Computational Methods of Optimization
Second Midterm(28th Oct, 2023)

Instructions:

- This is a closed book test. Please do not consult any additional material.
- Attempt all questions
- Total time is 70 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	5	Total
Points:	15	5	10	5	15	50
Score:						

1. Let $Q \in \mathcal{S}_+^d, h \in \mathbb{R}^d, c \in \mathbb{R}$ and $min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) \left(= \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + h^\top\mathbf{x} + c\right)$. Let $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \mathbf{u}^{(3)}$ be Q conjugate directions. It is further given that $\mathbf{u}^{(i)\top} Q\mathbf{u}^{(i)} = a_i, h^\top\mathbf{u}^{(i)} = \gamma_i a_i, i \in \{1, 2, 3\}$. Consdier implementing the Conjugate Direction Algorithm(CDA) for three iterations by using $\mathbf{u}^{(i)}$ in the ith iteration. Assume that the the starting point is $\mathbf{x}^{(0)} = 0$. Let $\mathbf{x}^{(3)}$ be the point after three iterations.

   (a) (6 points) Find $\mathbf{x}^{(3)}$ and $f(\mathbf{x}^{(3)})$. Your answer should be expressed in terms of $\mathbf{u}^{(i)}, \gamma_i, a_i$s.

   Solution: Noting that $\mathbf{x}^{(0)} = 0$, applying Expanding subspace theorem one concludes that

$$\mathbf{x}^{(3)} = argmin_{\mathbf{x}\in C} f(\mathbf{x}) \; C = \{\mathbf{x} = \sum_{i=1}^{3} \beta_i \mathbf{u}^{(i)} | \beta_i \in \mathbb{R}, i = \{1, 2, 3\}\}.$$

   For any $\mathbf{x} \in C$ it can be seen that

$$f(\mathbf{x}) = \sum_{i=1}^{3} \left(\frac{1}{2}\beta_i^2 \mathbf{u}^{(i)} Q\mathbf{u}^{(i)} + \beta_i(h^\top\mathbf{u}^{(i)})\right) + c = \sum_{i=1}^{3} g_i(\beta_i) + c$$

   where $g_i(\beta_i) = a_i\left(\frac{1}{2}\beta_i^2 + \beta_i\gamma_i\right)$. Minimizing such a function is same as minimizing each g_i, whose minimum is at $\beta_i^* = -\gamma_i$ and $g_i(\beta_i^*) = -\frac{1}{2}a_i\gamma_i^2$. Thus

$$\mathbf{x}^{(3)} = -\sum_{i=1}^{3} \gamma_i \mathbf{u}^{(i)}, \; f(\mathbf{x}^{(3)}) = -\frac{1}{2}\sum_{i=1}^{3} a_i\gamma_i^2 + c$$

   (b) Instead of CDA one decides to minimizes f along $\mathbf{v} = \sum_{i=1}^{3} \mathbf{u}^{(i)}$ starting from $\mathbf{x}^{(0)} = 0$. Let the minimum be attained at $\hat{\mathbf{x}}$.

      i. (5 points) Find $\hat{\mathbf{x}}$ and $f(\hat{\mathbf{x}})$.

      Solution: Note that $\hat{\mathbf{x}} = \beta^*\mathbf{v}$ where $\beta^* = argmin_{\beta\in\mathbb{R}} f(\beta\mathbf{v})$. By Q-conjugacy, $f(\beta\mathbf{v}) = c + g(\beta)\left(= \beta^2\sum_{i=1}^{3} a_i + \beta\sum_{i=1}^{3}\gamma_i a_i\right)$. The minimum is attained at $\beta^* = -\frac{\sum_{i=1}^{3}\gamma_i a_i}{\sum_{i=1}^{3} a_i}$ and the minimum value of g is $-\frac{(\sum_{i=1}^{3}\gamma_i a_i)^2}{\sum_{i=1}^{3} a_i}$. Thus

