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 (e.g. Problem 1(a) is worth $48/4 = 12$ points), 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* (which will be listed in a separate file) 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 9th**.

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. Practice with Free Energies (60 points)

Let μ be a probability measure on an arbitrary measurable space Ω, and let $H : \Omega \to [-C, C]$ be a bounded measurable function. The associated partition function, free energy, and Gibbs measure are:

$$Z(\beta) = \mathbb{E}^{\omega\sim\mu}[e^{\beta H(\omega)}], \qquad F(\beta) = \log Z(\beta), \qquad \mu_\beta(\mathrm{d}\omega) = \frac{e^{\beta H(\omega)}\mu(\mathrm{d}\omega)}{Z(\beta)}. \tag{1}$$

(Note: the lack of division by β in $F(\beta)$ is intentional, and is convenient for the first part below.)

(a) Show that F is a convex function of β, and

$$\frac{\mathrm{d}F(\beta)}{\mathrm{d}\beta} = \mathbb{E}^{\omega\sim\mu_\beta}[H(\omega)].$$

(Hint: Hölder's inequality may be helpful to show convexity.)

(b) For $\gamma \in \mathbb{R}$, define the super-level set

$$\Omega(\gamma_+) = \{\omega \in \Omega \ : \ H(\omega) \geq \gamma\}. \tag{2}$$

Give an integral formula for $F(\beta)$ in terms of the volumes $V(\gamma) = \mu(\Omega(\gamma_+))$.

Next, let $H_N : \Omega_N \to [-CN, CN]$ be a (non-random) sequence of Hamiltonians on probability spaces (Ω_N, μ_N), for $N \geq \mathbb{Z}_+$ and let $F_N(\beta) = \frac{1}{N}\log Z_N(\beta)$. Similarly to above, define the super-level set volumes

$$V_N(\gamma) = \mu_N\big(\{\omega \in \Omega_N \ : \ H_N(\omega) \geq \gamma N\}\big).$$

Assume in the parts below that

$$\lim_{N\to\infty} F_N(\beta) = \beta^2/2 \tag{3}$$

for each $\beta \in [0, \beta_0]$.

(c) For each $\beta \in (0, \beta_0)$, show that

$$\limsup_{N\to\infty} \frac{1}{N} \log V_N(\beta) \le -\beta^2/2.$$

(Hint: if $V_N(\beta)$ is too big, $F_N(\beta)$ will be too big for (3) to hold.)

(d) **(Added to help in the last part):** Similarly, for all $\alpha \ge \beta_0$, show that

$$\limsup_{N\to\infty} \frac{1}{N} \log V_N(\alpha) \le \frac{\beta_0^2}{2} - \beta_0 \alpha.$$

(Here we use the convention that $\log(0) = \infty$.)

(e) Show that in fact

$$\lim_{N\to\infty} \frac{1}{N} \log V_N(\beta) = -\beta^2/2, \quad \forall \beta \in (0, \beta_0).$$

(Hint: suppose that $\frac{1}{N} \log V_N(\beta) \le -\beta^2/2 - \varepsilon$ is too small. Using the upper bounds from the previous parts, show that the assumption (3) must fail for some other value β' near β. It may be helpful to observe that $V_N(\cdot)$ is a monotone function.)

Problem 2. Random Spherical Perceptron for $\kappa = 0$ (48 points)

In this problem, you will solve the random spherical perceptron in the case $\kappa = 0$. Recall that for $\kappa \in \mathbb{R}$ and $\alpha > 0$, the N-dimensional spherical perceptron is defined by the $M = \alpha N$ constraints

$$\langle \boldsymbol{\sigma}, \mathbf{g}_a \rangle \ge \kappa\sqrt{N}, \quad 1 \le a \le M.$$

Here the vectors $\mathbf{g}_1, \dots, \mathbf{g}_M \in \mathbb{R}^N$ are IID standard Gaussian (i.e. the NM coordinates are IID standard Gaussian). We say the point $\boldsymbol{\sigma} \in \mathcal{S}_N \subseteq \mathbb{R}^N$ with norm $\|\boldsymbol{\sigma}\| = \sqrt{N}$ is a *solution* if it obeys all of these constraints.

