UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

Faculty of Science and Engineering

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327

SEMESTER: Spring 2011

MODULE TITLE: Optimisation

DURATION OF EXAMINATION: 2 1/2 hours

LECTURER: Dr. J. Kinsella

GRADING SCHEME: 20% (Project)+80% (Exam)

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 E) 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 E. 8

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

2 "Nearly exact" trust region methods are defined in Appendices F, G and H.

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 G 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

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

where $\lambda_1 < \lambda_2 < \cdots < \lambda_n$ are the eigenvalues of B, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ 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 2 in Appendix H 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(e) When $q_1^T g = 0$, the preceding analysis is no longer valid. 1+3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?

3 The Fletcher-Reeves (FR) version of the non-linear conjugate gradient algorithm (Alg. 3) is given in Appendix I.

(a) Suppose that the algorithm is implemented with a step length α_k that satisfies the strong Wolfe conditions with $0 < c_2 < \frac{1}{2}$.
Prove that that the search directions p_k satisfy the inequalities: 16

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

(b) Show that it follows that the FR cgm generates descent directions. 2

(c) Explain the significance of the result. 1

(d) Although the FR cgm can be shown to converge, it tends to "stick". 2+2+2

(i) Show using (2) from part (a) that if $\dfrac{\|g_k\|}{\|p_k\|}$ is large then $\cos\theta_k \approx 0$ (where θ_k is defined in Appendix E).

(ii) Explain why this is not satisfactory.

(iii) Show that, if this behaviour occurs, it tends to persist as $p_{k+1} \approx p_k$.

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

(b) Given the DFP update formula:

$$\textbf{DFP}: \quad H_{k+1} = (I - \gamma_k y_k s_k^T) H_k (I - \gamma_k s_k y_k^T) + \gamma_k y_k y_k^T, \quad (3)$$

where

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

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. 3; 15

$$\textbf{DFP} - \textbf{Inverse} \quad J_{k+1} = J_k - \frac{J_k y_k y_k^T J_k}{y_k^T J_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}. \tag{4}$$

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 (18) as defined in App. L.

(a) Show that the penalty function $F^k(x)$ defined in (17) in App. K 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 (20a) and (20b) in App. N 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 (19) in App. M. 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 This question relates to the application of the KKT necessary conditions ((20a)–(20e) in App. N) to the Linear Program (LP) in Standard Form (SF)

$$\min c^T x, \text{ subject to } Ax = b, \ x \geq 0. \tag{5}$$

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 (5) take the form: 5

$$A^T\lambda + s = c \tag{5a}$$
$$Ax = b \tag{5b}$$
$$x \geq 0 \tag{5c}$$
$$s \geq 0 \tag{5d}$$
$$x_i s_i = 0, \ i = 1, \dots, n. \tag{5e}$$

(b) Show that the KKT conditions for the “dual” LP

$$\max b^T\lambda, \text{ subject to } A^T\lambda \leq c. \tag{6}$$

(when the dual LP is re-fornulated into SF) are the same as those for the primal problem (5). 6

(c) A vector x is a basic feasible point for the SF LP (5) above if it is feasible and if there exists a subset $\mathcal{B}$ of the index set $\{1, 2, \ldots, n\}$ 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 $\mathcal{N} = \{1, 2, \ldots, n\} \backslash \mathcal{B}$, 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. 4 in App. P) may be derived from the KKT conditions (5a)– (5e). 14

Appendix of Results

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

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha {p_k}^T g(x_k), \tag{6a}$$

$${p_k}^T g(x_k + \alpha p_k) \geq c_2 {p_k}^T g(x_k) \tag{6b}$$

where $g(x) \equiv \nabla f(x)$ and $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (6b) by

$$|{p_k}^T g(x_k + \alpha p_k)| \leq c_2 |{p_k}^T g(x_k)|. \tag{7}$$

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

$$\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0), \tag{7a}$$

$$\phi'(\alpha) \geq c_2 \phi'(0) \tag{7b}$$

where $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (7b) by

$$|\phi'(\alpha)| \leq c_2 |\phi'(0)|. \tag{8}$$

C **Algorithm 1 (Backtracking Line Search)**

begin
  Choose $\bar{\alpha} > 0$ and $\rho, c \in (0, 1)$
  $\alpha := \bar{\alpha}$
  while $\phi(\alpha) \geq \phi(0) + c\alpha\phi'(0)$ do
    $\alpha := \rho\alpha$
  end
  $\alpha_k := \alpha$
end

