UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

Faculty of Science and Engineering

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327

SEMESTER: Spring 2014

MODULE TITLE: Optimisation

DURATION OF EXAMINATION: $2\frac{1}{2}$ hours

LECTURER: Dr. J. Kinsella

GRADING SCHEME: 20% +80%

EXTERNAL EXAMINER: Prof. T. Myers

INSTRUCTIONS TO CANDIDATES: **Answer four questions correctly for full marks; 80%. See the Appendix at the end of the paper for some useful results.**

1 (a) Suppose that a function $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 on $\mathbb{R}^n$ and has gradient $\nabla f(x) \equiv g(x)$. If a vector $\mathbf{p}$ satisfies the condition $p^T g(x) < 0$, $\mathbf{p}$ is called a "descent direction" for f at x. Explain carefully the importance of this definition for optimisation algorithms — justifying any statement that you make. 4

(b) Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 and p_k is a descent direction at x_k. Assume that f is bounded below along the ray $\{x_k + \alpha p_k | \alpha > 0\}$. Show that if $0 < c_1 < c_2 < 1$ there exist step lengths α satisfying the Wolfe conditions Eqs. 13a and 13b. 12

(c) Draw a sketch to illustrate your proof. 4

(d) Show that if $c_1 > \dfrac{1}{2}$, a line search satisfying the Wolfe conditions would exclude the minimiser of a strictly convex quadratic objective function so the restriction $c_1 \leq \dfrac{1}{2}$ is reasonable. 5

Hint: write $\phi(\alpha) = \phi_0 + \alpha\phi_0' + \frac{1}{2}\alpha^2\phi_0''$, find the α–range determined by the first Wolfe condition and compare with α^*, the α–value that minimises $\phi(\alpha)$. Take $\phi_0'' > 0$ as given f strictly convex.

2 (a) Prove Zoutendijk's Theorem (Theorem 1 in Appendix C) 15

(b) Explain briefly the significance of the result. 1

(c) Suppose that search directions p_k are generated using a Newton-like method: $p_k = -B_k^{-1} g(x_k)$ where B_k is symmetric and positive definite. For any matrix A let $\|A\|$ be the matrix 2–norm, equal for real symmetric matrices to the absolute value of the largest eigenvalue of A. Show that if $\|B_k\|\|B_k^{-1}\| \leq M$ for all k then $\cos\theta_k \geq 1/M$ where θ_k is defined in Appendix C. 8

(d) Use Zoutendijk's Theorem above to show that in this case $\lim_{k\to\infty} \|g(x_k)\| = 0$. 1

3 (a) Quadratic convergence for a sequence of vectors $\{\mathbf{x}_k\}_{k=1}^{\infty} \in \mathbb{R}^n$ is defined in Appendix D.

(i) Show that quadratic convergence to a point x^* implies that the error $\epsilon_k \equiv \|x_k - x^*\|$ is less than or equal to $\frac{1}{M}(M\epsilon_0)^{2^k}$ where ϵ_0 is the starting value of the error. 4

(ii) If $M < 1/\epsilon_0$ what can you conclude? Explain carefully. 1

(b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be C^3 and let the point x^* satisfy $\nabla f(x^*) \equiv g(x^*) = 0$ and $\nabla^2 f(x^*) \equiv H(x^*)$ is positive definite. Let $R = \|\mathbf{x}_0 - x^*\|$ be the distance from the start point x_0 to x^*. Assume that R is small enough for the condition $R < \frac{1}{k_1 k_2}$ to hold for some $k_1, k_2 > 0$ defined by the two conditions:

(i) the inverse Hessian is bounded in norm: $\|H^{-1}(x)\| \leq k_1$

(ii) the remainder term in a second-order Taylor expansion of $g(x^*)$ centred at x is uniformly bounded: $\|g(x^*) - g(x) - H(x)(x^* - x)\| \leq k_2 \|x^* - x\|^2$

for each x inside the ball $B = \{\|x - x^*\| \leq R\}$.

