IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2017

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

Monday 12 December 2016, 10:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators not required

1 **Convexity** For each set, please answer: Is the set convex or nonconvex? For each function, please specify if the function is: (Strictly) Convex? (Strictly) Concave? Neither convex nor concave? Please clearly justify your mathematical reasoning.

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

a Is the set Ω convex?

$$\Omega = \left\{ \boldsymbol{x} \in \mathbb{R}^2 \;\middle|\; 3x_1^2 + \frac{1}{2}x_2^2 + 4x_1x_2 - x_1 + 4x_2 \leq 10 \right\}.$$

b Let $g_1, g_2 : \mathbb{R} \mapsto \mathbb{R}$ be two convex functions. Characterise the convexity of the maximum function $h : \mathbb{R} \mapsto \mathbb{R}$:

$$h(x) = \max\{g_1(x), g_2(x)\}.$$

c Consider the **soft-margin support vector machine** classifier:

$$\min_{\boldsymbol{w},b} f(\boldsymbol{w}, b) = \min_{\boldsymbol{w},b} \left[\frac{1}{n} \sum_{i=1}^{n} \max\left\{0,\, 1 - y_i \left(\boldsymbol{w}^\top \boldsymbol{x_i} + b\right)\right\} \right] + \lambda\, \|\boldsymbol{w}\|_2^2,$$

where $\boldsymbol{w} \in \mathbb{R}^p$ and $b \in \mathbb{R}$ are decision variables and $f : \mathbb{R}^{p+1} \to \mathbb{R}$ is the objective. We are given parameters $\lambda \in \mathbb{R}$ and n points of the form $(\boldsymbol{x}_i, y_i)$ where, for each $i \in \{1, \ldots, n\}$, $\boldsymbol{x}_i \in \mathbb{R}^p$ and $y_i \in \{-1, 1\}$.

i) Characterise the convexity of the function $g_i : \mathbb{R}^{p+1} \to \mathbb{R}$:

$$g_i(\boldsymbol{w}, b) = 1 - y_i \left(\boldsymbol{w}^\top \boldsymbol{x_i} + b\right).$$

ii) Characterise the convexity of the function $h_i : \mathbb{R}^{p+1} \to \mathbb{R}$:

$$h_i(\boldsymbol{w}, b) = \max\{0, g_i(\boldsymbol{w}, b)\}.$$

iii) Characterise the convexity of the function $\|\boldsymbol{w}\|_2^2$.

iv) Characterise the convexity of the minimisation problem corresponding to the soft-margin support vector machine classifier.

The three parts carry, respectively, 20%, 35%, and 45% of the marks.

2 First-order, gradient-based methods.

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

a **Inexact, backtracking line search** Consider unconstrained minimisation problem:

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

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is twice-continuously differentiable. Assume that we are solving using first-order, steepest-descent with backtracking line search.

Recall that backtracking line search initialises parameters c_α, c_β. Each iteration k starts with stepsize $\alpha_k = 1$ and while:

$$f\left(\boldsymbol{x}^{(k)} - \alpha_k \nabla f\left(\boldsymbol{x}^{(k)}\right)\right) > f(\boldsymbol{x}^{(k)}) - c_\alpha \alpha_k \left\|\nabla f\left(\boldsymbol{x}^{(k)}\right)\right\|_2^2,$$

sets $\alpha_k = c_\beta \alpha_k$. Otherwise, backtracking line search performs a gradient descent update: $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \nabla f\left(\boldsymbol{x}^{(k)}\right)$.

Please specify and justify bounds, i.e. a range of reasonable values, for c_α and c_β.

Recall: In C477, $\nabla f\left(\boldsymbol{x}^{(k)}\right) \in \mathbb{R}^n$ is a column vector.

b **Exact line search for quadratic functions** Consider unconstrained minimisation problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \min_{\boldsymbol{x}\in\mathbb{R}^n} \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + 2\boldsymbol{b}^\top \boldsymbol{x},$$

where $\boldsymbol{A} \succ \boldsymbol{0}$ is an $n \times n$ positive definite matrix and $\boldsymbol{b} \in \mathbb{R}^n$. Let $\boldsymbol{x}^{(0)} \in R^n$ be the starting point and let $\boldsymbol{d}^{(0)} \in \mathbb{R}^n$ be a descent direction of f at $\boldsymbol{x}^{(0)}$. Stepsize α is generated by exact line search:

$$\alpha = \arg\min_{\alpha \geq 0} \; f(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}^{(0)})$$

i) Please develop an explicit formula expressing α as an algebraic function of $\boldsymbol{A}$, $\boldsymbol{b}$, $\boldsymbol{x}^{(0)}$, and $\boldsymbol{d}^{(0)}$.

ii) Now assume that the descent direction $\boldsymbol{d}^{(0)}$ is chosen so that we are using steepest descent. Please develop an explicit formula expressing α as an algebraic function of $\boldsymbol{A}$, $\boldsymbol{b}$, and $\boldsymbol{x}^{(0)}$.

