# Computational Methods of Optimization
# First Midterm(7th Dec, 2020)

Time: 60 minutes

## Instructions

- Answer all questions
- See upload instructions in the form

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 $j$th 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. (10 points) Please indicate True(T) or False(F) in the space given after each question. All questions carry equal marks
   (a) Let $a < b$ where $a, b \in \mathbb{R}$ and $h : [a, b] \to \mathbb{R}$ be differentiable and satisfies $h(a) = h(b)$. Then $h$ has a critical point in $(a, b)$. **T**
   (b) Suppose the function defined in the previous question satisfies $|h(x) - h(y)| \leq 1|x - y|$ for all $x, y \in (a, b)$. There could exist a point in $(a, b)$ such that $h'(x) \geq 2$, where $h'(x)$ is derivative of $h$ at $x$. **F**
   (c) If $f$ is a coercive function then the global mimimum must lie at one of the critical points. Recall that a critical point is a point, $\mathbf{x}$, such that $\nabla f(\mathbf{x}) = 0$. **T**
   (d) Consider $g : \mathbb{R} \to \mathbb{R}, g(u) = u^2 - \frac{1}{3}u^3$.The function has a global minimum. **F**
   (e) The local maximum of $g$(defined in the previous question) is at $u = 0$. **F**

2. Let $f : S \subset \mathbb{R}^d \to \mathbb{R} \in \mathcal{C}^2$ function. Let $H(\mathbf{x})$ be the Hessian of $f$ with eigenvectors denoted by columns of $U \in \mathbb{R}^{d \times d}$
   (a) (2 points) Consider $S = \{\mathbf{x} | \mathbf{x} = \mathbf{x}^* + \mathbf{U}\mathbf{v}, \mathbf{v} \in \mathbb{R}^d\}$. $\bar{S}$, the complement of the set $S$, is not empty. True or False. Justify with reasons.

   **Solution:** False. The eigenvectors of $H(\mathbf{x}^*)$ form a basis of $\mathbb{R}^d$ and hence $S = \mathbb{R}^d$. Thus $\bar{S}$ is empty.

   (b) (4 points) The Hessian, $H(\mathbf{x}^*)$ at a stationary point $\mathbf{x}^*$ has one eigenvalue 0 and the rest positive. For any $\mathbf{x} \in S$ with $\mathbf{v} \neq 0$ $f(\mathbf{x}) > f(\mathbf{x}^*)$ is true. Prove or disprove

   **Solution:** Using second order Taylor expansion and noting that $\mathbf{x}^*$ is a critical point

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top H(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + o(\|\mathbf{x} - \mathbf{x}^*\|^2)$$

   we have $f(\mathbf{x}) - f(\mathbf{x}^*) = \sum_{i=1}^d \lambda_i v_i^2 + o(\|\mathbf{v}\|^2)$. $\lambda_i$ is an eigenvalue of $H$ and the last term is true because $\|\mathbf{x} - \mathbf{x}^*\| = \|\mathbf{v}\|$. For arbitrary $\mathbf{v}$, $f(\mathbf{x}) < f(\mathbf{x}^*)$ as the last term can be of any sign. Hence disproved.

   (c) (4 points) Suppose $H(\mathbf{x}^*)$ has negative and positive eigenvalues. Construct two distinct points, $\mathbf{x}^1$ and $\mathbf{x}^2$ in terms of $\mathbf{U}$ so that

$$f(\mathbf{x}^1) < f(\mathbf{x}^*) < f(\mathbf{x}^2)$$

   **Solution:** There exists a $\delta > 0$ the small oh term can be neglected. Let $j$ and $k$ be such that $\lambda_j < 0 < \lambda_k$. The points $\mathbf{x}^1 = \mathbf{x}^* + tu_j$ and $\mathbf{x}^2 = \mathbf{x}^* + tu_k$ for any non-zero $|t| \leq \delta$ satisfy

$$f(\mathbf{x}^1) = f(\mathbf{x}^*) + t^2\lambda_j < f(\mathbf{x}^*), \text{ and } f(\mathbf{x}^2) = f(\mathbf{x}^*) + t^2\lambda_k > f(\mathbf{x}^*)$$

   and hence they are desired points.

3. Let $f(x) = \sqrt{1 + x^2}, x \in \mathbb{R}$.

(a) (2 points) Find the $x^*$, global minimum of the problem

**Solution:**

$$f(x) \geq 1 = f(0), \text{ for all } x \in \mathbb{R}$$

(b) (4 points) Let $x^{(k)}$ be the output of the kth iteration of Newton's method applied to $f(x)$. Find a function $g$ such that

$$|x^{(k+1)} - x^*| \leq |g(x^{(k)} - x^*)|$$

**Solution:** Substituting $f'(x) = \frac{x}{f(x)}, f"(x) = \frac{1}{(1+x^2)^{3/2}}$ in Newton's method we obtain $g(z) = z - \frac{f'(z)}{f"(z)} = -z^3$.

