Computational Methods of Optimization
Second Midterm(Oct 20, 2024)

Time: 70 minutes

Instructions

- Answer all questions
- Please write your answer in the spaces provided. Answers written outside will not be graded.
- Rough work can be done in designated spaces.
- This is a closed book exam.

Name: ______________________________

Index: SRNO: Degree: Dept:

Question:	1	2	3	4	Total
Points:	10	10	10	15	45
Score:					

In the following, let $\mathbf{I} = [\mathbf{e}_1, \ldots, \mathbf{e}_d]$ 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 $\{1, 2, \ldots, n\}$ We will denote

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{R}^d}{\text{argmin}}\ f(\mathbf{x}) \quad f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + b^\top \mathbf{x} + c \tag{Quad}$$

where $Q \in \mathcal{S}_d, Q \succ 0$, $b \in \mathbb{R}^d$, and $c \in \mathbb{R}$. Let λ_i be the eigenvalue and $\mathbf{v}^{(i)}, \|\mathbf{v}^{(i)}\| = 1$ be the corresponding eigenvector of Q. Assume $\lambda_1 > \lambda_2 > \ldots, \lambda_d > a > 0$.

1. Consider (`Quad`). It is given that $c = 5, \lambda_2 = 5, b^\top \mathbf{v}^{(d)} = 2, b^\top \mathbf{v}^{(2)} = -10$. You have chosen to use the Conjugate Direction(CD) Algorithm. with $\mathbf{x}^{(0)} = 0$.

   (a) Let $\mathbf{v}^{(d)}$ be the first conjugate direction. After one iteration of the Algorithm one finds that $f(\mathbf{x}^{(1)}) = -5$.

   i. (1 point) Find $\mathbf{x}^{(1)}$. Your answer should be expressed in terms of $\mathbf{v}^{(d)}$. $\underline{\quad -10\mathbf{v}^{(d)} \quad}$

   ii. (2 points) Justify.

   Solution: By the question $\mathbf{x}^{(1)} = \alpha \mathbf{v}^{(d)}$. Note that $\nabla f(\mathbf{x}^{(0)}) = b$. By CD Algorithm stepsize is given by $\alpha = -\frac{\nabla f(\mathbf{x}^{(0)})^\top \mathbf{u}}{\mathbf{u}^\top Q\mathbf{u}}$ for a direction $\mathbf{u}$. The Decrease is given by $f(x^{(0)}) - f(\mathbf{x}^{(1)}) = \frac{1}{2}\alpha^2 \mathbf{u}^\top Q\mathbf{u}$. Putting everything together we obtain $10 = \frac{1}{2}\alpha^2 \lambda_d, \alpha = \frac{-2}{\lambda_d}$ which yields $\lambda_d = \frac{1}{5}$ and $\alpha = -10$.

   iii. (1 point) Is $\mathbf{v}^{(d)}$ a descent Direction? Justify.

   Solution: No. $\nabla f(\mathbf{x}^{(0)})^\top \mathbf{v}^{(d)} = b^\top \mathbf{v}^{(d)} = 2 > 0$.

   (b) State and solve a problem in two variables which will yield $\tilde{\mathbf{x}}$ defined below.

$$\tilde{\mathbf{x}} = \text{argmin}_{\ \mathbf{x} = \beta_1 \mathbf{v}^{(2)} + \beta_2 \mathbf{v}^{(d)}, \beta_1, \beta_2 \in \mathbb{R}} f(\mathbf{x})$$

   i. (2 points) State the problem. All constants in the problem should be evaluted.

   Solution:

$$f(\mathbf{x}) = c + (\beta_1 \mathbf{v}^{(2)} + \beta_2 \mathbf{v}^{(d)})^\top b + \frac{1}{2}(\lambda_2 \beta_1^2 + \lambda_d \beta_2^2)$$

$$\tilde{\mathbf{x}} = \beta_1^* \mathbf{v}^{(2)} + \beta_2^* \mathbf{v}^{(d)} \quad \beta_1^*, \beta_2^* = \underset{\beta_1 \in \mathbb{R}, \beta_2 \in \mathbb{R}}{\text{argmin}} \ -10\beta_1 + 2\beta_2 + \frac{1}{2}(5\beta_1^2 + \frac{1}{5}\beta_2^2)$$

   ii. (1 point) Solve the problem to determine $\tilde{\mathbf{x}} = \underline{\ 2\mathbf{v}^{(2)} - 10\mathbf{v}^{(d)}\ }$

   Rough Work

(c) Suppose the CD algorithm is continued from the first question with $\mathbf{v}^{(2)}$ as the next direction to obtain $\mathbf{x}^{(2)}$.

i. (1 point) For this choice find the value of $f(\mathbf{x}^{(2)}) =$ **-15**

ii. (2 points) Justify.