(a) Consider the M hyperplanes

$$U_a = \mathbf{g}_a^\perp \subseteq \mathbb{R}^N.$$

Argue that each region in $\mathbb{R}^N$ formed by these hyperplanes corresponds to a different value of the vector

$$\Big(\mathrm{sign}(\langle \boldsymbol{\sigma}, \mathbf{g}_1 \rangle), \mathrm{sign}(\langle \boldsymbol{\sigma}, \mathbf{g}_2 \rangle), \dots, \mathrm{sign}(\langle \boldsymbol{\sigma}, \mathbf{g}_M \rangle)\Big) \in \{\pm 1\}^M.$$

In other words, two vectors $\boldsymbol{\sigma}, \boldsymbol{\sigma}' \notin \bigcup_{a=1}^M U_a$ are connected by a continuous path in the complement of $\bigcup_{a=1}^M U_a$ if and only if they have the same sign pattern. (Hint: line segments will suffice as paths.)

(b) Let $R_N(\mathbf{g}_1, \dots, \mathbf{g}_M)$ denote the (random) number of N-dimensional regions formed in this way. Show that if $M \le N$, then almost surely

$$R_N(\mathbf{g}_1, \dots, \mathbf{g}_M) = 2^M.$$

(Hint: use linear independence and the previous part.)

(c) Note that R_N does not change if any subset of vectors $\mathbf{g}_a$ are negated. Using symmetry, deduce that for $\kappa = 0$, the probability for a solution $\boldsymbol{\sigma}$ to exist is exactly $\mathbb{E}[R_N(\mathbf{g}_1, \dots, \mathbf{g}_M)]/2^M$.

(d) Argue that

$$R_N(\mathbf{g}_1,\ldots,\mathbf{g}_M) - R_N(\mathbf{g}_1,\ldots,\mathbf{g}_{M-1})$$

is equal to the number of regions formed by the $(N-2)$ dimensional subspaces

$$U_1 \cap U_M, \quad U_2 \cap U_M, \ldots, \quad U_{M-1} \cap U_M$$

within U_M. Identify this as $R_{N-1}(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{M-1})$ for some vectors $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{M-1} \in U_M \simeq \mathbb{R}^{N-1}$, and show the vectors $\boldsymbol{v}_a$ are again IID standard Gaussian.

(e) Show by induction (e.g. on $N+M$) that $R_N(\mathbf{g}_1,\ldots,\mathbf{g}_M)$ is almost surely constant, and in fact equals

$$2\sum_{k=0}^{N-1}\binom{M-1}{k}.$$

(f) Deduce that for $\kappa = 0$, a spherical perceptron solution $\boldsymbol{\sigma} \in \mathcal{S}_N$ exists with high probability (i.e. tending to 1 as $N \to \infty$) for any fixed $\alpha < 2$. Conversely for $\alpha > 2$, show the probability for a solution to exist tends to 0 as $N \to \infty$.

Problem 3. Concentration and Spin Glass Free Energies (48 points)

This problem concerns p-spin models as in lecture, with Hamiltonian

$$H_N(\boldsymbol{\sigma}) = N^{-(p-1)/2} \sum_{1\le i_1,\ldots,i_p\le N} g_{i_1,\ldots,i_p}\boldsymbol{\sigma}_{i_1}\ldots\boldsymbol{\sigma}_{i_p} = N^{-(p-1)/2}\langle \boldsymbol{G}_N^{(p)}, \boldsymbol{\sigma}^{\otimes p}\rangle.$$

Here $\boldsymbol{G}_N^{(p)} \in \mathbb{R}^{N^p}$ has IID standard Gaussian coordinates $(g_{i_1,\ldots,i_p})_{1\le i_1,\ldots,i_p\le N}$. The Ising and spherical partition functions are defined by:

$$Z_N^{\mathrm{Is}}(\beta) = Z_N^{\mathrm{Is}}(H_N;\beta) = 2^{-N}\sum_{\boldsymbol{\sigma}\in\{\pm1\}^N} e^{\beta H_N(\boldsymbol{\sigma})};$$

$$Z_N^{\mathrm{sp}}(\beta) = \int_{\mathcal{S}_N} e^{\beta H_N(\boldsymbol{\sigma})}\mathrm{d}\boldsymbol{\sigma},$$

where the latter integral is with respect to the uniform distribution on the sphere

$$\mathcal{S}_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N \ : \ \|\boldsymbol{\sigma}\| = \sqrt{N}\}.$$

We set $F_N^{\mathrm{Is/sp}}(\beta) = \frac{1}{N}\log Z_N^{\mathrm{Is/sp}}(\beta)$. (Note the extra factor of $1/N$ compared to (1).)

