Imperial College
London

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May – June 2011

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science.

Optimisation

Date: Monday, 16 May 2011.   Time: 2.00pm.   Time allowed: 2 hours.

This paper has FOUR questions.

Candidates should use ONE main answer book.

Supplementary books may only be used after the main book is full.

Statistical tables will not be provided.

- DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO.

- Answer all the questions. Each question carries equal weight.

- Credit will be given for all questions attempted, but extra credit will be given for complete or nearly complete answers.

- Affix one of the labels provided to each answer book that you use, but DO NOT USE THE LABEL WITH YOUR NAME ON IT.

- Calculators may not be used.

M3/M4/M5 N3

1. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a smooth function and consider Newton's method to minimise f.

   (a) Explain carefully the underlying strategy for obtaining the next iterate x_{i+1} from the current iterate x_i. [It is not enough just to state the formula!] What condition on the Hessian matrix $\nabla^2 f(x_i) \equiv H(x_i)$ needs to hold in order for this strategy to make sense?

   (b) Suppose $x^\star \in \mathbb{R}^n$ satisfies $\nabla f(x^\star) \equiv g(x^\star) = 0$ with $H(x^\star)$ positive definite, and H also satisfies the Lipschitz condition

   $$\|H(x) - H(y)\| \leq \gamma(d) \|x - y\| \quad \forall x, y \in \overline{B}(x^\star, d)$$

   for all $d > 0$, where $\overline{B}(x^\star, d) \equiv \{x \in \mathbb{R}^n : \|x - x^\star\| \leq d\}$. Carry out the following steps to deduce that the Newton iterates converge quadratically to $x^\star$ from any starting value x_0 close enough to satisfy

   $$\|H(x^\star)^{-1}\| \, d_0 \, \gamma(d_0) < \tfrac{2}{3},$$

   where $d_0 \equiv \|x_0 - x^\star\|$.

   (i) Assuming the result that, if A, B are symmetric $n \times n$ matrices with A non-singular, then

   $$\|A - B\| < \frac{1}{\|A^{-1}\|}$$

   implies that B is also non-singular with

   $$\|B^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\|A - B\|},$$

   prove that $H(x)$ is non-singular, and hence positive definite, for $x \in \overline{B}(x^\star, d_0)$ with

   $$\|H(x)^{-1}\| < 3 \|H(x^\star)^{-1}\|.$$

   (ii) Use the Newton formula for x_{i+1} and the integral mean value theorem to prove that, if $x_i \in \overline{B}(x^\star, d_0)$, then

   (†) $\|x_{i+1} - x^\star\| < \tfrac{3}{2} \|H(x^\star)^{-1}\| \, \gamma(d_0) \|x_i - x^\star\|^2$.

   (iii) Use (†) to prove by induction that the Newton iterates remain in $\overline{B}(x^\star, d_0)$.

   (iv) Use (†) to prove that the Newton iterates converge to $x^\star$, and hence converge quadratically.

2. (a) Explain carefully <u>one</u> step of the Newton + trust region algorithm for the unconstrained optimisation problem

$$\min f(x),$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a smooth function, which may be applied even when the current Hessian matrix fails to be positive definite. Your explanation should include the following points.

(i) What are the algorithm parameters?

(ii) What is the input data at the beginning of the step, and what is the output data ready for the next step?

(iii) What is the constrained quadratic optimisation problem that provides a tentative increment? [You do <u>not</u> have to describe how to solve constrained quadratic optimisation problems.]

(iv) How does the algorithm decide whether to accept or reject the tentative increment, and what strategy is followed in each case?

(b) Consider the constrained quadratic optimisation problem

$$\min q(x) \equiv c - b^T x + \tfrac{1}{2}x^T G x \quad \text{subject to } \|x\| \leq \Delta,$$

where $q : \mathbb{R}^n \mapsto \mathbb{R}$ with $c \in \mathbb{R}$, $b \in \mathbb{R}^n$, G an $n \times n$ symmetric matrix, and $\Delta > 0$. Prove that x^* solves this problem if and only if $\exists \mu^* \geq 0$ such that

- $[G + \mu^* I] x^* = b$,
- the complementarity condition $\mu^* [\|x^*\| - \Delta] = 0$ holds,
- $G + \mu^* I$ is positive semi-definite.

In addition, if $G + \mu^* I$ is positive definite, show that x^* is the unique solution of the constrained quadratic optimisation problem.

[You are reminded of the following result, which you may use without proof.

If $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $c : \mathbb{R}^n \mapsto \mathbb{R}$ are smooth functions and $y^* \in \mathbb{R}^n$ is a local minimum for the inequality constrained problem

$$\min f(x)$$
$$\text{subject to } c(x) \geq 0,$$

then $\exists \lambda^* \geq 0$ such that

$$g(y^*) = \lambda^* g_c(y^*), \quad \lambda^* c(y^*) = 0$$
$$g_c(y^*)^T x = 0 \quad \Rightarrow \quad x^T [H(y^*) - \lambda^* H_c(y^*)] x \geq 0,$$

where $g(x) \equiv \nabla f(x)$, $H(x) \equiv \nabla^2 f(x)$, $g_c(x) \equiv \nabla c(x)$ and $H_c(x) \equiv \nabla^2 c(x)$.]

