IMPERIAL COLLEGE LONDON

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)
MAY–JUNE 2007

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

M3N3/M4N3 OPTIMISATION

Date: Thursday, 17th May 2007 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.

Statistical tables will not be available.

1. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a smooth function, which is bounded below, and consider the iteration

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i;$$

where the direction $\boldsymbol{p}_i \in \mathbb{R}^n$ satisfies $\boldsymbol{\nabla} f(\boldsymbol{x}_i)^T \boldsymbol{p}_i < 0$ and $\alpha_i > 0$ satisfies the two conditions

$$(\dagger) \qquad \begin{aligned} f(\boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i) &\leq f(\boldsymbol{x}_i) + \rho \alpha_i \boldsymbol{\nabla} f(\boldsymbol{x}_i)^T \boldsymbol{p}_i \\ \boldsymbol{\nabla} f(\boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i)^T \boldsymbol{p}_i &\geq \sigma \boldsymbol{\nabla} f(\boldsymbol{x}_i)^T \boldsymbol{p}_i \end{aligned}$$

for fixed parameters $0 < \rho < \sigma < 1$ independent of i.

(a) Illustrate the condition $\boldsymbol{\nabla} f(\boldsymbol{x}_i)^T \boldsymbol{p}_i < 0$ by sketching a typical graph of

$$\phi_i(\alpha) \equiv f(\boldsymbol{x}_i + \alpha \boldsymbol{p}_i) \quad \text{for } \alpha \geq 0.$$

(b) Re-write the two conditions of $(\dagger)$ in terms of $\phi_i(\alpha)$ and $\phi_i'(\alpha)$, and indicate clearly on your graph in (a) the values of $\alpha \geq 0$ which satisfy the <u>first</u> condition of $(\dagger)$.

(c) You are now asked to prove that it will always be possible to find an $\alpha_i > 0$ which satisfies $(\dagger)$.

(i) Explain why $\exists\, \widehat{\alpha}_i > 0$ such that the first condition of $(\dagger)$ is satisfied $\forall \alpha \in [0, \widehat{\alpha}_i]$.

(ii) Apply the point mean value theorem to $\phi_i(\widehat{\alpha}_i) - \phi_i(0)$ and carefully explain why this produces an $\alpha_i > 0$ satisfying <u>both</u> conditions of $(\dagger)$.

(d) Make the <u>two</u> additional assumptions that the iteration does not finitely terminate, i.e. $\boldsymbol{\nabla} f(\boldsymbol{x}_i) \neq \boldsymbol{0} \quad \forall i \geq 0$, and that the Hessian matrix $\mathsf{H}(\boldsymbol{x}) \equiv \nabla^2 f(\boldsymbol{x})$ satisfies the bound

$$(\ddagger) \qquad \|\mathsf{H}(\boldsymbol{x})\| \leq \mathrm{H}_{\max} \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

(i) Combine $(\ddagger)$ and the second condition in $(\dagger)$ to prove that

$$\alpha_i \geq \frac{\sigma - 1}{\mathrm{H}_{\max}} \frac{\boldsymbol{\nabla} f(\boldsymbol{x}_i)^T \boldsymbol{p}_i}{\|\boldsymbol{p}_i\|^2}.$$

(ii) Insert this result into the first condition of $(\dagger)$ to obtain

$$f(\boldsymbol{x}_{i+1}) \leq f(\boldsymbol{x}_i) - C \cos^2 \theta_i \|\boldsymbol{\nabla} f(\boldsymbol{x}_i)\|^2,$$

where $C > 0$ is a constant independent of i and

$$\cos \theta_i \equiv \frac{-\boldsymbol{\nabla} f(\boldsymbol{x}_i)^T \boldsymbol{p}_i}{\|\boldsymbol{\nabla} f(\boldsymbol{x}_i)\| \|\boldsymbol{p}_i\|}.$$

Write down the formula for C and explain why $\theta_i \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

(iii) Use the fact that f is bounded below to prove that

$$\sum_{i=0}^{\infty} \cos^2 \theta_i \|\boldsymbol{\nabla} f(\boldsymbol{x}_i)\|^2$$

is finite.

