STAT 291 **Spring 2024**

Problem Set 1

Instructions: many points are available in the problems below. 100 **points will count as a perfect score.** Subproblems are equally weighted unless stated otherwise, and you can earn credit for later subproblems without solving the previous ones. Points will be cumulative throughout the semester, so you will get credit for earning more than 100 points on this problem set if you choose to do so. You may also submit solutions to *extra problems* (listed in a separate file on the course website) at the end of any problem set. **Solutions may be handwritten or Latexed** and should be submitted in PDF format via Canvas or email. The due date for this assignment is **February 23rd**.

Collaboration with your classmates is encouraged. **Please identify everyone you worked with at the beginning of your solution PDF** (e.g. *Collaborators: Alice, Bob, and GPT4*). Your solutions should be written entirely by you, even if you collaborated to solve the problems.

The first person to report each mistake in this problem set (by emailing me and Yufan) will receive up to 5 extra points, depending on the mistake.

Problem 1. The Kac-Rice Formula (98 points)

1. Define the random polynomial $P : \mathbb{R} \to \mathbb{R}$ by $P(x) = \sum_{i=0}^{N} \sqrt{\binom{N}{i}}\, g_i x^i$, where $g_i \sim Normal(0,1)$ are i.i.d. standard Gaussians. Find the expected number of real zeros for such a polynomial.

2. Give formulas (possibly in terms of integrals or combinatorial sums) for:

   (a) The second moment of the number of zeros of P in $[-10, 10]$.

   (b) The expected number of zeros x satisfying $P'(x) \geq 10$.

For the rest of this problem, fix $p \geq 3$ and consider the tensor PCA Hamiltonian

$$\widehat{H}_{N,p}(\boldsymbol{x}) = H_{N,p}(\boldsymbol{x}) + \lambda N \left(\frac{\langle \boldsymbol{x}, \boldsymbol{\sigma} \rangle}{N} \right)^p.$$

3. Given $\boldsymbol{x} \in \mathcal{S}_N$, determine the joint law of

$$\left(\widehat{H}_{N,p}(\boldsymbol{x}), \nabla_{\mathrm{rad}} \widehat{H}_{N,p}(\boldsymbol{x}), \nabla_{\mathrm{sp}} \widehat{H}_{N,p}(\boldsymbol{x}), \nabla^2_{tan} \widehat{H}_{N,p} \right)$$

   in terms of $R = R(\boldsymbol{x}, \boldsymbol{\sigma})$.

4. Show that for fixed $\boldsymbol{x} \in \mathcal{S}_N$, the spectrum of the tangential Hessian $\nabla^2_{tan} \widehat{H}_{N,p}(\boldsymbol{x})$ converges in probability to the semicircle distribution in the bounded-Lipschitz metric, similarly to the case of the p-spin model. (Hint: Cauchy's eigenvalue interlacing theorem may be helpful.)

5. The previous parts suggest a variational formula for the expected number of critical points of $\widehat{H}_{N,p}$ with energy and overlap $(E, R) \in [a_1, b_1] \times [a_2, b_2]$ in a given rectangle, to leading exponential order. Provide such a formula, and discuss what changes from class, if any, are required to prove the upper bound. **In fact the corresponding formula remains true, and you may use it in the next part**.

6. Assume $\lambda \geq \lambda_0(p)$ is a sufficiently large constant. Show the expected number of critical points $\boldsymbol{y} \in \mathcal{S}_N$ of $\widehat{H}_{N,p}$ which satisfy $R(\boldsymbol{y}, \boldsymbol{\sigma}) \geq 1/2$ is $e^{o(N)}$. (Note that this is both an upper and lower bound!)

7. In the previous part, give a simpler argument (without the Kac–Rice formula) that the expected number of such critical points is at least $1 - o(1)$. (Hint: consider the global maximum of $\widehat{H}_{N,p}$.)

