Computational Methods of Optimization
First Midterm(Sep 23, 2023)

Time: 60 minutes

Instructions

- Answer all questions
- Write your full SR No and your full name(as per IISc records)
- See instructions in the board

Name: ____

SRNO: Degree: Dept:

Question:	1	2	3	4	5	Total
Points:	5	5	5	10	15	40
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. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ be such that $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$. It is given that $\mathbf{x} \notin S$ where the set $S = \{\mathbf{z} | \mathbf{z} = t\mathbf{y}, t \in \mathbb{R}\}$. It is given that $\|\mathbf{x} - \mathbf{y}\| = 1$. Find a point, $\hat{\mathbf{z}} \in S$ which is closest to $\mathbf{x}$.

   (a) (1 point) Set the above problem as an optimization problem.

   Solution: $\hat{\mathbf{z}} = t^* \mathbf{y}, \quad t^* = argmin_{t \in \mathbb{R}} f(t) \quad f(t) = \|\mathbf{x} - t\mathbf{y}\|$

   (b) (2 points) Is such a point unique? Justify

   Solution: Check that the minimum of f is strict, i.e. $f(t) > f(t^*)$. Thus t^* is unique and hence $\mathbf{z}$ is unique.

   (c) (2 points) Find $\hat{\mathbf{z}}$

   Solution: Optimal value of $t^* = \frac{1}{2}$.

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

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

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

   The value of $a \in \mathbb{R}$ is not known.

   (a) (5 points) Answer true or false. Please write T(for true) and F(for false). All questions carry 1 mark

   i. f is not in $\mathcal{C}^2$. __F__

   ii. For any fixed $a \neq 0$ the function f does not have a global minimum. __F__

   iii. The eigenvalues of the Hessian are real __T__

   iv. For any fixed value of $a \neq 0$ the function f has a saddle point. Note that Saddle points are points which are neither local minimum or maximum but the gradient is zero. __F__

   v. There exists $a \neq 0$ such that the set $DS(0)$, the set of Descent Direction at $\mathbf{x} = 0$, is empty. __F__

3. (5 points) For a $\mathcal{C}^2$ function, f, it is given that $\hat{\mathbf{x}}$ is a point whose gradient is zero. It is further given that for all $\mathbf{x} \in \mathbb{R}^d$ $\mathbf{I} \preceq H(\mathbf{x}) \preceq 2\mathbf{I}$ for the Hessian, $H(\mathbf{x})$. Find good upper and lower bound on $\frac{f(\mathbf{x}) - f(\hat{\mathbf{x}})}{\|\mathbf{x} - \hat{\mathbf{x}}\|^2}$

   Solution: By Taylor expansion $f(\mathbf{x}) = f(\hat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^\top H(\mathbf{z})(\mathbf{x} - \hat{\mathbf{x}})$ Furthermore $\|\mathbf{x} - \hat{\mathbf{x}}\|^2 \leq (\mathbf{x} - \hat{\mathbf{x}})^\top H(\mathbf{z})(\mathbf{x} - \hat{\mathbf{x}}) \leq 2\|\mathbf{x} - \hat{\mathbf{x}}\|^2$ Thus the lower bound is $\frac{1}{2}$ and upperbound is 1.

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

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

$$A = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}, b = [1, \beta]^\top, c = 1$$

The values of $a_1, a_2, \beta \in \mathbb{R}$ are not known but it is given that they are non-zero. It is given that gradient of f at $\hat{\mathbf{x}} = [1, 2]^\top$ is zero.

(a) (2 points) Find a_1, a_2, β such that $\hat{\mathbf{x}}$ is a saddle point? In case there are more than one set of values, all sets of values need to mentioned. Note that saddle points are points where the gradient is zero but they are neither minima or maxima

Solution: Note that $\nabla f(\mathbf{x}) = [a_1 x_1 - 1, a_2 x_2 - \beta]^\top = 0$ which is attained at $\mathbf{x} = [\frac{1}{a_1}, \frac{\beta}{a_2}]^\top$. This implies $a_1 = 1$ and $\beta = 2a_2$. For this to be a saddle point $a_2 < 0$. $[a_1, a_2, \beta]^\top = [1, t, 2t]^\top, t < 0$.

(b) (2 points) Find a_1, a_2, β such that $\hat{\mathbf{x}}$ is a global minimum and the minimum value is $-\frac{1}{2}$? In case there are more than one set of values, all sets of values need to mentioned. Use these set of values for the following questions.

Solution: As before, from the gradient is zero condition we have $a_1 = 1$ and $\beta = 2a_2$. For this to be a minimum $a_2 > 0$ and the optimum value is $f(\hat{\mathbf{x}}) = \frac{1}{2} + \frac{a_2}{2}2^2 - 1 - \beta 2 + 1 = -\frac{1}{2}$ $\frac{1}{2} - 2a_2 = -\frac{1}{2}$ which yields $a_2 = \frac{1}{2}$. $[a_1, a_2, \beta]^\top = [1, \frac{1}{2}, 1]^\top$.

(c) (2 points) Compute the Lipschitz constant of f? Recall that Lipschitz constant is defined as the smallest constant $L > 0$ such that

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \le L\|\mathbf{x} - \mathbf{y}\|$$

holds for all points $\mathbf{x}, \mathbf{y}$ in the domain of f.

Solution: $L = 1$, if $a_2 \le 1$, otherwise $L = a_2$. Hence answer is 1.

(d) (2 points) Evaluate the largest possible error made by the first order approximation of $f(\mathbf{z})$ computed at $\mathbf{x}^0 = [1, 1]^\top$ for any $\mathbf{z} \in B_1(\mathbf{x}^0)$. Recall that $B_r(\mathbf{y}) = \{\mathbf{x} | \|\mathbf{x} - \mathbf{y}\| \le r\}$.