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

3 **Optimality Conditions**

a Suppose that X is a non-empty convex set, and c is a scalar. Show that if there exist,

$$\lambda_j \geq 0 \quad j = 1, \ldots, m,$$

such that the following condition is satisfied,

$$\min_{\boldsymbol{x} \in X} \{f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x})\} \geq c,$$

then there does not exist an $\boldsymbol{x}$ to satisfy the following conditions,

$$\begin{aligned} f(\boldsymbol{x}) &< c, \\ g_j(\boldsymbol{x}) &\leq 0, \quad j = 1, \ldots, m, \\ x &\in X. \end{aligned}$$

b Consider the following optimisation problem,

$$\begin{aligned} f^\star = \min_{x,y} & \exp(-2x) + y^2 \\ & x + y = 1. \end{aligned}$$

Use the result in part (a) to show that $f^\star \geq 1$.

c Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $m \leq n$, and suppose that the rank of $\boldsymbol{A}$ is m. Let $\boldsymbol{x}^\star$ be defined as follows,

$$\begin{aligned} \boldsymbol{x}^\star = \min_{\boldsymbol{x}} \quad & \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2 \\ & \boldsymbol{A}\boldsymbol{x} = 0. \end{aligned}$$

(i) : Show that $(\boldsymbol{x}^\star - \boldsymbol{x}_0)^\top \boldsymbol{x}^\star = 0$.

(ii) : Find $\boldsymbol{x}^\star$ as a function of $\boldsymbol{A}$ and $\boldsymbol{x}_0$.

d Let $\boldsymbol{x}^\star$ be a local minimiser (or maximiser) of $f : \mathbb{R}^2 \to \mathbb{R}$ subject to $h(\boldsymbol{x}) = 0$ (note that $h : \mathbb{R}^2 \to \mathbb{R}$). Show that there exists a scalar $\lambda^\star \in \mathbb{R}$ such that,

$$\nabla f(x^\star) + \nabla h(x^\star) \lambda^\star = 0.$$

The four parts carry, respectively, 25%, 10%, 25%, and 40% of the marks.

4 **Projection Algorithms**

Consider the following linearly constrained problem,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t. } & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{aligned}$$

Where $f : \mathbb{R}^n \to \mathbb{R}$ is a twice differentiable strongly convex function, and the matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has rank m with $m < n$.

a Derive the gradient projection algorithm for this problem i.e. show that at iteration $k+1$, $\boldsymbol{x}_{k+1}$ is given by,

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k (\boldsymbol{I} - \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top)^{-1} \boldsymbol{A}) \boldsymbol{\nabla} f(\boldsymbol{x}_k),$$

where $\alpha_k > 0$ is the stepsize parameter. You may assume that $\boldsymbol{x}_k$ is feasible.

b Show that if $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}$ then all iterates of the gradient projection algorithm are feasible.

c Is the gradient projection algorithm a descent algorithm? Justify your answer.

d At iteration k, the Projected Newton Algorithm follows the iterative scheme below,

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{P} \boldsymbol{d}_k,$$

where α_k is a stepsize parameter and the direction d_k is given by the solution of the following problem,

$$\begin{aligned} \boldsymbol{d}_k \in \arg\min \ & \boldsymbol{d}^\top \nabla f(\boldsymbol{x}_k) + \frac{1}{2} \boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}_k) \boldsymbol{d} \\ \text{s.t. } & \boldsymbol{A}\boldsymbol{d} = 0 \end{aligned}$$

Derive an expression for $\boldsymbol{x}_{k+1}$ given that $\boldsymbol{x}_k$ is feasible. Is the direction $\boldsymbol{d}_k$ a feasible descent direction? Justify your answer.

e Describe possible modifications of the projected Newton algorithm if the function is non-convex.

The five parts carry, respectively, 15%, 10%, 15%, 45%, and 15% of the marks.