(a) Show that for the Ising p-spin model, one has almost surely

$$\left(\frac{1}{N}\max_{\boldsymbol{\sigma}\in\{\pm1\}^N} H_N(\boldsymbol{\sigma})\right) - \frac{\log 2}{\beta} \le F_N^{\mathrm{Is}}(\beta)/\beta \le \frac{1}{N}\max_{\boldsymbol{\sigma}\in\{\pm1\}^N} H_N(\boldsymbol{\sigma}). \tag{4}$$

Explain why this can be interpreted as commutation of $N\to\infty$ and $\beta\to\infty$ limits.

(b) It follows from Lecture 2 (**see added Section 2.1 in notes for day 2, to be discussed at the start of Lecture 5**) that for some $C = C(p) > 0$ independent of N, we have

$$\max_{\boldsymbol{\sigma}\in\mathcal{S}_N} \|\nabla H_N(\boldsymbol{\sigma})\| \le C\sqrt{N} \tag{5}$$

with probability $1 - e^{-N}$. In particular, almost surely, this bound holds for all but finitely many N (by the Borel–Cantelli lemma). Prove that if all but finitely many $H_1, H_2, \ldots$ obey (5), then

$$\lim_{\beta\to\infty}\lim_{N\to\infty}\left|\frac{1}{N}\max_{\|\boldsymbol{\sigma}\|=\sqrt{N}} H_N(\boldsymbol{\sigma}) - F_N^{\mathrm{sp}}(\beta)/\beta\right| = 0.$$

This qualitatively matches (4). (Hint: consider a ball of radius $\delta\sqrt{N}$ around the maximizer $\boldsymbol{\sigma} \in \mathcal{S}_N$, where δ tends to 0 slowly with β.)

(c) Recall from class that both $\max_{\|\boldsymbol{\sigma}\|=\sqrt{N}} H_N(\boldsymbol{\sigma})/N$ and $F_N^{\mathrm{sp}}(\beta)$ concentrate sharply around their expectations. Using this and the previous part, prove that

$$\lim_{\beta\to\infty}\lim_{N\to\infty}\mathbb{E}\left|\frac{1}{N}\max_{\|\boldsymbol{\sigma}\|=\sqrt{N}} H_N(\boldsymbol{\sigma}) - F_N^{\mathrm{sp}}(\beta)/\beta\right| = 0.$$

(d) In the famously non-rigorous replica method from physics, one computes free energies using the observation

$$\frac{1}{N}\log Z_N = \lim_{\varepsilon\to 0}\frac{Z_N^{\varepsilon} - 1}{N\varepsilon}. \tag{6}$$

The method first finds an asymptotic formula as $N \to \infty$ for $\frac{1}{N}\log\mathbb{E}[Z_N^k]$ for integers $k \ge 1$, and then formally sends $k \downarrow 0$ in this formula. While the latter step seems impossible to justify directly, another potential issue is the inconsistency of the order of limits. Namely the formula (6) sends $\varepsilon \to 0$ for fixed N instead of sending $N \to \infty$ for fixed ε. Using the concentration of the free energy, prove that for any constant β,

$$\lim_{\varepsilon\to 0}\lim_{N\to\infty}\left|\frac{1}{N}\mathbb{E}\log Z_N(\beta) - \mathbb{E}\left[\frac{Z_N(\beta)^{\varepsilon/N} - 1}{\varepsilon}\right]\right| = 0.$$

In other words, the interchange of limits is justified if ε is rescaled by a factor of N. You may use either the Ising or spherical model here; the proof should not strongly depend on this choice. (Hint: the main surprise is the upper bound on $\mathbb{E}[Z_N(\beta)^{\varepsilon/N}]$. Here you will really need the sub-Gaussian tail bound for $F_N(\beta)$, not just concentration with exponentially good probability.)

Problem 4: Posterior of Tensor PCA (96 points)

Recall from class that in tensor PCA, one generates $\boldsymbol{x} \in \mathcal{S}_N$ uniformly at random, and observes the signal

$$\boldsymbol{T}_N = \boldsymbol{G}_N^{(p)} + \lambda N^{-(p-1)/2}\boldsymbol{x}^{\otimes p}.$$

Here $\boldsymbol{G}_N^{(p)} \in \mathbb{R}^{N^p}$ is as in the previous problem, and $\lambda > 0$ is a "signal strength" not depending on N. Hence $\boldsymbol{T}_N \in \mathbb{R}^{N^p}$ has entries

$$(\boldsymbol{T}_N)_{i_1,\ldots,i_p} = g_{i_1,\ldots,i_p} + \lambda N^{-(p-1)/2} x_{i_1}x_{i_2}\ldots x_{i_p}.$$