D The Trust Region Method is based on the problem:

$$\min_{p \in \mathbb{R}^n} m(p) \equiv f_0 + g^T p + \frac{1}{2} p^T B p, \quad \text{such that } \|p\| \leq \Delta, \tag{9}$$

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 (14) 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

$$p^u = -\frac{g^T g}{g^T B g} g \tag{10}$$

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:

$$\tilde{p}(\tau) = \begin{cases} \tau p^U, & 0 \le \tau \le 1, \\ p^U + (\tau - 1)(p^B - p^U), & 1 \le \tau \le 2. \end{cases} \tag{11}$$

E **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. 6a and 6b 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*

$$\|g(x) - g(\bar{x})\| \le L\|x - \bar{x}\|, \quad \textit{for all } x, \bar{x} \in \mathcal{N}. \tag{12}$$

Then

$$\sum_{k \ge 0} \cos^2 \theta_k \, \|g(x_k)\|^2 < \infty, \tag{13}$$

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

F The Trust Region problem:

$$\min_{p \in \mathbb{R}^n} m(p) \equiv f_0 + g^T p + \frac{1}{2} p^T B p, \quad \text{such that } \|p\| \le \Delta, \tag{14}$$

The "nearly exact" method seeks to solve (14) 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.)

G **Theorem 2** *The vector* p^* *is a global solution of the problem* (14) *if and only if there is a scalar* $\lambda \ge 0$ *such that the following conditions are satisfied:*

$$(B + \lambda I)p^* = -g \tag{15a}$$

$$\lambda(\Delta - \|p^*\|) = 0 \tag{15b}$$

$$(B + \lambda I) \textit{ is positive semi-definite.} \tag{15c}$$

H **Algorithm 2 (Exact Trust Region)**

begin
Given $\lambda_0 > 0$, $\Delta > 0$, $\varepsilon > 0$
while $n < n_{max}$ AND 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

I **Algorithm 3 (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;
$x_{k+1} \leftarrow x_k + \alpha_k p_k;$
$g_{k+1} \leftarrow \nabla f_{k+1}$
$\beta_{k+1}^{FR} \leftarrow \frac{\|g_{k+1}\|^2}{\|g_k\|^2};$
$p_{k+1} \leftarrow -g_{k+1} + \beta_{k+1}^{FR} p_k;$
$k \leftarrow k + 1;$
end (while)
end

J 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

$$A'^{-1} \equiv (A + \Delta A)^{-1} = A^{-1} - A^{-1} R U^{-1} T^T A^{-1}, \tag{16}$$

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

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

$$F^k(x) = f(x) + \frac{k}{2}\|c(x)\|^2 + \frac{\alpha}{2}\|x - x^*\|^2, \tag{17}$$

where $\alpha > 0$ is arbitrary.

L The general equality constrained optinisation problem is:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad c_i(x) = 0, i \in \mathcal{E} \tag{18}$$

M The general inequality constrained optinisation problem is:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \begin{cases} c_i(x) = 0 & i \in \mathcal{E}, \\ c_i(x) \geq 0 & i \in \mathcal{I} \end{cases} \tag{19}$$

N 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. M) problem and that the LICQ holds at x^* (the active constraint gradients are linearly independant). 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^*)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{20a}$$

$$c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E}, \tag{20b}$$

$$c_i(x^*) \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{20c}$$

$$\lambda_i^* \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{20d}$$

$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E} \cup \mathcal{I} \tag{20e}$$

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

O The Second-Order necessary conditions for a point x^* with optimal multipliers λ^* to be a local solution of an equality-constrained minimisation problem are as follows: Suppose that x^* is a local solution of an equality-constrained problem as defined in App. L and that the LICQ (see App. N) constraint qualification is satisfied. Let λ^* be a Lagrange multiplier vector such that the first-order (KKT) necessary conditions in App. N are satisfied,

Then $w^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0$, for all vectors w such that $w \perp \nabla c_i(x^*)$, for each $i \in \mathcal{E}$.

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

Algorithm 4

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 $\mathcal{B}$.