E0 230 : Computational Methods of Optimization
Assignment 1 (Due: 29^{th} November)

Instructions:

- Attempt all questions
- For Numerical Answer questions, insert only the digits **with no leading or trailing spaces** in the MS Teams Form provided
- You can submit the MS Teams Form **only once**. No further changes can be made, so click on *Submit* wisely.
- Late Submissions will be penalised.

1. Consider a matrix $B \in \mathbb{R}^{m\times n}$. Assume $\sigma_1 > \sigma_2 > \cdots > \sigma_n$. Consider the following claims. Select **all** of the statements that hold.

   A. $\boldsymbol{\sigma_1 \leq \|B\|_F}$

   B. $\boldsymbol{\sigma_{k+1} \leq \frac{\|B\|_F}{\sqrt{k+1}}}$

   C. $\sigma_{k+1} \leq \frac{\|B\|_F}{k+1}$

Solution: Obviously, $\sigma_1 \leq \sqrt{\sum_i \sigma_i^2}$. So A is true. From Tutorial 1, we saw that

$$\underset{\hat{B}\, of\, rank\, k+1}{\arg\min} \|B - \hat{B}\|_F = B_{k+1} = \sum_{i\leq k+1} \sigma_i v_i u_i^T$$

Now, we know that

$$\begin{aligned} \|B_{k+1}\| &= \sqrt{\sigma_1^2 + \cdots + \sigma_{k+1}^2} \\ &\geq \sqrt{\sigma_{k+1}^2(k+1)} \\ &\geq \sqrt{k+1}\sigma_{k+1} \end{aligned}$$

But $\|B_{k+1}\|_F \leq \|B\|_F \Rightarrow \|B\|_F \geq \sqrt{k+1}\sigma_{k+1}$. Thus B is true.

2. Suppose a quadratic function(x^TQx) is expanded as:

$$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + 5x_3^2 + 2\varepsilon x_1x_2 - 2x_1x_3 + 4x_2x_3$$

   Determine the range of values ε can take so that $Q \succ 0$.

   A. $\varepsilon \in (-\infty, 0) \cap (\frac{4}{5}, +\infty)$ (i.e. No $Q \succ 0$ is possible for any ε)

   B. $\varepsilon \in (-1, 1)$

   C. $\boldsymbol{\varepsilon \in (-\frac{4}{5}, 0)}$

   D. $\varepsilon \in (-\infty, +\infty)$ (i.e. $Q \succ 0$ for all values of ε)

Solution: We represent $f(x)$ in quadratic form x^TQx where

$$Q = \begin{bmatrix} 1 & \varepsilon & -1 \\ \varepsilon & 1 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

To have $Q \succ 0$, all the leading principal minors of Q must be positive (*Sylvester's criterion*). The leading principal minors of Q are $\Delta_1 = 1, \Delta_2 = 1 - \varepsilon^2, and \Delta_3 = -5\varepsilon^2 - 4\varepsilon$. Solving inequalities $\Delta_1 > 0, \Delta_2 > 0$, and $\Delta_3 > 0$, we get $\varepsilon \in (-\frac{4}{5}, 0)$.

3. Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ be continuous. It is known that if $f(x)$ is coercive and first partial derivatives of $f(x)$ exist on all of $\mathbb{R}^n$, then global minimizers of $f(x)$ exist and is within the critical points of $f(x)$.

   Given $h(x, y) = x^4 - 4xy + y^4$. We wish to find the global minimum by locating the local minima (by computing the gradient and the hessian of h). Then,

   (a) h has global minimum within the local minima because:

A. Hessian is positive everywhere.

B. h is coercive.

C. h is bounded from below.

D. The global minimum does not exist within the local minima.

(b) Choose **all** of the statements that are true for h from the following choices:

A. $(0,0)$ is the only critical point and a local minimum.

B. $(0,0),(1,1),(-1,-1)$ are the complete set of real critical points.

C. $(0,0),(-1,-1)$ are local minimum.

D. Hessian at $(1,1),(-1,-1)$ are positive.

(c) The global minimum of h is **-2**

Solution: It is easy to see by taking derivative and checking the Hessian that the local minimum of the function is indeed at -2. We now prove h is coercive to justify that -2 is also the global minimum. Note that $h(x,y)$ can be written as,

$$h(x,y) = x^4 + y^4 \left(1 - \frac{4xy}{x^4+y^4}\right)$$

As $\|(x,y)\| = \sqrt{x^2+y^2} \to \infty$, the term $\frac{4xy}{x^4+y^4} \to 0$. Hence,

$$\lim_{\|(x,y)\|\to\infty} h(x,y) = \lim_{\|(x,y)\|\to\infty} (x^4+y^4) = +\infty$$

. Thus, h is coercive.

