Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2005

This paper is also taken for the relevant examination for the Associateship.

M3N3/M4N3 Optimisation

Date: Thursday 19th June 2005 Time: 10 am – 12 noon

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

Calculators may not be used.

1. (a) Give the definition of $\boldsymbol{x}^\star \in \mathbb{R}^n$ being a local minimum of $f : \mathbb{R}^n \mapsto \mathbb{R}$ and, if f is smooth, prove by contradiction that the gradient condition

$$(\dagger) \qquad \boldsymbol{\nabla} f(\boldsymbol{x}^\star) \equiv \boldsymbol{g}(\boldsymbol{x}^\star) = \mathbf{0}$$

is necessary for $\boldsymbol{x}^\star$ to satisfy this definition. What is the definition of $\boldsymbol{x}^\star$ being a strict local minimum of f?

(b) State the definition of a symmetric matrix $\mathsf{G} \in \mathbb{R}^{n\times n}$ being

- positive definite
- positive semi-definite.

By using an orthonormal set of eigenvectors for G, prove that these definitions are equivalent to

- all eigenvalues of G are strictly positive
- all eigenvalues of G are non-negative

respectively.

(c) If $\boldsymbol{x}^\star \in \mathbb{R}^n$ is a local minimum of a smooth $f : \mathbb{R}^n \mapsto \mathbb{R}$, prove by contradiction that, in addition to ($\dagger$), the Hessian matrix

$$\nabla^2 f(\boldsymbol{x}^\star) \equiv \mathsf{H}(\boldsymbol{x}^\star)$$

must be positive semi-definite. If $\mathsf{H}(\boldsymbol{x}^\star)$ is actually positive definite, prove that $\boldsymbol{x}^\star$ is a strict local minimum.

(d) By calculation, show that the gradient of $f : \mathbb{R}^2 \mapsto \mathbb{R}$ defined by

$$f(x_1, x_2) \equiv 100 \left(x_2 - x_1^2\right)^2 + (1 - x_1)^2$$

is zero at precisely one point and that the Hessian matrix is positive definite there.

2. 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 $\boldsymbol{x}_{i+1}$ from the current iterate $\boldsymbol{x}_i$. [It is not enough just to state the formula!] What condition on the Hessian matrix $\nabla^2 f(\boldsymbol{x}_i) \equiv \mathsf{H}(\boldsymbol{x}_i)$ needs to hold in order for this strategy to make sense?

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

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

for all $d > 0$, where $\overline{B}(\boldsymbol{x}^\star, d) \equiv \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{x}^\star\| \le d\}$. Carry out the following steps to deduce that the Newton iterates converge quadratically to $\boldsymbol{x}^\star$ from any starting value $\boldsymbol{x}_0$ close enough to satisfy

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

where $d_0 \equiv \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|$.

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

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

implies that B is also non-singular with

$$\|\mathsf{B}^{-1}\| \le \frac{\|\mathsf{A}^{-1}\|}{1 - \|\mathsf{A}^{-1}\|\|\mathsf{A} - \mathsf{B}\|};$$

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

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

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

$$(\dagger) \qquad \|\boldsymbol{x}_{i+1} - \boldsymbol{x}^\star\| < \tfrac{3}{2}\,\|\mathsf{H}(\boldsymbol{x}^\star)^{-1}\|\,\gamma(d_0)\,\|\boldsymbol{x}_i - \boldsymbol{x}^\star\|^2.$$

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

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

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

$$\min f(\boldsymbol{x}),$$

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

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

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

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

In addition, if $\mathsf{G} + \mu^\star \mathsf{I}$ is positive definite, show that $\boldsymbol{x}^\star$ 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 $\boldsymbol{y}^\star \in \mathbb{R}^n$ is a local minimum for the inequality constrained problem

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

then $\exists\, \lambda^\star \geq 0$ such that

$$\boldsymbol{g}(\boldsymbol{y}^\star) = \lambda^\star \boldsymbol{g}_c(\boldsymbol{y}^\star), \quad \lambda^\star c(\boldsymbol{y}^\star) = 0$$
$$\boldsymbol{g}_c(\boldsymbol{y}^\star)^T\boldsymbol{x} = 0 \quad \Rightarrow \quad \boldsymbol{x}^T\,[\mathsf{H}(\boldsymbol{y}^\star) - \lambda^\star \mathsf{H}_c(\boldsymbol{y}^\star)]\,\boldsymbol{x} \geq 0,$$

where $\boldsymbol{g}(\boldsymbol{x}) \equiv \boldsymbol{\nabla} f(\boldsymbol{x})$, $\mathsf{H}(\boldsymbol{x}) \equiv \nabla^2 f(\boldsymbol{x})$, $\boldsymbol{g}_c(\boldsymbol{x}) \equiv \boldsymbol{\nabla} c(\boldsymbol{x})$ and $\mathsf{H}_c(\boldsymbol{x}) \equiv \nabla^2 c(\boldsymbol{x})$.]

