UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

College of Informatics and Electronics

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327

SEMESTER: Spring 2006

MODULE TITLE: Optimisation

DURATION OF EXAMINATION: 2 1/2 hours

LECTURER: Dr. J. Kinsella

PERCENTAGE OF TOTAL MARKS: 80%

EXTERNAL EXAMINER: Prof. J. King

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) Given the one-dimensional "line" function

$$\phi(\alpha) = 1 - 2\alpha + \frac{3}{2}\alpha^2 - \frac{1}{3}\alpha^3$$

show that the strong Wolfe conditions (Appendix B) hold in the interval $[0 \cdot 63, 1 \cdot 74]$ (approximately). 10

(b) The function $\phi(\alpha)$ has a minimum at $\alpha = 1$. Is this consistent with the above result? 1

(c) For any line function ϕ, define $\Phi(\alpha) = \phi(\alpha) - (\phi(0) + c_1\alpha\phi'(0))$ and show that:

(i) For any α, $\Phi(\alpha) \leq 0$ is equivalent to the first Wolfe condition (12a) holding at α. 1

(ii) For any α, $\Phi'(\alpha) \geq 0$ implies that the second (weak) Wolfe condition (12b) holds at α. 2

(d) Use the results from (c) to show that if the first Wolfe condition fails for some $\alpha_0 > 0$ then the interval $(0, \alpha_0)$ contains values for α that satisfy the strong Wolfe conditions. (Assume that $\phi'(0) < 0$.) Illustrate your proof with a sketch graph of Φ. 10

(e) Explain briefly the significance of this result for step-size selection algorithms. 1

2 "Nearly exact" trust region methods are defined in Appendix 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, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ 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 $\|p(\lambda)\|^2$. 2

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

(i) $\|p(\lambda)\|^2$ has a vertical asymptote at $\lambda = -\lambda_1$. 1

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

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

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

(v) 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 $1/\|p(\lambda)\| = 1/\Delta$. 10

3 (a) Given the Preliminary form of the Linear Conjugate Gradient Method (Alg. 2 in Appendix G); show that: 22

$$p_k^T A p_i = 0, \quad \text{for } i = 0, \ldots, k-1. \tag{2}$$

(b) Explain the significance of this result for the performance of the algorithm. 1

(c) How must the algorithm be changed so that it may be used to solve a general non-linear minimisation problem? 2

4 Different versions of the NonLinear Conjugate Gradient Method differ from the FR CGM (Alg. 3 in Appendix H) 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. I) . In the following we will assume that the coefficient β_k satisfies

$$r_k \in [-c, 1], \quad \text{where } c = \frac{1 - c_2}{1 + c_2} \tag{3}$$

(a) Use the update formula $p_k = -g_k + \beta_k p_{k-1}$ (where β_k satisfies (3)) to show that: 5

$$g_k^T p_k = \beta_k^{DY} \left\{ p_{k-1}^T g_{k-1} + (r_k - 1) g_k^T p_{k-1} \right\}. \tag{4}$$

(b) Show that we can now express β_k as: 3

$$\beta_k = \xi_k \frac{g_k^T p_k}{g_{k-1}^T p_{k-1}}, \quad \text{where} \tag{5}$$

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

(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 (4)). 5

(d) Using the results derived above, prove that the method has the property that if

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

then the algorithm either stops at a stationary point ($\|g_k\| = 0$) or $\lim \inf_{k\to\infty} \|g_k\| = 0$. 12

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 $\xi_k^2 \leq 1$.

5 The BFGS method updates an estimate H_k of the Inverse Hessian in the form:

$$H_{k+1} = (I - \gamma_k s_k y_k^T) H_k (I - \gamma_k y_k s_k^T) + \gamma_k s_k s_k^T. \quad (6)$$

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

(a) Derive this update rule by solving the problem: 20

$$\min_H \|H - H_k\| \quad (6a)$$

$$\text{subject to} \quad H = H^T, Hy_k = s_k. \quad (6b)$$

Use the "weighted Frobenius norm": $\|A\|_W \equiv \|W^{\frac{1}{2}} A W^{\frac{1}{2}}\|_F$, where $\|C\|_F^2 \equiv \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2$. The weight W is an unspecified positive definite symmetric matrix satisfying $Ws_k = y_k$.

(b) Show that if H_k is positive definite, then the next iterate H_{k+1} is also (if the BFGS algorithm is used). 4

(c) Explain the significance of this result. 1

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

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

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

(a) 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 J) require that there must be a vector λ^* of Lagrange multipliers such that the following system of equations is satisfied: 5

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix} \quad (8)$$

(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

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} g \\ c \end{bmatrix} \quad (9)$$

where $c = Ax_0 - b$, $g = Gx_0 + d$ (the gradient of $q(x_0)$) and $p = x^* - x_0$.

