FINAL

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2017-2018

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)
MRes in Advanced Computing
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

Tuesday 12 December 2017, 10:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators not required

1 Search in One Dimension

For problem parts building on each other: If you cannot complete one part, just write how you would proceed given an answer to the previous part

a Optimality Conditions in One Dimension

i) Let $f(x) : \mathbb{R} \mapsto \mathbb{R}$ be a continuous function defined on the interval $[a, b] \subset \mathbb{R}$. Assume that $f'(x) < 0$ for all x in the open set (a, b). What can you conclude about the global and local extreme points, i.e. points x where $f(x)$ achieves a minimum or a maximum, and the corresponding values of f?

ii) Let $f(x) : \mathbb{R} \mapsto \mathbb{R}$ be a function defined on the interval $[-a, a] \subset \mathbb{R}$. Assume that $f'(x) < 0$ for all $-a < x < 0$ and $f'(x) > 0$ for all $0 < x < a$. Why can't you conclude anything about the global and local extreme points, i.e. x where $f(x)$ achieves a minimum or a maximum, and the corresponding values of f?

b Root-Finding in One Dimension

Root-Finding in One Dimension Consider the problem of finding the zero(s) of $g(x) = (e^{2 \cdot x} - 1)/(e^{2 \cdot x} + 1)$, $x \in \mathbb{R}$.

i) What are the value(s) of x for which $g(x) = 0$?

ii) Write down the algorithm for Newton's method to find a zero of this function $g(x)$. If you wish, simplify your calculations using the identity $\sinh(x) = (e^x - e^{-x})/2$.

iii) Find an initial condition x_0 such that Newton's method cycles. You need not explicitly calculate the initial condition; it suffices to provide an equation that the initial condition must satisfy.

iv) For what initial condition values does Newton's method converge?

The two parts carry, respectively, 40% and 60% of the marks.

2 Convexity & Optimisation

a Consider the function $f(x,y) : \mathbb{R}^2 \mapsto \mathbb{R}$ where $\beta \in \mathbb{R}$ is a scalar:

$$f(x,y) = x^2 + y^2 + \beta xy + x + 2y$$

Observe that the Hessian for this function is:

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & \beta \\ \beta & 2 \end{bmatrix}$$

i) Show that, when $-2 \leq \beta \leq 2$, $f(x,y)$ is convex. Also show that otherwise $f(x,y)$ is indefinite, i.e. neither convex nor concave.

ii) For each scalar $\beta \in \mathbb{R}$, what is the point (or set of points) satisfying the *First Order Necessary Condition* for $f(x,y)$?

iii) For each scalar $\beta \in \mathbb{R}$, please classify which points satisfying the FONC are local and/or global minima and/or maxima? Which are saddle points?

iv) For each scalar $\beta \in \mathbb{R}$, characterise the behaviour of a steepest descent, first order method with exact step size. For what values of β and initial conditions (x_0, y_0) does this algorithm converge? Are there conditions where the algorithm will converge, but not to a local minimum?

b Suppose $h : \mathbb{R}^m \mapsto \mathbb{R}$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, and $\boldsymbol{b} \in \mathbb{R}^m$. We define:

$$g_1(\boldsymbol{x}) = h(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}).$$

If h is convex, then so is g_1. In other words, a convex function composed with an affine mapping is also convex. Now consider the reverse composition, i.e. an affine mapping composed with a convex function:

$$g_2(\boldsymbol{x}) = a \cdot h(\boldsymbol{x}) - b,$$

where $a, b \in \mathbb{R}$. What are necessary and sufficient conditions for g_2 to be convex?

