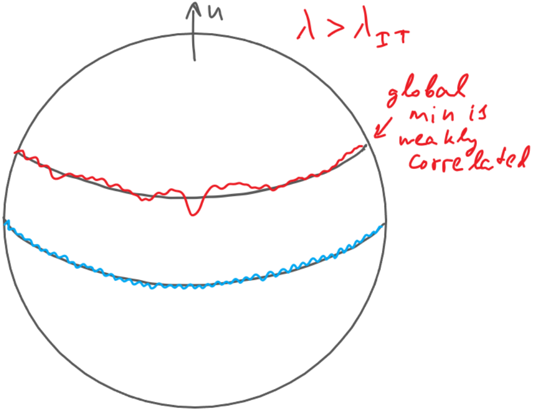
Complexity of MLE

Considering λ around $\lambda_{IT} = \lambda_{MLE}$, introduce $\lambda_c < \lambda_{IT}$



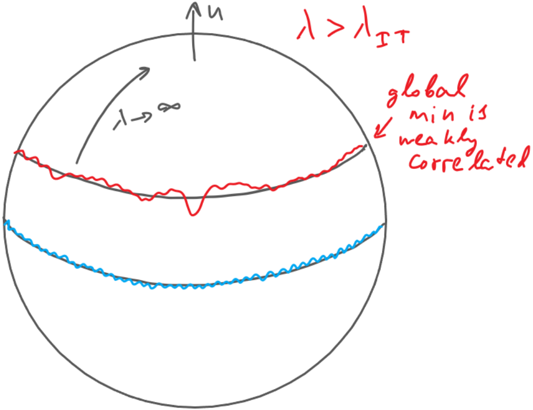
Complexity of MLE

Considering λ around $\lambda_{IT} = \lambda_{MLE}$, introduce $\lambda_c < \lambda_{IT}$



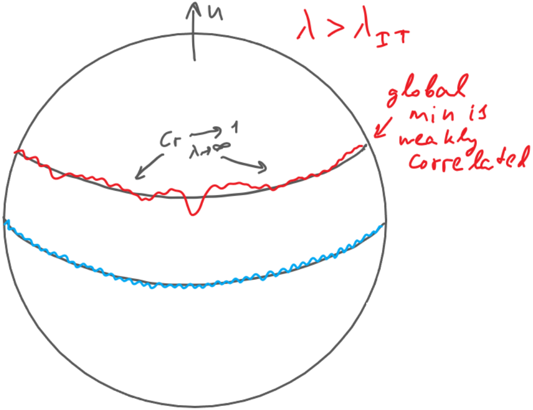
Complexity of MLE

Considering λ around $\lambda_{IT} = \lambda_{MLE}$, introduce $\lambda_c < \lambda_{IT}$



Complexity of MLE

Considering λ around $\lambda_{IT} = \lambda_{MLE}$, introduce $\lambda_c < \lambda_{IT}$



Gradient Descent with restarts

ML models are being trained from r multiple initialization for a fixed number of steps T.

- If λ is small :

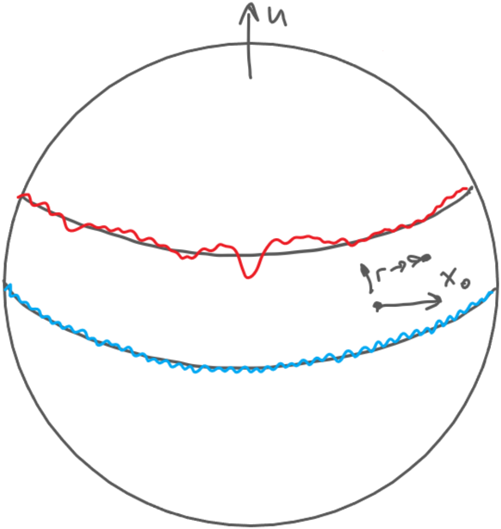
$$(T \text{ fixed}, r \to \infty) \text{ works better than } (T \to \infty, r \text{ fixed})$$

- If λ is large :

$$(T \to \infty, r \text{ fixed}) \text{ works better than } (T \text{ fixed}, r \to \infty)$$

What is the right way to pick T/r ?

Gradient Descent with restarts



Tight anticoncentration for $\langle x_0, u\rangle$

Let $(Z_1, \ldots, Z_N)$ be a standard Gaussian vector. Then $\langle x_0, u\rangle \sim \frac{Z_1}{\sqrt{Z_1^2+\ldots+Z_N^2}}$

$$\langle x_0, u\rangle^2 \sim Beta(1/2, (N-1)/2)$$

Summary

- Kac-Rice formula is a valuable tool for the analysis of a noisy landscape
- Tensor PCA is a rich problem class with a lot of structure, very close to the problems our group has been working on
- I'm working towards the specific trade-off between the number of trajectories and their length, but there is a myriad of open problems

Questions ?

Thank you for your attention !

Questions ?

References I

Robert J Adler, Jonathan E Taylor, et al. **Random fields and geometry**, volume 80. Springer, 2007.

Gerard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The landscape of the spiked tensor model. **Communications on Pure and Applied Mathematics**, 72(11) :2282–2330, 2019.

Gerard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Online stochastic gradient descent on non-convex losses from high-dimensional inference. **Journal of Machine Learning Research**, 22(106) :1–51, 2021.

Antonio Auffinger, Gérard Ben Arous, and Jiří Černỳ. Random matrices and complexity of spin glasses. **Communications on Pure and Applied Mathematics**, 66(2) :165–201, 2013.

Jean-Marc Azaïs and Mario Wschebor. **Level sets and extrema of random processes and fields**. John Wiley & Sons, 2009.

References II

Alexander S Wein, Ahmed El Alaoui, and Cristopher Moore. The kikuchi hierarchy and tensor pca. In **2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)**, pages 1446–1468. IEEE, 2019.