Show that, under these assumptions,

(i) that the sequence of $\mathbf{x}$ vectors generated by Newton's method ($x_{k+1} = x_k + p_k$ with $p_k = -H^{-1}(x_k)g(x_k)$) converges to x^*. 12

(ii) that the convergence is quadratic. 2

(c) Let x^* be a local minimum for f on for an open ball $\mathcal{B}(x^*)$ centred at x^* (of radius R, say). It can be shown that , $\forall x \in \mathcal{B}$,

$$\|x - x^*\| \leq \frac{\|\nabla f(x)\ \|}{m} \tag{1}$$

$$f(x) - f(x^*) \leq \frac{\|\nabla f(x)\ \|^2}{m}. \tag{2}$$

(i) Show that, as a consequence, if the stopping criterion is $\|\nabla f(x)\ \| \leq \varepsilon$ for some small quantity ε, then if the stop criterion on the gradient holds at two successive points x and x' that: 4

$$\|\Delta x\| \equiv \|x - x'\| \leq \frac{2\varepsilon}{m} \tag{3}$$

and

$$|\Delta f| \equiv |f(x) - f(x')| \leq \frac{2\varepsilon^2}{m}. \tag{4}$$

(ii) If I choose the stop criterion $|\Delta f| \leq K\varepsilon$ (where K is some constant such as 10 or $|f(x_0)|$ and $\varepsilon = \epsilon_M$ — machine epsilon), what corresponding stop criteria should I choose for $\|\nabla f\|$ and $\|\Delta x\|$ based on (3) and (4) in Q3.(c)(i)? 2

4 (a) Prove the Sherman-Morrison-Woodbury formula for the inverse of a matrix $A + \Delta A$ — stated in App. E. 6

(b) Given the DFP update formula:

$$\textbf{DFP}: \quad H_{k+1} = (I - \gamma_k y_k s_k^T) H_k (I - \gamma_k s_k y_k^T) + \gamma_k y_k y_k^T, \qquad (5)$$

where

$$\gamma_k = \frac{1}{y_k^T s_k}, s_k = x_{k+1} - x_k \quad \text{and} \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

show that the DFP-update formula may be written as $H_{k+1} = H_k + \Delta H_k$, with 6

$$\Delta H_k = RSR^T,$$

where

$$R = \begin{bmatrix} y_k & Hs_k \end{bmatrix}, \quad S = \gamma_k \begin{bmatrix} 1 + \gamma_k s_k^T H_k s_k & -1 \\ -1 & 0 \end{bmatrix}.$$

(c) Apply the SMW formula from part (a) to the formula $H_{k+1} = H_k + \Delta H_k$ from part (b) to derive the following equation for the update of the inverse Hessian approximation, J_k that corresponds to the DFP update of H_k in Eq. 5; 13

$$\textbf{DFP} - \textbf{Inverse} \quad J_{k+1} = J_k - \frac{J_k y_k y_k^T J_k}{y_k^T J_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}. \qquad (6)$$

Remember that for any 2×2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{(ad - bc)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

5 Consider the general equality-constrained problem (21) as defined in App. G.

(a) Show that the penalty function $F^k(x)$ defined in (20) in App. F has the property that the (unconstrained) minima x_k of F^k converge to a local minimum x^*. 10

(b) Prove the KKT first-order necessary conditions (23a) and (23b) in App. I for an **equality**-constrained problem . 10

(c) Finally; prove that when inequality constraints are introduced, the extra condition $\lambda_i^* \geq 0$ for all $i \in \mathcal{I}$ is necessary for x^* to be an optimal solution of (22) in App. H. Define the functions $c_i^-(x) = \min\{0, c_i(x)\}$, $j \in \mathcal{I}$. Obviously $c_i^-(x) \leq 0$, $j \in \mathcal{I}$.
Adapt (without repeating the details) your argument for part (b) to show that $\lambda_i^* = -\lim_{k\to\infty} kc_i^-(x_k), i \in \mathcal{I}$. 5

6 Consider the equality-constrained Quadratic Program (Q.P.):

$$\min_x q(x) = \frac{1}{2}x^TQx + x^Td, \tag{7a}$$

$$\text{subject to} \quad a_i^Tx = b_i, \quad i = 1, \ldots, k \tag{7b}$$

(a) Let A be the matrix s.t. the vectors $\{a_i\}_{i\in\mathcal{E}}$ are the columns of A^T. Writing the set of k equality constraints (7b) as the matrix equation $Ax - b = 0$, show that if x^* is a local minimum then the KKT conditions (Appendix I) require that there must be a vector λ^* of Lagrange multipliers such that the following system of equations is satisfied: 4

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ -\lambda^* \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix} \tag{8}$$

(b) If we write $x^* = x_0 + p$, where x_0 is any estimate of the solution and p the required step to the solution, show that (8) can be re-written as: 2

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} g \\ r \end{bmatrix} \tag{9}$$

where the residual $r = Ax_0 - b$, $g = Qx_0 + d$ (the gradient of $q(x_0)$) and $p = x^* - x_0$.

