Computational Methods of Optimization
First Midterm(Sep 9 , 2021)

Time: 60 minutes

Instructions

- Answer all questions
- See upload instructions in the form

Question:	1	2	3	4	5	6	Total
Points:	5	5	10	5	10	10	45
Score:							

In the following, assume that f is a $\mathcal{C}^1$ function defined from $\mathbb{R}^d \to \mathbb{R}$ unless otherwise mentioned. Let $\mathbf{I} = [e_1, \ldots, e_d]$ be a $d \times d$ matrix with e_j be the jth column. Also $\mathbf{x} = [x_1, x_2, \ldots, x_d]^\top \in \mathbb{R}^d$ and $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\top \mathbf{x}}$. Set of real symmetric $d \times d$ matrices will be denoted by $\mathcal{S}_d$. $[n]$ will denote the set $\{1, 2, \ldots, n\}$

1. Answer True or False
   (a) (1 point) The function $f(x) = x^3, x \in \mathbb{R}$ is convex **F**
   (b) (1 point) The function $f(x) = -\ln x, x > 0$ is convex **T**
   (c) (1 point) The function $f(x) = -x + 1, x \in \mathbb{R}$ is convex **T**
   (d) (1 point) The set $\{(x,t) | e^{-x} \leq t, t \geq 0, x \in \mathbb{R}\}$ is convex **T**
   (e) (1 point) The set $\{(x,t) | \mathbf{x}^\top A \mathbf{x} \leq t, A \in \mathcal{S}_d, trace(A) = 0, t \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^d\}$ is convex **F**.

2. (5 points) Is the set $S = \{(x,t) | -\ln x \leq t, x > 0, t \in \mathbb{R}\}$ convex? Give reasons.

Solution: Let $S = \{(x,t) | -\ln x \leq t, x > 0, t \in \mathbb{R}\}$ If $(x_1, t_1) \in S$ and $(x_2, t_2) \in S$ then we need to prove or disprove that $(x,t)^\top = \lambda_1 (x_1, t_1)^\top + \lambda_2 (x_2, t_2)^\top \in S$ for $\lambda_1, \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1$. From the statement of the question and noting that $-\ln x$ is a convex function the following holds $-\ln x_1 \leq t_1, -\ln x_2 \leq t_2$.

$$-\ln(\lambda_1 x_1 + \lambda_2 x_2) \leq -\lambda_1 \ln(x_1) - \lambda_2 \ln(x_2) \leq \lambda_1 t_1 + \lambda_2 t_2 = t$$

Hence $-\ln x \leq t$ and hence the set is convex.

3. Consider the function, $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(\mathbf{x}) = x_1^6 + x_2^6 - 96 x_1 x_2$$

where $\mathbf{x} = [x_1, x_2]^\top$. Let $f^* = f(\mathbf{x}^*) = min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x})$

(a) (2 points) find all critical points

Solution: $\nabla f(\mathbf{x}) = \begin{bmatrix} 6x_1^5 - 96x_2 \\ 6x_2^5 - 96x_1 \end{bmatrix}$
From direct substitution $\nabla f(\mathbf{x}) = 0$ occurs at $\mathbf{x} = [0,0]^\top, [2,2]^\top, [-2,-2]^\top$

(b) (2 points) find f^*

Solution: $f^* = -256$

(c) (2 points) find $\mathbf{x}^*$ (In case there are more than one you need to find all the points)

Solution: $\mathbf{x} = [2,2]^\top, [-2,-2]^\top$

(d) (4 points) Justify your answers about $f^*, \mathbf{x}^*$.

Solution: The function f is corecieve. Observe that $f(\mathbf{x}) = (x_1^6 + x_2^6)\left(1 - \frac{96 x_1 x_2}{x_1^6 + x_2^6}\right)$ As $\|\mathbf{x}\| \to \infty$ the term $\frac{96 x_1 x_2}{x_1^6 + x_2^6}$ goes to zero, and hence f tends to infinity. Since it is also $\mathcal{C}^1$, the optimum must lie at one or more of the critical points. Direct substitution gives $f([0,0]^\top) = 0, f([2,2]^\top) = f([-2,-2]^\top) = -256$

4. Consider minimization of the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as follows

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - b^\top \mathbf{x} + c$$

$$A = \begin{bmatrix} 1 & \frac{1}{2} \\ -\frac{1}{2} & 2 \end{bmatrix}, b = [1, 2], c = 1$$

Answer true or false.

