UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

Faculty of Science and Engineering

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327

SEMESTER: Spring 2009

MODULE TITLE: Optimisation

DURATION OF EXAMINATION: 2 1/2 hours

LECTURER: Dr. J. Kinsella

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

EXTERNAL EXAMINER: Prof. J. Flavin

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 2 in Appendix F) 15

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

(c) Suppose that the 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 as defined above. 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 C, D and E.

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 1 in Appendix D 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 $\|p(\lambda)\|^2$. 2

(c) Show that $\|p(\lambda)\|^2$ has the following properties for $\lambda \geq -\lambda_1$.

(i) $\|p(\lambda)\|^2 \to +\infty$ as $\lambda \to -\lambda_1$ from the right. 1

(ii) $\|p(\lambda)\|^2$ is monotone decreasing for $\lambda \geq -\lambda_1$. 1

(iii) $\|p(\lambda)\|^2 \to 0$ as $\lambda \to +\infty$. 1

(iv) The equation $\|p(\lambda)\|^2 = \Delta^2$ has a single root $\lambda = -\lambda^*$ in the interval $(-\lambda_1, \infty)$. 1

(d) Sketch the graph of $\|p(\lambda)\|^2$ for $\lambda \geq -\lambda_1$. What conclusions may be drawn about the existence of a solution to the equation $\|p(\lambda)\| = \Delta$? 2

(e) Show that Algorithm 1 in Appendix E correctly implements Newton's method for root-finding applied to the problem $1/\|p(\lambda)\| = 1/\Delta$. 11

3 Different versions of the NonLinear Conjugate Gradient Method differ from the FR CGM (Alg. 2 in Appendix G) in the definition of the coefficient β_k. For **any** β_k define $r_k = \frac{\beta_k}{\beta_k^{DY}}$ (β_k^{DY} defined in App. H) . In the following we will assume that the coefficient β_k satisfies

$$r_k \in [-c, 1], \quad \text{where } c = \frac{1-c_2}{1+c_2} \tag{2}$$

(a) Use the update formula $p_k = -g_k + \beta_k p_{k-1}$ (where β_k satisfies (2)) to show that: 4

$$g_k^T p_k = \beta_k^{DY} \left\{ p_{k-1}^T g_{k-1} + (r_k - 1) g_k^T p_{k-1} \right\}. \tag{3}$$

(b) Show that we can now express β_k as: 4

$$\beta_k = \xi_k \frac{g_k^T p_k}{g_{k-1}^T p_{k-1}}, \quad \text{where} \tag{4}$$

$$\xi_k = \frac{r_k}{1 + (r_k - 1) l_{k-1}} \quad \text{and} \quad l_{k-1} = \frac{g_k^T p_{k-1}}{g_{k-1}^T p_{k-1}}.$$

(c) Defining $\zeta_k = \frac{1+(r_k-1)l_{k-1}}{l_{k-1}-1}$, check that $g_k^T p_k = \zeta_k \|g_k\|^2$. (Hint: use (3)). 8

(d) Using the results derived above, prove that the method has the property that if the same assumptions as those of Zoutendijk's Theorem (Theorem 2 in Appendix F) hold — in particular that

- the weak Wolfe conditions hold (Appendix A)
- f is bounded below

then the algorithm either stops at a stationary point ($\|g_k\| = 0$) or $\liminf_{k\to\infty} \|g_k\| = 0$. 9

(You need not check that the method generates descent directions.)

Hint: start by deriving the equation

$$\|p_k\|^2 = \beta_k^2 \|p_{k-1}\|^2 - \|g_k\|^2 - 2{p_k}^T g_k$$

from the update formula $p_k = -g_k + \beta_k p_{k-1}$; then divide across by $(p_k^T g_k)^2$ and use the formulas for β_k and $p_k^T g_k$ in terms of ξ_k and ζ_k derived in parts (b) and (c) above. You may assume that $\xi_k^2 \le 1$.

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

(b) Given the DFP update formula:

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

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 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

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

5 Derive the KKT first-order necessary conditions — (23a)– (23e) in App. M for a general **inequality**-constrained problem (defined in App. L) and also the second-order necessary conditions (defined in App. N) for a general **equality**-constrained problem as defined in App. K

Assume that the penalty function $F_k(x)$ defined in (20) in App. J has the property that the (unconstrained) minima x_k of F_k converge to a local minimum x^*.