(c) Show that this block matrix equation can be reduced to the problem of solving

$$\left(AQ^{-1}A^T\right)\lambda^* = \left(AQ^{-1}g - r\right). \tag{10}$$

for λ^* and then solving

$$-Qp + A^T\lambda^* = g \tag{11}$$

for p. 2

(**See the following page for part (d) of this question**.)

(d) Solve the **Inequality**-constrained QP:

$$\min_{x\in\mathbb{R}^2} q(x) \text{ subject to } Ax \geq b, \tag{12}$$

with $q(x) = \frac{1}{2}x^T Qx + d^T x$ where $Q = \begin{bmatrix} 5 & -3 \\ -3 & 7 \end{bmatrix}$ and $d = \begin{bmatrix} 1 \\ -9 \end{bmatrix}$ with $A = \begin{bmatrix} -1 & 3 \\ 1 & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 9 \\ -5 \end{bmatrix}$ using the following steps.

(i) Using the fact that $Q^{-1} = \frac{1}{26}\begin{bmatrix} 7 & 3 \\ 3 & 5 \end{bmatrix}$ find x_u, the **unconstrained** minimum of $q(x)$, ($\nabla f(x_u) = 0$). 1

(ii) Confirm that the first constraint, $-x_1 + 3x_2 \geq 9$, is violated at x_u while the second, $x_1 + x_2 \geq -5$ is inactive at x_u. 1

(iii) Check that $x_0 = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$ is the vertex common to the two constraints — the intersection point of the two lines. (A rough sketch of the two lines is useful in the following.) 1

(iv) Calculate the gradient $g(x_0) = Qx_0 + d$. 1

(v) To solve the Inequality constrained problem, use the procedure in part (c) of this question to solve the **Equality**-constrained QP

$$\min_{x\in\mathbb{R}^2} q(x) \text{ subject to } \tilde{A} = \begin{bmatrix} -1 & 3 \end{bmatrix} \text{ and } \tilde{b} = \begin{bmatrix} 9 \end{bmatrix} \tag{13}$$

starting at the point x_0, using the following steps:

A. Explain why the second constraint ($x_1 + x_2 \geq -5$) may be dropped from the working set $\mathcal{W}$. 3

B. Write a numerical (not algebraic) matrix expression for λ. (**Do not perform the matrix arithmetic.**) Note that $r = 0$ at x_0 as both constraints are active at x_0. 1

C. Taking $\lambda = 64/17$, write a matrix expression for p. (**Again, do not perform the matrix arithmetic.**) 1

D. Taking $p = \frac{40}{17}\begin{bmatrix} 3 \\ 1 \end{bmatrix}$, check that $\tilde{A}p = 0$. 1

E. Calculate the new point $x_1 = x_0 + p$ to two decimal places. 1

F. Explain why $\tilde{A}x_1 = \tilde{b}$ (no arithmetic). 1

G. Is x_1 feasible for both constraints? 1

H. Is either constraint active at x_1? 1

I. Explain carefully why the point x_1 is optimal for the full Inequality constrained problem. 3

Appendix of Results

A The Wolfe conditions for the step length α in a line search require that

$$f(x_k + \alpha p_k) \le f(x_k) + c_1 \alpha {p_k}^T g(x_k), \quad (13a)$$

$${p_k}^T g(x_k + \alpha p_k) \ge c_2 {p_k}^T g(x_k) \quad (13b)$$

where $g(x) \equiv \nabla f(x)$ and $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (13b) by

$$|{p_k}^T g(x_k + \alpha p_k)| \le c_2 |{p_k}^T g(x_k)|. \quad (14)$$

B In terms of a "line" function $\phi(\alpha) \equiv f(x + \alpha p)$; the Wolfe conditions for the step length α in a line search require that

