Computational Methods of Optimization
Second Midterm(20th Oct, 2019)

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	5	Total
Points:	10	10	10	10	10	50
Score:						

1. Consider applying Conjugate gradient 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 $Q \succ 0$.

(a) (2 points) Let $E(\mathbf{x}_r) = f(\mathbf{x}_r) - f(\mathbf{x}^*)$ denote the difference between the function values evaluated at $\mathbf{x}_r$, the output of rth iteration of Conjugate gradient algorithm and $\mathbf{x}^*$ is the global minimum of f. State a relationship between $E(\mathbf{x}_r)$ and $E(\mathbf{x}_0)$ involving eigenvalues of Q

Solution:

$$E(\mathbf{x}_r) \leq max_i(1+\lambda_i P_{r-1}(\lambda_i))^2 E(\mathbf{x}_0)$$

$P_{r-1}(\lambda)$ is a r-1 th degree polynomial with real coefficients.

(b) (8 points) Suppose $b = \sum_{i=1}^r h_i \mathbf{e}_i$ where $\mathbf{e}_i$ is the eigenvector correspondig to the eigenvalue λ_i of Q and r is less than n. Assuming the starting point is at $\mathbf{x}^0 = 0$,estimate the number of iterations required to solve the problem. Justify your answer.

Solution: From the proof of convergence of conjugate gradient(CG) algorithm we know that

$$SP(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_k) = SP(\mathbf{g}_0, Q\mathbf{g}_0, \ldots, Q^k\mathbf{g}_0)$$

where SP denotes the linear span of vectors, $\mathbf{u}_i$ denote the conjugate directions obtained from the CG, and $\mathbf{g}_0 = \nabla f(\mathbf{x}_0)$.
For any matrix Q with real eigenvalues, λ_i with eigenvectors $\mathbf{e}_i$

$$f(Q)\mathbf{x} = \sum_{i=1}^r h_i^2 f(\lambda_i)\mathbf{e}_i \;\mathbf{x}^\top f(Q)\mathbf{x} = \sum -i = 1^r h_i^2 \lambda_i \;\mathbf{x} = \sum_{i=1}^r h_i \mathbf{e}_i$$

In the problem it is given that $\mathbf{x}_0 = 0$ and so $\mathbf{g}_0 = -b$. For any $l \geq 1$, $Q^l\mathbf{g}_0 = -Q^l b = -\sum_{i=1}^r h_i \lambda_i^l \mathbf{e}_i$.

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

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

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

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

$$E(\mathbf{x}_r) \leq max_{1\leq i\leq r}(1+\lambda_i P_{r-1}(\lambda_i))^2 \left(\frac{1}{2}\sum_{i=1}^r h_i^2 \lambda_i\right)$$

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

2. (10 points) Let f be defined in Question 1. Let $\mathbf{x}^*$ be the global minimum of the problem

$$min_{\mathbf{x}\in C} f(\mathbf{x})$$

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

Derive a relationship between $\nabla f(\mathbf{x}^*)$ and A.

Solution: There exists $\mathbf{u}^* \in \mathbb{R}^l$ such that $\mathbf{x}^* = \mathbf{x}_0 + A\mathbf{u}^*$. Define

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

The minimization can now be stated as

$$min_{\mathbf{u}\in\mathbb{R}^l} h(\mathbf{u})$$

This is a convex function and at optimality

$$\nabla h(\mathbf{u}) = 0$$

holds. Equivalently

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

This minimum is attained at $\mathbf{u}^*$ and it yields the relationship,

$$A^\top (\nabla f(\mathbf{x}_0) + A\mathbf{u}^*) = 0.$$

From the Definition of $\mathbf{u}^*$ we get the relationship

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

Note: This is the basis of expanding subspace theorem.

3. Consider minimizing a convex function $f : C \subset \mathbb{R}^d \to \mathbb{R}$ over the convex set C. The function maynot be $\mathcal{C}^1$.

(a) (5 points) Let $\mathbf{x}^*$ be a global minimum. Show that if $f(\mathbf{x}^*) < f(\mathbf{x}^0)$ then $\mathbf{x}^0$ cannot be a local minimum, i.e. there exists a point, $\mathbf{z}$, in every neighbourhood of $\mathbf{x}^0$ such that $f(\mathbf{z}) < f(\mathbf{x}^0)$

Solution: Consider the set

$$D = \{\mathbf{u} | \mathbf{u} = (1-\alpha)\mathbf{x}^* + \alpha \mathbf{x}^0, 0 < \alpha < 1\}.$$

For any $\mathbf{u} \in D$,

$$\begin{aligned} f(\mathbf{u}) &\leq (1-\alpha)f(\mathbf{x}^*) + \alpha f(\mathbf{x}^0) \\ &< (1-\alpha)f(\mathbf{x}^0) + \alpha f(\mathbf{x}^0) = f(\mathbf{x}^0) \end{aligned}$$

For every $\delta > 0$, the neighbourhood, $N_\delta(\mathbf{x}^0)$, intersects D. Any point, $\mathbf{z}$, in the intersection satisfy $f(\mathbf{z}) < f(\mathbf{x}^0)$.

