UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

Faculty of Science and Engineering

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327

SEMESTER: Spring 2013

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) Prove Zoutendijk's Theorem (Theorem 1 in Appendix D) 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 D. 8

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

2 (a) Let $f \in C^2(\mathbb{R}^n)$. Let x^* be a local minimum for f in an open ball $\mathcal{B}(x^*)$ centred at x^* (of radius R, say). Suppose that there is a number $m > 0$ such that for all x in the ball $\mathcal{B}(x^*)$

$$m\|d\|^2 \leq d^T H(x) d, \quad \forall d \in \mathbb{R}^n.$$

Show that $\forall x \in \mathcal{B}$,

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

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

Hints: For (1), Taylor's Theorem can be written as: $\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)\, p\, dt$. For (2), use $F(t) = f(x^* + t(x - x^*))$ for $t \in [0,1]$ and show that $F''(t) \geq 0$. 14

(b) Show using the results in (a) that if the stopping criterion $\|\nabla f(x)\| \leq \varepsilon$ on the gradient holds at x and x' then $\|\Delta x\| \equiv \|x - x'\| \leq \frac{2\varepsilon}{m}$ and $|\Delta f| \equiv |f(x) - f(x')| \leq \frac{2\varepsilon^2}{m}$. 5

(c) If I impose the stopping criterion $|\Delta f| \leq K\varepsilon$ on $|\Delta f|$ what stopping criteria should I impose on $\|\nabla f\|$ and $\|\Delta x\|$? 4

(d) Briefly explain how these results may be adapted to produce stop criteria for α, $\phi(\alpha)$ and $\phi'(\alpha)$ when performing a line search where $\phi(\alpha) \equiv f(x + \alpha p)$. (Take $x' = x + \alpha p$ and assume that $\|p\|$ is $O(1)$.) 2

3 "Nearly exact" trust region methods are defined in Appendices E, F and G.

If we define $p(\lambda) = -(B + \lambda I)^{-1} g$, we need to show that the equation $\|p(\lambda)\| = \Delta$ may be solved for λ — referring to Theorem 2 in Appendix F where necessary. Proceed as follows:

(a) Use the fact that a symmetric matrix B can be written $B = Q\Lambda Q^T$ to show that: 6

$$p(\lambda) = -Q(\Lambda + \lambda I)^{-1} Q^T g = -\sum_{j=1}^{n} \frac{q_j^T g}{\lambda + \lambda_j} q_j, \qquad (3)$$

where $\lambda_1 < \lambda_2 < \cdots < \lambda_n$ are the eigenvalues of B, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and the orthonormal matrix $Q = [q_1 q_2 \ldots q_n]$ where the column vectors q_i are the eigenvectors of B.

(b) Derive a formula for $\|p(\lambda)\|^2$. 1

(c) Sketch the graph of $\|p(\lambda)\|^2$ for $\lambda \geq -\lambda_1$ incorporating the properties that follow from the formula that you found in part (b). What conclusions may be drawn about the existence of a solution λ^* to the equation $\|p(\lambda)\| = \Delta$? 3+1

(d) Show that Algorithm 1 in Appendix G correctly implements Newton's method for root-finding applied to the problem $1/\|p(\lambda)\| = 1/\Delta$. 10

(e) When $q_1^T g = 0$, the preceding analysis is no longer valid. 1+3

(i) Explain why.

(ii) How can $p(\lambda^*)$ be computed in this case?

4 The Fletcher-Reeves (FR) version of the non-linear conjugate gradient algorithm (Alg. 2) is given in Appendix H.

(a) Suppose that the algorithm is implemented with a step length α_k that satisfies the strong Wolfe conditions with $0 < c_2 < \frac{1}{2}$ and that the norm of the gradient is bounded above. Assume Zoutendijk's Theorem (Theorem 1 in Appendix D) and the result that

$$-\frac{1}{1 - c_2} \leq \frac{p_k^T g_k}{\|g_k\|^2} \leq \frac{2c_2 - 1}{1 - c_2}, \text{for all } k = 0, 1, \ldots. \qquad (4)$$

and prove that the FR cgm has the global convergence property — i.e. that 23

$$\liminf_{k\to\infty} \|g_k\| = 0.$$

Hints: (each step uses some or all of those preceding it)

(i) Use a proof by contradiction — so assume that $\|g_k\|$ is bounded below, say by $\gamma > 0$. 0

(ii) First use Eq. 4 above to show that

$$\frac{1-2c_2}{1-c_2}\frac{\|g_k\|}{\|p_k\|} \le \cos\theta_k \le \frac{1}{1-c_2}\frac{\|g_k\|}{\|p_k\|}.$$

3

(iii) Then use Zoutendijk's Theorem to show that $\sum_{k=0}^{\infty}\frac{\|g_k\|^4}{\|p_k\|^4}$ converges. 3

