UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

College of Informatics and Electronics

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327

SEMESTER: Spring 2004

MODULE TITLE: Optimisation

DURATION OF EXAMINATION: 2 1/2 hours

LECTURER: Dr. J. Kinsella

PERCENTAGE OF TOTAL MARKS: 70%

EXTERNAL EXAMINER: Prof. J.D. Gibbon

INSTRUCTIONS TO CANDIDATES: **Answer four questions correctly for full marks; 70%.**

1 In this question you will analyse the convergence of the Steepest Descent Method (S.D.M.). Assume that the matrix Q in the following is symmetric and positive definite. Let Q have eigenvalues $\lambda_1 \leq \lambda_2 \cdots \leq \lambda_n$.

(a) First we derive some results for quadratic functions.

(i) Show that the minimum of a general quadratic function $f(x) = \frac{1}{2}x^T Qx - b^T x$ is at x_0 where x_0 satisfies $Qx_0 = b$. 2

(ii) Show that (apart from a constant term which may be ignored) $f(x)$ can be transformed into $\bar{f}(y) = \frac{1}{2}y^T Qy$ by the change of variables $x = y + x_0$. 2

(iii) Check that the unique minimum of $\bar{f}(x)$ is $x^* = 0$. 1

(iv) Show that the gradient and Hessian of $\bar{f}(x)$ are Qx and Q respectively. 1

(b) Now we derive the convergence rate result.

(i) Show that when the S.D.M. is applied to $\bar{f}(x)$ the norm of the error satisfies: 8

$\|x_{k+1}\|^2 \leq \kappa^2 \|x_k\|$, where $\kappa^2 = \max\left\{(1-\alpha_k\lambda_1)^2, (1-\alpha_k\lambda_1)^2\right\}$.

(Hint: the S.D.M. update can be written $x_{k+1} = x_k - \alpha_k \nabla f(x_k) = (I - \alpha_k Q)x_k$ where α_k is the step size.)

(ii) Show that the α-value that minimises $\max\{(1-\alpha_k\lambda_1)^2, (1-\alpha_k\lambda_1)^2\}$ is $\alpha = \frac{2}{\lambda_1+\lambda_n}$. 6

(iii) Finally show that setting α_k equal to $\frac{2}{\lambda_1+\lambda_n}$ in κ^2 gives the result: 4

$$\|x_{k+1}\| \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right) \|x_k\|.$$

(c) Explain the significance of the result for the performance of the S.D.M. 1

2 Given a function f from $\mathbb{R}^n$ to $\mathbb{R}$, a point x and a search direction p, define $\phi(\alpha) = f(x + \alpha p)$. Then the Wolfe conditions for an inexact line search are:

$$\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0), \quad (1a)$$
$$\phi'(\alpha) \geq c_2 \phi'(0). \quad (1b)$$

with $0 < c_1 < c_2 < 1$.

(a) Show that if p is a descent direction then $\phi'(0) < 0$. 2

(b) Define $\Phi(\alpha) = \phi(\alpha) - (\phi(0) + c_1 \alpha \phi'(0))$ and show that: 1

(i) For any α, $\Phi(\alpha) \leq 0$ is equivalent to the first Wolfe condition (1a) holding at α. 2

(ii) For any α, $\Phi'(\alpha) \geq 0$ implies that the second (weak) Wolfe condition (1b) holds at α.

(iii) $\Phi(0) = 0$. 1

(iv) $\Phi'(0) < 0$ 1

(c) 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 where (1b) is replaced by $|\phi'(\alpha)| \leq c_2 |\phi'(0)| \equiv -c_2 \phi'(0)$. (Assume $\phi'(0) < 0$.) You will find it helpful to draw a rough sketch graph of Φ. 15

(d) Explain briefly the significance of the result for step-size selection algorithms. 3

3 "Nearly exact" trust region methods seek to solve the problem

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

as accurately as possible. (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.)