4. Consider the following optimization problem.

$$x^* = \arg\min \frac{1}{2}x^T Q x - b^T x \tag{1}$$

where

$$Q = \begin{bmatrix} 2.3346 & 1.1384 & 2.5606 & 1.4507 \\ 1.1384 & 0.7860 & 1.2743 & 0.9531 \\ 2.5606 & 1.2743 & 2.8147 & 1.6487 \\ 1.4507 & 0.9531 & 1.6487 & 1.8123 \end{bmatrix} \text{ and } b = \begin{bmatrix} 0.4218 \\ 0.9157 \\ 0.7922 \\ 0.9595 \end{bmatrix}$$

We aim to solve this problem using gradient descent, initialized with

$$x_0 = \begin{bmatrix} -135.1150 \\ -4.5224 \\ 130.1168 \\ -5.6879 \end{bmatrix}.$$

(a) How many iterations are needed to reach a point z such that $\|z - x^*\| \leq 0.0001$ with a stepsize $\alpha = 0.2$? **10 ± 1**

(b) How many iterations are needed to reach a point z such that $\|z - x^*\| \leq 0.0001$ with a stepsize $\alpha = 0.07$? **14 ± 1**

(c) Repeat the above for stepsizes $\alpha_1 = 0.0728$ and $\alpha_2 = 0.2185$. How many iterations are needed to reach a point z such that $\|z - x^*\| \leq 0.0001$ with both the stepsizes? (Enter only one number) **14 ± 1**

(d) What is the α^*(correct to 4 significant digits) such that the exact solution is reached in one step with the given choice of x_0; that is, $x^* - x_0 = -\alpha^*(Qx_0 - b)$? **0.1457 ± 0.0001**

> **Solution:** $\alpha^* = .1457$.

(e) Is there a minimum stepsize for which convergence to the minimum is not guaranteed? If yes, what are the $10^{100000000}$th and $10^{100000000} + 1$th iterates for that stepsize?

- **A.** $\boldsymbol{\alpha = 2/\lambda_1, x^* + v_1, x^* - v_1}$
- B. $\alpha = 2/\lambda_1, x^* - v_1, x^* + v_1$
- C. No such stepsize

> **Solution:** First, recall that $x_0 = x^* + v_1$. Plugging this into the gradient descent update, we get $x_1 = (1 - \alpha\lambda)v_1 + x^*$, and, by induction, we get $x_k = x^* + (1 - \alpha\lambda)^k v_1$. Then, recall that
>
> $$\begin{aligned} E_k &= (x_k - x^*)^T Q(x_k - x^*) \\ &= (1 - \alpha\lambda)^{2k}. \end{aligned}$$
>
> This quantity converges if $\alpha \in (0, \frac{2}{\lambda})$. At $\alpha = \frac{2}{\lambda}$ $E_k = 1$ for all k. Substituting this value of α back into the update equation that we derived earlier, we get $x_{k+1} = x_* + (-1)^k v_1$, after which, see the solution for even and odd terms is $x * + v_1$ and $x^* - v_1$ respectively.

5. We've provided a file which takes a 2d vector as an input, and returns the function value, the gradient, and the Hessian of a 2d function. Consider the initial point $[x_0, y_0] = [1.2, 1.2]^T$. We aim to use iterative methods to solve $\arg\min f(x, y)$. In the following questions, report the minimum number of iterations required to reach a point z, where $\|\nabla f(z)\| \leq \epsilon$? (*Write* -1 *to indicate "Does not converge with this stepsize"*)

(a) Gradient descent with stepsize of 0.001, and $\epsilon = 0.1$ **5 ± 1**

(b) Gradient descent with stepsize of 0.001, and $\epsilon = 0.01$ **6146 ± 1**

(c) Gradient descent with stepsize of 0.001, and $\epsilon = 0.001$ **11972 ± 1**

(d) Gradient descent with stepsize of 0.001, and $\epsilon = 0.0001$ **17742 ± 1**

(e) Gradient descent with stepsize of 0.002, and $\epsilon = 0.1$ **70 ± 1**

(f) Gradient descent with stepsize of 0.002, and $\epsilon = 0.01$ **1518 ± 1**

(g) Gradient descent with stepsize of 0.002, and $\epsilon = 0.001$ **4370 ± 1**

(h) Gradient descent with stepsize of 0.002, and $\epsilon = 0.0001$ **7248 ± 1**

(i) Gradient descent with stepsize of 0.005, and $\epsilon = 0.1$ **-1**

(j) Gradient descent with stepsize of 0.005, and $\epsilon = 0.01$ **-1**

(k) Gradient descent with stepsize of 0.005, and $\epsilon = 0.001$ **-1**

(l) Gradient descent with stepsize of 0.005, and $\epsilon = 0.0001$ **-1**

(m) Newton's Method with $\epsilon = 0.1$ **3 ± 1**

(n) Newton's Method with $\epsilon = 0.01$ **3 ± 1**

(o) Newton's Method with $\epsilon = 0.001$ **4 ± 1**

(p) Newton's Method with $\epsilon = 0.0001$ **5 ± 1**

6. For finding the minimum of a function $f : \mathbb{R} \to \mathbb{R}$, four iterative algorithms calculate solutions in the following way (x_k is the solution in the k-th iteration):