(iv) Use the strong version of the second Wolfe condition (Eq. 9 in App. A) and the bound on $\cos\theta_k$ above to show that $|g_k^T p_{k-1}| \le \frac{c_2}{1-c_2}\|g_{k-1}\|^2$. 4

(v) Use the update rule for p_{k+1} in the FR CGM algorithm (Appendix H) (replacing $k+1$ by k) to show that

$$\|p_k\|^2 \le c_3\|g_k\|^2 + \beta_k^2\|p_{k-1}\|^2$$

with $c_3 = \frac{1+c_2}{1-c_2}$. 4

(vi) Finally iterate this inequality to obtain a contradiction with the result of hint (iii) — you can assume that $\|g_k\|$ is bounded above, say by $\bar{\gamma}$. 9

(b) Explain the significance of the result. 2

5 (a) Prove the Sherman-Morrison-Woodbury formula for the inverse of a matrix $A + \Delta A$ — stated in App. I. 5

(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 5

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

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; 15

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

Remember that for any 2×2 matrix

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

6 This question relates to the application of the KKT necessary conditions ((22a)–(22e) in App. M) to the Linear Program (LP) in Standard Form (SF)

$$\min c^T x, \text{ subject to } Ax = b,\ x \geq 0. \tag{7}$$

where c and x are vectors in $\mathbb{R}^n$, b is a vector in $\mathbb{R}^m$, and A is an $m \times n$ matrix.

(a) Show that the KKT conditions for the SF LP (Eq. 7 above) take the form: 5

$$A^T \lambda + s = c \tag{7a}$$
$$Ax = b \tag{7b}$$
$$x \geq 0 \tag{7c}$$
$$s \geq 0 \tag{7d}$$
$$x_i s_i = 0,\ i = 1, \dots, n. \tag{7e}$$

(b) Show that the KKT conditions for the "dual" LP

$$\max b^T \lambda, \text{ subject to } A^T \lambda \leq c. \tag{8}$$

(when the dual LP is re-fornulated into SF) are the same as those for the primal problem (7). 5

(c) Show that the optimal solutions of the primal and dual problems are equal — $c^T x^* = b^T \lambda^*$ and explain the significance of this result. 2

(See overleaf for Part (d) .)

(d) A vector x is a basic feasible point for the SF LP (7) above if it is feasible and if there exists a subset $\mathcal{B}$ of the index set $\{1, 2, \dots, n\}$ such that

- $\mathcal{B}$ contains exactly m indices ($m < n$).
- $i \notin \mathcal{B} \Rightarrow x_i = 0$ (i.e. the bound $x_i \geq 0$ is inactive only if $i \in \mathcal{B}$).
- The $m \times m$ matrix B defined by $B = [A_i]_{i \in \mathcal{B}}$ is non-singular (A_i is the i^{th} column of A)

Show that if we partition the vectors x, s and c according to the index sets $\mathcal{B}$ and $\mathcal{N} = \{1, 2, \dots, n\} \backslash \mathcal{B}$, using the notation

$$\begin{aligned} x_B &= [x_i]_{i \in \mathcal{B}}, & x_N &= [x_i]_{i \in \mathcal{N}} \\ s_B &= [s_i]_{i \in \mathcal{B}}, & s_N &= [s_i]_{i \in \mathcal{N}} \\ c_B &= [c_i]_{i \in \mathcal{B}}, & c_N &= [c_i]_{i \in \mathcal{N}} \end{aligned}$$

the Simplex algorithm (Alg. 3 in App. N) may be derived from the KKT conditions (7a)– (7e). 13

(Take $\mathcal{B} = \{1, 2, \dots, m\}$, $\mathcal{N} = \{m+1, \dots, n\}$ and $N = [A_i]_{i \in \mathcal{N}}$.)

Appendix of Results

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

$$
\begin{aligned}
f(x_k + \alpha p_k) &\leq & f(x_k) + c_1 \alpha {p_k}^T g(x_k), \qquad &(8a)\\
{p_k}^T g(x_k + \alpha p_k) &\geq & c_2 {p_k}^T g(x_k) \qquad &(8b)
\end{aligned}
$$

where $g(x) \equiv \nabla f(x)$ and $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (8b) by

$$|{p_k}^T g(x_k + \alpha p_k)| \leq c_2 |{p_k}^T g(x_k)|. \tag{9}$$

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

$$
\begin{aligned}
\phi(\alpha) &\leq & \phi(0) + c_1 \alpha \phi'(0), \qquad &(9a)\\
\phi'(\alpha) &\geq & c_2 \phi'(0) \qquad &(9b)
\end{aligned}
$$

where $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (9b) by

$$|\phi'(\alpha)| \leq c_2 |\phi'(0)|. \tag{10}$$

C The Trust Region Method is based on the problem:

$$\min_{p \in \mathbb{R}^n} m(p) \equiv f_0 + g^T p + \frac{1}{2} p^T B p, \quad \text{such that } \|p\| \leq \Delta, \tag{11}$$

