UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

College of Informatics and Electronics

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327

SEMESTER: Spring 2005

MODULE TITLE: Optimisation

DURATION OF EXAMINATION: 2 1/2 hours

LECTURER: Dr. J. Kinsella

PERCENTAGE OF TOTAL MARKS: 70%

EXTERNAL EXAMINER: Prof. J. King

INSTRUCTIONS TO CANDIDATES: **Answer four questions correctly for full marks; 70%. See the Appendix at the end of the paper for some useful results..**

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

(b) 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. 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

(c) Use Zoutendijk's Theorem above to show that in this case $\lim_{k\to\infty} \|g(x_k)\| = 0$. 2

2 The Dogleg Trust Region method is described in Appendix C.

(a) Show that if B is positive definite then

(i) $\|\tilde{p}(\tau)\|$ is an increasing function of τ and 5

(ii) $m(\tilde{p}(\tau))$ is a decreasing function of τ. 5

(b) Give a detailed argument which demonstrates that the path $\tilde{p}(\tau)$ intersects the trust region boundary $\|p\| = \Delta$ at exactly one point if $\|\tilde{p}(2)\| \equiv \|p_B\| \geq \Delta$ and nowhere otherwise. 2

(c) What value should be selected for p if $\|p_B\| \leq \Delta$? 3

(d) In the case where the vector p is chosen to be on the boundary, explain carefully how the appropriate value of the parameter τ is chosen. 5

(e) Finally, if $p_U = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $p_B = \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix}$ and $\Delta = 1.5$, find the appropriate value of τ. 5

3 Given the Preliminary form of the Linear Conjugate Gradient Method (Appendix D); prove 25

Theorem 1 *Suppose that the k^{th} iterate generated by the conjugate gradient method Alg. 1 is not the solution point x^*. The following properties hold:*

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

$$r_k{}^T r_i = 0, \quad \text{for } i = 0, \ldots, k-1, \tag{2}$$

4 Given the Dai-Yuan version (Appendix F) of the Non-linear Conjugate Gradient Method (Appendix E) prove that the method has the property that under the same assumptions as those of Zoutendijk's Theorem (Appendix B), in particular

- the weak Wolfe conditions hold (Appendix A)
- f is bounded below
- with the exception that we need not assume that the p_k are descent directions

we have that the algorithm either stops at a stationary point ($\|g_k\| = 0$) or $\liminf \|g_k\| = 0$. 25

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

$$H_{k+1} = (I - \gamma_k s_k y_k^T) H_k (I - \gamma_k y_k s_k^T) + \gamma_k s_k s_k^T. \tag{3}$$

Here $s_k = x_{k+1} - x_k$, $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ and $\gamma_k = s_k^T y_k$. Derive this update rule by solving the problem: 25

$$\min_H \|H - H_k\| \tag{3a}$$

$$\text{subject to} \quad H = H^T, H y_k = s_k. \tag{3b}$$

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 matrix satisfying $W s_k = y_k$.

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

$$\min_x q(x) = \frac{1}{2}x^T Gx + x^T d, \tag{4a}$$

$$\text{subject to} \quad a_i^T x = b, \quad i = 1, \dots, k \tag{4b}$$

(a) Writing the set of k equality constraints (4b) as the matrix equation $Ax - b = 0$, show that if x^* is a local minimum then the KKT conditions (Appendix G) require that there must be a vector λ^* of Lagrange multipliers such that the following system of equations is satisfied: 5

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix} \tag{5}$$

(b) If we write $x^* = x + p$, where x is an estimate of the solution and p the required step to the solution, show that (5) can be re-written as: 2

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} g \\ c \end{bmatrix} \tag{6}$$

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

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

$$\left(AG^{-1}A^T\right)\lambda^* = \left(AG^{-1}g - c\right). \tag{7}$$

for λ^* and then solving

$$-Gp + A^T\lambda^* = g \tag{8}$$

for p. 5

(d) Given $q(x) = \frac{1}{2}x_1^2 + 2x_1x_2 - x_2^2 + 3x_1 + 5x_2$. and the equality constraint $x_1 - 2x_2 = 3$, solve the QP using the method you derived in (c). 8

(e) Finally check that your solution satisfies the second-order conditions (Appendix H). 5

Appendix of Results

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

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha {p_k}^T g(x_k), \tag{8a}$$
$${p_k}^T g(x_k + \alpha p_k) \geq c_2 {p_k}^T g(x_k) \tag{8b}$$

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 **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. 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 $\mathcal{L} \equiv \{x : f(x) \leq f(x_0)\}$, 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*

$$\|g(x) - g(\bar{x})\| \leq L\|x - \bar{x}\|, \quad \textit{for all } x, \bar{x} \in \mathcal{N}. \tag{10}$$

Then

$$\sum_{k \geq 0} \cos^2 \theta_k \, \|g(x_k)\|^2 < \infty, \tag{11}$$

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

C The Trust Region Method is based on the problem:

$$\min_{p \in \mathbb{R}^n} m(p) \equiv f_0 + g^T p + \frac{1}{2} p^T B p, \quad \text{such that } \|p\| \leq \Delta, \tag{12}$$

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 (12) 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

$$p^U = -\frac{g^T g}{g^T B g} g \tag{13}$$

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:

$$\tilde{p}(\tau) = \begin{cases} \tau p^U, & 0 \leq \tau \leq 1, \\ p^U + (\tau - 1)(p^B - p^U), & 1 \leq \tau \leq 2. \end{cases} \tag{14}$$

D **Algorithm 1 (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

E The original Fletcher-Reeves Conjugate Gradient Method is given in the following algorithm:

Algorithm 2

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

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

$$\beta_{k+1}^{DY} = \frac{\|g_{k+1}\|^2}{p_k^T y_k}, \tag{15}$$

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

G 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

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{16a}$$
$$c_i(x^*) = 0, \quad i = 1, \ldots, k, \tag{16b}$$

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

H 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 $d^T \nabla c_i(x^*) = 0$.