Solution: One can use the Expanding subspace theorem to note that $\mathbf{x}^{(2)} = \tilde{\mathbf{x}}$ and

$$f(\mathbf{x}^{(2)}) = c - 20 = -15$$

Rough Work

2. Refer to (Quad). Let $\lambda_4 > 9 > \lambda_5$, $a = 1$, $d = 100$. It is expected that starting from an arbitrary point, $\mathbf{x}^{(0)}$, the Conjugate Gradient Algorithm provably finds a point, $\tilde{\mathbf{x}}$, such

$$E(\tilde{\mathbf{x}}) \leq rE(\mathbf{x}^0), \quad E(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^\top Q(\mathbf{x}-\mathbf{x}^*)$$

where r is a positive number less than 1.
Report your answers to first decimal place.

(a) (2 points) In how many iterations will we reach $r = 0.7$? Number of iterations= 5 .

(b) (2 points) If $\lambda_9 = 7$, what is the value of r after 10 iterations? $r =$ 0.6

(c) (2 points) Suppose you used partial conjugate gradient algorithm by restarting it after 5 iterations what is the obtained value of r after 10 iterations. $r =$ 0.4

(d) (4 points) Justify your answer.

Solution: Recall that for the kth iterate $\mathbf{x}^{(k)}$ satisfies

$$E(\mathbf{x}^{(k)}) \leq \bar{\lambda}E(\mathbf{x}^{(0)}), \quad \bar{\lambda} = max_{1\leq i\leq d}(q_k(\lambda_i))^2$$

for any kth degree polynomial, $q_k(t)$ such that $q_k(0) = 1$. Consider the polynomial $q_k(t) = (1 - \frac{t}{l+a})\prod_{i=1}^{k-1}\left(1 - \frac{t}{\lambda_i}\right)$ where $\lambda_{k+1} \leq l \leq \lambda_k$. If we put $k = 5, l = 9$ we obtain $\bar{\lambda} = 0.64$. If we put $k = 10, l = 7$ we obtain $\bar{\lambda} = 0.5625$. If we do partial conjugate gradient we obtain $r = (0.64)^2 \approx 0.37$.

Rough Work

3. Refer to (`Quad`). Let $G^{(k)} \succ 0, \delta_k = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ and $\gamma_k = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$ where $\mathbf{x}^{(k)}$ are the iterates of a Quasi Newton Algorithm of the form.

$$G^{(k+1)} = G^{(k)} + BAB^\top, B = [\delta_k \ \ G^{(k)}\gamma_k], \quad A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

It is given that

$$\delta_k^\top \gamma_k = 10 \quad \gamma_k^\top G^{(k)}\gamma_k = 2$$

(a) (2 points) Let $a = 0$. Determine values b, c such that the update satisfies the Quasi-Newton condition $G^{(k+1)}\gamma_k = \delta_k$
$b =$ $\frac{1}{2}$, $c =$ -3

(b) (3 points) Justify your answer

Solution: From the expression $G^{(k+1)}\gamma_k = \delta_k$, we have

$$(1 + b\delta_k^\top \gamma_k + c\gamma_k^\top G^{(k)}\gamma_k)G^{(k)}\gamma_k + (-1 + b\gamma_k G^{(k)}\gamma_k)\delta_k = 0$$

Since $a = 0$ $\ 1 = b\gamma_k^\top G^{(k)}\gamma_k, \quad 0 = 1 + b\delta_k^\top \gamma_k + c\gamma_k^\top G^{(k)}\gamma_k$

$$1 = 2b, -1 = 10b + 2c \implies b = \frac{1}{2}, c = -3$$

(c) (2 points) Let $b = 0$. Determine values a, c such that the update satisfies the Quasi-Newton condition $G^{(k+1)}\gamma_k = \delta_k$
$a =$ $\frac{1}{10}$, $c =$ $-\frac{1}{2}$

(d) (3 points) Between the choices where $a = 0$ and $b = 0$, which one should we prefer. Give brief justification.

