Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May–June 2006

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

M3N3/M4N3

Optimisation

Date: Wednesday, 31st May 2006 Time: 2 pm – 4 pm

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. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be the quadratic function

$$f(\boldsymbol{x}) \equiv c - \boldsymbol{b}^T\boldsymbol{x} + \tfrac{1}{2}\boldsymbol{x}^T\mathsf{G}\boldsymbol{x},$$

where $c \in \mathbb{R}$, $\boldsymbol{b} \in \mathbb{R}^n$ and $\mathsf{G} \in \mathbb{R}^{n\times n}$ is symmetric and positive definite. Explain why $\boldsymbol{\nabla} f$ has exactly one zero $\boldsymbol{x}^\star \in \mathbb{R}^n$. If a new vector norm is defined by

$$\|\boldsymbol{x}\|_{\mathsf{G}} \equiv \sqrt{\boldsymbol{x}^T\mathsf{G}\boldsymbol{x}} \qquad \forall \boldsymbol{x} \in \mathbb{R}^n,$$

show that

$$\|\boldsymbol{x} - \boldsymbol{x}^\star\|_{\mathsf{G}}^2 = \boldsymbol{g}(\boldsymbol{x})^T\mathsf{G}^{-1}\boldsymbol{g}(\boldsymbol{x}) \qquad \forall \boldsymbol{x} \in \mathbb{R}^n,$$

where $\boldsymbol{g} \equiv \boldsymbol{\nabla} f$, and

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^\star) = \tfrac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^\star\|_{\mathsf{G}}^2 \qquad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Explain why $\boldsymbol{x}^\star$ is the <u>strict</u> global minimum of f.

Consider the steepest descent algorithm with <u>exact line search</u> for finding $\boldsymbol{x}^\star$, i.e.

$$(\dagger) \quad \boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \alpha_i\boldsymbol{g}(\boldsymbol{x}_i),$$

and explain carefully what the formula for α_i is, and how it is derived. (A graph illustrating the line search would be helpful here.) Analyze the convergence of this algorithm by using the above vector norm to establish the following results.

(a) Use $(\dagger)$ to obtain

$$\|\boldsymbol{x}_i - \boldsymbol{x}^\star\|_{\mathsf{G}}^2 - \|\boldsymbol{x}_{i+1} - \boldsymbol{x}^\star\|_{\mathsf{G}}^2 = 2\alpha_i\boldsymbol{g}(\boldsymbol{x}_i)^T\mathsf{G}\left[\boldsymbol{x}_i - \boldsymbol{x}^\star\right] - \alpha_i^2\boldsymbol{g}(\boldsymbol{x}_i)^T\mathsf{G}\boldsymbol{g}(\boldsymbol{x}_i)$$

and hence

$$\|\boldsymbol{x}_i - \boldsymbol{x}^\star\|_{\mathsf{G}}^2 - \|\boldsymbol{x}_{i+1} - \boldsymbol{x}^\star\|_{\mathsf{G}}^2 = \frac{\left[\boldsymbol{g}(\boldsymbol{x}_i)^T\boldsymbol{g}(\boldsymbol{x}_i)\right]^2}{\boldsymbol{g}(\boldsymbol{x}_i)^T\mathsf{G}\boldsymbol{g}(\boldsymbol{x}_i)}.$$

(b) Use the result from (a) to show that

$$\|\boldsymbol{x}_{i+1} - \boldsymbol{x}^\star\|_{\mathsf{G}}^2 = \left\{1 - \frac{\left[\boldsymbol{g}(\boldsymbol{x}_i)^T\boldsymbol{g}(\boldsymbol{x}_i)\right]^2}{\left[\boldsymbol{g}(\boldsymbol{x}_i)^T\mathsf{G}\boldsymbol{g}(\boldsymbol{x}_i)\right]\left[\boldsymbol{g}(\boldsymbol{x}_i)^T\mathsf{G}^{-1}\boldsymbol{g}(\boldsymbol{x}_i)\right]}\right\}\|\boldsymbol{x}_i - \boldsymbol{x}^\star\|_{\mathsf{G}}^2.$$

(c) <u>Assuming</u> the Kantorovich inequality

$$\frac{\left[\boldsymbol{x}^T\boldsymbol{x}\right]^2}{\left[\boldsymbol{x}^T\mathsf{G}\boldsymbol{x}\right]\left[\boldsymbol{x}^T\mathsf{G}^{-1}\boldsymbol{x}\right]} \le \frac{4\lambda_{\max}\lambda_{\min}}{\left[\lambda_{\max} + \lambda_{\min}\right]^2} \qquad \forall \boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^n,$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalues of G respectively; explain why

$$0 < \frac{4\lambda_{\max}\lambda_{\min}}{\left[\lambda_{\max} + \lambda_{\min}\right]^2} \le 1.$$