(a) First prove the KKT first-order necessary conditions (23a) and (23b) in App. M for an **equality**-constrained problem as defined in App. L. 10

(b) Now extend these first-order necessary conditions to a general **inequality**-constrained problem (defined in App. L) — i.e. justify conditions (23d) and (23e) in App. M. 6

(c) Finally, derive the second-order necessary conditions (defined in App. N) for a general **equality**-constrained problem as defined in App. K. 9

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

$$\min_x q(x) = \frac{1}{2}x^T Q x + x^T d, \tag{7a}$$

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

(a) Let A be the matrix whose columns are the vectors $\{a_i\}_{i\in\mathcal{E}}$.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 M) 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. 4

(d) Given $q(x) = x^2 - 2xy + 2y^2 + 4x - 5y$ and the **inequality** constraint $3x - 2y \geq 3$, solve the QP using the method you derived in (c). (First check whether the unconstrained solution is feasible — if it is, you are finished; if not the inequality constraint must be active at the optimal point.) 15

Appendix of Results

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

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha {p_k}^T g(x_k), \tag{11a}$$
$${p_k}^T g(x_k + \alpha p_k) \geq c_2 {p_k}^T g(x_k) \tag{11b}$$

where $g(x) \equiv \nabla f(x)$ and $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (11b) by

$$|{p_k}^T g(x_k + \alpha p_k)| \leq c_2 |{p_k}^T g(x_k)|. \tag{12}$$

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

$$\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0), \tag{12a}$$
$$\phi'(\alpha) \geq c_2 \phi'(0) \tag{12b}$$

where $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (12b) by

$$|\phi'(\alpha)| \leq c_2 |\phi'(0)|. \tag{13}$$

C The Trust Region problem:

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

The "nearly exact" method seeks to solve (14) as accurately as possible using the results from Theorem 1 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.)

D **Theorem 1** *The vector* p^* *is a global solution of the problem* (14) *if and only if there is a scalar* $\lambda \geq 0$ *such that the following conditions are satisfied:*

$$(B + \lambda I) p^* = -g \tag{15a}$$
$$\lambda(\Delta - \|p^*\|) = 0 \tag{15b}$$
$$(B + \lambda I) \textit{ is positive semi-definite.} \tag{15c}$$

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

begin
  Given $\lambda_0 > 0, \Delta > 0, \varepsilon > 0$
  while $l < l_{max} \wedge \text{abs}(\|p_l(\lambda)\| - \Delta) > \varepsilon$ **do**
    Factor $B + \lambda^{(l)} I = R^T R$
    Solve $R^T R p_l = -g$, $R^T q_l = p_l$
    $\lambda^{(l+1)} := \lambda^{(l)} + \left(\frac{\|p_l\|}{\|q_l\|}\right)^2 \left(\frac{\|p_l(\lambda)\| - \Delta}{\Delta}\right)$
    $l := l + 1$
  end
end

F **Theorem 2 (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. 11a and 11b 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* $\mathcal{L} \equiv \{x : f(x) \leq f(x_0)\}$, *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*

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

Then

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

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

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

begin
  Given x_0.
  set $r_0 \leftarrow \nabla f_0, p_0 \leftarrow -r_0, k \leftarrow 0$;
  while $r_k \neq 0$ **do**
    $\alpha_k \leftarrow$ Result of line search along p_k;
    $x_{k+1} \leftarrow x_k + \alpha_k p_k$;
    $r_{k+1} \leftarrow \nabla f_{k+1}$
    $\beta^{FR}_{k+1} \leftarrow \frac{r^T_{k+1} r_{k+1}}{r_k{}^T r_k} \equiv \frac{\|r_{k+1}\|^2}{\|r_k\|^2}$;
    $p_{k+1} \leftarrow -r_{k+1} + \beta^{FR}_{k+1} p_k$;
    $k \leftarrow k + 1$;
  end (while)
end

H The Dai-Yuan (DY) version of the Conjugate Gradient Method replaces β^{FR} in Alg. 2 (App, G) by

$$\beta_{k+1}^{DY} = \frac{\|g_{k+1}\|^2}{p_k^T y_k}, \tag{18}$$

where $y_k \equiv g_{k+1} - g_k$

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

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

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

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**

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

where $\alpha > 0$ is arbitrary.

K The general equality constrained optinisation problem is:

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

L The general inequality constrained optinisation problem is:

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

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

N 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:

Theorem 3 (Second-Order N. C.'s — E. Constr.) *Suppose that x^* is a local solution of an equality-constrained problem as defined in App. K and that the LICQ constraint qualification is satisfied. Let λ^* be a Lagrange multiplier vector such that the first-order (KKT) necessary conditions in App. M 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 $i \in \mathcal{E}$.