Math 281C **2020 Spring Final** **June 9, 2020**

Name: ______________________ **PID:** ____________________

Do not turn the page until told to do so.

1. No calculators, tablets, phones, or other electronic devices are allowed during this exam.
2. Read each question carefully and answer each question completely.
3. Show all of your work. No credit will be given for unsupported answers, even if correct.
4. If you are unsure of what a question is asking for, **do not hesitate to ask an instructor or course assistant for clarification**.
5. This exam has 3 pages.

Question	*Points Available*	*Points Earned*
1	50	
2	50	
TOTAL	100	

1. [50 points] Let $\theta \in \mathbb{R}^p$ and define

$$f(\theta) = \mathbb{E}\{F(\theta; X)\} = \int_{\mathcal{X}} F(\theta; x) \mathrm{d}P(x),$$

where $F(\cdot; x)$ is convex in its first argument (in θ) for all $x \in \mathcal{X}$, and P is a probability distribution. We assume $F(\theta; \cdot)$ is integrable for all θ. Recall that a function h is convex if

$$h(t\theta + (1-t)\theta') \le th(\theta) + (1-t)h(\theta') \quad \text{for all} \quad \theta, \theta' \in \mathbb{R}^p, \quad t \in [0,1].$$

Let $\theta_0 \in \arg\min_\theta f(\theta)$, and assume that f satisfies the following ν-strong convexity property:

$$f(\theta) \ge f(\theta_0) + \frac{\nu}{2}\|\theta - \theta_0\|^2 \quad \text{for all } \theta \text{ satisfying } \|\theta - \theta_0\| \le \beta,$$

where $\beta > 0$ is some constant. We also assume that $F(\cdot; x)$ is L-Lipschitz with respect to the norm $\|\cdot\|$, the Euclidean norm in $\mathbb{R}^p$.

Let $X_1, \ldots, X_n$ be an iid sample from P, and define $f_n(\theta) = (1/n)\sum_{i=1}^n F(\theta; X_i)$. Let

$$\hat{\theta}_n \in \arg\min_\theta f_n(\theta).$$

(a) Show that for any convex function $h : \mathbb{R}^p \to \mathbb{R}$, if there is some $r > 0$ and a point θ_0 such that $h(\theta) > h(\theta_0)$ for all θ such that $\|\theta - \theta_0\| = r$, then $h(\theta') > h(\theta_0)$ for all θ' with $\|\theta' - \theta_0\| > r$.

(b) Show that f and f_n are convex.

(c) Show that θ_0 is unique.

(d) Let

$$\Delta(\theta, x) = \{F(\theta; x) - f(\theta)\} - \{F(\theta_0; x) - f(\theta_0)\}.$$

Show that $\Delta(\theta, X)$ (with $X \sim P$) is $4L^2\|\theta - \theta_0\|^2$-sub-Gaussian. [We say a random variable X with mean μ is ν^2-sub-Gaussian if $\log \mathbb{E}e^{\lambda(X-\mu)} \le \lambda^2\nu^2/2$ for all $\lambda \in \mathbb{R}$.]

(e) Show that for some constant $\sigma < \infty$, which may depend on the parameters of the problem (you should specify this dependence in your solution),

$$\mathbb{P}\left(\|\hat{\theta}_n - \theta_0\| \ge \sigma \cdot \frac{1+t}{\sqrt{n}}\right) \le Ce^{-t^2}$$

for all $t \ge 0$, where $C < \infty$ is a numerical constant. [Hint: The quantity $\Delta_n(\theta) := (1/n)\sum_{i=1}^n \Delta(\theta, X_i)$ may be helpful, as may be the bounded differences inequality.]

2. [50 points] In the phase retrieval problem, the goal is to recover a signal $\theta^* \in \mathbb{R}^p$ based on noisy observations of the magnitudes of inner products $\langle X_i, \theta^* \rangle$ with a sample of n vectors $X_1, \ldots, X_n \in \mathbb{R}^p$. In physical detectors, we observe a number of photons $Y_i \in \mathbb{N}$ (here $\mathbb{N}$ denotes the collection of all non-negative integers) that scale roughly with $\langle X_i, \theta^* \rangle^2$. This association is usually characterized via a Poisson regression model, that is, the distribution of Y_i given X_i is

$$Y_i | X_i \sim \text{Poisson}(\langle X_i, \theta^* \rangle^2).$$

Recall that $Y \sim \text{Poisson}(\lambda)$ if the probability mass function of Y is

$$p_\lambda(k) = \frac{e^{-\lambda} \lambda^k}{k!}, k = 0, 1, \ldots.$$

Consider the (conditional) expectation of negative log-likelihood

$$\varphi_i(\theta) = \mathbb{E}_{\theta^*}\{-\log p_{\langle X_i, \theta \rangle^2}(Y_i)\},$$

where the expectation is taken over $Y_i \sim \text{Poisson}(\langle X_i, \theta^* \rangle^2)$.

(a) Suppose that $Y \sim \text{Poisson}(\lambda_0)$ for some $\lambda_0 > 0$. Show that

$$\mathbb{E}\{-\log p_\lambda(Y)\} - \mathbb{E}\{-\log p_{\lambda_0}(Y)\} \geq \frac{1}{4} \min\left\{ |\lambda - \lambda_0|, \frac{(\lambda - \lambda_0)^2}{\lambda_0} \right\}.$$

(b) Let $g : \mathbb{R}^p \to \mathbb{R}$ be a twice-differentiable convex function and satisfy $\nabla^2 g(\theta) \succeq \lambda I_p$ (I_p is the $p \times p$ identity matrix) for all θ satisfying $\|\theta - \theta_0\| \leq c$. Show that

$$g(\theta) \geq g(\theta_0) + \nabla g(\theta_0)^\intercal (\theta - \theta_0) + \frac{\lambda}{2} \min\{\|\theta - \theta_0\|^2, c\|\theta - \theta_0\|\}.$$

(c) Show that

$$\varphi_i(\theta) - \varphi_i(\theta^*) \geq \frac{1}{4} \min\left\{ |\langle X_i, \theta - \theta^* \rangle \langle X_i, \theta + \theta^* \rangle|, \frac{|\langle X_i, \theta - \theta^* \rangle \langle X_i, \theta + \theta^* \rangle|^2}{\langle X_i, \theta^* \rangle^2} \right\}.$$

(d) Suppose that $X_i \in \mathbb{R}^p$ are random vectors satisfying

$$\mathbb{P}\big(|\langle X_i, v \rangle| \geq \epsilon \|v\|_2\big) \geq 1 - \epsilon \quad \text{and} \quad \mathbb{E}\langle X_i, \theta^* \rangle^2 \leq M^2 \|\theta^*\|_2^2$$

for all $\epsilon \geq 0$ and all vectors $v \in \mathbb{R}^d$. Show that for (numerical) constants c_0, c_1, for any $\delta \in (0, 1)$, if

$$\sqrt{\frac{p + \log(1/\delta)}{n}} \leq c_0,$$

then with probability at least $1 - \delta$,

$$\frac{1}{n} \sum_{i=1}^n \{\varphi_i(\theta) - \varphi_i(\theta^*)\} \geq c_1 \min\left\{ d(\theta, \theta^*) \cdot \max\{\|\theta\|_2, \|\theta^*\|_2\}, \frac{d^2(\theta, \theta^*)}{M^2} \right\}$$

holds simultaneously for all $\theta \in \mathbb{R}^p$, where $d(\theta, \theta^*) := \min_{s \in \{1, -1\}} \|\theta + s\theta^*\|_2$ is the distance (ignoring sign) between θ and θ^*.