Use the result from (b) to deduce

$$\|\boldsymbol{x}_{i+1} - \boldsymbol{x}^\star\|_{\mathsf{G}}^2 \le \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^2\|\boldsymbol{x}_i - \boldsymbol{x}^\star\|_{\mathsf{G}}^2.$$

2. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a smooth uniformly convex function: i.e. $\exists\, \mathrm{H}_{\min} > 0$ such that

$$\begin{aligned} f(\boldsymbol{y}) - f(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{x})^T[\boldsymbol{y} - \boldsymbol{x}] &\geq \tfrac{1}{2}\mathrm{H}_{\min}\|\boldsymbol{y} - \boldsymbol{x}\|^2 \\ [\boldsymbol{g}(\boldsymbol{y}) - \boldsymbol{g}(\boldsymbol{x})]^T[\boldsymbol{y} - \boldsymbol{x}] &\geq \mathrm{H}_{\min}\|\boldsymbol{y} - \boldsymbol{x}\|^2 \\ \boldsymbol{z}^T\mathsf{H}(\boldsymbol{x})\boldsymbol{z} &\geq \mathrm{H}_{\min}\|\boldsymbol{z}\|^2 \end{aligned}$$

for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n$, where $\boldsymbol{g}(\boldsymbol{x}) \equiv \boldsymbol{\nabla} f(\boldsymbol{x})$ and $\mathsf{H}(\boldsymbol{x}) \equiv \nabla^2 f(\boldsymbol{x})$ are respectively the gradient vector and Hessian matrix of f at $\boldsymbol{x}$.

(a) (i) Prove that $f(\boldsymbol{x}) \to \infty$ as $\|\boldsymbol{x}\| \to \infty$, and hence deduce that f has a global minimum, at $\boldsymbol{x}^\star$ say.

(ii) Given that $\boldsymbol{g}(\boldsymbol{x}^\star) = \boldsymbol{0}$, prove that $\boldsymbol{g}$ has no other zero and that $\boldsymbol{x}^\star$ is the strict global minimum of f.

(b) For a given starting value $\boldsymbol{x}_0 \in \mathbb{R}^n$, let

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i, \qquad \text{where } \boldsymbol{p}_i \equiv -\mathsf{H}(\boldsymbol{x}_i)^{-1}\boldsymbol{g}(\boldsymbol{x}_i),$$

be a safeguarded Newton iteration: with the steplength $\alpha_i > 0$ chosen to satisfy

$$(\dagger) \qquad \begin{aligned} f(\boldsymbol{x}_i) - f(\boldsymbol{x}_{i+1}) &\geq \rho\alpha_i \boldsymbol{g}(\boldsymbol{x}_i)^T\mathsf{H}(\boldsymbol{x}_i)^{-1}\boldsymbol{g}(\boldsymbol{x}_i) \\ \boldsymbol{g}(\boldsymbol{x}_{i+1})^T\boldsymbol{p}_i &\geq \sigma \boldsymbol{g}(\boldsymbol{x}_i)^T\boldsymbol{p}_i \end{aligned}$$

for some fixed $0 < \rho < \sigma < 1$ independent of i. Establish the convergence of the iteration through the following steps.

(i) Explain why there exists $\mathrm{H}_{\max} > 0$ such that

$$(\ddagger) \qquad f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0) \Rightarrow \|\mathsf{H}(\boldsymbol{x})\| \leq \mathrm{H}_{\max}.$$

(ii) Use the first inequality in (†) to prove that $\{\alpha_i \boldsymbol{g}(\boldsymbol{x}_i)^T\mathsf{H}(\boldsymbol{x}_i)^{-1}\boldsymbol{g}(\boldsymbol{x}_i)\}$ converges to zero as $i \to \infty$.