(b) (5 points) The function f is said to be strictly convex if for any $\alpha \in (0, 1)$

$$f(\alpha \mathbf{y} + (1-\alpha)\mathbf{x}) < (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y})$$

holds for every $\mathbf{x}, \mathbf{y} \in C$. Prove or Disprove that there could exist two distinct points $\mathbf{x}^*, \mathbf{y}^*$ such that $f(\mathbf{x}^*) = f(\mathbf{y}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in C$.

Solution: Suppose $\mathbf{x}^*$ and $\mathbf{y}^*$ are both global minimum. In other words they satisfy

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \quad f(\mathbf{y}^*) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in C. \tag{1}$$

Construct a set

$$D = \{\mathbf{u} | \mathbf{u} = (1-\alpha)\mathbf{x}^* + \alpha \mathbf{y}^*, 0 < \alpha < 1\}.$$

For any $\mathbf{z} \in D$ we have $f(\mathbf{z}) < (1-\alpha)f(\mathbf{x}^*) + \alpha f(\mathbf{y}^*) = f(\mathbf{x}^*) = f(\mathbf{y}^*)$. This contradicts (1) and hence there cannot exist two points $\mathbf{x}^*, \mathbf{y}^*$ such that they are both global minima.

4. Consider the following model of the relationship between observation, $\mathbf{o} \in \mathbb{R}^d$, and response, $r \in \mathbb{R}$.

$$r = \mathbf{w}^\top \mathbf{o}$$

The parameter of the model, $\mathbf{w}$, is unknown and need to be determined. Suppose n pairs of $(\mathbf{o}_i, r_i), i = 1, \dots, n$ are given to us. One could take the Least squares approach to determine $\mathbf{w}^*$ by solving the following problem

$$\mathbf{w}^* = argmin_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^{n} (r_i - \mathbf{w}^\top \mathbf{o}_i)^2$$

(a) (5 points) Compute $\mathbf{w}^*$ and express your answer in terms of the matrix $\mathbf{O} = [\mathbf{o}_1, \dots, \mathbf{o}_n]$, and the vector $\mathbf{r} = [r_1, r_2, \dots, r_n]^\top$.

Solution:

$$f(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^{n} (r_i - \mathbf{w}^\top \mathbf{o}_i)^2 = \frac{1}{n} \|\mathbf{r} - \mathbf{O}^\top \mathbf{w}\|^2$$

Check that the hessian of f is

$$H = \frac{2}{n} \mathbf{O}\mathbf{O}^\top$$

and hence positive definite. This is a convex function and optimality is attaimed at

$$\nabla f(\mathbf{w}) = 0, \mathbf{O}\mathbf{O}^\top \mathbf{w} = \mathbf{O}\mathbf{r}$$

and hence $\mathbf{w}^* = \left(\mathbf{O}\mathbf{O}^\top\right)^{-1} \mathbf{O}\mathbf{r}$

(b) (5 points) Compute one iteration of Newton Method starting from $\mathbf{w}^{(0)} = 0$. State any assumption you make.

Solution:

$$\nabla f(\mathbf{w}^{(0)}) = \frac{2}{n}\mathbf{O}(\mathbf{O}^\top \mathbf{w}^{(0)} - \mathbf{r}) = -\frac{2}{n}\mathbf{r}$$

One newton iteration can be stated as follows

$$\mathbf{w}^{(1)} = \mathbf{w}^{(0)} - H^{-1}\nabla f(\mathbf{x}^{(0)}) = \left(\mathbf{O}\mathbf{O}^\top\right)^{-1}\mathbf{r}$$

The matrix $\mathbf{O}\mathbf{O}^\top$ need to be positive definite.

5. Let $f(\mathbf{x})$ be defined in Question 1.

   (a) (3 points) Derive the rank two Quasi-Newton update?

Solution:

(b) (2 points) State one iteration of rank two quasi-newton update using the exact line search?

Solution:

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

$$\alpha_k = \frac{\mathbf{g}_k H_k \mathbf{g}_k}{\mathbf{g}_k H_k Q H_k \mathbf{g}_k}$$

(c) (5 points) For f are the updates always Positive semidefinite? Repeat the same question if inexact line search is used. Give reasons?

Solution:
(**Partial answer**) Check that positive semidefiniteness holds if

$$\delta_k^\top \gamma_k \geq 0, \quad \delta_k = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \gamma_k = \mathbf{g}_{k+1} - \mathbf{g}_k$$

Because of Exact line search we know that $\mathbf{g}_{k+1}^\top(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} = 0$, and since $\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ is a feasible descent direction so $\mathbf{g}_k^\top \delta_k < 0$, we have

$$\delta_k^\top \gamma_k \geq 0.$$

For inexact line-searches the same holds if Wolfe condition is satisfied.

Rough Sheet 1

Rough Sheet 2

Rough Sheet 3

Rough Sheet 4