PAPER C477

COMPUTATIONAL OPTIMISATION

Monday 16 March 2020, 15:00
Duration: 120 minutes
Post-processing time: 30 minutes
Answer THREE questions

Paper contains 4 questions

1 **The Newton Algorithm**

a Suppose that $A \in \mathbb{R}^{n \times n}$ is a non-singular matrix i.e. A^{-1} exists. Derive the directions generated by the Newton algorithm on the two problems:

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\min_{y \in \mathbb{R}^n} f(Ay).$$

b Let d_x denote the Newton direction of the first problem above and d_y denote the Newton direction of the second. Show that when $x = Ay$

$$d_x = Ad_y.$$

c Consider the following function: $f(x) = x^r$ ($x \in \mathbb{R}$). Show that there exists a real number $0 < \delta < 1$, and a natural number r (i.e $r = 1, 2, \ldots$) such that f is convex, has a unique global minimum and that the Newton algorithm applied to this function satisfies the following:

$$|x_{k+1} - x^\star| \geq (1 - \delta)|x_k - x^\star|$$

Where x_k denotes the k^{th} iterate of the Newton algorithm (you may assume that $x_0 > 0$) and $x^\star$ denotes the optimal solution.

The three parts carry, respectively, 25%, 25%, and 50% of the marks.

2 **Constrained Optimality Conditions**

Consider the following Constraint Least Squares problem,

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 \qquad \text{(CLS)}$$
$$\text{s.t. } \|x\|_2^2 \leq \alpha.$$

Where $A \in \mathbb{R}^{m\times n}$ is a matrix that has full column rank, $b \in \mathbb{R}^m$, and $\alpha > 0$ is a scalar.

a State (but do not solve) the necessary and sufficient optimality conditions for (CLS).

b Suppose that,

$$\|(A^\top A)^{-1}A^\top b\|_2^2 \leq \alpha.$$

Find a point that satisfies the necessary and sufficient optimality optimality conditions of (CLS).

c Suppose that,

$$\|(A^\top A)^{-1}A^\top b\|_2^2 > \alpha.$$

Show that in this case there exists a pair $(x^\star, \lambda^\star) \in (\mathbb{R}^n, \mathbb{R})$ satisfies the optimality conditions. In particular $x^\star$ is given by,

$$x^\star = (A^\top A + \lambda^\star I)^{-1}A^\top b,$$

where $\lambda^\star$ is the solution to the following one-dimensional equation,

$$\|(A^\top A + \lambda^\star I)^{-1}A^\top b\|_2^2 - \alpha = 0.$$

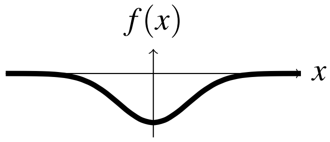
The three parts carry, respectively, 35%, 25%, and 40% of the marks.

3 **Convexity**

a Consider the function $f(x) : \mathbb{R} \mapsto \mathbb{R}$:

$$f(x) = -e^{-x^2}.$$

The function $f(x)$ looks like this:



Show that $f(x)$ is convex on the open interval $x \in \left(-1/\sqrt{2}, 1/\sqrt{2}\right)$. Show $f(x)$ is both nonconvex and nonconcave on $x \in \mathbb{R}$. Show that the only stationary point is $x = 0$, $f(x) = -1$.

b Let S be a nonempty, convex set in $\mathbb{R}^n$. A function $f(\mathbf{x}) : S \mapsto \mathbb{R}$ is *pseudo-convex* if, for all $\mathbf{x}_1, \mathbf{x}_2 \in S$:

$$[\nabla f(\mathbf{x}_1) \cdot (\mathbf{x}_2 - \mathbf{x}_1) \geq 0] \text{ implies that } [f(\mathbf{x}_2) \geq f(\mathbf{x}_1)].$$

Show that a convex function is also pseudo-convex. Using an example, show that a pseudo-convex function is not necessarily convex.

Hint Try the function in Part a on the interval $x \in [0, \infty)$.

c Show that $f(x) = -e^{-x^2}$ on the interval $x \in [0, \infty)$ has its global mimimum at $x = 0$, $f(x) = -1$.

The three parts carry, respectively, 40%, 50%, and 10% of the marks.

4 First-Order, Gradient-Based Methods

Consider minimizing the *Leon function*:

$$\min_{x_1,x_2} f(x_1,x_2) = \min_{x_1,x_2} 100(x_2 - x_1^3)^2 + (1-x_1)^2.$$

The Leon function has gradient:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 600 \cdot x_1^5 - 600 \cdot x_1^2 \cdot x_2 - 2 \cdot (1-x_1) \\ 200(x_2 - x_1^3) \end{bmatrix},$$

and Hessian:

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 3000 \cdot x_1^4 - 1200 \cdot x_1 \cdot x_2 + 2 & -600 \cdot x_1^2 \\ -600 \cdot x_1^2 & 200 \end{bmatrix}.$$

The Leon function has its unique global minimum at $(x_1,x_2) = (1,1)$.

a Suppose that we have implemented a first-order, gradient based method. The current iterate is $\mathbf{x}^{(k)} = (x_1^k, x_2^k) = (0,0)$. Show that the direction of steepest descent with $\|\mathbf{d}\|_2 = 1$ is $\mathbf{d} = [1,0]^T$. For steepest descent with an exact step size strategy, show that the optimal step size α is the solution to:

$$600\alpha^5 + 2\alpha - 2 = 0.$$

Numerically, $\alpha \approx 0.30$.

b The Leon function is ill-conditioned, so the optimal step size α in Part a is fairly small. What is the condition number at the k^{th} iterate, $\mathbf{x}^{(k)} = (0,0)$?

c Now consider a *scaled* gradient method with a diagonal scaling matrix $\mathbf{D}_k$:

$$\mathbf{D}_k = \begin{bmatrix} \left(\nabla^2 f(0,0)\right)_{11}^{-1} & 0 \\ 0 & \left(\nabla^2 f(0,0)\right)_{22}^{-1} \end{bmatrix}$$

Assuming that we are still using a steepest descent direction and an exact step size strategy, show that the new optimal step size is the solution to:

$$9.375\alpha^3 + \frac{1}{2}\alpha - 1 = 0.$$

Numerically, $\alpha \approx 0.60$ and the new step size is larger with the scaled gradient method.

The three parts carry, respectively, 35%, 30%, and 35% of the marks.