where f_0 is a fixed scalar, g a fixed vector in $\mathbb{R}^n$, B a fixed $n \times n$ matrix and Δ a fixed positive scalar. The "dogleg" method finds an approximate solution to (16) by replacing the (unknown) curved trajectory for $p^*(\Delta)$ with a path consisting of two line segments. The first line segment runs from the starting point to the unconstrained minimiser along the steepest descent direction defined by

$$p^U = -\frac{g^T g}{g^T B g} g \tag{12}$$

while the second line segment runs from p^U to $p^B \equiv -B^{-1} g$. We can define the trajectory as a path $\tilde{p}(\tau)$ parameterised by τ as follows:

$$\tilde{p}(\tau) = \begin{cases} \tau p^U, & 0 \leq \tau \leq 1, \\ p^U + (\tau - 1)(p^B - p^U), & 1 \leq \tau \leq 2. \end{cases} \tag{13}$$

D **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. 8a and 8b 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*

$$\|g(x) - g(\bar{x})\| \le L\|x - \bar{x}\|, \quad \textit{for all } x,\ \bar{x} \in \mathcal{N}. \tag{14}$$

Then

$$\sum_{k\ge 0} \cos^2\theta_k \, \|g(x_k)\|^2 < \infty, \tag{15}$$

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

E The Trust Region problem:

$$\min_{p\in\mathbb{R}^n} m(p) \equiv f_0 + g^T p + \frac{1}{2} p^T B p, \quad \text{such that } \|p\| \le \Delta, \tag{16}$$

The "nearly exact" method seeks to solve (16) as accurately as possible using the results from Theorem 2 below. (Here f_0 is a fixed scalar, g a fixed vector in $\mathbb{R}^n$, B a fixed $n \times n$ symmetric matrix and Δ a fixed positive scalar.)

F **Theorem 2** *The vector* p^* *is a global solution of the problem* (16) *if and only if there is a scalar* $\lambda \ge 0$ *such that the following conditions are satisfied:*

$$(B + \lambda I)p^* = -g \tag{17a}$$

$$\lambda(\Delta - \|p^*\|) = 0 \tag{17b}$$

$$(B + \lambda I) \text{ is positive semi-definite.} \tag{17c}$$

G **Algorithm 1 (Exact Trust Region)**

begin
  Given $\lambda_0 > 0$, $\Delta > 0$, $\varepsilon > 0$
  while $n < n_{max}$ AND $\text{abs}(\|p_n(\lambda)\| - \Delta) > \varepsilon$ do
    Factor $B + \lambda^{(n)} I = R^T R$
    Solve $R^T R p_n = -g$, $R^T q_n = p_n$
    $\lambda^{(n+1)} := \lambda^{(n)} + \left(\frac{\|p_n\|}{\|q_n\|}\right)^2 \left(\frac{\|p_n(\lambda)\| - \Delta}{\Delta}\right)$
    $n := n + 1$
  end
end

H **Algorithm 2 (FR-CGM)**

begin
  Given x_0.
  set $g_0 \leftarrow \nabla f_0, p_0 \leftarrow -g_0, k \leftarrow 0$;
  while $g_k \neq 0$ do
    $\alpha_k \leftarrow$ Result of line search along p_k;
    $x_{k+1} \leftarrow x_k + \alpha_k p_k$;
    $g_{k+1} \leftarrow \nabla f_{k+1}$
    $\beta^{FR}_{k+1} \leftarrow \frac{\|g_{k+1}\|^2}{\|g_k\|^2}$;
    $p_{k+1} \leftarrow -g_{k+1} + \beta^{FR}_{k+1} p_k$;
    $k \leftarrow k + 1$;
  end (while)
end

I 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

$$A'^{-1} \equiv (A + \Delta A)^{-1} = A^{-1} - A^{-1} R U^{-1} T^T A^{-1}, \tag{18}$$

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

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

$$F^k(x) = f(x) + \frac{k}{2}\|c(x)\|^2 + \frac{\alpha}{2}\|x - x^*\|^2, \tag{19}$$

where $\alpha > 0$ is arbitrary.

K The general equality constrained optinisation problem is:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad c_i(x) = 0, i \in \mathcal{E} \tag{20}$$

L The general inequality constrained optinisation problem is:

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

M 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. L) 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^*)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{22a}$$
$$c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E}, \tag{22b}$$
$$c_i(x^*) \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{22c}$$
$$\lambda_i^* \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{22d}$$
$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E} \cup \mathcal{I} \tag{22e}$$

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

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

Algorithm 3

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 $x_B^+ = x_B - dx_q^+$, $x_N^+ = (0, \dots, 0, x_q^+, 0, \dots, 0)^T$;
Change $\mathcal{B}$ by adding q and removing the basic variable corresponding to column p of $\mathcal{B}$.