We are given the following:

Theorem 1 *The vector p^* is a global solution of the problem* (2) *if and only if there is a scalar $\lambda \geq 0$ such that the following conditions are satisfied:*

$$(B + \lambda I)p^* = -g \quad (3a)$$
$$\lambda(\Delta - \|p^*\|) = 0 \quad (3b)$$
$$(B + \lambda I) \text{ is positive semi-definite.} \quad (3c)$$

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. Proceed as follows:

(a) Use the fact that a symmetric matrix B can be written $B = Q \Lambda Q^T$ to show that: 8

$$p(\lambda) = -Q(\Lambda + \lambda I)^{-1} Q^T g = -\sum_{j=1}^{n} \frac{q_j^T g}{\lambda + \lambda_j} q_j, \qquad (4)$$

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 $\lambda = -\lambda_1$. 2

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

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

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

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

(d) Sketch the graph of $\|p(\lambda)\|^2$ for $\lambda \geq -\lambda_1$ — illustrating the above properties. 3

(e) We could use Newton's method for root-finding (solve $F(x) = 0$ using $x_{n+1} = x_n - F(x_n)/F'(x_n)$) to solve the equation $\|p(\lambda)\| = \Delta$. Explain why this is not satisfactory and why the equation $1/\|p(\lambda)\| = 1/\Delta$ is a better choice. 2

4 The Fletcher-Reeves (FR) version of the non-linear conjugate gradient algorithm is:

Algorithm 1

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

(a) Suppose that Alg. 1 is implemented with a step length α_k that satisfies the strong Wolfe conditions with $0 < c_2 < \frac{1}{2}$. Then the method generates descent directions p_k that satisfy the following inequalities: 22

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

(b) Explain the significance of the result. 3

5 The DFP method updates an estimate B_k of the Hessian in the form:

$$B_{k+1} = (I - \gamma_k y_k s_k^T) B_k (I - \gamma_k s_k y_k^T) + \gamma_k y_k y_k^T. \tag{5}$$

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:

$$\min_B \|B - B_k\| \tag{5a}$$

$$\text{subject to} \quad B = B^T, B s_k = y_k. \tag{5b}$$

Here $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$. You will 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 $W y_k = s_k$.

(a) Given the choice $B = K^T A K + P$, where $K = (I - \gamma_k s_k y_k^T)$ and $P = \gamma_k y_k y_k^T$;

(i) Check that $K s_k = 0$ and $P s_k = y_k$. 2

(ii) Show that B satisfies the "secant equation" $B s_k = y_k$ for all choices of the symmetric matrix $n \times n$ A. 1

(iii) Check that this matrix B is always symmetric. 1

(b) Show that for any $n \times n$ matrix C, $\|C\|_W^2 = \mathrm{Trace}(WCWC)$ where $\mathrm{Trace}(A)$ is the sum of its diagonal elements. 8

(c) Now, writing $F(A) = \mathrm{Trace}(WCWC)$ where $C = K^T A K + P - B_k$, show that $\frac{\partial F(A)}{\partial A_{\alpha\beta}} = 0$ implies that $KWCWK^T = 0$. 8

(d) Finally show that the latter matrix equation can be simplified to $KW(A - B_k)WK^T = 0$ using the identities for P,K and W above — allowing us to conclude that $A = B_k$. 5

6 The 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

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{6a}$$
$$c_i(x^*) = 0, \quad \text{for all} \quad i, \tag{6b}$$

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

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

$$\min_x q(x) = \frac{1}{2} x^T G x + x^T d, \tag{7a}$$
$$\text{subject to} \quad a_i^T x = b, \quad i = 1, \ldots, k \tag{7b}$$

(a) Writing the set of k equality constraints (7b) as the matrix equation $Ax - b = 0$, show that the KKT conditions require that there must be a vector λ^* of Lagrange multipliers such that the following system of equations is satisfied: 5

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

(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 (8) can be re-written as: 5

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

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

(c) Take $q(x) = 3x_1^2 + 2x_1x_2 + 4x_2^2 + x_1 - 2x_2$. Calculate the gradient $g = \nabla q(x)$ and the Hessian $G = \nabla^2 q(x)$. 3

(d) Given the equality constraint $x_1 + x_2 = 1$, find A and b. 2

(e) Use (9) to solve the Q.P.: $\min q(x)$ subject to $Ax = b$. Use $(0, 0)^T$ as your starting point x. 10