   1. $x_k = 1 + \frac{1}{k}$
   2. $x_k = 1 + (0.5)^{2^k}$
   3. $x_k = 1 + \frac{1}{k!}$
   4. $x_k = x_{k-1} - 0.5^k, k \geq 1$ (initialised with $x_0 = 2$)

   (a) What is the nearest integer to which these algorithms converge? (*Hint: Find* $\lfloor x_k \rfloor$ *after large number of iterations*) __**1**__

   (b) Suppose each of the algorithms is executed 10 times. Plot the sequence of solutions obtained on a single graph. What is the correct order in which these algorithms converge (slowest to fastest)?

   A. $1 < 2 < 3 < 4$
   B. $4 < 1 < 2 < 3$
   C. $4 < 3 < 2 < 1$
   D. $\mathbf{1 < 4 < 3 < 2}$

Solving part (a) and (b) shows that the four algorithms converge to a point, and each of the algorithms converge at a different speed. In general, to understand the **rate of convergence**, we use the following to represent how quickly the error $e_k = x_k - x^*$ converges to zero:

$$\lim_{n\to\infty} \frac{|e_{k+1}|}{|e_k|^p} = \lim_{n\to\infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|^p} = \mu$$

Here $p \geq 1$ is called the order of convergence, the constant μ is the rate of convergence. This expression may be better understood when it is interpreted as $|e_{k+1}| = \mu |e_k|^p$ when $n \to \infty$. Obviously, the larger p and the smaller μ, the more quickly the sequence converges. Specially, consider the following cases:

- If $p = 1$ and
  - if $\mu = 1$, the convergence is **sublinear**
  - if $0 < \mu < 1$, the convergence is **linear** with the rate of convergence of μ.
  - if $\mu = 0$, the convergence is **superlinear**
- If $p = 2$, $|e_{k+1}| = \mu |e_k|^2$, $(\mu > 0)$, the convergence is **quadratic**.

   (c) What is the rate of convergence of Algorithms 1,2,3,4.

   A. Linear, Sublinear, Superlinear, Quadratic
   B. Sublinear, Quadratic, Superlinear, Linear
   C. Superlinear, Quadratic, Sublinear , Linear
   D. Sublinear, Linear, Superlinear, Quadratic

Solution: First, express x_k in Algorithm 4 only in terms of k. Note that,

$$\begin{aligned} x_{k+1} &= x_k - 0.5^{k+1} \\ &= x_{k-1} - 0.5^k - 0.5^{k+1} \\ &\vdots \\ &= x_0 - (0.5 + \cdots + 0.5^k + 0.5^{k+1}) \\ &= x_0 - 1 + 0.5^{k+1} && \text{(sum of G.P. Series)} \\ &= 1 + 0.5^{k+1} \end{aligned}$$

Clearly, $\lim_{k\to\infty} x_k$ for all 4 sequences is 1. Therefore, the minimizer is, $x^* = 1$.

1: $\lim_{k\to\infty} \frac{|x_{k+1}-x^*|}{|x_k-x^*|} = \lim_{k\to\infty} \frac{|x_{k+1}-1|}{|x_k-1|} = \lim_{k\to\infty} \frac{\frac{1}{k+1}}{\frac{1}{k}} = \lim_{k\to\infty} \frac{1}{1+\frac{1}{k}} = 1$ (***Sublinear***)

2: $\lim_{k\to\infty} \frac{|x_{k+1}-1|}{|x_k-1|^2} = \lim_{k\to\infty} \frac{0.5^{2^{k+1}}}{0.5^{2^k \cdot 2}} = 1$ (***Quadratic***)

3: $\lim_{k\to\infty} \frac{|x_{k+1}-1|}{|x_k-1|} = \lim_{k\to\infty} \frac{\frac{1}{(k+1)!}}{\frac{1}{k!}} = \frac{1}{k+1} = 0$ (***Superlinear***)

4: $\lim_{k\to\infty} \frac{|x_{k+1}-1|}{|x_k-1|} = \lim_{k\to\infty} \frac{0.5^{k+1}}{0.5^k} = 0.5$ (***Linear***)

Assume x^* is the minimum value of the function, as obtained in part (a) of this problem. In the following questions, report the minimum number of iterations required to reach point z such that $z \leq x^* + \epsilon$. (*Use $z \leq x^* + \epsilon$ as the stopping criteria, and not $z - x^* \leq \epsilon$ for better numerical stability.*)