(iv) Explain why $\lim_{i \to \infty} \boldsymbol{\nabla} f(\boldsymbol{x}_i) = \boldsymbol{0}$ if <u>either</u> $\boldsymbol{p}_i \equiv -\boldsymbol{\nabla} f(\boldsymbol{x}_i) \quad \forall i \geq 0$ <u>or</u> $\exists\, \widehat{\theta} \in (0, \frac{\pi}{2})$ such that $\theta_i \in [-\widehat{\theta}, \widehat{\theta}] \quad \forall i \geq 0$.

2. (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

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

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 $\mathsf{H}_c(\boldsymbol{x}) \equiv \nabla^2 c(\boldsymbol{x})$.]

3. Consider the linear equality constrained problem

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

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a smooth function, $\mathsf{A} \in \mathbb{R}^{m \times n}$ with $1 \leq m < n$ and $\boldsymbol{d} \in \mathbb{R}^m$.

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

(b) State carefully the orthogonal subspace decomposition of $\mathbb{R}^n$ depending on $\mathcal{R} \subseteq \mathbb{R}^n$, the range-space of A^T, and $\mathcal{N} \subseteq \mathbb{R}^n$, the null-space of A.

(c) Derive the first derivative necessary condition for $\boldsymbol{x}^\star \in \mathbb{R}^n$ to be a local minimum of $(\dagger)$: explaining your condition both in terms of the orthogonal subspace decomposition and by using Lagrange multipliers. Why will the Lagrange multipliers only be unique if A has full rank?

(d) If $\widehat{\boldsymbol{x}} \in \mathbb{R}^n$ satisfies $\mathsf{A}\widehat{\boldsymbol{x}} = \boldsymbol{d}$, use the orthogonal subspace decomposition to describe the solutions $(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{\mu}}) \in \mathbb{R}^n \times \mathbb{R}^m$ of

$$\begin{pmatrix} \mathsf{I} & \mathsf{A}^T \\ \mathsf{A} & 0 \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{p}} \\ \widehat{\boldsymbol{\mu}} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\nabla} f(\widehat{\boldsymbol{x}}) \\ \boldsymbol{0} \end{pmatrix}.$$

Explain why this linear system will always have a solution with unique $\widehat{\boldsymbol{p}}$, but $\widehat{\boldsymbol{\mu}}$ will only be unique if A has full rank.

(i) If $\widehat{\boldsymbol{p}} = \boldsymbol{0}$, explain why $\widehat{\boldsymbol{x}}$ satisfies the first derivative necessary condition for a local minimum of $(\dagger)$.

(ii) If $\widehat{\boldsymbol{p}} \neq \boldsymbol{0}$, explain carefully why

$$\boldsymbol{p} \equiv \frac{\widehat{\boldsymbol{p}}}{\|\widehat{\boldsymbol{p}}\|} \quad \text{solves} \quad \left\{ \begin{array}{c} \min \boldsymbol{\nabla} f(\widehat{\boldsymbol{x}})^T \boldsymbol{p} \\ \mathsf{A}\boldsymbol{p} = \boldsymbol{0} \\ \|\boldsymbol{p}\| = 1 \end{array} \right\}$$

and why it is sensible to call $\widehat{\boldsymbol{p}}$ the steepest descent direction at $\widehat{\boldsymbol{x}}$ for $(\dagger)$.

4. (a) Consider the linear inequality constrained problem

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

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a smooth function, $\mathsf{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{d} \in \mathbb{R}^m$. Suppose $\widehat{\boldsymbol{x}} \in \mathbb{R}^n$ satisfies $\boldsymbol{d} - \mathsf{A}\widehat{\boldsymbol{x}} \geq \mathbf{0}$ and $\widetilde{\mathsf{A}} \in \mathbb{R}^{t\times n}$, with $0 < t < n$, is the submatrix of A corresponding to the active constraints at $\widehat{\boldsymbol{x}}$. Assuming $\widetilde{\mathsf{A}}$ has full rank, explain why

$$\mathsf{P} \equiv \mathsf{I} - \widetilde{\mathsf{A}}^T \left[\widetilde{\mathsf{A}}\widetilde{\mathsf{A}}^T\right]^{-1} \widetilde{\mathsf{A}} \in \mathbb{R}^{n\times n}$$

is the orthogonal projection of $\mathbb{R}^n$ onto the null-space of $\widetilde{\mathsf{A}}$, and define

$$\widehat{\boldsymbol{z}} \equiv \mathsf{P}\boldsymbol{\nabla} f(\widehat{\boldsymbol{x}}) \in \mathbb{R}^n.$$