(c) (4 points) Using the above relationship find largest $a$ so that the Newton's method is most effective for any $x^{(0)} \in (x^* - a, x^* + a)$.

**Solution:** $|x^{(0)}| \geq 1$ the method diverges otherwise it converges. Hence $a = 1$.

4. Consider minimizing the function

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

over $\mathbf{x} \in \mathbb{R}^d$ with $Q \in \mathcal{S}_d^+, \mathbf{b} \in \mathbb{R}^d$ using the steepest descent iterates,

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$$

executed with exact step-size selection strategy. Let $f^* = f(\mathbf{x}^*)$ be the global minimum. Let $g(\mathbf{y}) = f(A\mathbf{x})$ where $A$ is a $d \times d$ natrix. Consider applying steepest descent procedure to $g$ i.e.

$$\mathbf{y}^{(k+1)} = \mathbf{y}^{(k)} - \alpha_k \nabla g(\mathbf{y}^{(k)})$$

and $g^*$ is the global minimum of $g$ attained at $\mathbf{y}^*$.

(a) (4 points) State the Hessian of $f$ and $g$. Derive the relationship between $\mathbf{x}^*$ and $\mathbf{y}^*$

**Solution:**

$$g(\mathbf{y}) = \frac{1}{2}\mathbf{y}^\top A^\top QA\mathbf{y} - b^\top A\mathbf{y}$$

Hence Hessian of $g$ is $A^\top QA$ but Hessian of $f$ is Q. Note that $\nabla g(\mathbf{y}^*) = A^\top QA\mathbf{y}^* - A^\top b = 0$ which is same as $A\mathbf{y}^* = \mathbf{x}^*$.

(b) (3 points) What is the convergence rate of the steepest descent procedure for $g$?

**Solution:** Let $\kappa = \frac{\mu_1}{\mu_d}$ be the condition number of $A^\top QA$ where $\mu_1$ and $\mu_d$ are the largest and the smallest eigenvalues. The rate for $g$ is $\left(\frac{\kappa-1}{\kappa+1}\right)^2$.

(c) (3 points) What is the best value of $A$?

**Solution:** The best value of $A$ is obtained when $A^\top QA = \mathbf{I}$. This is attained at $A^\top A = Q^{-1}$.

5. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a $C^{(2)}$ function. Consider the Quasi-newton update

**a.** $s^{(k)} = -G^{(k)}\nabla f(\mathbf{x}^{(k)})$'

**b.** $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k s^{(k)}$

**c.** $G^{(k+1)} = G^{(k)} + A^{(k)}EA^{(k)\top}$

where $\gamma_k = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$, $\delta_k = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$, and $A^{(k)} = [\delta_k \quad G^{(k)}\gamma^{(k)}]$ and $E$ is a matrix.

(a) (1 point) What is your name

**Solution:**

(b) (4 points) Find $E$ corresponding to DFP updates.

**Solution:** DFP Update is

$$G^{(k+1)} = G^{(k)} + \frac{1}{\delta_k^\top \gamma_k}\delta_k\delta_k^\top - \frac{1}{\gamma_k G^{(k)}\gamma_k}G^{(k)}\gamma_k\gamma_k^\top G^{(k)}$$

Hence $E = \begin{bmatrix} c_1 & 0 \\ 0 & -c_2 \end{bmatrix}$ where $c_1 = \frac{1}{\delta_k^\top \gamma_k}$ and $c_2 = \frac{1}{\gamma_k^\top G^{(k)}\gamma_k}$

(c) (5 points) Show that for exact line search DFP updates yield positive definite matrices.

**Solution:** The DFP update

$$G^{(k+1)} = G^{(k)} + \frac{1}{\delta_k^\top \gamma_k}\delta_k\delta_k^\top - \frac{1}{\gamma_k^\top G^{(k)}\gamma_k}G^{(k)}\gamma_k\gamma_k^\top G^{(k)}$$

*The proof for* $G^{(k)} - \frac{1}{\gamma_k G^{(k)}\gamma_k}G^{(k)}\gamma_k\gamma_k^\top G^{(k)}$ *was done in class* (**2 points**). We need to show that $\delta_k^\top \gamma_k > 0$. For exact line search

$$\nabla f(\mathbf{x}^{(k+1)})^\top \delta_k = 0$$

Thus $\delta_k^\top \gamma_k = -\nabla f(\mathbf{x}^{(k)})^\top \delta_k \geq 0$ as $\delta_k = \alpha_k s_k$ where $s_k$ is a descent direction.