(a) (1 point) The Hessian matrix at any $\mathbf{x}$ is A <u>**F**</u>

(b) (1 point) f is in $\mathcal{C}^2$ <u>**T**</u>

(c) (1 point) f have global minima. <u>**T**</u>

(d) (1 point) f is convex <u>**T**</u>

(e) (1 point) The set $\{\mathbf{x}|f(\mathbf{x}) \leq f(0)\}$ is not convex <u>**F**</u>

5. Consider minimizing the function, $f : \mathbb{R}^3 \to \mathbb{R}$

$$f(\mathbf{x}) = \frac{1}{2}x_1^2 + x_2^2 + \frac{1}{2}x_3^2 - x_1 - 2x_2 - x_3 - 1$$

(a) (4 points) Find smallest value of L such that

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x})\|^2$$

holds for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$

Solution: By Taylor's Theorem

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^\top Q(\mathbf{y} - \mathbf{x})$$

The Hessian, Q is diagonal, with $Q_{11} = Q_{33} = 1, Q_{22} = 2$. Since $\mathbf{u}^\top Q\mathbf{u} \leq \lambda_{max}\|\mathbf{u}\|^2$ for any $\mathbf{u} \in \mathbb{R}^3$ and using $\lambda_{max} = 2$ it follows that

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) + \frac{2}{2}\|\mathbf{y} - \mathbf{x})\|^2$$

(b) (6 points) Let L be defined as above. Consider a scheme of the form $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha\nabla f(\mathbf{x}^{(k)})$. Using the above inequality find the range of positive values of α so thatr $f(\mathbf{x}^{(k+1)}) \leq f(\mathbf{x}^{(k)})$. The lower bound of the range is <u>0</u> and upper bound on the range <u>1</u>. The value of α^*, the α value which gives maximum decrease is given by <u>$\frac{1}{2}$</u>
Justify your answers

Solution: Substituting $\mathbf{x}^{(k+1)}$ in the previous question $f(\mathbf{x}^{(k+1)}) \leq f(\mathbf{x}^{(k)}) - (\alpha - \alpha^2\frac{L}{2})\|\nabla f(\mathbf{x}^{(k)})\|^2$. Thus $f(\mathbf{x}^{(k+1)}) \leq f(\mathbf{x}^{(k)})$ whenever $0 \leq \alpha \leq \frac{2}{L} = 1$. The value α^* is obtained so that $\alpha - \alpha^2\frac{L}{2}$ is maximized.

6. Let f ve defined as in the previous question. Suppose you started from $\mathbf{x}^{(0)} = [0, 0, 0]^\top$. Consider implementing the steepest descent procedure with exact line search

(a) (2 points) What is the gradient of f in the first iteration?

Solution:

$$\nabla f(\mathbf{x}^{(0)} = [-1, -2, -1]^\top$$

(b) (4 points) What is the stepsize in the first iteration

Solution:

$$g_0 = [-1, -2, -1]^\top$$

$$\alpha = \frac{\|g_0\|^2}{g_0^\top Q g_0} = \frac{1 + 2^2 + 1}{1 + 2.2^2 + 1} = 0.6$$

(c) (4 points) In how many iterations, T, we will reach a point $\mathbf{x}^{(T)}$ such that $E(x^{(T)}) \leq 10^{-3} E(x^{(0)})$.

Solution: In exact line search $E(x^{(k+1)})[\leq \rho^2 E(x^{(k)})$ holds for all $k \geq 0$ where $\rho = \frac{r-1}{r+1} = \frac{1}{3}$ and r is the ratio of the largest and the smallest eigenvalue of Q. Consequently $E(x^{(T)})[\leq \rho^{2T} E(x^{(0)})$ and thus $T \geq \lceil \frac{\ln 10^{-3}}{2 \ln \frac{1}{3}} \rceil = 4$ is needed for $\mathbf{x}^{(T)}$ to achieve the desired accuracy.