$$\hat{\mathbf{x}} = -\frac{\sum_{i=1}^{3}\gamma_i a_i}{\sum_{i=1}^{3} a_i}\mathbf{v} \quad f(\hat{\mathbf{x}}) = c - \frac{(\sum_{i=1}^{3}\gamma_i a_i)^2}{\sum_{i=1}^{3} a_i}.$$

      ii. (4 points) Express $f(\hat{\mathbf{x}}) - f(\mathbf{x}^{(3)})$ in terms of $\gamma_i, a_i, i \in \{1, 2, 3\}$. Determine the sign of the expression and explain your answer in the context of CDA. Can you find conditions on γ_i such that the difference between the function values are zero?

      Solution: By substitution

$$f(\hat{\mathbf{x}}) - f(\mathbf{x}^{(3)}) = \frac{1}{2}\sum_{i=1}^{3} a_i\gamma_i^2 - \frac{(\sum_{i=1}^{3}\gamma_i a_i)^2}{\sum_{i=1}^{3} a_i}.$$

      By Cauchy Schwartz inequality $(\sum_{i=1}^{3}\gamma_i a_i)^2 \leq \sum_{i=1}^{3}(\gamma_i^2 a_i)(\sum_{i=1}^{3} a_i)$. Hence the desired term is positive. CDA computes the minimum over the appropriate subspaces in each iteration while minimization over $\mathbf{v}$ is only a subset of the subspaces considered. It is 0 if equality holds if $\gamma_i\sqrt{a_i} = \alpha\sqrt{a_i}$ for some α. Thus the condition is all $\gamma_i = \alpha$ for some α.

2. Answer True or False

(a) (1 point) Newton method does not generate descent directions **F** .

(b) (1 point) The computational effort in computing one iterate of Newton method is same as one iteration of Gradient Dewcent. **F**

(c) (1 point) For a $\mathcal{C}^3$ function Newton's method is not applicable **F**

(d) (1 point) Newton's method does not have a stepsize **T**

(e) (1 point) Rank one Quasi-newton method is faster then Newton's method for Convex Quadratic programs. **T**

3. Consider minimizing $f : \mathbb{R}^d \to \mathbb{R}$, a $\mathcal{C}^2$ function, by the following iterates

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

is a Quasi Newton update with

$$G^{(k+1)} = G^{(k)} + ADA^\top, \quad A = [\delta_k, G^{(k)}\gamma_k], \quad D = \begin{bmatrix} \frac{a}{\delta_k^\top \gamma_k}, & 0 \\ 0 & \frac{b}{\gamma_k^\top G^{(k)} \gamma_k} \end{bmatrix}$$

where $\delta_k = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ and $\gamma_k = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$. Assume $\delta_k^\top \gamma_k > 0$ for all k.

(a) (4 points) For what choices of a, b does the iteration gives a valid Quasi newton condition.

Solution: Since $G^{(k+1)}\gamma_k = \delta_k$, and using the update equation one finds

$$DA^\top \gamma_k = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

Hence $G^{(k+1)}\gamma_k = G^{(k)}\gamma_k + a\delta_k + bG^{(k)}\gamma_k = \delta_k$ By inspection $a = 1, b = -1$ is the desired solution.

(b) (2 points) For such a choice of a, b is D positive definite?

Solution: No. It is indefinite

(c) (4 points) For such a choice of a, b is $G^{(k+1)}$ positive definite? Give reasons. Assume $G^{(k)}$ positive definite.

Solution: There exists B such that $G^{(k)} = BB^\top$. Note that for the choice of a, b

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

For any $\mathbf{z} \in \mathbb{R}^d$ we note that

$$\mathbf{z}^\top G^{(k+1)} \mathbf{z} = \mathbf{z}^\top BB^\top \mathbf{z} - \frac{1}{\|B^\top \gamma_k\|^2}(\mathbf{z}^\top BB^\top \gamma_k)^2 + \frac{1}{\delta_k^\top \gamma_k}(\mathbf{z}^\top \delta_k)^2$$

. Choose $\mathbf{u} = B^\top \mathbf{z}, \mathbf{v} = B^\top \gamma_k$ we have $\|\mathbf{u}\|^2 - \frac{1}{\|\mathbf{v}\|^2}(\mathbf{u}^\top \mathbf{v})^2 \geq 0$ Since the remaining term is positive the claim follows.