(i) If $\widehat{\boldsymbol{z}} \neq \mathbf{0}$, prove that $\widehat{\boldsymbol{x}} + \varepsilon\widehat{\boldsymbol{z}}$ is feasible for ($\dagger$) when $|\varepsilon|$ is sufficiently small and also prove that $f(\widehat{\boldsymbol{x}} + \varepsilon\widehat{\boldsymbol{z}}) < f(\widehat{\boldsymbol{x}})$ for $\varepsilon < 0$ and $|\varepsilon|$ is sufficiently small.

(ii) If $\widehat{\boldsymbol{z}} = \mathbf{0}$ and

$$\mathbf{0} \leq \left[\widetilde{\mathsf{A}}\widetilde{\mathsf{A}}^T\right]^{-1} \widetilde{\mathsf{A}}\boldsymbol{\nabla} f(\widehat{\boldsymbol{x}}) \in \mathbb{R}^t,$$

carefully verify that $\widehat{\boldsymbol{x}}$ satisfies the KKT conditions for ($\dagger$).

(b) Consider the two quadratic linear inequality constrained problems

$$\begin{array}{c} \min \boldsymbol{b}^T\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^T\mathsf{G}\boldsymbol{x} \\ \text{subject to } \boldsymbol{d} - \mathsf{A}\boldsymbol{x} \geq \mathbf{0} \end{array} \quad \text{and} \quad \begin{array}{c} \min \boldsymbol{c}^T\boldsymbol{u} + \frac{1}{2}\boldsymbol{u}^T\mathsf{H}\boldsymbol{u} \\ \text{subject to } \boldsymbol{u} \geq \mathbf{0}: \end{array}$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^n$, $\mathsf{G} \in \mathbb{R}^{n\times n}$ is symmetric and invertible, $\mathsf{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{d} \in \mathbb{R}^m$, $\boldsymbol{u} \in \mathbb{R}^m$,

$$\mathsf{H} \equiv \mathsf{A}\mathsf{G}^{-1}\mathsf{A}^T \in \mathbb{R}^{m\times m} \quad \text{and} \quad \boldsymbol{c} \equiv \mathsf{A}\mathsf{G}^{-1}\boldsymbol{b} + \boldsymbol{d} \in \mathbb{R}^m.$$

Write down the KKT conditions for each of these problems (using $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ for the respective Lagrange multipliers). If $(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)$ satisfies the KKT conditions for the first problem, verify that

$$(\boldsymbol{u}^\star, \boldsymbol{\mu}^\star) \equiv (\boldsymbol{\lambda}^\star, \boldsymbol{d} - \mathsf{A}\boldsymbol{x}^\star)$$

satisfies the KKT conditions for the second problem. On the other hand, if $(\boldsymbol{u}^\star, \boldsymbol{\mu}^\star)$ satisfies the KKT conditions for the second problem, verify that

$$(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star) \equiv \left(-\mathsf{G}^{-1}\left[\mathsf{A}^T\boldsymbol{u}^\star + \boldsymbol{b}\right], \boldsymbol{u}^\star\right)$$

satisfies the KKT conditions for the first problem.

5. Consider the nonlinear inequality constrained problem

$$\min f(\boldsymbol{x}) \equiv \left(x_1 - \tfrac{9}{4}\right)^2 + (x_2 - 2)^2$$

$$\text{subject to} \quad \begin{cases} x_2 - x_1^2 \geq 0, \\ 6 - x_2 - x_1 \geq 0, \end{cases} \quad x_1 \geq 0.$$

(a) Graph the constraints and draw contours for the objective function.

(b) Write down the KKT conditions and verify that these are satisfied at the point $\boldsymbol{x}^\star \equiv \left(\frac{3}{2}, \frac{9}{4}\right)^T$.

(c) Use your graph to present a geometrical interpretation of the KKT conditions at $\boldsymbol{x}^\star$.

(d) Explain why the objective function is strictly convex and the feasible set is convex. Use these results to carefully prove that $\boldsymbol{x}^\star$ is <u>the strict global</u> constrained minimum for the above problem.