IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2018-2019

MEng Honours Degree in Electronic and Information Engineering Part IV
MEng Honours Degree in Mathematics and Computer Science Part IV
MEng Honours Degrees in Computing Part IV
MSc in Advanced Computing
MSc in Computing Science (Specialist)
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the
Associateship of the City and Guilds of London Institute

PAPER C477

COMPUTATIONAL OPTIMISATION

Friday 14th December 2018, 10:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators not required

1 **Projected Gradient Method**

Consider the optimisation problem,

$$\begin{aligned} \min \ & f(x) \\ \text{s.t} \ & x \in \Omega. \end{aligned}$$

Where $f(x) = \frac{1}{2}x^\top Qx$, and $Q \in \mathbb{R}^{n\times n}$ is symmetric and positive definite. The constraint set is defined as follows, $\Omega = \{x \in \mathbb{R}^n \mid \|x\|_2^2 = 1\}$.

At iteration k, the *projected gradient algorithm* for this problem is,

$$x_{k+1} = \Pi[x_k + \alpha d_k].$$

Where $x_k \in \Omega$ is the current point, $d_k = -\nabla f(x_k)$, and α is a fixed step size. The operator Π is called a *projection operator* and is defined as follows,

$$\Pi[x] = \arg\min_{z\in\Omega} \|z - x\|_2^2$$

a Derive a formula for the update equation of the algorithm (i.e. write down an explicit formula for x_{k+1} as a function for x_k, Q and the fixed step size α). You may assume that the argument in the projection operator to obtain x_k is never zero.

b Show that for $0 < \alpha < 1/\lambda_{\max}$ (where $\lambda_{\max}$ is the largest eigenvalue of Q), the fixed step size projected gradient algorithm (with step size α) converges to an optimal solution, provided that x_0 is not orthogonal to the eigenvector of Q corresponding to the smallest eigenvalue. You may assume that the eigenvalues of Q are distinct.

The two parts carry, respectively, 30% and 70% of the marks.

2 KKT Optimality Conditions

a Consider the problem,

$$\min\ x_1^2 + (x_2+1)^2$$
$$\text{subject to } x_2 \geq \exp(x_1)$$

Let $x^\star$ denote the optimal solution of the problem above. Write down the KKT condition that must be satisfied by $x^\star$. In addition show that $x_2^\star = \exp(x_1^\star)$ and that $-2 < x_1^\star < 0$.

b Consider the problem,

$$\min\ c^\top x + 8$$
$$\text{subject to } \frac{1}{2}\|x\|_2^2 \leq 1,$$

where $c \in \mathbb{R}^n$, $c \neq 0$. Suppose that $x^\star = \alpha\mathbf{1}$ is an optimal solution of the problem, where $\alpha \in \mathbb{R}$ and $\mathbf{1} = [1,\ldots,1]^\top$, and that the corresponding objective function value is 4. Show that $\|x^\star\|_2^2 = 2$ and find α and c (they may depend on n).

c Consider the problem,

$$\min\ f(x)$$
$$\text{subject to } h(x) = 0,$$

You may assume that the functions used in the definitions above satisfy the assumptions of the KKT theorem.

We can convert the above into an equivalent optimisation problem,

$$\min\ f(x)$$
$$\text{subject to } \frac{1}{2}\|h(x)\|_2^2 \leq 0. \tag{In-Eq}$$

Write down the KKT conditions for both problems and explain why the KKT theorem cannot be applied to problem (In-Eq).

The three parts carry, respectively, 30%, 30%, and 40% of the marks.

3 **Convexity**

a Consider the function $f(x) : \mathbb{R} \mapsto \mathbb{R}$ where $p \in \mathbb{R}$ is a scalar:

$$f(x) = |x|^p.$$

Show that, when $p \geq 1$, $f(x)$ is convex. Show that, when $0 < p < 1$, $f(x)$ is neither convex nor concave. Further, show that, when $0 < p < 1$, $f(x)$ is concave on the interval $(-\infty, 0]$ and is also concave on the interval $[0, \infty)$.

b Let S be a nonempty, convex set in $\mathbb{R}^n$. A function $f(\mathbf{x}) : S \mapsto \mathbb{R}$ is called *quasi-convex* if, for all $\lambda \in [0,1]$ and $\mathbf{x}_1, \mathbf{x}_2 \in S$:

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \max[f(\mathbf{x}_1), f(\mathbf{x}_2)].$$

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

Hint Try the function in Part a.

c Consider the function $f(x,y) : \mathbb{R}^2 \mapsto \mathbb{R}$ where $0 < p < 1$ is a scalar:

$$f(x,y) = -x^p y^{1-p}$$

Show that $f(x,y)$ is convex for $x, y > 0$.

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

4 **First-Order, Gradient-Based Methods**

Consider minimizing the *Rosenbrock function*:

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

which has gradient:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} -400 \cdot (x_2 - x_1^2) \cdot x_1 - 2 \cdot (1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}.$$

The Rosenbrock 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 $\mathbf{d} = [1,0]^T$. If we choose steepest descent with an exact step size strategy, show that the optimal step size α is the solution to:

$$400\alpha^3 + 2\alpha - 2 = 0.$$

Numerically, $\alpha \approx 0.16$.

b The Rosenbrock function is infamous for being ill-conditioned, so you may have noticed in Part a that the optimal step size α is fairly small. What is the condition number at the k^{th} iterate, $\mathbf{x}^{(k)} = (x_1^k, x_2^k) = (0,0)$?

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:

$$25\alpha^3 + \frac{1}{2}\alpha - 1 = 0.$$

Numerically, $\alpha \approx 0.32$ and we see that the new step size is larger with the scaled gradient method.

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