(d) Algorithm 1 with $\epsilon = 0.1$ $\mathbf{10 \pm 1}$

(e) Algorithm 1 with $\epsilon = 0.01$ $\mathbf{100 \pm 1}$

(f) Algorithm 1 with $\epsilon = 0.001$ $\mathbf{1000 \pm 1}$

(g) Algorithm 1 with $\epsilon = 0.00001$ $\mathbf{10000 \pm 1}$

(h) Algorithm 2 with $\epsilon = 0.1$ $\mathbf{2 \pm 1}$

(i) Algorithm 2 with $\epsilon = 0.01$ $\mathbf{3 \pm 1}$

(j) Algorithm 2 with $\epsilon = 0.001$ $\mathbf{4 \pm 1}$

(k) Algorithm 2 with $\epsilon = 0.00001$ $\mathbf{5 \pm 1}$

(l) Algorithm 3 with $\epsilon = 0.1$ $\mathbf{4 \pm 1}$

(m) Algorithm 3 with $\epsilon = 0.01$ $\mathbf{5 \pm 1}$

(n) Algorithm 3 with $\epsilon = 0.001$ $\mathbf{7 \pm 1}$

(o) Algorithm 3 with $\epsilon = 0.00001$ $\mathbf{9 \pm 1}$

(p) Algorithm 4 with $\epsilon = 0.1$ $\mathbf{4 \pm 1}$

(q) Algorithm 4 with $\epsilon = 0.01$ $\mathbf{7 \pm 1}$

(r) Algorithm 4 with $\epsilon = 0.001$ $\mathbf{10 \pm 1}$

(s) Algorithm 4 with $\epsilon = 0.00001$ $\mathbf{17 \pm 1}$

7. Bisection Method, which is based on Intermediate Value Theorem, is generally used to find roots of a one-dimensional function. In this problem, we will use Bisection Method to find the minimum of a differentiable one-dimensional function by finding the roots of its derivative within an interval $[a, b]$. The pseudo-code for Bisection Method is as follows:

   - Input: Interval $[a, b]$, tolerance *tol*, maximum iterations N_{max}
   - Initialize: $x_{left} = a, x_{right} = b$
   - Inside loop:
     - $x_{mid} = \frac{x_{left}+x_{right}}{2}$
     - if $f(x_{mid}) = 0$ or $\frac{x_{right}-x_{left}}{2} \leq tol$ or N_{max} is reached then, **Stop** and return (x_{mid}, number of iterations)
     - if $sign(f(x_{mid})) \cdot sign(f(x_{left})) < 0$ then $x_{right} = x_{mid}$
     - if $sign(f(x_{mid})) \cdot sign(f(x_{left})) > 0$ then $x_{left} = x_{mid}$
   - Output: number of iterations, x_{mid}

(a) Implement the following function, $f(x)$, and its derivative:

$$f(x) = x^4 - 14x^3 + 60x^2 - 70x$$

Plot $f(x)$ vs $x \in [0, 7]$, with distance between two x values as 0.001. How many stationary points of f(x) exist in this interval? **3** (Optional: Also plot $f^{'}(x)$ vs x for the same interval. Can you identify the stationary points from this plot? Notice how the signs of derivatives are reversed in the neighborhood of stationary points.)

From the plot above, note that a single minimum exists within $[0, 3]$. Since $f : (a, b) \to \mathbb{R}$ is differentiable and unimodal, we can use Bisection Method to find the minimum of $f(x)$ by finding the root of $f'(x)$. Implement the Bisection Method as described in the pseudocode above to find the root of $f'(x)$ in the specified intervals and report the minimum number of iterations for the following input combinations :

(b) Starting Interval: $[0, 2]$, Tolerance=0.01 $\mathbf{8 \pm 1}$

(c) Starting Interval: $[0, 2]$, Tolerance=0.0001 $\mathbf{15 \pm 1}$

(d) Starting Interval: $[0, 3]$, Tolerance=0.001 $\mathbf{12 \pm 1}$

(e) What is the minimizer of $f(x)$ (correct to three decimal places) if the Bisection Method is executed for 10 times starting with interval $[0, 2]$?

A. 0.770
B. 0.775
C. 0.779
D. 0.781

(f) What is the minimizer of $f(x)$ (correct to three decimal places) if the Bisection Method is executed for 20 times starting with interval $[0, 2]$?

A. 0.770
B. 0.775
C. 0.779
D. 0.781

Solution:

We can also calculate number of iterations analytically.

$$\begin{aligned} n &= \left\lceil \log_2 \frac{b-a}{tolerance} \right\rceil && \text{(where [a,b] is the initial range given)} \\ &= \left\lceil \log_2 \frac{2}{0.01} \right\rceil && \text{(substituting given values from part (b))} \\ &= 8 \end{aligned}$$