(c) Let A have full row rank and let Z be the $n \times (n-m)$ matrix whose columns form a basis for the null space of A. ($AZ = 0$.) Show that, if the "reduced-Hessian" matrix $Z^T G Z$ is positive definite, the KKT matrix

$$K = \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \tag{10}$$

is non-singular. 5

(d) Briefly explain the significance of this result. 1

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

$$\left(AG^{-1}A^T\right)\lambda^* = \left(AG^{-1}g - c\right). \tag{11}$$

for λ^* and then solving

$$-Gp + A^T\lambda^* = g \tag{12}$$

for p. 2

(f) Given $q(x) = x^2 + 0.5\,y^2 + 1.5\,z^2 + x + 3\,y + 5\,z$ and the equality constraint $2x + 4y = 1$, solve the QP using the method you derived in (e). 10

Appendix of Results

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

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha {p_k}^T g(x_k), \tag{12a}$$
$${p_k}^T g(x_k + \alpha p_k) \geq c_2 {p_k}^T g(x_k) \tag{12b}$$

where $g(x) \equiv \nabla f(x)$ and $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (12b) by

$$|{p_k}^T g(x_k + \alpha p_k)| \leq c_2 |{p_k}^T g(x_k)|. \tag{13}$$

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

$$\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0), \tag{13a}$$
$$\phi'(\alpha) \geq c_2 \phi'(0) \tag{13b}$$

where $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (13b) by

$$|\phi'(\alpha)| \leq c_2 |\phi'(0)|. \tag{14}$$

C The Trust Region problem:

$$\min_{p \in \mathbb{R}^n} m(p) \equiv f_0 + g^T p + \frac{1}{2} p^T B p, \quad \text{such that } \|p\| \leq \Delta, \tag{15}$$

The "nearly exact" method seeks to solve (15) 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* (15) *if and only if there is a scalar $\lambda \geq 0$ such that the following conditions are satisfied:*

$$(B + \lambda I) p^* = -g \tag{16a}$$
$$\lambda(\Delta - \|p^*\|) = 0 \tag{16b}$$
$$(B + \lambda I) \textit{ is positive semi-definite.} \tag{16c}$$

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. 12a and 12b 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*

$$\|g(x) - g(\bar{x})\| \leq L\|x - \bar{x}\|, \quad \text{for all } x, \bar{x} \in \mathcal{N}. \tag{17}$$

Then

$$\sum_{k \geq 0} \cos^2 \theta_k \, \|g(x_k)\|^2 < \infty, \tag{18}$$

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

G **Algorithm 2 (CGM—Preliminary Version)**

begin
Given x_0.
set $r_0 \leftarrow Ax_0 - b, p_0 \leftarrow -r_0, k \leftarrow 0$;
while $r_k \neq 0$ **do**
$\alpha_k \leftarrow -\frac{r_k^T p_k}{p_k^T A p_k}$; **minimises residual along** p_k
$x_{k+1} \leftarrow x_k + \alpha_k p_k$;
$r_{k+1} \leftarrow Ax_{k+1} - b$;
$\beta_{k+1} \leftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k}$;
$p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k$; **so** p_{k+1} **conjugate to** p_k
$k \leftarrow k + 1$;
end (while)
end

H **Algorithm 3 (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;
$x_{k+1} \leftarrow x_k + \alpha_k p_k$;
$r_{k+1} \leftarrow \nabla f_{k+1}$
$\beta_{k+1}^{FR} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k{}^T r_k} \equiv \frac{\|r_{k+1}\|^2}{\|r_k\|^2}$;
$p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1}^{FR} p_k$;
$k \leftarrow k + 1$;
end (while)
end

I The Dai-Yuan (DY) version of the Conjugate Gradient Method replaces β^{FR} in Alg. 3 by

$$\beta_{k+1}^{DY} = \frac{\|g_{k+1}\|^2}{p_k^T y_k}, \tag{19}$$

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

J The first-order KKT necessary conditions for a point x^* with optimal multipliers λ^* to be a local solution of an equality-constrained minimisation problem

$$\min f(x) \text{ subject to } c_i(x) = 0, i = 1, \ldots, k$$

are

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{20a}$$
$$c_i(x^*) = 0, \quad i = 1, \ldots, k, \tag{20b}$$

where $\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^{k} \lambda_i c_i(x)$

K The second-order KKT necessary conditions for a point x^* with optimal multipliers λ^* to be a local solution of the above equality-constrained minimisation problem are that $d^T \nabla^2 f(x^*) d \geq 0$ for all stationary directions d (directions that satisfy $d^T \nabla c_i(x^*) = 0$.