Continuous Assessment for EE6138 Optimization for Electrical Engineering (Advanced)

Semester 2, 20/21

This continuous assessment contains 2 questions with 100 marks in total. Answer all the questions. Please submit your answer in one single PDF file (with your name indicated as the file name) on the class website (under the folder "Student Submission") by **11:59pm April 18 2021** (a firm deadline with no extension allowed). Please show your working in all the answers and include all the used **Matlab codes** in your submitted PDF file.

1. [40 marks] Consider the following unconstrained problem,

$$\texttt{minimize} \quad f(x) = -\sum_{i=1}^{m} \log(1 - a_i^T x) - \sum_{i=1}^{n} \log(1 - x_i^2),$$

with variable $x \in \mathbf{R}^n$, and $\mathbf{dom}\ f = \{x | a_i^T x < 1, i = 1, \cdots, m, |x_i| < 1, i = 1, \cdots, n\}$. Note that this is the problem for computing the analytic center of the set of linear inequalities:

$$a_i^T x < 1, \quad i = 1, \cdots, m, \quad |x_i| \leq 1, \quad i = 1, \cdots, n.$$

(a) Show the gradient and Hessian of $f(x)$ are, respectively,

$$\nabla f(x) = \sum_{i=1}^{m} \frac{a_i}{1 - a_i^T x} - \left(\frac{1}{1 + x_1} - \frac{1}{1 - x_1}, \cdots, \frac{1}{1 + x_n} - \frac{1}{1 - x_n} \right)^T,$$

$$\nabla^2 f(x) = \sum_{i=1}^{m} \frac{a_i a_i^T}{(1 - a_i^T x)^2} + \mathrm{diag}\left(\frac{1}{(1 + x_1)^2} + \frac{1}{(1 - x_1)^2}, \cdots, \frac{1}{(1 + x_n)^2} + \frac{1}{(1 - x_n)^2} \right),$$

where $\mathrm{diag}(z)$ denotes a diagonal matrix with its main diagonal given by vector z.

[10 marks]

(b) Use *gradient method* to solve this problem. Consider the case of $m = 300$ and $n = 200$. Note that the parameters a_i's are stored in the matrix $A = [a_1, \cdots, a_m]^T$, which is given in the attached A.txt file. Please read matrix A in matlab using the command "A=dlmread('A.txt')" and then solve the problem based on this given A. Choose $x^{(0)} = 0$ as your initial point, and $\|\nabla f(x^{(k)})\|_2 \leq 10^{-3}$ as the stopping criterion for gradient method. Use the backtracking line search with each of the following four groups of parameters: $\alpha = 0.01$, $\beta = 0.1$; $\alpha = 0.01$, $\beta = 0.5$; $\alpha = 0.2$, $\beta = 0.1$; $\alpha = 0.2$, $\beta = 0.5$.

(b.1) Find the optimal value p^* of this problem obtained by gradient method.

(b.2) Plot $f(x^{(k)}) - p^*$ versus iteration for the given 4 sets of backtracking parameters in one figure and comment on the effect of backtracking parameters α and β on the total number of iterations required for convergence.

(b.3) Plot the step size $t^{(k)}$ versus iteration for the case of $\alpha = 0.01$, $\beta = 0.5$.

[15 marks]

(c) Use *Newton method* to solve this problem. Consider again the case of $m = 300$ and $n = 200$ and use the same matrix A as in part (b). Choose $x^{(0)} = 0$ as your initial point, and $\lambda(x^{(k)})^2 \leq 10^{-8}$ as the stopping criterion for Newton method. Set $\alpha = 0.01$ and $\beta = 0.5$ for the backtracking line search.

(c.1) Find the optimal value p^* of this problem obtained by Newton method.

(c.2) Plot $f(x^{(k)}) - p^*$ versus iteration and comment on the quadratic local convergence observed.

(c.3) Plot the step size $t^{(k)}$ versus iteration.

[15 marks]

2. [60 marks] Consider the following two-way partition problem,

$$\begin{aligned} (\text{P1}): \ \texttt{minimize} \quad & x^T W x \\ \texttt{subject to} \quad & x_i^2 = 1, \quad i = 1, \cdots, n, \end{aligned}$$

with variable $x \in \mathbf{R}^n$. We assume, without loss of generality, that $W \in \mathbf{S}^n$ and satisfies $W_{i,i} = 0$, $1 \le i \le n$. We denote the optimal value of (P1) as p^*, and the corresponding optimal solution as x^* (Note the $-x^*$ is also an optimal solution).

Problem (P1) is non-convex in general and its Lagrange dual problem can be shown to be the following semi-definite program (SDP),

$$\begin{aligned} (\text{P2}): \ \texttt{maximize} \quad & -1^T \nu \\ \texttt{subject to} \quad & W + \operatorname{diag}(\nu) \succeq 0, \end{aligned}$$

with variable $\nu \in \mathbf{R}^n$. Due to the non-convexity of (P1), the duality gap between (P1) and (P2) is in general non-zero.

The dual problem of (P2) is another SDP given by,

$$\begin{aligned} (\text{P3}): \ \texttt{minimize} \quad & \operatorname{Tr}(WX) \\ \texttt{subject to} \quad & X \succeq 0, \\ & X_{i,i} = 1, \quad i = 1, \cdots, n, \end{aligned}$$

with variable $X \in \mathbf{S}^n$. Both (P2) and (P3) are convex problems and satisfy the Slater's condition; thus, they have a zero duality gap.