(a) Given $\boldsymbol{T}_N$, define[1]

$$\widetilde{H}_N(\boldsymbol{\sigma}) = N^{-(p-1)/2}\langle \boldsymbol{T}_N, \boldsymbol{\sigma}^{\otimes p}\rangle = N^{-(p-1)/2}\langle \boldsymbol{G}_N^{(p)}, \boldsymbol{\sigma}^{\otimes p}\rangle + \lambda N\left(\frac{\langle \boldsymbol{x}, \boldsymbol{\sigma}\rangle}{N}\right)^p$$
$$= H_N(\boldsymbol{\sigma}) + \lambda N\left(\frac{\langle \boldsymbol{x}, \boldsymbol{\sigma}\rangle}{N}\right)^p.$$

Show that the posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{T}_N$ is the Gibbs measure μ_λ for Hamiltonian $\widetilde{H}_N : \mathcal{S}_N \to \mathbb{R}$.

(b) Let

$$Band_{\boldsymbol{x}}(q) = \{\boldsymbol{\sigma} \in \mathcal{S}_N \;:\; \langle \boldsymbol{\sigma}, \boldsymbol{x}\rangle/N = q\}$$

and note that $Band_{\boldsymbol{x}}(q)$ has $(N-2)$-dimensional volume proportional to $(1-q^2)^{N/2}$ (up to lower order factors). Let $\mu_{\boldsymbol{x},q}$ denote the uniform distribution on this band. Explain why, **assuming** that the limit

$$F_\lambda(q) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}^{\boldsymbol{G}_N^{(p)}} \log \int e^{\lambda H_N(\boldsymbol{\sigma})} \mathrm{d}\mu_{\boldsymbol{x},q}(\boldsymbol{\sigma}) \tag{7}$$

exists for each $q \in (-1,1)$ and $\lambda \geq 0$, it is natural to expect the limiting free energy for tensor PCA is

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\log\int_{\mathcal{S}_N} e^{\lambda\widetilde{H}_N(\boldsymbol{\sigma})}\mathrm{d}\boldsymbol{\sigma} = \max_{q\in(-1,1)} F_\lambda(q) + q^p + \frac{\log(1-q^2)}{2}. \tag{8}$$

(c) Using continuity and concentration properties of H_N, prove that (8) indeed follows rigorously from the assumption that the limit $F_\lambda(q)$ exists. (Hint: approximate $\mathcal{S}_N$ by a large constant number of bands, and use the bound in (5) on $\sup_{\boldsymbol{x}\in\mathcal{S}_N} \|\nabla H_N(\boldsymbol{x})\|$ to justify this approximation.)

In the next problems, we will find the optimal mean-squared error in tensor PCA. You may **assume** below that the limit (7) defining $F_\lambda(q)$ exists for all (λ, q), that p is odd, and that the maximum in (8) is attained uniquely at $q = q_*(\lambda)$.

(d) Let $\boldsymbol{\sigma}$ be a sample from the posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{T}_N$. Show that $\langle \boldsymbol{\sigma}, \boldsymbol{x}\rangle/N$ converges in probability to q_* as $N \to \infty$.

(e) Let $\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)}$ be independent samples from the posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{T}_N$. Show that averaged over *all the randomness* (including uniform $\boldsymbol{x} \in \mathcal{S}_N$ and Gaussian $\boldsymbol{G}_N^{(p)}$), the pair $(\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)}) \in \mathcal{S}_N^2$ has the same distribution as $(\boldsymbol{\sigma}^{(1)}, \boldsymbol{x})$. (Hint: this is a very general property of posterior sampling.)

(f) Let $\boldsymbol{y} = \mathbb{E}[\boldsymbol{x} \mid \boldsymbol{T}_N]$ denote the posterior mean of $\boldsymbol{x}$ given $\boldsymbol{T}_N$. Show that

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\big[\|\boldsymbol{y}\|^2\big] = q_*; \qquad \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\big[\|\boldsymbol{x}-\boldsymbol{y}\|^2\big] = 1 - q_*.$$

The left-hand side defines the **asymptotically optimal mean-squared error** in tensor PCA. (Hint: let $\widehat{\boldsymbol{y}}$ be the average of K IID posterior samples $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(K)}$, for K large but independent of N. Argue that $\mathbb{E}[\|\widehat{\boldsymbol{y}} - \boldsymbol{y}\|^2]/N$ tends to 0 with K, and use the previous parts to study $\|\widehat{\boldsymbol{y}}\|^2$.)

[1] Here the inner product is the usual one in $\mathbb{R}^{N^p}$, defined by summing over p-tuples $(i_1, \ldots, i_p) \in \{1, 2, \ldots, N\}^p$.

Problem 5. Survey (4 points)

Rate each of the four 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.