Computational Methods of Optimization
Second Midterm(7th Oct, 2021)

Instructions:

- This is a closed book test. Please do not consult any additional material.
- Attempt all questions
- Total time is 90 mins.
- Answer the questions in the spaces provided. Answers outside the spaces provided will not be graded.
- Rough work can be done in the spaces provided at the end of the booklet

Name: ____________________

SRNO: Degree: Dept:

Question:	1	2	3	4	Total
Points:	10	10	10	5	35
Score:					

1. Consider a function $f : \mathbb{R}^d \times \mathbb{R}$ be $\mathcal{C}_L^1$ with $L = 2$. We apply gradient descent algorithm with an arbitrary *descent* direction. We apply Goldstein stepsize selection method with parameter $\rho = \frac{1}{4}$.

   (a) (1 point) Inexact line-search requires the knowledge of L. **F**.

   (b) (1 point) Steepest descent direction with inexact line search converges to a critical point **T**.

   (c) (2 points) Let $\bar{\alpha}_k$ be the lower-bound on the stepsize α_k chosen by Goldstein conditon. Which of the following must be true

   A. $\bar{\alpha}_k > \frac{2\rho}{L}$ for the steepest descent direction

   B. $\bar{\alpha}_k = \frac{2\rho}{L}$ **for the steepest descent direction**

   C. $\bar{\alpha}_k = \frac{2\rho}{L}$ for any arbitrary descent direction

   D. $\bar{\alpha}_k > \frac{2\rho}{L}$ for any arbitrary descent direction

   (d) (4 points) Show that there exists a constant $c(\rho, L)$ such that for every iteration the decrease in function value can be stated as

$$f(\mathbf{x}^{(k)}) - f(\mathbf{x}^{(k+1)}) \leq c(\rho, L) \frac{1}{\|\mathbf{u}\|^2} \left(\nabla f(\mathbf{x}^{(k)})^\top \mathbf{u} \right)^2$$

   where $\mathbf{u}$ is a descent direction.

   Solution: Derivation in class notes. We find that $c(\rho, L) = 2\frac{\rho^2}{L}$

   (e) (2 points) Find $c(\rho, L)$?

   (e) $2 \times \left(\frac{1}{4}\right)^2 \frac{1}{2} = \frac{1}{16}$

   Note: Answer in fraction

2. Consider the following problem

$$\min_{\mathbf{x} \in C} f(\mathbf{x}) \left(= \frac{1}{2} \mathbf{x}^\top Q \mathbf{x} - b^\top \mathbf{x} \right)$$

$$C = \{\mathbf{z} | \mathbf{z} = \mathbf{x}_0 + A\mathbf{u}, \quad \mathbf{u} \in \mathbb{R}^k\}, A \in \mathbb{R}^{d \times k}$$

   It is given that $Q \succ 0$.

   (a) (4 points) Can you transfoem this problem into an un-constrained minimization problem of the form

$$\min_{\mathbf{y} \in \mathbb{R}^l} h(\mathbf{y}) \left(= \frac{1}{2} \mathbf{y}^\top \tilde{Q} \mathbf{x} - \tilde{b}^\top \mathbf{y} + c \right)$$

   State $\tilde{Q}, \tilde{b}$, and c.

Solution: Using substitution we can write

$$\min_{\mathbf{x}\in C} f(\mathbf{x}) = min_{\mathbf{y}\in\mathbb{R}^l} f(\mathbf{x}_0 + A\mathbf{y})$$

This is an unconstrained problem.
Taylor expanding around $\mathbf{x}_0$ yields

$$h(\mathbf{y}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top A\mathbf{y} + \frac{1}{2}\mathbf{y}^\top A^\top QA\mathbf{y}$$

This immediately yields

$$c = f(\mathbf{x}_0), \quad \tilde{b} = A^\top \nabla f(\mathbf{x}_0), \quad \tilde{Q} = A^\top QA$$

(b) (1 point) The above problem in $\mathbf{y}$ is not convex __**F**__

(c) (2 points) Justify your answer

Solution: The above problem in $\mathbf{y}$ is convex as the Hessian is $\tilde{Q} = A^\top QA$ is positive definite. For any $\mathbf{y} \in \mathbb{R}^l$ define $\mathbf{z} = A\mathbf{y}$ and hence

$$\mathbf{y}^\top \tilde{Q}\mathbf{y} = \mathbf{z}^\top Q\mathbf{z} > 0$$

for all $\mathbf{y} \neq 0$.