(a) Use *barrier method* to solve (P2). Let the logarithmic barrier function of (P2) be denoted by

$$\Phi(\nu) = -\log\det(W + \operatorname{diag}(\nu)).$$

Then define $f(\nu) = t\mathbf{1}^T\nu + \Phi(\nu)$, where $t > 0$ and $\mathbf{1}$ is a vector with all the elements equal to 1. It can be shown that the gradient and Hessian of $f(\nu)$

with a given t are, respectively,

$$\nabla f(\nu) = t\mathbf{1} - \mathrm{Diag}((W + \mathrm{diag}(\nu))^{-1}),$$
$$\nabla^2 f(\nu) = (W + \mathrm{diag}(\nu))^{-1} \circ (W + \mathrm{diag}(\nu))^{-1},$$

where $\mathrm{Diag}(X)$ denotes a vector comprised of the main diagonal of a square matrix X; and $\circ$ denotes the element-wise product of two matrices. Consider the case of $n = 20$, and the matrix W given in the attached W.txt file. Please read matrix W in matlab using the command "W=dlmread('W.txt')" and then solve (P2) based on this given W. In each iteration of the barrier method, use Newton method to minimize $f(\nu)$. Let σ denote the minimum eigenvalue of the given W. To guarantee that $W + \mathrm{diag}(\nu) \succeq 0$, use $\nu^{(0)} = (1 - \sigma, \cdots, 1 - \sigma)$ as your initial point. Furthermore, use $\lambda(x^{(k)})^2 \leq 10^{-8}$ as the stopping criterion for Newton method, and $n/t \leq 10^{-6}$ for the barrier method.

(a.1) Find the optimal value d^* of (P2) obtained by barrier method.

(a.2) Let $\nu^{(k)}$ denote the optimal solution obtained after the kth iteration of the barrier method. Plot $-\frac{f(\nu^{(k)})}{t} - d^*$ versus barrier iterations.

[20 marks]

(b) Find the optimal dual solution to (P2) or the optimal solution to (P3), denoted by X^*. What is the rank of X^* that you obtained?

[10 marks]

Note that (P3) can be interpreted as a relaxation of the two-way partitioning problem (P1) without the constraint $\mathrm{rank}(X) = 1$. If the optimal solution to (P3), i.e., X^* obtained in part (b), has rank one, it must have the form $X^* = x^*(x^*)^T$, where x^* is the optimal solution to (P1). Otherwise, the optimal value of (P3) or that of (P2), d^*, provides a lower bound to that of (P1). In this case, to obtain a feasible solution to (P1), we need to find an approximation of X^* with rank one, i.e.,

$\hat{X} = \hat{x}\hat{x}^T$, and then $\hat{x}$ is a feasible solution to (P1). Consider the following three heuristic algorithms for obtaining such an approximation.

i. *simple partitioning.* Given the solution X^* of (P3), let v denote its dominant eigenvector associated with the largest eigenvalue. We then set $\hat{x} = \text{sign}(v)$.

ii. *randomized method.* Given the solution X^* of (P3), we generate K independent samples $x^{(1)}, \cdots, x^{(K)}$ from a normal distribution on $\mathbf{R}^n$, with zero mean and covariance matrix X^*. For each sample we obtain an approximated solution $\hat{x}^{(k)} = \text{sign}\left(x^{(k)}\right)$. We then take the best among $\hat{x}^{(k)}$'s, i.e., the one with the lowest value of (P1) as $\hat{x}$.

iii. *greedy refinement.* Since for any given partition x, i.e., $x_i \in \{-1, 1\}$, $i = 1, \cdots, n$, the objective value of (P1) changes if we move element i from one set to the other, i.e., changing x_i to $-x_i$. Thus, we can move the element in x which gives the largest reduction in the objective value of (P1), and repeat this procedure until we find an $\hat{x}$ for which no reduction in the objective value can be obtained by moving an element from one set to the other.

With d^* obtained in part (a) and X^* obtained in part (b), do the following parts.

(c) Find the optimal solution x^* and optimal value p^* of (P1) by an *exhaustive search.* Compare your obtained p^* with d^* to check whether (P1) and (P2) have a zero duality gap or not (or in other words, whether the rank-relaxed SDP (P3) is tight for (P1) or not).

[5 marks]

(d) Find the feasible solution $\hat{x}$ obtained by the *simple partitioning*, and the corresponding value of (P1) with $x = \hat{x}$. Compare your obtained objective value with the lower bound d^* as well as the optimal value p^* for (P1).

[5 marks]

(e) Find the feasible solution $\hat{x}$ obtained by *randomized method* with $K = 100$, and the corresponding value of (P1) with $x = \hat{x}$. Compare your obtained objective value with the lower bound d^* as well as the optimal value p^* for (P1).

[10 marks]

(f) For *greedy refinement*, try this algorithm with three initial points: $x = 1$, the solution obtained in (d), and the solution obtained in (e). For each of these three initial points, find the feasible solution $\hat{x}$ obtained by greedy refinement algorithm, and the corresponding value of (P1) with $x = \hat{x}$. Compare your obtained objective values with the lower bound d^* as well as the optimal value p^* for (P1).

[10 marks]

– END –