4. (5 points) Consider the following problem

$$min_{x\in\mathbb{R}}\frac{1}{2}(x-a)^2, \quad b \le x \le c$$

Consider three points $x = b, x = c$ and for any $\hat{x}$ such that $b < \hat{x} < c$. Let $a > c$. For each of the points find the set of Feasible Directions, $FD(x)$, and the set of Descent Directions, $DD(x)$, From these info. deduce the global optimal point.

Solution: In 1-Dimension there are only two directions, 1 and -1. $FD(b) = \{1\}, DD(b) = \{1\}$. $FD(c) = \{-1\}, DD(b) = \{1\}$. $FD(\hat{x}) = \{-1, 1\}, DD(\hat{x}) = \{-1, 1\}$. For point $x = c$ we see that $FD(\hat{x}) \bigcap DD(\hat{x})$ is empty. Thus there is no Feasible Descent Direction and since the problem is convex, $x = c$ is the optimal point.

5. Recall that $P_C(\mathbf{z})$ is the projection of a point $\mathbf{z}$ on a convex set C.

(a) (6 points) Find $P_C(z)$ for any $\mathbf{z} \in \mathbb{R}$ where $C = [a, b]$. Your answer must work for any z and should be expressed in terms of z, a, b. Justify your answer using KKT conditions.

Solution: $P_C(z) = argmin_{x\in C}\frac{1}{2}(x-z)^2$ The KKT conditions are

$$x - z = \lambda_1 - \lambda_2, \lambda_1(x-a) = \lambda_2(b-x) = 0, \lambda_1, \lambda_2 \ge 0, a \le x \le b$$

. This is a convex problem and hence any KKT point yields a global minimum. There are three cases. If $z < a$, check that $x = a, \lambda_1 = a - z, \lambda_2 = 0$ is a KKT point and hence $P_C(z) = a$ If $z > b$, check that $x = b, \lambda_1 = 0, \lambda_2 = z - b$ is a KKT point and hence $P_C(z) = a$ If $a \le z \le b$, check that $x = z, \lambda_1 = 0, \lambda_2 = 0$ is a KKT point and hence $P_C(z) = z$

(b) (5 points) Find $P_C(\mathbf{z})$ for any $\mathbf{z} \in \mathbb{R}^d$ where $C = [0, 1]^d$. Assume that all coordinates of $\mathbf{z}$ are less than 0. Your answer must work for any $\mathbf{z}$.

Solution: $P_C(\mathbf{z}) = argmin_{\mathbf{x}\in C}\frac{1}{2}\|\mathbf{x} - \mathbf{z}\|^2$. Observing that the constraints and the objective, one obtains

$$(P_C(\mathbf{z}))_i = P_C(z_i) = argmin_{0\le x_i\le 1}\frac{1}{2}(x_i - z_i)^2$$

Using the above question we have

$$(P_C(\mathbf{z}))_i = 0, \forall i \in [d]$$

(c) (4 points) Find $max_{\mathbf{x}\in C}\ \mathbf{z}^\top\mathbf{x}$ where $\mathbf{z}$ and C are defined above.

Solution: By property of the projection we have $(P_C(\mathbf{z}) - \mathbf{z})^\top\mathbf{x} \ge (P_C(\mathbf{z}) - \mathbf{z})^\top P_C(\mathbf{z})$ for all $\mathbf{x} \in C$. Since $P_C(\mathbf{z}) = 0$ the above can be written as $-\mathbf{z}^\top\mathbf{x} \ge 0$ Hence the maximum is 0 and attained at $\mathbf{x} = 0$.