Problem 2. Contiguity at Exponential Scale (98 points)

Recall that if P, Q are probability measures on a common space, $Q \ll P$ ("Q is *absolutely continuous* with respect to P") if for any measurable set E, if $P(E) = 0$ then $Q(E) = 0$. The classical extension of this notion to sequences of probability distributions is *contiguity*. If $(Q_N, P_N)_{N\geq 1}$ are probability measures on a sequence of common spaces Ω_N, we write $P_N \triangleleft Q_N$ ("P_N is contiguous with respect to Q_N") if for any sequence of events $(E_N)_{N\geq 1}$,

$$\lim_{N\to\infty} Q_N(E_N) = 0 \quad \Longrightarrow \quad \lim_{N\to\infty} P_N(E_N) = 0.$$

1. Assume that $P_N \ll Q_N$ for each N, i.e. the Radon–Nikodym derivative $\frac{\mathrm{d}P_N}{\mathrm{d}Q_N}$ is defined.

   (a) Show that if
   $$\frac{\mathrm{d}P_N}{\mathrm{d}Q_N} \leq C$$
   holds Q_N-almost surely for an N-independent constant C, then contiguity $P_N \triangleleft Q_N$ follows.[1]

This problem investigates a flexible notion of *contiguity at exponential scale*. Say events $(E_N)_{N\geq 1}$ are Q_N*-exponentially unlikely* if $Q_N(E_N) \leq Ce^{-cN}$ holds for N-independent constants $C, c > 0$; similarly for P_N. We say "P_N is contiguous with respect to Q_N at exponential scale" if any sequence of Q_N-exponentially unlikely events is also P_N-exponentially unlikely (possibly for different constants C, c).

2. **Suppose** that for all $\varepsilon > 0$, the sequence of events
   $$\frac{\mathrm{d}P_N}{\mathrm{d}Q_N} \geq e^{\varepsilon N}$$
   are P_N-exponentially unlikely, with constants C, c possibly depending on ε. **Prove** that P_N is contiguous with respect to Q_N at exponential scale.

3. For a p-spin model as in class, suppose β is such that $\lim_{N\to\infty} \mathbb{E}F_N(\beta) = \beta^2/2$. Show that $\left|\mathbb{E}F_N(\beta) - \frac{\beta^2}{2}\right| \geq \varepsilon$ is exponentially unlikely.

We now use contiguity at exponential scale to connect the p-spin model with tensor PCA in the replica-symmetric phase. It may help to use/recall the first part of Homework 1, problem 4 below. Consider the following two methods to sample a p-spin Hamiltonian together with a point $\boldsymbol{\sigma} \in \mathcal{S}_N$:

- Generate a p-spin Hamiltonian $H_{N,p}(\boldsymbol{x})$ as usual, and then sample $\boldsymbol{\sigma} \sim \mu_\beta$ from its Gibbs measure. Output $(H_{N,p}, \boldsymbol{\sigma})$.

- Sample $\boldsymbol{\sigma} \in \mathcal{S}_N$ uniformly at random, generate a p-spin Hamiltonian $H_{N,p}(\boldsymbol{x})$ as usual, and let
  $$\widehat{H}_{N,p}(\boldsymbol{x}) = H_{N,p}(\boldsymbol{x}) + \beta N R(\boldsymbol{\sigma}, \boldsymbol{x})^p.$$
  Output $(\widehat{H}_{N,p}, \boldsymbol{\sigma})$.

[1] See Le Cam's first lemma for a much more precise description of contiguity.

We let P_N be the law of the former, and Q_N the law of the latter.

4. Show that $\frac{\mathrm{d}Q_N}{\mathrm{d}P_N} = \frac{Z_N(\beta)}{\mathbb{E}^{P_N}[Z_N(\beta)]}$, where $Z_N(\beta)$ is the partition function of the first argument. (Hint: both methods are "trying" to sample from the density $\exp\left(-\|\boldsymbol{G}_N^{(p)}\|_2^2/2 + \beta N^{-(p-1)/2}\langle \boldsymbol{G}_N^{(p)}, \boldsymbol{\sigma}^{\otimes p}\rangle\right)$. But one of them is doing so incorrectly.)

5. For β such that $\lim_{N\to\infty} \mathbb{E}F_N(\beta) = \beta^2/2$, conclude that P_N is contiguous with respect to Q_N at exponential scale.

6. For β as in the previous part, prove the free energy of the tensor PCA Hamiltonian $\widehat{H}_{N,p}$ also converges in probability to $\beta^2/2$ as $N \to \infty$. (Hint: similarly to $H_{N,p}$, you can show the free energy of $\widehat{H}_{N,p}$ also sharply concentrates. It might help to use the rotational symmetry in distribution of the p-spin model to argue the location of $\boldsymbol{\sigma}$ is irrelevant, before adding in the Gaussian noise.)