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

3 Lagrange Multipliers & Optimality Conditions

a Consider the following optimisation problem,

$$\begin{aligned} \min_{\boldsymbol{x}} \ & f(\boldsymbol{x}) \\ \text{s.t}\ & \boldsymbol{h}(\boldsymbol{x}) \le 0, \end{aligned}$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $h = [h_1, \dots, h_m]^T$, and $m \le n$. Suppose that the function f is convex, and that the feasible region is also convex. State the KKT necessary and sufficient conditions for optimality of the problem above. If a point $(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)$ satisfies the KKT conditions, is the point $\boldsymbol{x}^\star$ a unique global minimum for this problem? Justify your answer.

b Consider the following optimisation problem,

$$\begin{aligned} \min_{x_1, x_2} \ & x_1 + x_2 \\ \text{s.t}\ & (x_1 - 1)^2 + x_2^2 - 1 = 0, \\ & (x_1 - 2)^2 + x_2^2 - 4 = 0. \end{aligned}$$

Show that this problem has no Lagrange multipliers.

c Consider the following optimisation problem,

$$\begin{aligned} \min_{x_1, x_2, x_3} \ & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{s.t}\ & x_1 + x_2 + x_3 = 1. \end{aligned}$$

Solve this problem using the KKT optimality conditions. If a point $(x_1^\star, x_2^\star, x_3^\star, \lambda^\star)$ satisfies the optimality conditions, then is it a global or local minimum?

d Suppose that the objective function of the problem in part (c) is changed to: $-\frac{1}{2}(x_1^2 + x_2^2 + x_3^2)$ (the constraint remains the same). Show that in this case the problem has no finite solution.

The four parts carry, respectively, 25%, 25%, 20%, and 30% of the marks.

4 **The Newton Method with an approximate Hessian**

Consider the following unconstrained optimisation problem,

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

Where $f:\mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function. Suppose that the problem above is solved using the following variant of the Newton method:

(Step 0) : **Initialisation:** Given a starting point $\boldsymbol{x}_0$ in the domain of f, set $k = 0$

(Step 1) : **Compute direction:** Let $\Delta\boldsymbol{x}_k$ be the solution of $\boldsymbol{B}_k\Delta\boldsymbol{x}_k = -\nabla \boldsymbol{f}(\boldsymbol{x}_k)$.

(Step 2) : **Update:** Determine a scalar step size $t_k > 0$ and update the solution,

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t_k\Delta\boldsymbol{x}_k$$

(Step 3) : If the convergence criteria are not satisfied, set $k := k+1$ and go to 1.

The matrix $\boldsymbol{B}_k$ is chosen as follows:

$$\boldsymbol{B}_k = \begin{cases} \nabla^2 f(\boldsymbol{x}_k) & \text{if } \nabla^2 f(\boldsymbol{x}_k) \succ 0 \\ \boldsymbol{I} & \text{otherwise} \end{cases}$$

Where $\boldsymbol{I}$ is the n-dimensional identity matrix.

a Show that $\Delta\boldsymbol{x}_k$ is a descent direction at $\boldsymbol{x}_k$.

b Specify an appropriate convergence criterion for the problem above. Justify your choice.

c Given $\Delta\boldsymbol{x}_k$, let $\alpha \in (0, 0.5)$, $\beta \in (0,1)$ and suppose that t_k is chosen according to the following step size strategy:

$$\text{While } f(\boldsymbol{x}_k + t_k\Delta\boldsymbol{x}_k) > f(\boldsymbol{x}_k) + \alpha t_k\langle\nabla f(\boldsymbol{x}_k), \Delta\boldsymbol{x}_k\rangle \text{ set } t_k \text{ to } \beta t_k$$

Show that $f(\boldsymbol{x}_k + t_k\Delta\boldsymbol{x}_k) \leq f(\boldsymbol{x}_k) + \alpha t_k\langle\nabla f(\boldsymbol{x}_k), \Delta\boldsymbol{x}_k\rangle$, for some $0 < t < b$ and find an explicit expression for b.

Hint: You may use the following inequality (known as the Descent Lemma),

$$f(\boldsymbol{x}_k + t_k\Delta\boldsymbol{x}_k) - f(\boldsymbol{x}_k) \leq t_k\langle\nabla f(\boldsymbol{x}_k), \Delta\boldsymbol{x}_k\rangle + \frac{M}{2}t_k^2\|\Delta\boldsymbol{x}_k\|^2.$$

Where $M > 0$ is a scalar.

d Suppose that it is easy to compute the eigenvalues of the Hessian of f. Briefly describe an alternative way to compute a suitable matrix B_k in Step 1.

The four parts carry equal marks.