(iii) Use the second inequality in (†), together with (‡) to bound

$$[\boldsymbol{g}(\boldsymbol{x}_{i+1}) - \boldsymbol{g}(\boldsymbol{x}_i)]^T \boldsymbol{p}_i,$$

to show that $\{\alpha_i\}$ cannot approach zero; therefore proving that

$$\{\boldsymbol{g}(\boldsymbol{x}_i)^T\mathsf{H}(\boldsymbol{x}_i)^{-1}\boldsymbol{g}(\boldsymbol{x}_i)\} \to 0 \quad \text{as } i \to \infty.$$

(iv) Use (‡) to show that $\{\boldsymbol{g}(\boldsymbol{x}_i)\} \to \boldsymbol{0}$ as $i \to \infty$ and then use the uniform convexity of f to prove that $\{\boldsymbol{x}_i\} \to \boldsymbol{x}^\star$ as $i \to \infty$.

[You may assume without proof that it is always possible to satisfy (†).]

3. Consider the norm-constrained quadratic optimisation problem

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

where $q : \mathbb{R}^n \mapsto \mathbb{R}$ with $c \in \mathbb{R}$, $\boldsymbol{b}$ is a non-zero vector in $\mathbb{R}^n$, G is an $n \times n$ symmetric matrix, and $\Delta > 0$. You are reminded that $\boldsymbol{x}^\star$ solves $(\dagger)$ if and only if $\exists\, \mu^\star \ge 0$ such that

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

If G has eigenvalues and corresponding orthonormal eigenvectors

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{n-1} \ge \lambda_n \quad \text{and} \quad \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{n-1}, \boldsymbol{u}_n,$$

with

$$\boldsymbol{b} \equiv \sum_{j=1}^{n} \beta_j \boldsymbol{u}_j,$$

explain why $\|\boldsymbol{x}(\mu)\|$, where

$$\boldsymbol{x}(\mu) \equiv [\mathsf{G} + \mu \mathsf{I}]^{-1} \boldsymbol{b},$$

is strictly monotonically decreasing for $\mu > -\lambda_n$ and why $\lim_{\mu \to \infty} \|\boldsymbol{x}(\mu)\| = 0$.

(a) If $\lambda_n > 0$, draw a graph of $\|\boldsymbol{x}(\mu)\|$ for $\mu \ge 0$ and describe how $(\dagger)$ has exactly one solution in either of the cases

- $\|\boldsymbol{x}(0)\| < \Delta$
- $\|\boldsymbol{x}(0)\| \ge \Delta$.

(b) If $\lambda_n \le 0$ with $\beta_n \ne 0$, draw a graph of $\|\boldsymbol{x}(\mu)\|$ for $\mu > -\lambda_n$ and explain carefully why $(\dagger)$ has exactly one solution.

(c) If $\lambda_n \le 0$ and $\underline{\lambda_{n-1} > \lambda_n}$ with $\beta_n = 0$, draw a graph of $\|\boldsymbol{x}(\mu)\|$ for $\mu > -\lambda_n$ and explain carefully why $(\dagger)$ may have one solution or may have two solutions.

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

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

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a smooth function, $\mathsf{A} \in \mathbb{R}^{m\times n}$ 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 (†) is that $\exists\, \boldsymbol{\lambda}^\star \in \mathbb{R}^m$ such that

$$\boldsymbol{\nabla} f(\boldsymbol{x}^\star) + \mathsf{A}^T\boldsymbol{\lambda}^\star = \boldsymbol{0}.$$

Hence write down a system of $m+n$ equations that the unknowns $\boldsymbol{x}^\star$ and $\boldsymbol{\lambda}^\star$ must satisfy, in order for $\boldsymbol{x}^\star$ to be a local minimum for (†). If, in addition, f is a convex function, prove that a solution of this system implies that $\boldsymbol{x}^\star$ is a global minimum for (†). [You may use any property of convex functions without proof.]

For the particular quadratic problem $m = 2$, $n = 3$, $f(\boldsymbol{x}) \equiv x_1^2 + x_2^2 + x_3^2$,

$$\mathsf{A} = \begin{pmatrix} 1 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{d} = \begin{pmatrix} 4 \\ -2 \end{pmatrix},$$

set up and solve the 5×5 system of linear equations for $\boldsymbol{x}^\star$ and $\boldsymbol{\lambda}^\star$.

5. Consider 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.

(a) Define what is meant by a local minimum for ($\dagger$).

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

(c) For the particular problem

$$f(\boldsymbol{x}) \equiv x_1 + x_2 \quad \text{and} \quad \boldsymbol{c}(\boldsymbol{x}) \equiv \begin{pmatrix} 2 - x_1^2 - x_2^2 \\ x_2 \end{pmatrix}$$

with $n = 2$ and $m = 2$:

(i) write down the KKT conditions and find all the solutions,

(ii) display the feasible set on a graph and use the objective function to determine the constrained minimum.