4. Define carefully what is meant by both a global and a local minimum for the linear equality constrained quadratic problem

$$(\dagger) \qquad \begin{array}{c} \min q(\boldsymbol{x}) \equiv c - \boldsymbol{b}^T\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^T\mathsf{G}\boldsymbol{x} \\ \text{subject to } \mathsf{A}\boldsymbol{x} = \boldsymbol{d}; \end{array}$$

where $q : \mathbb{R}^n \mapsto \mathbb{R}$ is constructed from $c \in \mathbb{R}$, $\boldsymbol{b} \in \mathbb{R}^n$ and a symmetric $n \times n$ matrix G, and the constraint depends on $\mathsf{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{d} \in \mathbb{R}^m$. Assuming the orthogonal subspace decomposition

$$\mathbb{R}^n = \mathcal{R} \oplus \mathcal{N},$$

where $\mathcal{R} \subseteq \mathbb{R}^n$ denotes the range-space of A^T and $\mathcal{N} \subseteq \mathbb{R}^n$ denotes the null-space of A, prove by contradiction that a necessary condition for $\boldsymbol{x}^\star \in \mathbb{R}^n$ to be a local minimum for $(\dagger)$ is that

$$\mathsf{G}\boldsymbol{x}^\star - \boldsymbol{b} \in \mathcal{R}.$$

Explain why this implies that, if $\boldsymbol{x}^\star \in \mathbb{R}^n$ is a local minimum for $(\dagger)$, then $\exists\, \boldsymbol{\lambda}^\star \in \mathbb{R}^m$ such that $(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star) \in \mathbb{R}^n \times \mathbb{R}^m$ is a solution of the system

$$(\S) \qquad \begin{pmatrix} \mathsf{G} & \mathsf{A}^T \\ \mathsf{A} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{d} \end{pmatrix}.$$

Prove that it is impossible for $(\dagger)$ to have a local minimum if

$$(\ddagger) \qquad \boldsymbol{z} \in \mathcal{N} \quad \Rightarrow \quad \boldsymbol{z}^T\mathsf{G}\boldsymbol{z} \geq 0$$

does <u>not</u> hold; on the other hand, if $(\ddagger)$ does hold, show that any solution $(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)$ of $(\S)$ means that $\boldsymbol{x}^\star$ is a global minimum for $(\dagger)$.

For the particular quadratic problem

$$\min \sum_{i=1}^{n} i x_i^2 \quad \text{subject to} \quad \sum_{i=1}^{n} x_i = K,$$

where K is a given positive constant, set up and solve $(\S)$, an $(n+1) \times (n+1)$ system for $\boldsymbol{x}^\star \in \mathbb{R}^n$ and $\lambda^\star \in \mathbb{R}$. Verify that $(\ddagger)$ holds for this particular problem and, for $n = 3$ and $K = 10$, numerically evaluate the global constrained minimum.

5. Define what is meant by a global minimum for the nonlinear inequality constrained problem

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

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $\boldsymbol{c} : \mathbb{R}^n \mapsto \mathbb{R}^m$ are smooth functions. What is a strict global minimum of $(\dagger)$, and explain why there can be at most one?

$\boldsymbol{x}^\star \in \mathbb{R}^n$ is said to satisfy the KKT conditions if $\exists\, \boldsymbol{\lambda}^\star \in \mathbb{R}^m$ such that

$$\begin{gathered} \boldsymbol{g}(\boldsymbol{x}^\star) = \sum_{i=1}^{m} \lambda_i^\star \boldsymbol{g}_i(\boldsymbol{x}^\star), \\ \boldsymbol{\lambda}^\star \geq \mathbf{0}, \quad \boldsymbol{c}(\boldsymbol{x}^\star) \geq \mathbf{0}, \\ \lambda_i^\star c_i(\boldsymbol{x}^\star) = 0 \quad i = 1, \ldots, m. \end{gathered}$$

Here $\boldsymbol{g}(\boldsymbol{x}^\star) \equiv \boldsymbol{\nabla} f(\boldsymbol{x}^\star) \in \mathbb{R}^n$ is the gradient vector of f at $\boldsymbol{x}^\star$ and $\boldsymbol{g}_i(\boldsymbol{x}^\star) \equiv \boldsymbol{\nabla} c_i(\boldsymbol{x}^\star) \in \mathbb{R}^n$ is the gradient vector of the i^{th} component of $\boldsymbol{c}$ at $\boldsymbol{x}^\star$.

(a) Using the gradient of f, state the definitions of f being a convex or strictly convex function.

(b) State the definition of $\boldsymbol{c}$ being a concave function in two equivalent ways; one using a linear interpolating polynomial and the other using the $m \times n$ matrix whose rows are the gradients of the components of $\boldsymbol{c}$. [You do not have to prove the equivalence.]

(c) Prove that if $\boldsymbol{c}$ is a concave function then the feasible set

$$\mathcal{S} \equiv \left\{ \boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{c}(\boldsymbol{x}) \geq \mathbf{0} \right\}$$

is convex.

(d) If f is convex and $\boldsymbol{c}$ is concave, prove that $\boldsymbol{x}^\star \in \mathbb{R}^n$ satisfying the KKT conditions is a sufficient condition for $\boldsymbol{x}^\star$ to be a global minimum of $(\dagger)$.

(e) If f is strictly convex and $\boldsymbol{c}$ is concave, prove that $\boldsymbol{x}^\star \in \mathbb{R}^n$ satisfying the KKT conditions is a sufficient condition for $\boldsymbol{x}^\star$ to be the strict global minimum of $(\dagger)$.