7. For β as in the previous parts, let $\boldsymbol{\sigma}, \widetilde{\boldsymbol{\sigma}} \sim \mu_\beta^{H_{N,p}}$ be IID samples from the p-spin Gibbs measure. Show that if p is odd, then for any $\varepsilon > 0$, the event

$$\mathbb{P}[R(\boldsymbol{\sigma}, \widetilde{\boldsymbol{\sigma}}) \le -\varepsilon]$$

is exponentially unlikely. (Hint: compare $(H_{N,p}, \boldsymbol{\sigma})$ with $(\widehat{H}_{N,p}, \boldsymbol{\sigma})$ sampled from the second method above. By using the tensor PCA formulation and reflection symmetry in the direction $\boldsymbol{\sigma}$, show that the free energy at positive overlap $R \ge \varepsilon$ with $\boldsymbol{\sigma}$ is strictly larger than the free energy at overlap $-R$ with $\boldsymbol{\sigma}$.)

Problem 3. Extreme Critical Points of Spherical p-spin Models (60 points)

1. **(10 points)** Recall from class that $\sup_{\boldsymbol{x}\in\mathcal{S}_N} \|\nabla H_{N,p}(\boldsymbol{x})\| \le C\sqrt{N}$ with probability at least $1 - e^{-N}$. Show that for other constants $C' = C'(p)$ and $C''(p) = C''(p)$,

$$\begin{aligned} \sup_{\boldsymbol{x}\in\mathcal{S}_N} \|\nabla^2 H_{N,p}(\boldsymbol{x})\|_{op} &\le C', \\ \sup_{\boldsymbol{x}\in\mathcal{S}_N} \|\nabla^3 H_{N,p}(\boldsymbol{x})\|_{op} &\le C'/\sqrt{N} \end{aligned} \tag{1}$$

hold with the same probability.

2. **(25 points)** Recall from class that $E_0 = E_0(p)$ is the energy threshold above which the expected number of critical points drops to $e^{-\Omega(N)}$. Show there exists $\varepsilon = \varepsilon(p)$ such that with probability $1 - e^{-c(\varepsilon,p)N}$, **all** critical points $\boldsymbol{x} \in \mathcal{S}_N$ with $H_{N,p}(\boldsymbol{x})/N \ge E_0 - \varepsilon$ are local maxima, with all eigenvalues of $\nabla^2_{\mathrm{sp}} H_{N,p}(\boldsymbol{x})$ less than $-\varepsilon$. (Hint: use the technique from Lecture 7.)

3. **(25 points)** Show that for small $\delta \le \delta_0(p)$, with probability $1 - e^{-c(\delta,p)N}$, the following holds for **all** points $\boldsymbol{x} \in \mathcal{S}_N$ with $H_{N,p}(\boldsymbol{x})/N \ge E_0 - \delta$. There exists $\boldsymbol{y} \in \mathcal{S}_N$ which is a local maximum of $H_{N,p}$ and satisfies $\|\boldsymbol{x} - \boldsymbol{y}\| \le O(\delta\sqrt{N})$.
(Hint: consider gradient flow $(\boldsymbol{x}_t)_{t\ge 0}$ on $\mathcal{S}_N$ starting from $\boldsymbol{x}_0 = \boldsymbol{x}$, and let $\boldsymbol{y}$ be an accumulation point of $\{\boldsymbol{x}_n\}_{n\in\mathbb{Z}_+}$. Show that $\boldsymbol{y}$ is a critical point with larger energy, and use the previous parts to show that $\boldsymbol{x}$ could not have started far from $\boldsymbol{y}$. It may be helpful to Taylor expand $H_{N,p}$ near $\boldsymbol{y}$ to second order and bound the error via (1).)

Problem 3. Survey (4 points)

Rate each of the three problems above that you worked on from 1 to 5 based on:

- Interestingness (1 for boring, 5 for exciting)
- Difficulty (1 for too easy, 5 for too hard)
- Learning (1 for almost nothing, 5 if you learned a lot).

Optionally, you are encouraged to elaborate on your ratings, and to share any other comments you have regarding this problem set or the recent lectures. Suggestions on potential improvements for future weeks or years of the course are especially appreciated.