Solution: The error of the first order approximation of $f(\mathbf{y})$ computed at $f(\mathbf{x})$ is given by $f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x})$. As derived in class, $|f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x})| \le \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2$. If $\|\mathbf{y} - \mathbf{x}\| \le r$ then the error is given by $\frac{L}{2}r^2$. Putting all values we obtain $\frac{1}{2}$.

(e) (2 points) Starting from $\mathbf{x}_0 = 0$ we implement the following iterations $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)})$. The stepsize α is constant across iterations. What is the range of α such that after every iteration $f(\mathbf{x}^{(k+1)}) \le f(\mathbf{x}^{(k)}), k \ge 0$ is guaranteed?

Solution: $\alpha \in (0, \frac{2}{L}) = (0, 2)$

5. Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q^2\mathbf{x} - \mathbf{h}^\top\mathbf{x} + c$ where $h \in \mathbb{R}^d, c \in \mathbb{R}, Q \succ 0$ and $Q \in \mathcal{S}_d$.

   (a) It is given that $Q = \sum_{i=1}^d \lambda_i \mathbf{u}_i\mathbf{u}_i^\top$ where $\mathbf{u}_i \in \mathbb{R}^d, \mathbf{u}_i^\top\mathbf{u}_j = 0, i \neq j, \|\mathbf{u}_i\| = 1$ and $1 \leq \lambda_i \leq 3$ for all $i \in [d]$.

      i. (2 points) Compute an upperbound on ρ where

$$f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^*) \leq \rho\left(f(\mathbf{x}^{(k)}) - f(\mathbf{x}^*)\right),$$

      $\mathbf{x}^{(k)}$ is the k-th iterate of steepest descent direction using exact stepsize and $\mathbf{x}^*$ be the global minimum.

      Solution: As given, $\mathbf{u}_i$ are eigenvectors and λ_i are the eigenvalues of Q. Since we know that $\rho = \left(\frac{\lambda_{max}-\lambda_{min}}{\lambda_{max}+\lambda_{min}}\right)^2$, where λ_{max} and λ_{min} are the eigenvalues of the Hessian. The Hessian is Q^2 and hence $\lambda_{min} = 1$ and $\lambda_{max} = 9$ Hence $\rho \leq (\frac{9-1}{9+1})^2 = 0.64$.

      ii. (4 points) Suppose $\mathbf{x}^{(0)} = -5\mathbf{u}_1 + Q^{-2}\mathbf{h}$. Compute $f(\mathbf{x}^{(0)}) - f(\mathbf{x}^*)$ assuming that $\lambda_1 = 2$. The answer should be a number.

      Solution: Note that $\mathbf{x}^* = Q^{-2}\mathbf{h}$ and $f(\mathbf{x}^{(0)}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x}^{(0)} - \mathbf{x}^*)Q^2(\mathbf{x}^{(0)} - \mathbf{x}^*)$. Given data $\mathbf{x}^{(0)} - \mathbf{x}^* = -5\mathbf{u}_1$ and Since u_1 is an eigenvector of Q the above expression simplifies to $(-5)^2\lambda_1^2/2 = 25.4/2 = 50$

   (b) Consider inexact line search on f starting at $\mathbf{x}^{(0)}$.

      i. (2 points) Find the gradient of f at $\mathbf{x}^{(0)}$.

      Solution: $\nabla f(\mathbf{x}^{(0)}) = Q^2(\mathbf{x}^{(0)} - \mathbf{x}^*) = Q^2(-5\mathbf{u}_1) = -5(\lambda_1)^2\mathbf{u}_1 = -20\mathbf{u}_1$

      ii. (3 points) Let $\mathbf{u} \in \mathbb{R}^d$ be a unit vector and be a Descent direction of f at $\mathbf{x}^{(0)}$. Let $\mathbf{u} = \mathbf{u}_1 + \beta_2\mathbf{u}_3$ where $\beta_2 \in \mathbb{R}$ is not known. For what value of β_2 is $\mathbf{u}$ a descent direction.

      Solution:

$$\nabla f(\mathbf{x}^{(0)})^{\mathbf{u}} < 0 \implies -20\mathbf{u}_1^(\mathbf{u}_1 + \beta_2\mathbf{u}_3) < 0, implies - 20 < 0$$

      Hence for all $\beta_2 \in \mathbb{R}$, $\mathbf{u}$ is a Descent Direction. However since $\mathbf{u}$ is a unit vector, the only possible value of $\beta_2 = 0$.

      iii. (4 points) For a Descent direction $\mathbf{u} \in \mathbb{R}^d$. Find $\bar{\alpha}$ such that

$$g(\alpha) \leq g(0) + \rho_1 g'(0), \quad g(\alpha) = f(\mathbf{x}^{(0)} + \alpha\mathbf{u})$$

      holds for all $\alpha \in (0, \bar{\alpha})$. Take $\rho_1 = 0.2$. Give steps to justify your answer.

      Solution:

$$g'(0) = -20, g(\alpha) - g(0) = \alpha g'(0) + \frac{\alpha^2}{2}g"(0) = -20\alpha + \frac{1}{2}(-20)^2(2^2)\alpha^2 = -20\alpha + 800\alpha^2$$

$$-20\alpha + 800\alpha^2 \leq \rho_1 g'(0)\alpha = -4\alpha$$

$$\alpha(-16 + 800\alpha) \leq 0, \quad \implies 0 \leq \alpha \leq \frac{16}{800}$$