$$\phi(\alpha) \le \phi(0) + c_1 \alpha \phi'(0), \quad (14a)$$

$$\phi'(\alpha) \ge c_2 \phi'(0) \quad (14b)$$

where $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (14b) by

$$|\phi'(\alpha)| \le c_2 |\phi'(0)|. \quad (15)$$

C **Theorem 1 (Zoutendijk)** *Consider any iteration of the form* $x_{k+1} = x_k + \alpha_k p_k$*, where* p_k *is a descent direction and* α_k *satisfies the Wolfe conditions Eqs. 13a and 13b in Appendix A above. Suppose that* f *is bounded below in* $\mathbb{R}^n$ *and that* f *is* C^1 *in an open set* $\mathcal{N}$ *containing the level set* $\mathcal{L} \equiv \{x : f(x) \le f(x_0)\}$*, where* x_0 *is the starting point. Also assume that* $g(x)$*, the gradient of* f*, is Lipschitz continuous on* $\mathcal{N}$*, i.e. there exists a constant* L *such that*

$$\|g(x) - g(\bar{x})\| \le L \|x - \bar{x}\|, \quad \textit{for all } x, \bar{x} \in \mathcal{N}. \quad (16)$$

Then

$$\sum_{k \ge 0} \cos^2 \theta_k \, \|g(x_k)\|^2 < \infty, \quad (17)$$

where θ_k *is the angle between* p_k *and the steepest descent direction* $-g(x_k)$*.*

D A sequence $\{x_k\}_{k=1}^{\infty}$ that converges to a point x^* has quadratic convergence if for some positive constant M;

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} \le M, \quad \text{for all } k \text{ sufficiently large.} \quad (18)$$

E The Sherman-Morrison-Woodbury formula states that if a square non-singular matrix A is updated by

$$A' = A + \Delta A, \quad \text{where } \Delta A = RST^T$$

where R, T are $n \times p$ matrices for $1 < p < n$ and S is $p \times p$ then

$$A'^{-1} \equiv (A + \Delta A)^{-1} = A^{-1} - A^{-1}RU^{-1}T^TA^{-1}, \tag{19}$$

where $U = S^{-1} + T^TA^{-1}R$.

F Let x^* be a local minimum of the equality-constrained optimisation problem defined in App. G . For each positive integer k, define a **penalty function**

$$F^k(x) = f(x) + \frac{k}{2}\|c(x)\|^2 + \frac{\alpha}{2}\|x - x^*\|^2, \tag{20}$$

where $\alpha > 0$ is arbitrary.

G The general equality constrained optinisation problem is:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad c_i(x) = 0, i \in \mathcal{E} \tag{21}$$

H The general inequality constrained optinisation problem is:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \begin{cases} c_i(x) = 0 & i \in \mathcal{E}, \\ c_i(x) \geq 0 & i \in \mathcal{I} \end{cases} \tag{22}$$

I The first-order KKT necessary conditions for a point x^* with optimal multipliers λ^* to be a local solution of an inequality-constrained minimisation problem are as follows: Suppose that x^* is a local solution of a general constrained optimisation (as in App. H) problem and that the LICQ holds at x^* (i.e. the active constraint gradients are linearly independent). Then there is a Lagrange multiplier vector λ^*, with components $\lambda_i^*, i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at (x^*, λ^*)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{23a}$$
$$c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E}, \tag{23b}$$
$$c_i(x^*) \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{23c}$$
$$\lambda_i^* \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{23d}$$
$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E} \cup \mathcal{I} \tag{23e}$$

where $\mathcal{L}(\mathbf{x}, \lambda) \equiv f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$.

J The (revised) Simplex algorithm may be stated as:

Algorithm 1

Given $\mathcal{B}, \mathcal{N}, x_B = B^{-1}b \geq 0, x_N = 0$;
Solve $B^T\lambda = c_B$ for λ
$s_N = c_N - N^T\lambda$; (pricing)
if $s_N \geq 0$
then STOP; (optimal point found)
fi
Select $q \in \mathcal{N}$ with $s_q < 0$ as the entering index
Solve $Bd = A_q$ for d;
if $d \leq 0$
then STOP; (problem is unbounded)
fi
Calculate $x_q^+ = \min_{i|d_i>0}(x_B)_i/d_i$ and use p to denote the minimizing i;
Update $x_B^+ = x_B - dx_q^+$, $x_N^+ = (0, \ldots, 0, x_q^+, 0, \ldots, 0)^T$;
Change $\mathcal{B}$ by adding q and removing the basic variable
corresponding to column p of $\mathcal{B}$.