Solution: We prefer the second update as the obtained G would be positive definite.(Why?) No guarantee exists for $a = 0$.

Rough Work

4. Consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x}) \quad \text{subject to } \|\mathbf{x}-\mathbf{a}\|^2 \leq r^2$$

where $\mathbf{a}\in\mathbb{R}^d, r\in\mathbb{R}$ and f is defined in (`Quad`) with $Q=\mathbf{I}$.

(a) (1 point) State the Lagrangian of the problem, $\mathcal{L}(\mathbf{x},\lambda)$.

Solution: $\mathcal{L}(\mathbf{x},\lambda) = \frac{1}{2}\|\mathbf{x}\|^2 + b^\top\mathbf{x} + c + \lambda(\|\mathbf{x}-\mathbf{a}\|^2 - r^2)$

(b) (2 points) State the KKT conditions of the problem.

Solution: $(1+2\lambda)\mathbf{x} + b - \lambda\mathbf{a} = 0$

$$\lambda(\|\mathbf{x}-\mathbf{a}\|^2 - r^2) = 0, \lambda \geq 0$$

$$\|\mathbf{x}-\mathbf{a}\|^2 \leq r^2$$

(c) (4 points) Find $\hat{\mathbf{x}}$, a KKT point for the problem. Express your answer in the form of $\hat{\mathbf{x}} = \mathbf{a}+\mathbf{v}$. Assume that $\|b+\mathbf{a}\|^2 > r^2$. Compute the value of λ.

Solution:

$$\hat{\mathbf{x}} = -\frac{1}{1+2\lambda}b + \mathbf{a}\frac{2\lambda}{1+2\lambda} = \mathbf{a} - \frac{1}{1+2\lambda}(\mathbf{a}+b)$$

$$\|\hat{\mathbf{x}}-\mathbf{a}\| = \frac{1}{1+2\lambda}\|-b-\mathbf{a}\| \leq r,\ \lambda > 0$$

Since $\|-b-\mathbf{a}\|^2 > r^2$, it implies that $\lambda > 0$ and hence

$$r = \|\hat{\mathbf{x}}-\mathbf{a}\| = \frac{1}{1+2\lambda}\|b+\mathbf{a}\|$$

Thus $\hat{\mathbf{x}} = \mathbf{a} - \frac{r}{\|b+\mathbf{a}\|}(b+\mathbf{a})$ is a KKT point. The value of $\lambda = \frac{1}{2}(\frac{\|b+\mathbf{a}\|}{r} - 1)$.

(d) Consider the projection problem

$$\min_{\mathbf{x}\in\mathbb{R}^d} \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|^2 \quad \text{subject to } \|\mathbf{x}-\mathbf{a}\|^2 \le r^2$$

i. (3 points) Find a KKT point for this problem.

Solution: Following the previous result

$$\hat{\mathbf{x}} = \mathbf{a} + \frac{r}{\|\mathbf{z}-\mathbf{a}\|}(\mathbf{z}-\mathbf{a})$$

is a KKT point

ii. (5 points) Is the obtained KKT point the optimal point? A self contained argument based on discussions of topics in the class needs to be provided.

Solution: Denoting $P_S(\mathbf{z})$ the projection onto the given set, we note the property that

$$(P_S(\mathbf{z})-\mathbf{z})^\top(\mathbf{x}-P_S(\mathbf{z})) \ge 0$$

For the KKT point, $\mathbf{a}+\frac{r}{\|\mathbf{z}-\mathbf{a}\|}(\mathbf{z}-\mathbf{a})$, and for any feasible $\mathbf{x}=\mathbf{a}+r\mathbf{v}$, $\|\mathbf{v}\|\le 1$, we have

$$r(\frac{r}{\|\mathbf{z}-\mathbf{a}\|}-1)(\mathbf{z}-\mathbf{a})^\top(\mathbf{v}-\frac{1}{\|z-\mathbf{a}\|}(z-\mathbf{a}))\ge 0$$

. This is due to the inequality $(\mathbf{z}-\mathbf{a})^\top\mathbf{v} \le \|\mathbf{z}-\mathbf{a}\|$ and $r < \|\mathbf{z}-\mathbf{a}\|$.

Rough Work

Rough Work

Rough Work

Rough Work