This homework is due at **2pm, March 4, 2019**.

Problem 1

Let $f : \mathbb{R}^p \to \mathbb{R}$. Its *gradient* is a p-dimensional vector given by

$$\nabla f(x) := \left(\frac{\partial f}{\partial x^{(1)}}(x), \dots, \frac{\partial f}{\partial x^{(p)}}(x) \right), \quad \forall x \in \mathbb{R}^p,$$

where $x^{(i)}$ denotes the i-th entry of the vector x. Its *Hessian* is a matrix in $\mathbb{R}^{p\times p}$ given by

$$\left[\nabla^2 f(x)\right]^{(i,j)} := \frac{\partial^2 f}{\partial x^{(i)} \partial x^{(j)}}(x), \quad \forall x \in \mathbb{R}^p,$$

for all $1 \le i, j \le p$, where $\left[\nabla^2 f(x)\right]^{(i,j)}$ denotes the (i,j)-th entry of the matrix $\nabla^2 f(x)$.

Define

$$g(x) := \frac{1}{2} \langle x, Ax \rangle, \quad \forall x \in \mathbb{R}^p,$$

for some symmetric matrix $A \in \mathbb{R}^{p\times p}$.

1. (10 points) **Show that**

$$\nabla g(x) = Ax, \quad \forall x \in \mathbb{R}^p.$$

   Solution. Notice that

$$g(x) = \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} x^{(i)} A^{(i,j)} x^{(j)}.$$

   Then, we have

$$\frac{\partial g}{\partial x^{(k)}}(x) = \frac{1}{2} \left[\sum_{j=1}^{p} A^{(k,j)} x^{(j)} + \sum_{i=1}^{p} x^{(i)} A^{(i,k)} \right] = \frac{1}{2} \left[2 \sum_{j=1}^{p} A^{(k,j)} x^{(j)} \right] = (Ax)^{(k)}, \quad \forall 1 \le k \le p.$$

   The second last equality follows from symmetry of the matrix A.

2. (15 points) **Show that**

$$\nabla^2 g(x) = A, \quad \forall x \in \mathbb{R}^p.$$

   Solution. By the derivation above, we write

$$\frac{\partial^2 g}{\partial x^{(i)} \partial x^{(j)}}(x) = \frac{\partial}{\partial x^i} \left(\sum_{k=1}^{p} A^{(j,k)} x^{(k)} \right) = A^{(j,i)} = A^{(i,j)}.$$

   The last equality follows from symmetry of the matrix A.

Problem 2

Let $\mathscr{X} \subset \mathbb{R}^{p\times p}$ be the set of positive semi-definite matrices of unit trace.

1. (10 points) **Show that for each $X \in \mathscr{X}$, we can write**

$$X = \sum_{i=1}^{p} \lambda_i \Pi_i,$$

   where $\lambda_1, \dots, \lambda_p$ are non-negative real numbers summing up to one, and Π_i are rank-1 matrices of unit trace.

   Solution. This is simply the eigenvalue decomposition of the matrix X, where $\lambda_1, \dots, \lambda_p$ are the eigenvalues.

Problem 2 continued on the next page...

2. (15 points) Let $Y \in \mathbb{R}^{p \times p}$ which may not be in $\mathscr{X}$. Define $f(X) := \mathrm{Tr}(YX)$, the trace of YX. Define

$$f^\star = \min_X \left\{ f(X) \,\middle|\, X \in \mathscr{X} \right\}.$$

Show that there exists some rank-1 matrix $X^\star \in \mathscr{X}$, such that $f^\star = f(X^\star)$.

Solution. Let $X^\star$ be a minimizer of f on $\mathscr{X}$. Then, we have

$$X^\star = \sum_{i=1}^p \lambda_i \Pi_i^\star,$$

where λ_i's sum up to 1, and Π_i's are rank-1 matrices. Suppose that

$$f(\Pi_i) = \mathrm{Tr}(Y\Pi_i) > f^\star, \quad \forall 1 \le i \le p.$$

Then, we have

$$f^\star = \mathrm{Tr}(YX^\star) = \mathrm{Tr}\left(Y \sum_{i=1}^p \lambda_i \Pi_i \right) = \sum_{i=1}^p \lambda_i \mathrm{Tr}(Y\Pi_i) > f^\star,$$

a contradiction, showing that at least one of the Π_i's is a minimizer.

Problem 3

Let ξ be a non-negative real-valued random variable, and η be a real-valued random variable.

1. (10 points) **Show that**

$$\mathsf{P}(\xi \ge t) \le \frac{\mathsf{E}\,\xi}{t}, \quad \forall\, t > 0,$$

where $\mathsf{P}(\xi \ge t)$ denotes the probability that $\xi \ge t$, and $\mathsf{E}\,\xi$ denotes the expectation of ξ.

Solution. This is called *Markov's inequality*. Search for it on Wikipedia to find a proof.