(d) Suppose a point $\tilde{\mathbf{x}} = \mathbf{x}_0 + A\tilde{\mathbf{y}}$ satisfies

$$A^\top \nabla f(\tilde{\mathbf{x}}) = 0$$

i. (1 point) $\tilde{\mathbf{x}}$ is a global minimum __**T**__

ii. (2 points) Justify your answer.

Solution: Since $h(\mathbf{y})$ is convex in $\mathbf{y}$ and $\mathbf{y}$ satisfying $\nabla h(\mathbf{y}) = 0$ is global optimal and hence $\tilde{\mathbf{x}}$ is global optimal.

$$\nabla h(\tilde{\mathbf{y}}) = A^\top \nabla f(\mathbf{x}_0) + A^\top QA\tilde{\mathbf{y}}$$
$$= A^\top (\nabla f(\mathbf{x}_0) + QA\tilde{\mathbf{y}}) = A^\top \nabla f(\tilde{\mathbf{x}}) = 0$$

3. Consider applying Conjugate Gradient(CG) method for solving the following problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \left(= \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} - b^\top \mathbf{x}\right)$$

where $Q \in \mathbb{R}^{n\times n}$ is symmetric and $Q \succ 0$. Let $\lambda_1 > \lambda_2 > \cdots, > \lambda_d > 0$ be the eigenvalues of Q.

(a) (2 points) Starting at the initial condition $\mathbf{x}_0 = 0$, Find the point after first iteration of Conjugte gradient algorithm.

Solution:

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha \mathbf{u}, \quad \mathbf{u} = -(Q\mathbf{x}_0 - b) = b, \quad \alpha = \frac{b^\top b}{b^\top Q b}$$

$$\mathbf{x}_1 = -\frac{\|b\|^2}{b^\top Q b} b$$

(b) Apply CG to $b = h_1\mathbf{e}_1 + h_2\mathbf{e}_2$ where $\mathbf{e}_1$ and $\mathbf{e}_2$ are eigenvectors correspondig to the eigenvalues λ_1 and λ_2. Assume $\mathbf{x}_0 = 0$.

i. (2 points) In how many iterations would the method converge

i. **2**

ii. (5 points) Justify your answer.

Solution: We note that

$$Qb = \lambda_1 h_1 e_1 + \lambda_2 h_2 e_2, \quad b^\top f(Q) b = f(\lambda_1) h_1^2 + f(\lambda_2) h_2^2$$

for any matrix polynomial of degree $k \geq 1$. From CG algorithm

$$E(\mathbf{x}_{k+1}) = min_{P_k} \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}^*) Q[I + QP_k(Q)]^2 (\mathbf{x}_0 - \mathbf{x}^*)$$

Since $\mathbf{x}_0 = 0$, then

$$E(\mathbf{x}_{k+1}) = min_{P_k} \frac{1}{2} b^\top Q[I + QP_k(Q)]^2 b$$

$$E(\mathbf{x}_{k+1}) = min_{P_k} \frac{1}{2} \sum_{i=1}^{2} h_i^2 \lambda_i (1 + \lambda_i P_k(\lambda_i))^2$$

We choose a polynomial $T(\lambda) = \prod_{i=1}^2 \left(1 - \frac{\lambda}{\lambda_i}\right)$. Note that there exists a $P_1(\lambda)$ such that $\lambda P_2(\lambda) = T(\lambda) - 1$ As a consequence,

$$E(\mathbf{x}_2) \leq max_{1 \leq i \leq 2} (1 + \lambda_i P_1(\lambda_i))^2 \left(\frac{1}{2} \sum_{i=1}^{2} h_i^2 \lambda_i \right)$$

By construction of $T(\lambda)$ it follows that $E(\mathbf{x}_2) = 0$ and hence the algorithm converges in 2 steps.

iii. (1 point) In how many iterations will the algorithm converge if $\mathbf{x}_0$ is not zero but arbitrary.

iii. **d**

4. Answer True or False

(a) (1 point) If we reset the conjugate gradient method after every iteration we will recover Newton method **F**.

(b) (1 point) To prove convergence of Newton method the function needs to be in $\mathcal{C}^3$. **F**

(c) (1 point) For a convex quadratic program Newton method will converge in half the number of iterations when compared with steepest descent with exact line search **F**

(d) (1 point) Rank one Quasi-newton method yields Rank one matrices **F**

(e) (1 point) Rank two Quasi-newton method yields Conjugate Gradient directions for convex quadratic programs_**T**_