STAT 280: OPTIMIZATION
SPRING 2022
PROBLEM SET 3

In the following, we write $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{R}_{++} = (0, \infty)$. So $\mathbb{R}^n_+ = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_i \geq 0\}$ and $\mathbb{R}^n_{++} = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_i > 0\}$.

As usual, no taking partial derivative, no quoting nonsense like Jacobi formula; everything here requires nothing more than definition and chain rule. You may freely quote results from previous homework and lecture notes.

1. Let $A_0, A_1, \dots, A_n \in \mathbb{S}^m$ and $\Omega = \{\mathbf{x} \in \mathbb{R}^n : A_0 + x_1 A_1 + \cdots + x_n A_n \succ 0\}$.

(a) Find the gradient of $f : \Omega \to \mathbb{R}$,

$$f(\mathbf{x}) = \det(A_0 + x_1 A_1 + \cdots + x_n A_n).$$

(b) Find the Hessian of $f : \Omega \to \mathbb{R}$,

$$f(\mathbf{x}) = \log\det(A_0 + x_1 A_1 + \cdots + x_n A_n).$$

Recall that we have already found the gradient of this function in the lectures.

(c) Find the gradient and Hessian of $f : \Omega \to \mathbb{R}$,

$$f(\mathbf{x}) = \operatorname{tr}\big((A_0 + x_1 A_1 + \cdots + x_n A_n)^{-1}\big).$$

(d) Find the gradient of $f : \Omega \to \mathbb{R}$,

$$f(\mathbf{x}) = (B\mathbf{x} + \mathbf{c})^{\mathsf{T}}(A_0 + x_1 A_1 + \cdots + x_n A_n)^{-1}(B\mathbf{x} + \mathbf{c})$$

where $B \in \mathbb{R}^{m \times n}$ and $\mathbf{c} \in \mathbb{R}^m$.

2. Decide which of the following sets are convex. Prove your answers.

$$\begin{aligned} \mathrm{GL}(n) &= \{X \in \mathbb{R}^{n \times n} : \det(X) \neq 0\}, \\ \mathbb{S}^n_{++} &= \{X \in \mathbb{S}^n : X \succ 0\}, \\ \Omega_1 &= \{\mathbf{x} \in \mathbb{R}^n : A_0 + x_1 A_1 + \cdots + x_n A_n \succ 0\}, \\ \Omega_2 &= \{X \in \mathbb{R}^{m \times n} : X^{\mathsf{T}} A X + B^{\mathsf{T}} X + X^{\mathsf{T}} B + C \succ 0\}, \end{aligned}$$

where Ω_1 is as defined in Problem **1** and Ω_2 is as defined in Homework **2**, Problem **4**(f).

3. Compute the Hessians of the following functions and decide if they are convex, concave, or neither on their respective domains.

(a) $f : \mathbb{R} \times \mathbb{R}_{++} \to \mathbb{R}$ defined by

$$f(x, y) = \frac{x^2}{y}.$$

(*Hint*: Write $\nabla^2 f(x, y)$ as a rank-one matrix).

(b) $f : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$ defined by

$$f(\mathbf{x}, y) = \frac{\mathbf{x}^{\mathsf{T}}\mathbf{x}}{y}.$$

Date: May 1, 2022 (Version 1.0); due: May 14, 2022.

(c) $f : \mathbb{R}^n \times \mathbb{S}^n_{++} \to \mathbb{R}$ defined by

$$f(\mathbf{x}, Y) = \mathbf{x}^\mathsf{T} Y^{-1} \mathbf{x}.$$

(d) $f : \mathbb{S}^n_{++} \to \mathbb{R}$ defined by

$$f(X) = \log\det(X) - \log\operatorname{tr}(X).$$

(e) $f : \Omega \to \mathbb{R}$ defined by

$$f(\mathbf{x}) = \frac{\|A\mathbf{x} + \mathbf{b}\|^2}{\mathbf{c}^\mathsf{T}\mathbf{x} + d}$$

where $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$, $d \in \mathbb{R}$, and $\Omega = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^\mathsf{T}\mathbf{x} + d > 0\}$.

4. (a) Find the Hessian of the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(\mathbf{x}) = \log(e^{x_1} + \cdots + e^{x_n})$. Show that for any $\mathbf{v} \in \mathbb{R}^n$,

$$\mathbf{v}^\mathsf{T}\nabla^2 f(\mathbf{x})\mathbf{v} = \frac{1}{(e^{x_1} + \cdots + e^{x_n})^2}\left[\left(\sum\nolimits_{i=1}^n e^{x_i}\right)\left(\sum\nolimits_{i=1}^n v_i^2 e^{x_i}\right) - \left(\sum\nolimits_{i=1}^n v_i e^{x_i}\right)^2\right].$$

Hence or otherwise, deduce that f is a convex function.

(b) Find the Hessian of the function $g : \mathbb{R}^n_{++} \to \mathbb{R}$ defined by $g(\mathbf{x}) = (x_1 \cdots x_n)^{1/n}$. Show that for any $\mathbf{v} \in \mathbb{R}^n$,

$$\mathbf{v}^\mathsf{T}\nabla^2 g(\mathbf{x})\mathbf{v} = -\frac{g(\mathbf{x})}{n^2}\left[n\sum\nolimits_{i=1}^n \frac{v_i^2}{x_i^2} - \left(\sum\nolimits_{i=1}^n \frac{v_i}{x_i}\right)^2\right]$$

Hence or otherwise, deduce that g is a concave function.

(c) Find the Hessian of the function $h : \mathbb{R}^n_{++} \to \mathbb{R}$ defined by

$$h(\mathbf{x}) = \frac{1}{1/x_1 + \cdots + 1/x_n}.$$

By emulating what we did in the previous two parts or otherwise, decide if h is convex, concave, or neither on $\mathbb{R}^n_{++}$.

(d) Find the Hessian of the function $\varphi : \mathbb{R}^n_{++} \to \mathbb{R}$ defined by $\varphi(\mathbf{x}) = \log h(\mathbf{x})$. Decide if φ is convex, concave, or neither on $\mathbb{R}^n_{++}$.

5. (a) Show that the negative log function $-\log : \mathbb{R}_{++} \to \mathbb{R}$ is strictly convex, i.e.,

$$\log(tx + (1-t)y) > t\log x + (1-t)\log y$$

for any $x, y \in \mathbb{R}_{++}$ and any $t \in (0, 1)$.

(b) Prove the generalized arithmetic-geometric mean inequality

$$a^t b^{1-t} \le ta + (1-t)b$$

for any $a, b \in \mathbb{R}_+$ and $t \in [0, 1]$ (note that $t = 1/2$ gives us the usual arithmetic-geometric mean inequality). Deduce the Hölder inequality: for $p > 1$ and $1/p + 1/q = 1$,

$$\sum_{i=1}^n |x_i y_i| \le \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}\left(\sum_{i=1}^n |y_i|^q\right)^{1/q}$$

for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (note that $p = q = 2$ gives us the Cauchy–Schwartz inequality).

(c) Show that $(\sin\theta)^{\sin\theta} < (\cos\theta)^{\cos\theta}$ for all $\theta \in (0, \pi/4)$.