2. (15 points) Suppose that η is a subgaussian random variable of parameter $\lambda > 0$, i.e.,

$$\mathsf{E} \exp\left(\lambda \eta^2\right) \le 2.$$

Show that

$$\mathsf{P}\left(|\eta| \ge t\right) \le 2 \exp\left(-\lambda t^2\right), \quad \forall\, t \ge 0.$$

Solution. By Markov's inequality, we write

$$\mathsf{P}\left(|\eta| \ge t\right) = \mathsf{P}\left(\eta^2 \ge t^2\right) = \mathsf{P}\left(\mathrm{e}^{\lambda\eta^2} \ge \mathrm{e}^{\lambda t^2}\right) \le \frac{\mathsf{E}\,\mathrm{e}^{\lambda t^2}}{\mathrm{e}^{\lambda t^2}} = 2\mathrm{e}^{-\lambda t^2}.$$

Problem 4

Let $\xi_1, \dots, \xi_n$ be a sequence of independent Bernoulli random variable of expectation $\theta^\natural \in [0,1]$, i.e.,

$$\mathsf{P}(\xi_i = 1) = 1 - \mathsf{P}(\xi_i = 0) = \theta^\natural, \quad 1 \le i \le n.$$

The maximum-likelihood estimator of θ is given by any minimizer of the function

$$f_n(\theta) := -\frac{1}{n} \sum_{i=1}^n \left[\xi_i \log\theta + (1 - \xi_i)\log(1-\theta) \right],$$

on the interval $[0,1]$.

Problem 4 continued on the next page...

1. (10 points) **Show that indeed, the maximum-likelihood estimator, which we denote by $\hat{\theta}_n$, is given by**

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n \xi_i.$$

Solution. We write

$$\frac{\mathrm{d}f_n}{\mathrm{d}\theta}(\theta) = -\frac{1}{n}\sum_{i=1}^n \left[\frac{\xi_i}{\theta} - \frac{1-\xi_i}{1-\theta} \right] = -\frac{1}{n}\sum_{i=1}^n \left[\frac{\xi_i - \theta}{\theta(1-\theta)} \right],$$

showing that

$$\frac{\mathrm{d}f_n}{\mathrm{d}\theta}(\hat{\theta}_n) = 0.$$

Moreover, we have

$$\frac{\mathrm{d}^2 f_n}{\mathrm{d}\theta^2}(\theta) = -\frac{1}{n}\sum_{i=1}^n \left[\frac{-\xi_i}{\theta^2} - \frac{2-\theta-\xi_i}{(1-\theta)^2} \right] \geq 0.$$

Then, by Taylor's theorem, we obtain

$$\begin{aligned} f_n(\theta) &= f_n(\hat{\theta}_n) + \frac{\mathrm{d}f_n}{\mathrm{d}\theta}(\hat{\theta}_n)(\theta - \hat{\theta}_n) + \int_0^1 \int_0^t \frac{\mathrm{d}^2 f_n}{\mathrm{d}\theta^2}(\hat{\theta}_n + \tau(\theta - \hat{\theta}_n))(\theta - \hat{\theta}_n)^2 \,\mathrm{d}\tau \,\mathrm{d}t \\ &\geq f_n(\hat{\theta}_n), \quad \forall \theta \in \,]0,1[, \end{aligned}$$

showing that $\hat{\theta}_n$ is a minimizer of f_n on $]0,1[$. Notice that f_n is not defined (or f_n equals $+\infty$) at the endpoints of the interval $[0,1]$.

2. (15 points) **Show that as $n \to +\infty$, it holds that $\hat{\theta}_n$ converges to $\theta^\natural$ in probability, i.e.,**

$$\lim_{n\to+\infty} \mathsf{P}\left(\left| \hat{\theta}_n - \theta^\natural \right| \geq t \right) = 0, \quad \forall t > 0.$$

Do not simply cite the law of large numbers.

Solution. Define

$$\eta_i := \xi_i - \mathsf{E}\xi_i = \xi_i - \theta^\natural, \quad \forall 1 \leq i \leq n.$$

Notice that $\eta_i \in [-2,2]$. Then, we have

$$\mathsf{E}\eta_i^2 \leq 4, \quad \forall 1 \leq i \leq n,$$

and hence

$$\mathsf{E}\left[\frac{1}{n}\sum_{i=1}^n \eta_i \right]^2 = \frac{1}{n^2}\left[\sum_{i=1}^n \mathsf{E}\eta_i^2 + \sum_{i\neq j} \mathsf{E}\left(\eta_i\eta_j\right) \right] = \frac{1}{n^2}\left[\sum_{i=1}^n \mathsf{E}\eta_i^2 + \sum_{i\neq j} \mathsf{E}\eta_i \mathsf{E}\eta_j \right] = \frac{1}{n^2}\sum_{i=1}^n \eta_i^2 \leq \frac{4}{n}.$$

By Markov's ienquality, we write

$$\begin{aligned} \mathsf{P}\left(\left| \hat{\theta}_n - \theta^\natural \right| \geq t \right) &= \mathsf{P}\left(\left| \frac{1}{n}\sum_{i=1}^n \eta_i \right| \geq t \right) \\ &= \mathsf{P}\left(\left| \frac{1}{n}\sum_{i=1}^n \eta_i \right|^2 \geq t^2 \right) \\ &\leq \frac{\mathsf{E}\left[\frac{1}{n}\sum_{i=1}^n \eta_i \right]^2}{t^2} \\ &\leq \frac{4}{t^2 n}, \quad \forall t > 0. \end{aligned}$$

The desired equality follows.