3. Consider the linear equality constrained problem

$$(\dagger) \qquad \begin{aligned} &\min f(\boldsymbol{x}) \\ &\text{subject to } A\boldsymbol{x} = \boldsymbol{d}; \end{aligned}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a smooth function, $\boldsymbol{d} \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$ has full rank with $0 < m < n$.

(a) If the constraint set is denoted

$$S \equiv \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{d}\},$$

prove that an equivalent formulation is

$$S \equiv \{\widehat{\boldsymbol{x}} + Z\boldsymbol{v} : \forall \boldsymbol{v} \in \mathbb{R}^{n-m}\},$$

where $\widehat{\boldsymbol{x}} \in \mathbb{R}^n$ is a feasible point for $(\dagger)$ and the columns of $Z \in \mathbb{R}^{n \times (n-m)}$ span the null-space of A.

(b) Use this equivalent formulation to explain how $(\dagger)$ may be replaced by an equivalent smaller unconstrained problem

$$(\ddagger) \quad \min \widehat{f}(\boldsymbol{v}),$$

where $\widehat{f} : \mathbb{R}^{n-m} \mapsto \mathbb{R}$.

(i) Determine the gradient vectors and Hessian matrices for $\widehat{f}$ in terms of Z and the gradient vectors and Hessian matrices respectively of f.

(ii) Hence state the second derivative necessary and sufficient conditions for $\boldsymbol{x}^* \in S$ to be a local minimum for $(\dagger)$.

(iii) If $\boldsymbol{x}_i \in S$, write down the formula for $\boldsymbol{x}_{i+1}$ when applying Newton's method to $(\ddagger)$. What condition on the Hessian matrix at $\boldsymbol{x}_i$ must hold, in order for this formula to make sense?

(c) For the particular problem with $n = 3$ and $m = 2$,

$$\min f(\boldsymbol{x}) \equiv x_1^2 + \tfrac{2}{3}x_2^3 + \tfrac{2}{3}x_3^3$$

subject to

$$\begin{aligned} x_1 - x_2 + x_3 &= 6 \\ x_1 - 3x_2 - x_3 &= 2 \end{aligned},$$

check that $\boldsymbol{x}_0 = (4, 0, 2)^T$ is a feasible solution. Use the orthogonal factorisation method to calculate Z, and hence the projected gradient vector and projected Hessian matrix at $\boldsymbol{x}_0$. Verify that this projected Hessian matrix is positive definite and carry out <u>one</u> step of Newton's method to obtain $\boldsymbol{x}_1$.

[You may use the fact that the orthogonal factorisation is $A = LQ$, where

$$L \equiv \begin{pmatrix} \sqrt{3} & 0 & 0 \\ \sqrt{3} & 2\sqrt{2} & 0 \end{pmatrix} \quad \text{and} \quad Q \equiv \begin{pmatrix} 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ 0 & -1/\sqrt{2} & -1/\sqrt{2} \\ -2/\sqrt{6} & -1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} .]$$

M3N3/M4N3

4. Consider the nonlinear inequality constrained problem

$$(\dagger) \qquad \begin{array}{c} \min f(\boldsymbol{x}) \\ \text{subject to } \boldsymbol{c}(\boldsymbol{x}) \geq 0; \end{array}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $\boldsymbol{c} : \mathbb{R}^n \mapsto \mathbb{R}^m$ are smooth functions.

(a) State carefully the KKT conditions, which are necessary for $\boldsymbol{x}^\star \in \mathbb{R}^n$ to be a local minimum of ($\dagger$); using $\boldsymbol{g}(\boldsymbol{x}^\star) \equiv \nabla f(\boldsymbol{x}^\star) \in \mathbb{R}^n$ to denote the gradient vector of f at $\boldsymbol{x}^\star$ and $\boldsymbol{g}_i(\boldsymbol{x}^\star) \equiv \nabla c_i(\boldsymbol{x}^\star) \in \mathbb{R}^n$ to denote the gradient vector of the ith component of $\boldsymbol{c}$ at $\boldsymbol{x}^\star$.

(b) Consider the following particular example of ($\dagger$) with $n = 2$ and $m = 4$:

$$(\ddagger) \qquad f(\boldsymbol{x}) \equiv \left(x_1 - \tfrac{3}{2}\right)^2 + \left(x_2 - \tfrac{1}{2}\right)^4 \quad \text{and} \quad \boldsymbol{c}(\boldsymbol{x}) \equiv \begin{pmatrix} 1 - x_1 - x_2 \\ 1 - x_1 + x_2 \\ 1 + x_1 - x_2 \\ 1 + x_1 + x_2 \end{pmatrix}.$$

(i) Draw the feasible set in $\mathbb{R}^2$ for these inequality constraints.

(ii) Write down the KKT conditions for ($\ddagger$).

(iii) By carefully considering <u>all</u> the allowed possibilities for zero Lagrange multipliers, determine the unique solution $\boldsymbol{x}^\star \in \mathbb{R}^2$ for these KKT equations. Make sure you state which constraints are active at $\boldsymbol{x}^\star$ and give the values for the corresponding Lagrange multipliers.